Defining, Assessing and Measuring Generic Competences

A thesis submitted in fulfilment of the requirements of the degree of
Doctor of Philosophy
in the School of Education
Faculty of Education, Humanities, Law and Theology
Flinders University of South Australia

David D Curtis
B.Sc. (Hons), Grad. Dip. Inst. Uses of Computers,
M.Ed.St., M.A., Ed.D.

February 2010

Declaration

I certify that this thesis does not incorporate without acknowledgment any material previously submitted for a degree or diploma in any university; and that to the best of my knowledge and belief it does not contain any material previously published or written by another person except where due reference is made in the text.

Abstract

This thesis reports an investigation into generic skills, a class of skills that appear to be broadly applicable to many work, social and civic contexts. Two major generic skills schemes were proposed in Australia, namely the key competencies (Australian Education Council. Mayer Committee, 1992) and the employability skills initiative (Australian Chamber of Commerce and Industry & Business Council of Australia, 2002). The implementation of these schemes is reviewed and several difficulties in their implementation are identified. The most significant issues are thought to be the definition and assessment of these skills.

The issue of definition occurs for generic skills as a class of constructs and arises in relation to each skill proposed as generic. Generic skills could be perceived as representations of either general intelligence or as particular kinds of intelligence. They could also be seen as components of competence. The representation of generic skills as aspects of competence, involving the deployment of cognitive and metacognitive processes, appears to be a fruitful approach to the investigation of generic skills.

In addition to defining generic skills as a class of constructs, each generic skill requires definition. For the research reported in this thesis, one commonly recognised generic skill, problem solving, is selected for investigation. Problem solving is defined as a set of processes that are deployed in identifying, defining, planning, executing, monitoring and evaluating problems and their solutions.

The second major issue identified in the implementation of generic skills schemes is assessment. A body of literature on assessment is reviewed. Assessment is found to serve two major sets of purposes, namely summative and formative. A variety of methods has been used in the assessment of generic skills, most of these methods having been designed for the summative assessment of generic skills achievement. There would appear to be a role for assessment methods that seek to enhance generic skills performance, and this is a focus of the research reported here.

Two studies are undertaken into the assessment of problem solving. In the first, the definition of problem solving, based upon notions of competence, is used to develop and validate a problem solving assessment instrument. The instrument is used as one element of a particular assessment process. In this process, students assess their own problem solving performance on routine assessment tasks that they undertake within their courses. They submit their work, including both the

substantive course-related tasks and their assessment of their problem solving performance on that task. Their self-assessment is validated by their lecturer and they receive feedback on that assessment. The results of the first study indicate that the problem-solving assessment instrument, based on a cognitive theory of problem solving, does provides a valid basis for the assessment and measurement of problem solving performance, although some improvements to the instrument are foreshadowed.

In the second study, a revised version of the problem solving assessment instrument developed in the first study is used. In this study, students use the problem solving assessment tool on a series of course-related assessment tasks over an academic year, receiving feedback on each assessment. The purpose of this study is to test the proposition that repeated assessment and feedback cycles might lead to improved problem solving performance. Evidence for such improvement is reported.

It is concluded that existing course-related activities can be used as vehicles for the development of students' problem solving skills. The development of generic skills (problem solving in this instance) would appear to depend upon two elements of an assessment regime. First, the assessment target needs to be defined in terms of an underlying construct that is operationalised through an assessment tool that focuses student attention on its key elements. Second, the development of problem solving proficiency is related to repeated assessment and feedback cycles, that is, to the implementation of a formative assessment approach.

Acknowledgments

I am indebted to many people who have supported and encouraged the investigation reported in this thesis.

Professor John Keeves, my principal supervisor, has been a mentor and an inspiration for more than a decade. He has given generously of his time and expertise and I join a large group of students from around the globe who are in John's debt.

Professor Mike Lawson, co-supervisor in this research, has influenced my learning and thinking for more than two decades. He alerted me to the importance of psychology as a foundation discipline in the study of numerous education issues and his influence has continued through this research.

I would also like to acknowledge the support of a generous scholarship, the Premier's Award for Post-graduate Research into Lifelong Learning. This scholarship was administered through the Centre for Lifelong Learning and Development under the leadership of Professor Denis Ralph. The encouragement provided by Denis, staff and other post-graduate candidates of the Centre has been especially valued.

I am pleased to acknowledge the support from teaching staff and students who participated in the studies. Mr Rob Denton of Torrens Valley Institute of Technical and Further Education was enthusiastic in facilitating the first of the two studies reported in this thesis. He and his colleagues and students were important contributors to this study. Ms Sharmil Randhawa of Flinders University and Ms Lyn Villis of the South East Institute of Technical and Further Education facilitated the second of the studies reported in this thesis. To those teaching staff and their students I am very grateful.

Contents

Tables

Figures

Acronyms

Names and titles are used in full at their first reference. Subsequently, the following acronyms are used.

ABS	The Australian Bureau of Statistics
ACCI	The Australian Chamber of Commerce and Industry
ACER	The Australian Council for Educational Research
AiG	The Australian Industry Group
ANTA	Australian National Training Authority
AVCC	Australian Vice-Chancellors Committee (now Universities Australia)
BCA	The Business Council of Australia
BHERT	Business Higher Education Round-Table
CBI	The Confederation of British Industry
CRESST	Center for Research in Evaluation, Standards, and Student Testing (USA)
CTT	Classical Test Theory
DEET	Department of Employment, Education, Training
DEETYA	Department of Employment, Education, Training and Youth Affairs
DEEWR	Department of Education, Employment and Workplace Relations
DeSeCo	Definition and Selection of Competencies (OECD project)
DEST	Department of Education, Science and Training
DETE	The (South Australian) Department of Education, Training and Employment
DETYA	(Australian) Department of Education, Training and Youth Affairs
GSA	Graduate Skills Assessment
IRT	Item Response Theory
MCEETYA	Ministerial Council on Education, Employment, Training and Youth Affairs
NCVER	The National Centre for Vocational Education Research
NQC	National Quality Council
NTB	National Training Board
NTQC	National Training Quality Council
OECD	Organisation for Economic Cooperation and Development
PISA	Programme for International Student Assessment (OECD program)
SACE	The South Australian Certificate of Education
SCANS	Secretary's Commission on Achieving Necessary Skills (USA)
TAFE	Technical and Further Education (often, College of...)
TST	True Score Theory
UNESCO	United Nations Educational, Scientific and Cultural Organisation
USA	United States of America

Chapter 1: Introduction

This thesis reports on an investigation into generic skills in the vocational education and training (VET) sector and the higher education sector in Australia. In common with many other countries, Australian governments and other organisations have generated sets of generic skills and have endeavoured to implement them within the major education systems – the school, VET and higher education sectors. The history of generic skills is reviewed, key constraints to their successful implementation are identified, and solutions are sought to overcome these difficulties.

A lifelong learning perspective

This research was conceived as a study of the generic skills that would be required to underpin a capacity for lifelong learning and before recent initiatives to reinvigorate activity on generic skills in Australia. Much of the recent impetus for the renewed activity on generic skills, in Australia and some other countries, has come from employer organisations and has been directed to the skills needed to gain and maintain employment (Allen Consulting Group, 1999; Australian Chamber of Commerce and Industry & Business Council of Australia, 2002; Australian Education Council. Mayer Committee, 1992).[1] Following these major initiatives, the aim of the study was to develop and evaluate a method for enhancing, assessing, measuring and reporting generic skills achievement, as those skills were defined within an employability framework. However, the lifelong learning view has been maintained and has provided an important perspective for articulating these constructs in ways that may have broader appeal than preparation for work alone.

Lifelong learning is itself a broad and evolving concept. Learning beyond formal schooling has a long tradition for the professions and many trades and crafts. Universities were identified as being among the very few social institutions that have persisted for a millennium despite major revolutions and global conflicts and have served the needs of the 'church' and the 'higher professions' (Coaldrake & Stedman, 1998). Their roles have included both professional vocational preparation and a liberal education designed to enhance the intellectual virtues of

[1] The report of the Australian Education Council. Mayer Committee (1992) is hereafter referred to as the Mayer Committee report.

disciplined effort of mind, respect for rational inquiry and intellectual honesty (Crittenden, 1997). Continuing learning for certain technical and trade vocations and was promoted through guilds and mechanics institutes, although few vestiges of these institutions have persisted (Ray, 2001).

The year 1972 is recognised as a turning point in lifelong learning and indirectly, generic skills, because it marked the publication of the Faure report (Faure et al., 1972). The Faure Commission was established to address serious social problems that had become apparent in Western Europe, especially in France, in the late 1960s. Changes in the organisation of work had occurred and had led to high levels of unemployment, in particular, among young people. The form and purposes of education that had been experienced by displaced workers was recognised as a contributor to the difficulties then being experienced in the labour market. Education that had been geared to the needs of declining industries was not helpful to those who were displaced, as it did not prepare them for alternative forms of employment, and it was not helpful to the emerging industries that were unable to recruit employees with the required skills profiles. The social disruption that followed required a political solution.

The Faure report urged the adoption of a set of three related educational strategies. One would lead to personal fulfilment for individuals apparent in the Commission's declaration that "…the physical, intellectual, emotional, and ethical integration of the individual into a complete person is a broad definition of the fundamental aim for education" (p. 156). A second goal related to the engagement of individuals with their communities was articulated in the assertion that democratic education "…must become a preparation for the real exercise of democracy" (p. 102). The report also canvassed a third range of strategies to enable adults to engage in 'recurrent education', to move between work and education, and for a greater involvement of companies in the education of employees. Thus, the Faure Commission recognised three clear purposes for education.

McKenzie (1983) identified recurrent education as a response to two perceived problems. The first was that increasing skill demands could result in young people being retained in a growing education system and that would incur substantial costs before they would be available to make a contribution to economic growth. The second concern was that older workers, having completed their pre-work education at a time when skill demands were lower, would be less able to compete with younger and more educated individuals in a changing work climate. Recurrent education would meet the needs of both groups. Its promise might encourage some younger people to enter the workforce sooner and to rely on later educational opportunities to top up their skills, while for older workers recurrent education would equip them with the skills in demand that had not been available during their initial formal education.

However, recurrent education, while focusing on the needs of individuals as employees and of industry, ignored the first two goals espoused by the Faure Commission, namely personal development and social engagement. Recurrent education was also concerned with individuals' interactions with the formal education system. Recent conceptions of lifelong learning have recognised roles for formal, informal and non-formal learning in education. Formal education was provided by institutions such as schools, colleges and universities and led to the award of qualifications. Non-formal education was organised and systematic, but carried out outside the formal institutions and did not lead to the award of qualifications. Informal education was lifelong experiential learning – the acquisition of values, knowledge, attitudes, and skills from experience and from resources available in the environment including workplaces, and was independent and self-directed (Tuijnman, 1996).

An alternative to recurrent education as a solution to the skills problems that had led to the establishment of the Faure Commission was suggested by Mertens (1974). He proposed that, in addition to vocationally-specific skills, a set of *Schlüsselqualificationen* (key skills) should be developed through initial formal education. These skills, which included basic vocational skills and the broadening skills of learning and information seeking, would enable individuals to adjust to the changing skill requirements of occupations. Mertens also recognised the importance of individual development, social engagement and workplace preparation in formal education. Mertens' contribution may be seen as the beginning of a focus on generic skills that has occurred in very many developed economies.

The three strategies proposed in the Faure report reflect three prominent themes in current literature on lifelong learning, namely, the development and fulfilment of the individual, the capacity for social engagement and democratic participation, and preparation for participation in work and economic security. These themes are common in the literature on lifelong learning, expressed for example in Chapman and Aspin (2001), who argued that the objectives of lifelong learning depend upon:

> The provision of educational opportunities throughout life that adhere to such principles and policy objectives as: economic efficiency and advance; social

> justice, social inclusion and democratic participation; and personal growth and fulfilment. (p. 1)

In summary, the literature on lifelong learning has emerged from and recognised change as a characteristic of the skills requirements of developed Western economies. How education and training systems accommodate the demands of changing skills requirements is a challenge. This accommodation may be accomplished by forms of continuing or recurrent education, or by changes to the skills that are taught in initial education to include generic skills, or a combination of both strategies. The literature on lifelong learning has also identified three broad goals for all education, namely the fulfilment of individuals, their engagement in social and democratic processes, and their participation in productive economic activity.

In this thesis, the focus is on generic skills as a class of constructs. Reference is made in reviewing generic skills schemes to their roles in personal development, social engagement and workforce preparation. However, neither lifelong learning nor the broad goals of generic skills are investigated in this thesis.

The focus of the thesis

The study was conceived in the late 1990s. The author had been involved in the implementation of generic skills (graduate qualities) in an Australian university. The key competencies initiative had emerged in the early 1990s, but by the late 1990s, its influence had waned. Two peak Australian business organisations, the Australian Chamber of Commerce and Industry (ACCI) and the Business Council of Australia (BCA) managed a program to re-invigorate interest in generic skills. One aspect of this work, a literature review and framework development, was undertaken by the Australian Council for Educational Research and the author was invited to contribute to this work (Curtis & McKenzie, 2002).

At the outset, it appeared that it would be necessary to develop assessment tools that could be used to discover the level of generic skills among students at key stages of their education – in schools, in vocational education and training settings, and in higher education. However, during the literature review for the ACCI and BCA project, a new possibility emerged. In the schemes that had been reviewed, and in which assessment had been considered, assessment was directed at discovering the stock of generic skills in the population under consideration, whether that was school-leavers or higher education graduates. However, the thinking of those involved in the Faure report and that of Mertens (see above) had shown that past approaches to education had not produced an adequate level of generic skills in a sufficient proportion of the population. This situation suggested that simply assessing the stock of generic skills would not lead directly to a solution to the problem a shortage of generic skills. A solution to the problem of a deficiency of these skills suggested the need to develop means for enhancing the stock of generic skills.

National and international assessment programs, for example the Graduate Skills Assessment (GSA, Australian Council for Educational Research, 2001b) and the Programme for International Student Assessment (PISA, OECD, 2003) directly assess the stock of skills. Indirectly, they contribute to the enhancement of knowledge and skill because differences in student performance between countries or jurisdictions or institutions, coupled with some knowledge of the characteristics of the education practices of those entities, enables policy makers and practitioners to adjust aspects of education provision in under-performing units. A direct mechanism for enhancing the stock of skills might complement national and international assessments.

Guiding propositions

In order to enhance the stock of generic skills, two sets of propositions are advanced. The first set contributes to an articulation of a clear conception of them. Three propositions about conceptions of generic skills are advanced. First, it is found useful to distinguish between relatively fixed abilities and competences and to regard generic skills as competences. Competences are based on cognitive processes that are deployed by individuals as they activate relevant knowledge and procedures or perceive the need to acquire new knowledge. An implication of regarding generic skills as competences is that they are amenable to development by drawing attention to their component cognitive processes. Second, competences are regarded as latent traits. They may not be observed directly but the extent of their presence may be inferred by an individual’s performance on set tasks. Third, effective performance, among novices, depends upon the deployment of the processes that underlie competence. A distinction is made between novices and experts because there is strong evidence that expert performance depends upon a highly developed knowledge base that is not available to novices (Chi, Glaser, & Rees, 1982). An implication that arises from these three propositions is that the tasks that are designed to reveal generic skills and presented to students must afford them the opportunity to develop and demonstrate the processes that constitute competence.

Articulating a clear conception of each of the generic skills is part of the solution to the problem of their development through education and training systems. It is argued in this thesis that particular methods of assessing generic skills achievement can lead to their enhanced development. Four propositions relating to the assessment of generic skills are developed. First, assessment can lead to enhanced learning. While assessment methods that are designed as a basis for reporting skills achievement are valued, for example in selection and certification contexts, other assessment methods may lead directly to the development of target constructs. Second, several approaches to assessment that involve the provision of informative feedback to learners on their performance in an iterative cycle of learning, assessment and feedback can enhance learning and, therefore, performance. In addition, engaging learners actively in the assessment process, through for example, self- and peer assessment, may contribute to gains in performance. Third, effective assessment requires performance standards to be defined on a theoretically sound basis and communicated to teachers and learners and to other interested parties, for example potential employers. Fourth, the development and assessment of generic skills can be conducted efficiently in the context of existing courses.

Together, the propositions about conceptions of generic skills and methods for their assessment lead to a set of four research questions.

Research questions

First, what are the dimensions of generic skills? This question is addressed through a review of generic skills schemes and their implementation. Early in the investigation of generic skills, it became apparent that, despite considerable resources being invested in generic skills schemes in Australia and elsewhere, the schemes had not enjoyed the degree of penetration into education and training systems that had been anticipated. Among the factors that contribute to the success, or lack of it, in implementing generic skills schemes two, namely their definition and their assessment, appear to be critical.

Second, can a foundation for generic skills be found in psychology? This question relates to the first set of definitional propositions outlined above. This question is framed as a basis for articulating a set of arguments about generic skills as a class of constructs and leads into an analysis of literature on human capabilities. It is through this analysis that defining these ‘skills’ as competences arises. In answering this research question, the first of the two critical success factors, their definition, is addressed. Many skills are recognised as being generic and it would not be feasible to investigate them all in depth. Therefore, in order to limit the scope of the research, problem solving is examined in detail.

Third, can valid and reliable indicators of performance in generic skills be identified? This question relates to the second set of assessment propositions outlined above. While this question reflects an early, and perhaps naive, intention of the investigation – to develop assessment instruments for a selection of generic skills, the realisation that enhancing these skills is as important as assessing them creates a new role for their assessment and underscores its importance. This question goes to the assessment of generic skills, but framing the question in terms of validity and reliability requires assessment to be grounded in sound conceptions of the target skills. That is, addressing the assessment issue depends upon resolving the definitional one. Further, this question requires an evaluation of assessment options and, if assessment is to lead to reliable reports of performance, the assessed performance scores must comply with the requirements of measurement. The response to this question is developed through the first of two studies presented in this thesis.

Four, can problem-solving performance be enhanced by repeated assessment and feedback cycles? The response to this question is an investigation into a practical implication that arises from research questions two and three (see above). In responding to the third research question and evaluating assessment options, it is hypothesised that learning, assessment and feedback might form a repeated sequence and that iterations through this sequence may lead to enhanced performance. The response to this question is developed through the second of two research studies reported in this thesis.

The nature of the thesis

The structure of the thesis and how the various propositions and research questions are addressed in it are outlined below. Before moving to that, a brief reflection on the nature of the thesis is offered. The research presented in the thesis is a combination of a qualitative investigation of generic skills and their implementation through various schemes, and two quantitative studies. The qualitative review of generic skills and their implementation represents the beginning of a narrative that seeks to connect the purposes for recognising generic skills, through aspects of their development historically, to a situation that requires considered action. The narrative leads to the first of the two quantitative studies. It is not hypothesis-driven; rather, it is a demonstration that, given a need to develop a method for assessing and reporting problem-solving performance, one particular approach is shown to work. This is almost certainly not the only approach that may work, and it may not be the ‘best’, by whatever criteria that judgment might be reached. Thus, the first quantitative study represents a continuation of the narrative. The second quantitative study is hypothesis-driven. It posits that, if the claims made for learning, assessment and feedback cycles work, improved performance should follow iterations through such cycles. The thesis may thus be regarded as narrative enquiry (Clandinin & Connelly, 2000) connecting a policy objective with a pragmatic outcome.

Structure of the thesis

The thesis is structured around an investigation of generic skills. Since Mertens' (1974) introduction of the term 'key skills' many similar proposals have emerged in many countries. In this thesis, the term 'generic skills' is used as a label for those skills that have been proposed to meet the continuously changing demands of new forms of work in the labour markets of advanced economies. However, terminology is recognised as a problem in this field and it is discussed specifically in Chapter 2. Generic skills have been proposed by working groups and committees appointed by governments, inter-governmental agencies and industry bodies. Each proposal has resulted in sets of skills and methods for their implementation being identified. These proposals for generic skills are referred to as generic skills schemes.

The structure of the thesis is shown in Figure 1 and the content and purpose of each chapter is described below.

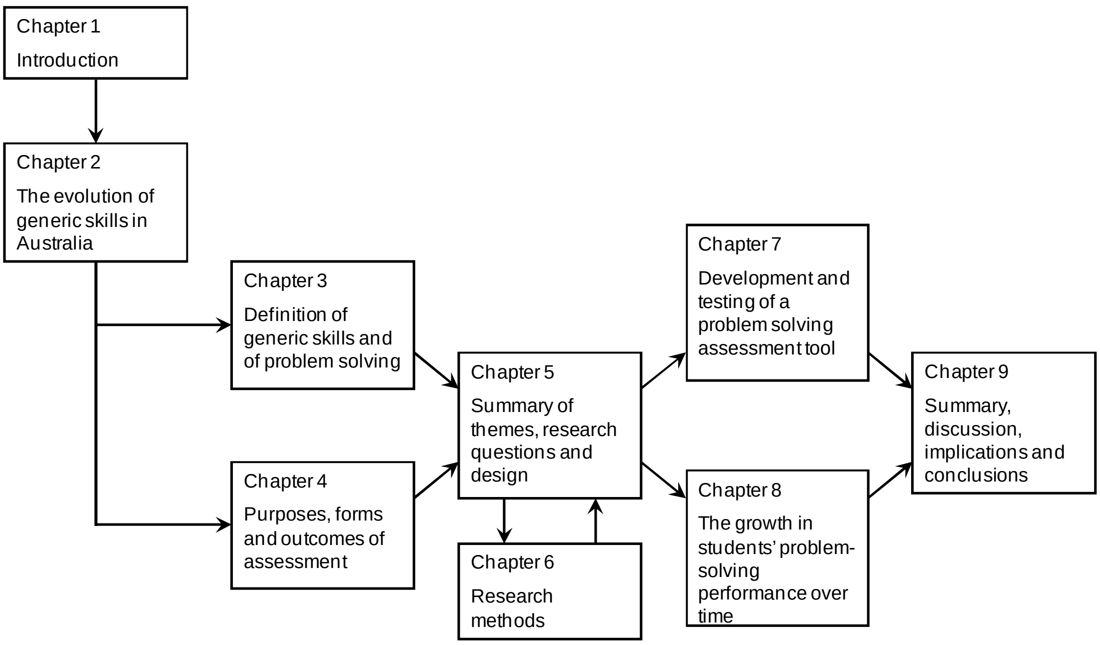


Figure 1: Structure of the thesis

Chapter 2: Generic skills schemes

The emergence of generic skills schemes in Australia is discussed on Chapter 2 and several key issues are identified as requiring resolution. The emphasis in this chapter is on developments in Australia, but those developments have been informed by proposals generated in other countries and reference is made to overseas developments where they appear to have influenced thinking in Australia and where they can be used to reflect on generic skills in Australia.

In Chapter 2, attention is paid to the recognition of generic skills in Australia through numerous reports beginning with the report of the Quality of Education Review Committee (1985) and ending with the report by the Australian Chamber of Commerce and Industry and the Business Council of Australia (ACCI & BCA, 2002). (Further developments have occurred since that time and they are acknowledged in the final chapter of the thesis).

Among the issues identified, two emerge as central to the successful implementation of generic skills. They are (a) defining and (b) assessing those skills that are required to enable individuals to contribute to and benefit from the world of work. Although the emphasis in recent Australian generic skills proposals, such as those developed by the Australian Industry Group (Allen Consulting Group, 1999) and by the Australian Chamber of Commerce and Industry and the Business Council of Australia (ACCI & BCA, 2002), is on employability, it is suggested that the three spheres of individuals' lives (personal development, social engagement and economic participation) are deeply interconnected and that the skill sets required for any one of these domains is also implicated in the others. Where differences in desired skill sets occur between these domains, it is likely to be a matter of emphasis rather than a substantially different set of skills being required.

Chapter 3: Finding a coherent basis for defining generic skills

In order to address the first of the two major issues identified in Chapter 2, namely the definition of generic skills, Chapter 3 presents a search for such a definition. The search occurs in two phases. First, it is argued that generic skills as a class of constructs require definition and then that each generic skill needs to be defined. In the literature of cognitive psychology, two constructs are identified as candidates for understanding generic skills as a class of related skills, namely intelligence and competence.

Intelligence has been the subject of theory and investigation in psychology for more than a century. A review of previous theoretical positions and the re-analysis of many of the original data sets was presented by Carroll (1993). His synthesis and its subsequent modification by Gustaffson (1997) proved to be a high point in the investigation of intelligence. Other conceptions of intelligence are also investigated, included a theory of successful intelligence proposed by Sternberg and his colleagues (2000).

As an alternative to intelligence, the concept of competence is also explored. Many theorists have contributed to that concept and the synthesis provided by Weinert (1999) as a background document to the OECD-sponsored DeSeCo (Definition and Selection of Competencies) project (Rychen & Salganik, 2000) proved to be a valuable summary. In Chapter 3 it is argued that competence is a more useful way of conceptualising generic skills for the purpose of defining them as a class of constructs and as a basis for describing each of them as assessable variables on which achievement can be measured and reported.

The ACCI and BCA (2002) generic skills scheme, which has been endorsed for use in the school and vocational education and training sectors in Australia, included eight ‘key skills’ and 13 personal attributes. Attempting to investigate each of the eight key skills is beyond the scope of this thesis, and attention is restricted to problem solving. Along with communication and teamwork, problem solving was one of the skills that were recognised in almost all generic skills schemes, and it is selected as the focus for subsequent sections of the thesis.

Two dominant theoretical approaches to problem solving, the so-called situative position and an information processing theory, are reviewed. Within the information processing position, problem solving can be understood as a set of processes that may be activated as problems are encountered and that they provide a framework that can be used by non-expert individuals to scaffold their problem solving attempts. Non-experts are the focus because there is evidence that expert problem solvers proceed to solutions using their highly conditioned knowledge, knowledge that non-experts do not have.

The Chapter concludes with a set of propositions about generic skills and problem solving that guide subsequent phases of the investigation.

Chapter 4: Addressing the challenge of assessment

Assessment is the second of the two major issues identified in Chapter 2 as critical for the successful implementation of generic skills. Assessment is discussed in Chapter 4.

Assessment is found to have several purposes and a distinction is made commonly between formative and summative assessment. Summative assessment is found to be concerned with reporting student achievement while formative assessment is concerned with diagnosing short-comings in students' knowledge states and with directing instruction aimed at improving students' achievement. This distinction proves to be important in this thesis as a focus on formative assessment emerges. Two issues related to formative assessment, namely feedback and self-assessment, are canvassed.

In any assessment, it is necessary to establish a basis for judging the quality of students' performance. Several approaches to the judgment of performance, including norm-referenced, criterion-referenced, standards-referenced and construct-referenced assessment, are reviewed. For other than norm-referenced assessment, a principled basis for establishing performance standards is desired and Bloom's Taxonomy and the SOLO Taxonomy are investigated.

Criteria by which assessment methods can be evaluated are reviewed. These criteria include validity, reliability, objectivity and feasibility and they are discussed in Chapter 4.

Methods by which generic skills may be assessed are reviewed. They are classified into five broad groups, namely standardised assessment, common assessment tasks, performance assessment, teacher-group judgment, and portfolio construction. Each of these methods has been used to assess generic skills achievement, and examples of the application of these approaches are presented. The evaluation of these methods against criteria for assessment is summarised in a table.

The chapter ends with the statement of seven propositions about assessment that inform the subsequent conduct of the investigation into generic skills.

Chapter 5: A summary of the problems surrounding generic skills and methods for addressing them

Chapter 5 is an interlude in the reporting of this investigation into generic skills. It serves to summarise the findings from Chapters 2, 3 and 4 on the implementation of generic skills in Australia, approaches to defining generic skills and possible methods for assessing them. This chapter also demonstrates the interdependence of the two critical issues of definition and assessment.

Arising from this summary, four questions that guide the research are posed:

- What are the characteristics of generic skills?
- Can a foundation for generic skills be found in cognitive psychology?
- Can valid and reliable measures of performance in generic competences be developed?

- Can problem solving performance be enhanced by repeated assessment and feedback cycles?

The first two of these questions address the issue of the definition of generic skills that was identified as one of the requirements for successful implementation of generic skills schemes. They are answered through analyses of the literature on generic skills, assessment and cognitive psychology.

Questions 3 and 4 address the second major requirement for the successful implementation of generic skills schemes, the issue of assessment. These questions are answered through two empirical investigations that are reported in Chapters 7 and 8.

Chapter 6: Research methods

In order to investigate the third and fourth research questions, two empirical (but not experimental) studies are designed. They depend upon the application of a range of statistical methods. Two methods are described in detail.

The term ‘measurement’ is used somewhat loosely in much social science research to refer to the collection of numerical data (Embretson & Hershberger, 1999; Michell, 1997). In this thesis, a more restricted use of the term is preferred. In order to achieve scientific measurement, the Rasch method is used (Rasch, 1960, 1980). The first part of Chapter 6 is devoted to a discussion of measurement and this is followed by an examination of the Rasch measurement model and several of its derivatives, in particular the partial credit model (Masters, 1982). This method is central to the construction of a scale on which to measure problem solving performance – the subject of Chapter 7.

The fourth research question concerns possible gains in student achievement that may follow iterative instruction, assessment and feedback cycles. The measurement of change within individuals over time involves some statistical challenges, especially the problem of auto-correlation between repeated observations made on the same individuals. This difficulty is addressed through the application of multilevel modelling, and the account provided in this chapter is based on the work of Bryk and Raudenbush (1992) and Singer and Willett (2003).

Chapter 7: Developing and testing a problem solving assessment instrument

The proposition is developed in Chapters 3 and 4 that the measurement of problem solving performance depends upon having a clearly defined construct. Further, that construct must be capable of being operationalised to yield indicators of that performance and that performance standards can be developed and applied to the indicators.

Chapter 7 reports on a study in which a problem solving assessment tool is developed and refined and is used in two courses in the vocational education and training sector. The problem solving assessment instrument is based on the set of five problem solving processes described in Chapter 3. The Rasch measurement model is used in the analysis of data collected through the administration of the problem-solving assessment instrument. The analyses suggest that the instrument is useful in assessing problem solving performance and in generating measures of that performance.

Chapter 8: Investigating the influence of assessment on learning

In the review of assessment methods (Chapter 4) several propositions arose about the potential of formative methods of assessment to lead to improved performance by individuals. In particular, the provision of feedback to learners following their performance is thought to contribute to learning. A second proposition, that engaging students meaningfully in the assessment process through self-assessment might lead to enhanced performance is identified.

Chapter 8 reports on a study designed to investigate the change in students' problem solving performance over time. In order to establish an independent measure of student performance, the Tertiary Skills Assessment, an instrument developed by the Australian Council for Educational Research, was used (ACER, 2000). This multiple-choice assessment instrument measured interpersonal understanding, critical thinking and problem solving.

The iterative cycle of problem solving assessment is undertaken using a modified version of the assessment tool whose development is described in Chapter 7. This tool is used in a process in which students undertook routine course assessment tasks, assessed their own problem solving performance using the assessment tool, submitted their work for assessment, and received feedback on both the substantive content of the assigned task and on their problem solving performance. This process was repeated on three occasions during the academic year. Data gathered through this process were analysed using the Rasch measurement model to generate measures of problem solving performance on each assessment occasion. The repeated measures of problem solving performance were analysed using multilevel methods (Singer & Willett, 2003) to investigate the proposition that student performance can be enhanced through the iterative assessment and feedback cycle.

Chapter 9: Summarising findings and exploring implications

In Chapter 9 a summary of the investigation is presented. This summary is organised around the four questions that framed the research. The summary includes the key findings from the review of generic skills schemes and their implementation in Australia. It also includes the interim conclusions reached about how generic skills and, in particular, problem solving might be represented. Key findings from the review of assessment are also presented.

The results of the two studies, the development of the problem solving assessment instrument (Chapter 7) and its use in investigating the influence of iterative cycles of assessment and feedback on student problem solving performance, are summarised.

Limitations to the design of the two studies are discussed and suggestions are advanced for further investigations that may improve the generalisability of the findings.

The approach to the assessment of problem solving developed in this investigation is contrasted with approaches taken in two international projects (the Adult Literacy and Life Skills survey and the Programme for International Student Assessment). Reports of these investigations emerged after the research for this investigation had been conducted. It is suggested that the approaches are complementary as the purposes of the international studies are summative, and support the evaluation of educational programs, while the second study reported in this thesis serves a formative purpose.

A final note is offered, reflecting on the continuing search within Australia for solutions to the policy difficulties encountered in assessing and reporting on the achievement of generic skills among young Australians.

Chapter 2: The Evolution of Generic Skills in Australia

In this chapter, the pressures to develop generic skills schemes in Australia are examined. Conceptions of generic skills in Australia have been influenced by similar schemes in other countries, especially the United Kingdom and the United States and some attention is given to schemes from those countries. A review of those schemes is presented in Appendix 2. Further, certain characteristics of generic skills, for example their presumed transferability, appear to have been influenced by developments in Europe, particularly in Germany and France. Consequently, aspects of generic skills schemes in those countries are also discussed briefly.

Understanding the factors that have led to the establishment of the schemes may help to elucidate the conceptual bases for them that have led to the selected sets of skills, their implementation methods and attendant limitations. It is argued that the intended concepts captured in generic skills schemes are fundamental to any attempt to develop, assess and report on the achievement of these skills.

Many terms are used to label generic skills and some confusion in the application of these terms is apparent in the literature. An attempt is made to clarify the terms used and the concepts that are conveyed by them.

The main section of this chapter reviews the history of generic skills schemes development and implementation in Australia to achieve three objectives:

- to identify the processes that were used to select the particular skills or attributes that were included in the scheme;
- to examine descriptions of the skills and attributes that were included; and
- to investigate ways in which the schemes were implemented in education and training programs, including reasons for failures in their implementation.

These analyses lead to the identification of four pairs of issues that, it is argued, need to be addressed if the relatively abstract constructs that constitute generic skills schemes are to be embedded in the routine practices of education and training programs. While the emerging Australian context is central to this discussion, it is apparent that this context is influenced by progress in generic skills development elsewhere and subsequent discussion draws upon Australian and overseas developments.

The impetus for generic skills

In Chapter 1, a lifelong learning perspective on the impetus for generic skills is outlined. Lifelong learning has three broad goals, namely economic efficiency; social and democratic participation; and personal growth and fulfilment (Chapman & Aspin, 2001, p. 1). Taking a hierarchy of needs perspective, it appears that individuals are more likely to pursue personal growth and to engage with their communities if they benefit from economic prosperity through employment. While these three broad goals are deeply inter-related, the greatest influence on the development of generic skills has been the economic imperative. The demands of work, however, are changing.

The changing requirements of work

Generic skills schemes have been promoted for many years. Streumer and Bjorkquist (1998, see esp. pp. 250-251) traced the emergence of generic skills to the use of the term *schlüsselqualifikationen* (key skills) to Mertens (1974).[2] Several factors, including changes in labour market requirements and a concern about youth unemployment, are implicated in the growing interest in generic skills in Europe.

The report *Learning to be* (Faure et al., 1972) was a response to social and labour market dislocation in Europe, especially France, but the report identified pressures for educational change in developing and developed countries and recognised that a once-only experience of education could not prepare citizens with the knowledge and skills they would need throughout their lives. The changing context included globalisation and technological innovation, and one of the Faure Report's prescriptions was for learning throughout life:

> …the aim of education in relation to employment and economic progress should be not so much to prepare young people and adults for a specific, lifetime vocation, as to 'optimise' mobility among the professions and afford a permanent stimulus to the desire to learn and to train oneself. (Faure et al., 1972, pp. xxxi-xxxii)

The Faure report also identified two features of education systems that were without precedent. First, that education "…is now engaged in preparing men [*sic*] for a type of society which does not yet exist" (p. 13) and second, that "…some societies are beginning to reject many of the products of institutionalized

---

[2] But see 'Definition of terms used to describe generic skills' below for a more detailed discussion of terminology.

education" (p. 14). These observations appear to have emerged from analyses of the causes of the unrest that occurred in France during 1968 when students, upon graduating from schools and other institutions, were experiencing great difficulty in finding work.

Mertens (1974) argued that the traditional separation of vocational education (in *berufschulen*) from academic secondary education in *gymnasien*, had led to a skills divide. The *gymnasien* had provided a broad general education but vocational education had become too narrowly focused and was not providing a broad preparation for the changing context of work. His solution was to provide four categories of skill under his umbrella term S*chlüsselqualifikationen*. These included basic skills, 'horizontal' skills (information seeking and learning), broadening elements (that were applicable across a range of vocations) and knowledge of general subject matter and conceptual systems. Such concerns were not limited to Europe: proposals to develop generic skills schemes emerged in many countries, including Australia.

In Australia, The Quality of Education Review Committee (1985) was asked to examine:

> …the attainment of a satisfactory standard by the great majority of students at successive stages of the general curriculum, with particular reference to communications, literacy and numeracy [and] an improved relationship between secondary education and employment and tertiary education opportunities and requirements. (p. 204)

This remit was further elaborated, requiring the Committee to provide advice on means for "attaining higher basic skill standards" for primary school students, and for "the attainment of appropriate standards relevant to subsequent employment opportunities and improved preparation for tertiary education" (p. 204) for secondary students. The terms of reference also referred to the "increasingly competitive, including internationally competitive environment" (p. 205) of Australian industries into which school leavers would move.

In the United States, Carnevale (1991) argued that the basis of economic competitiveness had expanded from a reliance on productivity and price in the old industrial economy to include quality, variety, customisation, convenience and timeliness. He argued that to meet these emerging standards of competition, revised forms of work organisation would be required and that enhanced skill sets would be demanded of 'new economy' workers. In addition to basic skills, the new skill sets would include learning to learn, communication, teamwork and problem solving skills. At the same time, Reich (1991) observed the same set of forces reshaping Western economies. He used a very broad classification of workers based on their skills, namely routine production, in-person service and symbolic-analytic workers. He warned that employment for routine production workers was at risk because corporations could choose to relocate the means of production to low wage countries. Symbolic analysts, by contrast, are mobile as their skills are carried with them and can be deployed wherever there is a demand for them. Several generic skills schemes, including workplace know-how (The Secretary's Commission on Achieving Necessary Skills, 1991) and the Work Key system (McLarty & Vansickle, 1997), were proposed in the United States following the publication of these two influential books. Similar schemes, core skills and key skills, had been developed in the United Kingdom (Dearing, 1996; National Curriculum Council, 1990).

The Mayer Committee (1992) noted the growing pressures on Australian industry to become more competitive in a changing global economy.

> Australia's economic success and hence our standard of living depends on a workforce and a work environment that is capable of matching, or improving on world best practice. Workplaces must be more competitive. They must be committed to continuity of service and quality of outcome, setting and meeting deadlines, and responding to the needs and wishes of clients, individually and collectively. To meet these commitments the focus of work and how work is organised will change. (p. 3)

The Committee went on to outline the need for a "multi-skilled, flexible and adaptable workforce" and in order to achieve this, individuals would need "a strong foundation of knowledge, skills and understanding" (p. 3). In addition to a general education and specific vocational skills, employees would require a set of broad skills that would enable them to apply their general education to the world of work. The Committee articulated a set of seven key competencies intended to achieve this objective.

The Australian Industry Group (Allen Consulting Group, 1999, see esp. pp. ii-iv) recognised various pressures on the competitiveness of Australian industry. These included increasing global competition and the consequent need to be more innovative, greater use of technology, and a move towards service and an integration of service and manufacturing functions. They noted the influences of global competition on Australian industry.

> The global knowledge–based economy is with us. Important policy and institutional changes, notably declining barriers to trade, the creation of a multilateral trade regime and the dismantling of capital controls, have fostered the rise of globalisation. Advances in technology have sharply reduced the costs associated with transport and (particularly) communications, further promoting the growth in trade, financial integration and transfer of technology. (p. 7)

They summarised the key response as one requiring the development of human capital resources, which included:

> a shift in demand to higher skills at all occupational levels;
>
> an increasing premium is being placed on generic skills, both 'hard' (notably IT skills) and 'soft' (e.g. problem–solving, team skills, willingness and ability to adapt) to be developed prior to recruitment;
>
> industry is also focusing on the key skills required for the productivity of the enterprise and is restructuring its workforce in order to maximise use of skills, for example through the implementation of self-managed work teams and multiskilling; (Allen Consulting Group, 1999, p. v)

A similar view was expressed in 'Employability Skills for the Future' (ACCI & BCA, 2002). Having reviewed several reports to government, they concluded that Australia's capacity to participate in the "global knowledge economy" (p. 7) had to be built, and that education and training had to play a key role in that capacity building by equipping the future workforce with high level skills.

The case for changes in the skills requirements of work and changing work organisation is made in a comprehensive analysis by Friedman (2005). He identified 10 major influences on changing patterns of work. Many of these changes have been enabled by the integration of information and communication technologies. He identified outsourcing, supply-chaining and off-shoring as being among these influences and he illustrated them with examples. In various ways, work is subdivided into components that require different skill levels. An enterprise can then allocate these work components in ways that ensure the work is completed at the lowest net cost, some of it in-house and some of it by suppliers who may be located in another country. The challenge to governments in more advanced economies is to ensure that the workforce (or at least a substantial proportion of it) is equipped with the high-level skills that will enable them to undertake the highest value-added components of the work. These mechanisms of change only work because enough governments have permitted the relatively free flow of goods and services, and this has occurred through international agreements on trade, for example, through the many rounds of the General Agreement on Tariffs and Trade (GATT) and, since 1995, the World Trade Organisation (WTO).

Impacts of the changes described by Friedman have been identified in Australia. Houghton and Sheehan (2000, see p. 4) showed that the knowledge intensity of traded goods had been growing steadily since the mid-1970s to the mid-1990s. Growth in employment and wages in knowledge intense jobs had also occurred over a similar period (p. 5). The demand for individuals with knowledge and skills required a greater focus on generic skills in education. Houghton and Sheehan (2000) suggested:

> What flexible organisations need most from education systems is not so much investment in the production of skilled but narrowly defined specialists, or a lot of investment in vocational training; but much more investment in the production of people with broad-based problem solving skills and with the social and inter-personal communication skills required for teamwork, along with the skills and attitudes required for flexibility. (p. 21)

Generic skills schemes have also been developed to contribute to the measurement of higher order skills as outcomes of education systems. The DeSeCo project (see below) was conceived in order to develop measures of educational outcomes other than those captured in discipline specific achievements, for example in the areas of literacy and mathematics. In addition to recognising the need for skills to enable global competition and to use new technologies effectively, Sweden also adopted an equity focus in promoting generic skills. There, intergenerational inequality in access to education had been a concern, as older citizens had experienced low attainment compared with the now high levels of school and post-school educational participation. Generic skills were developed through adult education programs as a way of ensuring greater age-related equity (Abrahamsson, 1999, see pp. 118-121).

It is too simple to suggest that the same set of forces has driven the emergence of generic skills in all countries. In addition to competition and technology, which appear to influence all developed economies, local factors, such as a focus on equity, also apply. In Australia, the impetus for generic skills is seen by some as an outgrowth of a move to overhaul an overly prescriptive work-role classification system through award restructuring (Roe, 2003).

The analyses cited above all reflect a relatively common theme – that economic competition is growing and becoming global – with some variations. This movement has been developing for some time and is driven by economic policies that include reduced trade and capital barriers embraced by many developed countries; increased ease of transportation of people, goods and services; and the deployment of new technologies. Increased global competition through reduced trade barriers has led to low-skill work being relocated to countries with low wage costs. This has led to reduced demand for low-skill workers in advanced economies and pressure to develop high-skill employment opportunities. This trend was noted by the Australian Bureau of Statistics (2006a, see pp. 5-6), finding that, because of changes to industry requirements and the application of technologies in the workplace, industry was demanding higher levels of post-school qualifications and was paying high wages premiums to well-qualified employees. Workforces must respond to this reality by developing higher skill levels and by becoming more adaptable. A key solution espoused in many countries is the development of broadly applicable skills. The nature of some of the proposed skills sets is discussed below. However, before discussing the particular sets of skills that have been proposed, the terms used to describe them warrant consideration.

Definition of terms used to describe generic skills

The definitions of skills that are regarded as generally applicable to many occupational and vocational roles and to other dimensions of people's lives need to be understood at two levels. At a macro level, the schemes that include these broadly applicable skills provide one view of the constructs that are being conveyed when terms like generic or key skills are being used. At a second level, the particular skills and attributes that are included within generic skills schemes need to be defined and elaborated, and provide an insight into the intentions of those who developed the various schemes.

In the following discussion, the term 'generic skills' is used to describe both the range of skills and the schemes in which they are included although, as it may become apparent, it is not certain that all labels were intended to describe the same sets of constructs.

Labels for generic skills

Many terms are used to describe the sets of characteristics and abilities that people need to develop and demonstrate through education and training and other experiences, but that transcend the particular discipline areas in which they are developed and that are applicable to a wide range of contexts, including work.

These characteristics and abilities are variously referred to as skills, competencies, qualities, or attributes. These descriptors are modified by a range of qualifiers to indicate the breadth or purpose of their application. It is not always clear whether these different terms reflect slight variants of the same basic concept or whether they are intended to signify alternative conceptions. For example, the word 'skill' may be used to convey a spectrum of meanings from a narrow ability to perform a very restricted task to a description of high level accomplishment. Further, there are common uses of terms as well as discipline-specific meanings, and it is occasionally difficult to know whether a common or a technical meaning is intended.

These difficulties have been compounded by various translations of terms from one language (and tradition) to others. The lack of a shared understanding about the meanings of terms used has added to the difficulties experienced in attempting to develop, assess and report on the achievement of generic skills that are being ascribed increasing importance in the changing context of working and living in advanced Western societies. A selection of some of the terms in use is presented in Table 1.

Table 1: Terms commonly used to describe the generic skills learners are expected to acquire

Qualifier	**Descriptor**
Core	Skills
Key	Competencies
Necessary	Competences
Essential	Attributes
Generic	Characteristics
Transversal	Qualities
Transferable	Outcomes
Graduate	Capabilities
Employment related	Abilities
Employability	
Lifelong learning	
Critical cross-field	

Some authors have used terms rather loosely so that ‘generic skills’ is taken to encompass the full gamut of these characteristics while others are rather more careful about the terms. For example, in discussing the achievement of generic skills as a quality indicator in higher education, Clanchy and Ballard (1995) were critical of a failure to discriminate between ‘competences’ and ‘competencies’ and Cummings, Ho and Bunic (1997) preferred the term ‘qualities’ over ‘competencies’ claiming that it includes knowledge and attitudes rather than just skills. Because of the potential for confusion, some discussion and clarification of the terminology is warranted.

The qualifiers

The qualifiers ‘core’, ‘key’, ‘necessary’ and ‘essential’ all seem to convey the sense that the entities being discussed are requirements for all people, irrespective of the level and nature of the work or other activities that they might undertake, and that there are minimal standards that all must achieve. The terms ‘generic’, ‘cross-field’ and ‘transversal’ carry the implication that the entities under discussion are applicable across all or many occupational types and levels. These skills or characteristics can be observed commonly in successful performances in a variety of fields. A frequent assertion is that such skills are also applicable to “life in general” beyond work contexts.[3]

[3] The phrase ‘and life in general’ is appended to references to the descriptions of generic skills in the workplace in many places in the Mayer Committee report (1992).

Occasionally, the assertion that such entities are common in a variety of contexts is also taken to indicate that they are inherently transferable. This assumption was made explicit in the definition of generic skills, namely "the transferable skills which can be used across occupational groups" (National Skills Task Force, 1998, p. 15) and in the Mayer Committee report (Mayer Committee, 1992, p. 7). The term 'transferable' is contentious. Transfer may be positive or negative, but it is assumed usually to be positive and this term is used to describe situations in which a skill acquired in one context leads to enhanced performance in another. The ascription of the term 'transferable' to generic skills represents an implicit claim, or perhaps an assumption, that these skills, having been developed in one context, will be deployed in others. The assertion of transferability requires substantiation before it is applied routinely in describing generic skills. The issue of transfer is discussed more fully in Chapter 3.

The term 'graduate' as a qualifier is used by universities to draw attention to the attributes that their students are supposed to achieve. There is considerable common ground in the attributes claimed for university graduates and those espoused for school leavers and vocational education and training (VET) graduates. For example, under the Australian Technology Network (ATN) universities project (Bowden, Hart, King, Trigwell, & Watts, 2000), attributes include communication skills, critical and creative thinking, problem-solving, and teamwork skills. Such skills are also found in the Key Competencies described by the Mayer Committee (Mayer Committee, 1992) and are skills expected of school leavers and VET sector graduates.

Of course, it is appropriate for universities to seek to distinguish their graduates from school leavers or VET graduates, since universities are quite selective in their entry requirements and the cost of university education to both individuals and to the community is relatively high. Thus, universities may choose to use graduate attributes as an indicator of the value, additional to discipline specific knowledge and skill, created through their education programs. However, the distinction between generic skills claimed by schools and vocational training on the one hand and by universities on the other may be based more appropriately on levels of achievement than on qualitative differences in the skills of their graduates. It is possible that university graduates may have acquired additional generic skills, but it seems likely that graduates may have developed to a higher level a set of core competences that are common to school leavers and VET sector graduates. This suggests that, while universities may wish to use the term 'graduate' as a qualifier, a more inclusive term is required for general use.

The adjectives ‘employment-related’ and ‘employability’ suggest that the entities being discussed are of particular interest to individuals in relation to their work and to employers. However, one of the issues raised in the context of Australia’s emerging knowledge economy is that such skills are important to people in several dimensions of their lives and to suggest that these skills are important only in their work is to understate their scope and significance.

The introduction of the term ‘lifelong learning’ has been recent in the discussion of generic skills, although the lifelong learning movement has a substantial history (Delors, 1996; Faure et al., 1972; McKenzie, 1983; OECD, 1996; Tuijnman, 1986; UNESCO Institute for Education, 1997). It does reflect a new emphasis on the need for people to be adaptable and flexible and to be able to learn new skills throughout their lives (OECD, 1996). The significance of this term is that, upon leaving formal education, whether at the end of compulsory schooling or after completing an advanced qualification, all people are expected to have achieved certain common skills, they have the capacity to continue to enhance these skills, and to acquire new ones. This suggests that there may be a need to extend the list of skills to include a capacity and willingness to continue to learn. An emphasis on lifelong learning is evident in the ‘Adelaide Declaration on the National Goals for Schooling in the Twenty-First Century’ (Ministerial Council on Education Employment Training and Youth Affairs, 1999) and learning is identified as one of the eight key skills of the Employability Skills Framework (ACCI & BCA, 2002).

The descriptors

‘Attributes’, ‘qualities’, and ‘characteristics’ refer to capabilities of individuals, although the term ‘characteristics’ is also used to describe the requirements of particular jobs. These descriptors are broader than others listed in Table 1. For this reason, they may be less attractive than say ‘competences’, but they have the advantage that they are used to include attributes from the affective as well as the cognitive domain.

The related terms ‘skills’ and ‘competencies’ have been used extensively in discussions that have followed the release of the Mayer Report. A brief aside on the context of the Mayer Committee’s deliberations is required. The composition of the Mayer Committee suggests that its focus was mainly on schooling and vocational education and on subsequent vocational rather than professional occupations: the higher education sector was not directly represented on the Committee, although it was represented on a sub-committee. The interests and methods of the vocational education sector are rather different from the higher education system. The Australian vocational education system is ‘industry driven’ and ‘competency based’. These terms are used to describe the ways in which vocational training is structured. Training packages are documents that list the skills (competencies) that are required for each qualification. Each training package lists several qualifications, and for each qualification, it lists the units of competency that an individual must complete in order to be awarded the relevant qualification. Each unit of competency describes what an individual can do, and for each unit there is a range statement, describing the contexts in which the competency should be observed, and an evidence guide that indicates the performances that should be observed in order to assert that the competency has been achieved. Assessment is dichotomous: the competency is either ‘achieved’ or ‘not yet achieved’. The vocational education system is industry led in that, for each industry grouping (of which there are ten), there is an Industry Skills Council and it is responsible for developing training packages for its industries. It determines what units of competency are required for the occupations in its industry. The use of competency based assessment is of particular concern in implementing generic skills schemes in the VET sector.

In Australian literature, a particular meaning for the term ‘competency,’ provided by the National Training Board (1991) has been accepted widely. In fact, the Board provided two rather different definitions:

> A competency comprises the specification of the knowledge and skill and the application of that knowledge and skill, within an occupation or industry level to the standard of performance required in employment. (p. 7)

and

> The concept of competency focuses on what is expected of an employee in the workplace rather than on the learning process; it embodies the ability to transfer and apply skills and knowledge to new situations and environments. This is a broad concept of competency in that all aspects of work performance, and not only narrow task skills, are included. (p. 18)

The second definition is much closer to that of competence (see Chapter 3) as it envisaged a capacity to perform and manage tasks, to adapt to contingencies and to transfer knowledge and skills. However, as Griffin and Gillis (1999, p. 1) noted, competency based assessment in the VET sector has focused on the first of the National Training Board’s definitions and has “ignored the remaining components in both the practice and in training of assessors.”

The narrow definition would appear to have pervaded the VET system. Recently, competency has been defined as:

> The broad concept of industry competency concerns the ability to perform particular tasks and duties to the standard of performance expected in the workplace. Competency requires the application of specified skills and

> knowledge relevant to effective participation in an industry sector or enterprise. (National Quality Council, 2007, p. 30)

In practice, occupational competencies are defined in training packages through elements of competency and range statements that delimit the contexts in which the competencies are observed. Thus, the concept of competency has taken on a narrow meaning consistent with the first definition proposed by the National Training Council.

When the Mayer Committee set out to define ‘competence’ (pp. 6-7), it endorsed the National Training Board’s second statement (1991, p. 18 quoted above) on ‘competency’. That is, it began to conflate the two terms competence and competency (for which the plural forms are respectively competences and competencies). In subsequent discussion, the Mayer Report used the plural ‘competencies’, and therefore contributed to the conceptual confusion that surrounds these two important terms. Thus, while preferring the broad concept of ‘competence’, the Mayer Committee embraced a narrow definition of competency – which subsequently is found to be at odds with usage of the term in vocational training – and used the term ‘key competencies’ as the label for the generic skills they sought to promote.

A particular problem in the VET sector, which is committed to competency-based assessment, is that generic skills are complex constructs. It is difficult to imagine that some level of ‘communicating ideas and information’ could be established as discriminating ‘competent’ from ‘not yet competent’ performance. But much effort has been invested in competency based assessment in the sector and there has been resistance to allowing some skills to be graded while requiring all vocational skills to be assessed using a competency based model.

Opposition has been expressed to the use of the concept of competency as a basis for describing outcomes of general education, especially in higher education (Bowden & Masters, 1993), but also in secondary and vocational education (Stanley, 1993). In their critique of attempts to introduce generic skills into the higher education context, Clanchy and Ballard (1995) commented that:

> It is hard to resist the impression that a lot of these terms are being used interchangeably as though their (in fact quite real) differences were of no consequence. (p. 157)

There is a view that advanced levels of performance on cognitively complex tasks are not amenable to disaggregation into discrete competencies (Hager, Holland, & Beckett, 2002). Complex performance is the result of having a body of knowledge, being able to recognise when it is appropriate to enact that knowledge, being able to activate that knowledge, to use it to guide actions, and to monitor the results of those actions. In addition, affective and conative factors shape performance and therefore there is a case for using terms that do not exclude such influences. However, in their discussion, the Mayer Committee acknowledged that their criteria for key competencies “preclude the inclusion of values and attitudes” (Mayer Committee, 1992, p. 13) and so, despite an expressed desire from industry and other groups for their incorporation and the Committee’s own acknowledgment of their importance, these affective dimensions of performance were excluded.

The term ‘competence’ is distinguished from ‘competency’. In Australia and the United Kingdom, competency has become a label for a specific and observable behaviour that demonstrates an ability to perform a particular task under prescribed conditions. While that ability may be indicative of a wider and deeper capability, competency appears to be reductionist, and in a climate characterised by rapid change and a degree of uncertainty about future requirements where flexibility and adaptability are valued characteristics, describing and assessing individuals’ abilities in terms of competencies may be far too limiting.

Weinert (1999) identified, in addition to common usage, nine different technical meanings that attach to the term ‘competence’. One was a general concept most closely related to traditional psychometric conceptions of intelligence and which was the product of a set of fundamental processes (p. 6). Another view arose from Chomsky’s conception of competence as an innate capacity to acquire and produce language, a latent capacity that waits to be shaped by experience, and through rule induction, to enable novel productions in specific domains. Weinert also included motivational constructs to produce what he termed an ‘action competence’ (p. 9). This model recognised the interactions of ability, motivation and context that determine performance in practice. Weinert recognised key competences as being the result of a search for competences that are “context-independent ... [and] that are equivalent in their use and effectiveness across different institutions, different tasks, and under varying demand conditions.” (p. 11). The various shades of meaning for ‘competence’ outlined by Weinert make the term a very desirable one for describing the sets of abilities that are required by individuals in order to equip them for the expected changes in work and work organisation.

Resolving problems around terminology in this field is a difficult undertaking in English and is made more complex when terms and concepts are imported from other languages and educational traditions. The added complexity is illustrated for two terms that have been translated from French and German into English.

The French term *compétences transversales* has been translated often as 'transferable skills' (for example, see E. Smith & Comyn, 2003, p.16). The term *compétence* may be taken to mean "1. competence, competency, jurisdiction, or powers" or "2. competence, ability, proficiency, or skill" (Harrap's Standard French and English Dictionary. Part 1. French-English, 1962, p. 175). The second of these meanings appears to reflect the most likely intention of the French use of the term, although conflation of the terms ability and skill illustrate the confusion that can occur when common meanings are used to convey technical subtleties. Of greater importance is the term *transversale* which is translated as "transversal, transverse, or cross..." and the examples given include cross-streets, cross-girders and cross-beams (p. 885). The notion of cutting across or spanning appears to be the most meaningful interpretation, and the translation as transferable, with its baggage, appears to be unwarranted. The nearest meaning that might be captured in English is the term used in South Africa for generic skills, namely 'critical cross-field outcomes'. This captures the notion of a set of skills that are applicable across a range of occupational types and levels. The translation 'transferable skills' conveys an untested claim, avoids a significant debate about the nature of generic skills, and fails to engage with a substantial body of literature that questions whether, to what extent and under what conditions transfer does occur.

The German term *schlüsselqualifikationen* has been translated as 'key qualifications' (Nijhof, 1998; E. Smith & Comyn, 2003, p.16). There is little doubt that *schlüssel* means key, and this may imply, as it does in English, that the referent is a main or important construct or even that it may unlock or provide access to unspecified goods (Borthwick, 1993). However, while the transliteration of *qualifikation* may be qualification, Wildhagen's Dictionary lists "capacity and ability" as alternatives (1972, p. 987). Although in English the term 'qualification' is unlikely to be equated with skill, ability or competence or other words that are used to convey that class of concepts, the literal use of the term 'qualification' may lead to an inference that the German use of *schlüsselqualifikationen* suggests that the achievement of a certain level of performance on a particular set of generic skills ought to lead to the award of a certificate or other 'qualification'. The Key Skills Qualification was introduced in England and Wales, and in Australia, a Certificate I in Generic Work-skills was proposed and some work was undertaken to implement it in several industries (ANTA & Ratio Pty Ltd, 2004; Australian Parliament. House of Representatives Standing Committee on Education and Training, 2004). The debate about whether the achievement of a certain level of performance should be acknowledged in this way has been an active one, and it is better to debate the issue on its merits, rather than to allow inferences drawn as the result of literal translations subtly to avoid those discussions.

An informative concept that has been canvassed in the debate is that of ‘competence’. The meanings ascribed to it by Weinert (1999) and by Streumer and Bjorkquist (1998) give the notion of broadly applicable skills some conceptual authority. However, some of the schemes that have been proposed do not warrant this appellation, and the terms ‘competency’ and ‘competence,’ despite their very different meanings, especially in the Australian context, have been confused very commonly. For these reasons, the term ‘competence’ is restricted in subsequent discussion of broadly applicable skills schemes.[4]

An interim position on terminology

The term ‘generic skills’ is used in subsequent discussion of broadly applicable skills schemes. Such a broad term may suggest a lack of clarity of purpose or the absence of a common purpose for the various skills schemes. However, there are substantial commonalities in purpose, as indicated in the reasons advanced for the introduction of generic skills schemes. The discussion above suggests the need for a more rigorous approach to the use of terms, and this issue is addressed in Chapter 3. Here, the term ‘generic skills’ is used to encompass the collection of schemes that have been proposed to describe the capabilities that people need in order to participate successfully in the emerging economy and society. The term ‘generic skills’ is chosen because the term itself is the most generic available and carries the least semantic baggage. The term ‘generic’ is taken to mean ‘generally applicable’ and the word ‘skill’ is used in its broadest sense referring to ‘the knowledge and ability required to perform some task (itself broad and unspecified) well’. In later chapters of this document, two of the meanings of competence identified by Weinert (1999), specifically the psychometric meaning and the production-generating meaning (after Chomsky), are given greater weight than are alternative constructions.

In order to explore the meanings that are attached to the particular skills that are included within generic skills schemes, an illustrative selection of those schemes now warrants attention.

Review of major generic skills schemes

Generic skills schemes have been developed in many countries. A list of countries and the schemes that they have introduced is shown in Table 2. In some countries, more than one scheme has been developed, either sponsored by different

[4] The concept of competence is revisited in Chapter 3.

organisations or because the original scheme has been modified as a result of experience with the first.

These schemes represent taxonomies of skills, to varying levels of complexity, and as taxonomies, they are informative about the theoretical bases – most of which are tacit, that formed the foundations for the development of these schemes.

Only a few of these schemes are reviewed in detail, largely because the Australian situation requires close examination as the research reported in this thesis has been conducted within the Australian context. Certain other schemes provide useful contrasts to the Australian schemes and they are included for this reason.

The emergence of generic skills in Australia

The first substantial national commitment to generic skills was taken with the development of key competencies. Three major reports that led to the definition and description of key competencies are generally recognised to be the Karmel (Quality of Education Review Committee, 1985), Finn (Australian Education Council, 1991), and Mayer (Mayer Committee, 1992) reports. The Carmichael Report is also significant for the contribution that it made in establishing the structural framework for Vocational Education and Training (Employment and Skills Formation Council, 1992). After the release of the Mayer Committee report and the endorsement of its recommendations by the Ministerial Council on Education, Training and Youth Affairs (MCEETYA), a considerable body of work was undertaken in the school and vocational education and training sectors on the implementation of this scheme. In about 1996, for various reasons, implementation of the key competencies initiative stalled. In the 1999, the Australian Industry Group commissioned a report that addressed skills formation for Australian industry and it began to reawaken interest in generic skills (Allen Consulting Group, 1999). Later, two other peak employer bodies, the Business Council of Australia and the Australian Chamber of Commerce and Industry, also pursued the promotion of generic skills (ACCI & BCA, 2002). The Australian Industry Group subsequently became involved in the debate and endorsed the Employability Skills Framework (Allen Consulting Group, 2006b). These developments are reviewed below.

Table 2: Generic skills schemes by country

Country	Generic skills scheme[a]	Reference
Australia	Key competencies; Employability skills	(Mayer Committee, 1992); (ACCI & BCA, 2002)
Austria	*Schlüsselqualifikationen* [Key skills]	(Mertens, 1974; Piskaty, Elsik, Blumberger, & Thonabauer, 2000)
Canada	Employability skills profile; Essential skills; Employability skills 2000+; *Compétences transversales* [Cross field skills] (Quebec)	(Conference Board of Canada, 1992, 2000a)
Denmark	Process independent qualifications (skills)	(OECD, 2001)
England, Wales	Core skills; Key skills	(Turner, 2002; Werner, 1995)
Finland	Framework for evaluating educational outcomes	(Hämäläinen & Jakku-Sihvonen, 2000)
France	*Compétences transversales* [Cross field skills]	(Trier, 2001)
Germany	*Schlüsselqualifikationen* [Key skills]	(Mertens, 1974)
Italy	*Competenze trasversali* [Transversal competencies]	(Trinchero, 2006)
New Zealand	Essential skills	NZ Ministry of Education (1993)
Norway	Core curriculum	(Norwegian Board of Education, 1997)
OECD[b] (DeSeCo project)	Key competencies	(Rychen & Salganik, 2001)
Scotland	Core skills	(Turner, 2002; Werner, 1995)
Singapore	Critical enabling skills	(Singapore Workforce Development Agency, 2003)
South Africa	Critical cross-field outcomes	(South African Qualifications Authority, 1997)
Switzerland	Trans-disciplinary goals	(Trier, 2001)
United States	Workplace know-how; 21st Century literacy	(21st Century Workforce Commission, 2000a; The Secretary's Commission on Achieving Necessary Skills, 1991)

a Some of these terms were identified in English language translations of original documents, and some terms are subject to the difficulties that arise in translation.

b While the OECD is not a country, the project which it established to investigate generic skills is a particularly important one and so it is included along with country schemes.

Quality of Education Review Committee

The Quality of Education Review Committee was established at a time of increasing youth unemployment and growing participation in post-compulsory education. Among its terms of reference, the Committee was required to provide advice on means for “the attainment of appropriate standards relevant to subsequent employment opportunities and improved preparation for tertiary education” for secondary students (1985, pp. 204-5). In addition to making recommendations on basic skills achievement, including mathematics, science and technology, the Committee recommended that emphasis be placed on communication skills and the world of work (Recommendation 10, p. 203). This report drew attention to the need to focus on outcomes of schooling rather that curriculum and other inputs, and it directed attention to more general skill areas, communication and work readiness, that became the focus of later recommendations of the Finn and Mayer Committees.

The Finn Review Committee

The Finn Review (Australian Education Council, 1991) was asked, among other very wide ranging terms of reference, to report on "appropriate national curriculum principles designed to enable all young people … to develop key competencies" (p. 2). Thus, the terms of reference of the Finn Committee introduced the term 'key competencies', although the term was not defined in them.

This Committee undertook its work at a time of major social, educational, and employment-related policy change. The youth labour market had declined and a much greater proportion of young people were remaining at school. Retention to Year 12 had increased from 35 per cent in 1980 to 75 per cent in 1990 (Australian Bureau of Statistics, 2006b). The senior school curriculum, however, was focused on those people who were planning to proceed to higher education and a need to provide a broader set of options for young people was recognised. Major policy themes that the Finn Committee identified included:

- a desire for a better educated and more highly skilled society with an interest in lifelong learning;
- the need to reassert the importance of vocational education and training and to raise its status relative to academic education; and
- an emphasis on education and training outcomes, that is the achievement of competencies. (p. 12)

The Finn Committee sought to strengthen the vocational orientation of secondary schooling, but within a comprehensive model of schooling that met the needs of all young people. To this end the Committee defined "areas of competence" that were to be "related to a young person's initial and lifelong employability" (p. 54).

The Committee drew attention to changes in the skill demands of industry and of rapid change in the Australian economy as a result of structural economic change nationally and international competition. It noted that "the most successful forms of work organisation are those which encourage people to be multi-skilled, creative and adaptable" (p. 6). Because of changing technologies and changing economic circumstances, they argued that "the ability to continue learning and acquiring new or higher level skills will be fundamental." Consequently "the emphasis of our training system has to be both on the acquisition of the specific skills for the job/trade and on flexibility" and that flexibility "requires a strong grounding in generic, transferable skills" (p. 55).

The Committee further noted a recognition by employers that students required “a foundation of basic skills and a range of broad skills and attributes which are generally relevant to the world of work without being occupation- or industry-specific” (p. 6).

The Committee recommended that emphasis be given to six “key areas of competence”:

- Language and Communication
- Mathematics
- Scientific and Technological Understanding
- Cultural Understanding
- Problem Solving
- Personal and Interpersonal. (Australian Education Council, 1991, p. 58)

The Committee then recommended that:

> All post-compulsory education and training programs for the 15-19-age cohort should include, within their overall expected outcomes, appropriate levels of competence in the six Key Areas. (p. 58)

Given the breadth of other tasks that the Committee was required to address, it recommended that an expert group be established to undertake more detailed work on defining and assessing the initial list of proposed key competencies. The work required of that group was to elaborate the basic concept of key competencies, to operationalise it for the school and training sectors, to specify levels of achievement and to recommend arrangements for assessing and reporting on student achievement. That group was chaired by Eric Mayer and reported in 1992.

The Mayer Committee

Stakeholders

Committee membership is raised because it appears that stakeholder interests influence the skills that are identified as necessary, how those skills are organised into a taxonomy, and mechanisms suggested for their implementation.

The Mayer Committee had a membership of 29 persons and a small secretariat. The chairperson, Mr Eric Mayer, had been Chief Executive Officer of National Mutual, a very large Australian financial services company. Other members were drawn from state and territory school education departments and Technical and Further Education (TAFE) departments. The Commonwealth was represented through a senior officer from the Department of Employment, Education and Training and officers from two Commonwealth created statutory authorities. The Australian Vice-Chancellors’ Committee nominated a member, as did the Business Council of Australia, one of Australia’s peak employer organisations.

Teacher unions, and the union movement more generally, were also represented. In addition, the Mayer Committee established two reference groups, one for industry and one for higher education. A detailed list of Committee members is included in an appendix to the Mayer Report (Mayer Committee, 1992, pp. 127-130). The Committee comprised educators, and industry and union leaders.

Methods

Among other tasks, the Mayer Committee was required to:

- survey work under way in the school and vocational education and training sectors in the areas of language and communication and mathematics and to advise on the feasibility of brining it together to develop useful national profiles in these areas of competence; and
- [advise on] the feasibility of a similar exercise in relation to each of the other areas of competence. (Mayer Committee, 1992, p. 77)

The Mayer Committee used its own expertise, consulted with industry and with educators in the school and VET sectors, and to a lesser extent with the higher education sector, and finally undertook a validation exercise which involved further consultations with industry.

The extensive involvement of the school and VET sectors reflected a concern at the time with post-compulsory education and training, mainly for 15 to 19 year-olds, and with the pathways available to them in moving from compulsory education to employment or further study.

Definitions

The Mayer Committee accepted the National Training Board's definition of competence.

> The concept of competence adopted by the National Training Board includes these elements: "it embodies the ability to transfer and apply skills and knowledge to new situations and environments. This is a broad concept of competency in that all aspects of work performance, not only narrow task skills, are included." (Mayer Committee, 1992, p. 7 citing the National Training Board, 1991)

The conflation of competence and competency at the point of defining their terms is notable in that the Committee, although very thorough in most aspects of its deliberations, did not recognise the emerging inconsistency between these terms. In addition, the definition assumed transferability of key competencies. This matter will be revisited in the discussion of cognitive abilities in Chapter 3.

The Mayer Committee required that for a construct to qualify as a key competency, it should conform to the following conditions.

> Key Competencies are competencies essential for effective participation in the emerging patterns of work and work organisation. They focus on the capacity to apply knowledge and skills in an integrated way in work situations. Key Competencies are generic in that they apply to work generally rather than being specific to work in particular occupations or industries. This characteristic means that the Key Competencies are not only essential for participation in work, but are also essential for effective participation in further education and in adult life more generally. (Mayer Committee, 1992, p. 7)

The Committee summarised their requirements for key competencies by saying that they must:

- be essential to preparation for employment;
- be generic to the kinds of work and work organisation emerging in the range of occupations at entry levels within industry, rather than be occupation- or industry-specific;
- equip individuals to participate effectively in a wide range of social settings, including workplaces and adult life more generally;
- involve the application of knowledge and skill;
- be able to be learned; and
- be amenable to credible assessment.

The Committee's criteria for accepting a construct as a key competency were crafted carefully and were followed in establishing the final list of seven key competencies. In establishing the above criteria, the Mayer Committee was more careful than others have been in defining, operationally, what the Committee meant by a 'key competency'.

In discussing values and attitudes and other personal qualities, the Committee said:

> Both the principles and characteristics the Committee has used to construct the set of key competencies preclude the inclusion of values and attitudes. (Mayer Committee, 1992, p. 13)

It is on this point that differences between the Mayer key competencies and comparable schemes developed elsewhere, for example the Secretary's Commission on the Achievement of Necessary Skills (SCANS) workplace know-how, emerged. Other schemes, but most notably the United States and Canadian ones, included generic skills that were based upon attitudes and dispositions. The Mayer Committee also noted that in their submissions, industry and community groups had advocated the inclusion of attitudinal and dispositional characteristics (Mayer Committee, 1992, Appendix 3, pp. 89-90).

One of the key areas of competence recommended by the Finn Review Committee, cultural understanding, was considered and tentatively proposed by the Mayer Committee, but eventually was not included as a key competency.

The Key Competencies Framework

The key competencies that were recommended by the Mayer Committee are shown in Table 3.

Table 3: The final seven key competencies

Key Competencies	Descriptors
Collecting, analysing and organising information	The capacity to locate information, sift and sort the information in order to select what is required and present it in a useful way, and evaluate both the information itself and the sources and methods used to obtain it.
Communicating ideas and information	The capacity to communicate effectively with others using a whole range of spoken, written, graphic and other non-verbal means of expression.
Planning and organising activities	The capacity to plan and organise one's own work activities, including making good use of time and resources, sorting out priorities and monitoring performance.
Working with others and in teams	The capacity to interact effectively with other people both on a one-to-one basis and in groups, including understanding and responding to the needs of others and working effectively as a member of a team to achieve a shared goal.
Using mathematical ideas and techniques	The capacity to use mathematical ideas, such as number and space, and techniques, such as estimation and approximation, for practical purposes.
Solving problems	The capacity to apply problem-solving strategies in purposeful ways, both in situations where the problem and the desired solution are clearly evident, and in situations requiring critical thinking and a creative approach to achieve an outcome.
Using technology	The capacity to apply technology, combining the physical and sensory skills needed to operate equipment with the understanding of scientific and technological principles needed to explore and adapt systems.

(Mayer Committee, 1992, pp. 8-9)

Stanley (1993) was critical of the skills that the Committee recommended. He contrasted the Finn Committee's use of the phrase 'areas of competence' with the Mayer Committee's term 'key competencies'. The use of the broader term by the Finn Committee provided scope for the constructs of interest to take on more inclusive meanings. By adopting a more restrictive definition, the Mayer Committee produced a list of constructs that were interpreted rather narrowly. Stanley also drew attention to the confusion between 'competence' and 'competency'. He expressed concern at the acceptance of competency-based education and training, which influenced the recommendations of the Mayer Committee. However, this decision had been taken before the Mayer Committee began its deliberations.

Assessment, Levels of Performance and Reporting

Section 1.5 of the Mayer report, 'Assessing and reporting achievement of the key competencies' (Mayer Committee, 1992, pp. 41-56), dealt extensively with both assessment and reporting issues. It recommended nationally consistent assessment and reporting of individual achievement of the key competencies (p. 42). It also recommended that key competencies be assessed at a particular performance level "in at least two different contexts" (p. 49), thus recognising the need to demonstrate the generic nature of these abilities.

The Committee then considered how key competencies achievement might be reported and recommended reporting at the individual level through a 'record of performance' using a common format (p. 51). The Committee also recommended reporting at an aggregate national level that was to be based upon statistical sampling of individual records of achievement, and that the performance of equity and target groups should be a specific focus of this approach (p. 55). The Committee recognised the different purposes served by reporting at these distinct levels. Individual reports were envisaged to "…provide individuals with evidence of their achievement to present to a third party, such as another education and training provider or a potential employer" (p. 50). The purposes served by national reporting were intended to "…contribute to meeting needs for public accountability about the nature of the education and training being provided to young people and about the effectiveness and efficiency of that provision" (p. 53).

In their terms of reference, the Mayer Committee had been asked to define appropriate levels of achievement for the key competencies. Some submissions to the Committee had argued for a single benchmark level for beginning employees, an approach that is consistent with competency-based education and training. The Mayer Committee recommended the establishment of three performance levels for each key competency. These levels were described as:

- Performance Level 1 describes the competence needed to undertake activities efficiently and with sufficient self-management to meet the explicit requirements of the activity and to make judgements about the outcome against established criteria.
- Performance Level 2 describes the competence needed to manage activities requiring the selection, application and integration of a number of elements, and to select from established criteria to judge quality of process and outcome.
- Performance Level 3 describes the competence needed to evaluate and reshape processes, to establish and use principles in order to determine appropriate ways of approaching activities, and to establish criteria for judging quality of process and outcomes. (Mayer Committee, 1992, p. 18)

Performance levels attracted substantial and conflicting criticism during the consultation phase. Some felt that there were too many levels, others that there were too few. Much of the criticism was directed at the descriptions of the levels: some felt that the levels were described in terms that were too abstract and that could not readily be operationalised (Mayer Committee, 1992, pp. 89-90).

These performance levels proved to be one of the barriers to the effective implementation of the key competencies initiative. The performance levels were not well understood, and some confusion was apparent between these performance levels and the first three levels of the Australian Qualifications Framework (AQF) (Down, 2000). The confusion was not surprising as most qualification offered in the VET sector were at level I, II or III of the AQF. The descriptions of these levels were common across each of the key competencies, and rather than reflecting increasing performance in each particular area of competence, they suggested increasing self-management skill and in so doing, conflated self-management with the particular skill represented by each key competency.

Summary

In summary, the publication of what became known as the *Mayer Report* was a critical moment in the schools and education and training sectors in Australia. It also had an impact upon the higher education sector, which began to address generic skills from about that time, although the uptake was slow and patchy. (This matter is discussed in the section 'Generic skills in Australian higher education'). Despite evidence of some conceptual confusion in its uses of the terms competence and competency, the Committee did define what it was prepared to admit as key competencies carefully, and remained true to its conditions for acceptance of key competencies. The Committee was criticised for excluding attitudes and values, and for the performance levels that it chose, in both their number and their descriptions.

There was support for a range of other skills, including learning to learn, and some support for the addition of a category of the basic skills of literacy and numeracy (Mayer Committee, 1992, pp. 86-95). This, and Stanley's (1993) criticism of the shift from 'areas of competence' to the Mayer Committee's representation of key competencies as narrow constructs, suggest the possibility of a skills taxonomy that is more complex than a simple list. The Committee was relatively silent on how these skills should be taught and assessed, and their assessment became one of the barriers to their effective implementation in the schools and education and training sectors.

Implementation of key competencies

Before examining recent efforts to develop and promote generic skills schemes in Australia, an investigation of the work done to implement key competencies in the schools, vocational education and training sectors is warranted. The key competencies initiative had been directed at the schools and vocational education and training sectors. Alternative generic skills schemes were generated within higher education in Australia and they will be considered separately.

The school sector

A recommendation of the Mayer Report was that a project be established "…to further validate and develop benchmarks for the performance levels of the key competencies" (Mayer Committee, 1992, p. 59). It also recommended "…that States and Territories should field test nationally-consistent assessment and reporting of achievement of the Key Competencies to develop strategies for managing implementation and identify resource implications of the proposed arrangements." (p. 63). Across Australia in the schools sector and the education and training sector, between 70 and 80 individual projects were funded. Few individual reports are available. However, project outcomes were analysed and summarised in a consolidated report (MCEETYA Transition from Schools Taskforce. Working Group on Key Competencies, 1996).

In analysing the results of individual projects, the Working Group addressed four aspects of the implementation of key competencies, namely curriculum; teaching and learning; assessment and reporting; and broad conceptual issues. The last of these will be appraised first, because they are fundamental to an appreciation of the other three sets of matters. In reviewing the pilot projects, the Working Group noted ambiguities in the ways in which the concept of key competencies was understood, and quoted a project participant who said "…the definitions were simultaneously too broad and too narrow, were tightly focused and not focused enough" (p. 145). They went on to add that this ambiguity reflected a diversity of goals attached to key competencies by different stakeholders. Those who see key competencies as employment focused outcomes may regard the definitions as too broad, while others who perceive them as 'life skills' may believe that the definitions are too restrictive. This analysis suggests the lack of a consensus of what key competencies are meant to deliver. The Working Group report itself was quite ambiguous about the consistency with which key competencies were understood. They said "At present, there is no consensus about the constructs that we are trying to assess" (MCEETYA Transition from Schools Taskforce. Working Group on Key Competencies, 1996, p. 142), but later declared: "One of the more positive outcomes of the pilot experience was that the deep conceptual differences that characterised schools of thought in much academic discussion of the competency concept were absent [among pilot project participants]." (p. 175). The conclusion, that conceptual uncertainty exists about both what key competencies are and about the particular competencies that were identified, appears to be unavoidable.

The curriculum issues raised by the Working Group are not the primary focus of interest of the current research study. However, because curriculum documents specify what is to be taught, learned and assessed, a brief reprise on the Working Group's comments about analyses of curriculum documents is warranted. The Group commented in 1996 that "Most current curriculum documentation pre-dates recent work on the Key Competencies" (p. 71) and that without explicit reference to key competencies in curriculum documents, little official support for the initiative could be inferred. An extensive review of curriculum documents was undertaken by Australasian Curriculum Assessment and Certification Authorities (ACACA) agencies in most states and territories and the conclusion was that most key competencies are implicit in existing curriculum documents, although to varying extents (Australasian Curriculum Assessment and Certification Authorities, 2003). The consistency of the various mapping activities was questioned, but this finding became important because it led to one of the assessment options, the inferred model (see below), that was widely endorsed for assessing key competencies. Similar mapping exercises have been undertaken more recently in relation to the employability skills initiative. Some unfortunate consequences of this model are discussed in Chapter 4.

Revisions of school curriculum documents have occurred in most states and territories since 1996 and key competencies have been explicitly addressed in most. In New South Wales, McGaw (1997) recommended: "That all syllabus documents explicitly identify the ways in which particular key competencies are expected to be developed by students taking the course" (Recommendation 15, McGaw, 1997, p. 67). In the South Australian Curriculum Standards and Accountability (SACSA) Framework, key competencies were included, although they were described only under the Enterprise and Vocational Education strand (South Australia. Department of Education Training and Employment, 2001, General Introduction, p. 17).

The Mayer Report had provided little if any guidance on how key competencies might be taught, and it was believed that the recommended field testing would lead to an exploration of teaching and learning issues and of consequent resource requirements. The Working Group reported generally very favourable comments by teachers who had been involved in the pilot projects and indicated that the initiative was consistent with 'good teaching and learning practices' (MCEETYA Transition from Schools Taskforce. Working Group on Key Competencies, 1996).

For example, teachers reported perceiving a need to be more reflective of their work, to adopt alternative teaching and learning methodologies, and integrate processes and outcomes of education (p. 104). Other teachers reported that they found no need to change practices and that they were able to integrate key competencies into their practices seamlessly (Dellitt, 1993, see p. 64). In their summary of teaching issues, the Working Group said:

> …many teachers found that the benefits of changing teaching style in increased student motivation and in enjoyment of the less didactic teaching role outweighed the increases in workload associated with materials preparation and assessment. (MCEETYA Transition from Schools Taskforce. Working Group on Key Competencies, 1996, p. i)

The reference to workload in the passage quoted above was one of many made in this 1996 report. The Working Group was prescient, because soon after the publication of this favourable evaluation of the acceptance of key competencies, workload, especially in relation to assessment and reporting, became a barrier to the implementation of this initiative in the schools sector.

The Working Group discussed assessment and reporting of key competencies achievement at length and concluded “Issues surrounding assessment have been the most contentious of the issues associated with the use of the Key Competencies” (p. 187). Two particular problems emerged from trials of several assessment approaches. The first was the lack of a consensus about the purposes of key competencies as an educational reform and the lack of conceptual clarity about what each of the key competencies were. The second major problem was the lack of any common understanding about the meanings of the three performance levels that had been proposed by the Mayer Committee.

The Working Group considered interim reports from three significant projects. One project had trialled consolidated teacher judgments to establish levels of students’ achievements of the key competencies (McCurry & Bryce, 1997). Another had tested the production of student portfolios as vehicles in which students could present evidence of their achievement of the key competencies (National Industry Education Forum, 2000). A third, similar, project in Queensland had used purpose-designed booklets (Key Competencies Assessment Booklets or KeyCABs) for students to record their key competencies achievements during work placements (Queensland Department of Education, 1997). Other projects in Tasmania and Queensland had identified opportunities for key competencies to be assessed within school subjects and had established criteria for that assessment (MCEETYA Transition from Schools Taskforce. Working Group on Key Competencies, 1996, see pp. 131-132). The McCurry and Bryce (1997) report is interesting because it revealed that teachers were able consistently to discriminate eight levels of performance in judging the key competencies achievements of students within their schools. By contrast, most teachers reported great difficulty with the three levels recommended by the Mayer Committee (MCEETYA Transition from Schools Taskforce. Working Group on Key Competencies, 1996, see pp. 139-141). Thus, the central problem appears not to be one of how many levels, but on what evidentiary basis desired performance is described.

Four models for eliciting evidence of key competencies performance were recognised:

- Inferred model: achievement of the key competencies is inferred from performances in elements of the curriculum in which key competencies are implicit.
- Parallel model: Key competencies achievement is judged by teachers based on students work in classrooms and in extra- or co-curricular activities (see, McCurry & Bryce, 1997).
- Separate assessment model: Specific tasks are established to reflect key competencies and students are assessed on those tasks. (See Common Assessment Tasks in Chapter 4).
- Integrated model: achievement of key competencies is assessed on tasks that students undertake for existing subject assessment activities (MCEETYA Transition from Schools Taskforce. Working Group on Key Competencies, 1996, see pp. 149-150).

Three options for reporting achievement of key competencies were offered (p. 151). Further discussion of these models is presented in Chapter 4.

In summary, the key competencies initiative was not consistently understood by teachers. However, the teachers involved in the pilot projects reported favourably on their capacities to incorporate key competencies, as they interpreted them, into their teaching practices. Assessment was found to be time-consuming, and in particular, the performance levels proposed by the Mayer Committee were not well understood by teachers.

In reviews of curriculum change in Australian States and Territories from 1986 to 1996 (Lokan, 1997), most state and territory reports referred to difficulties in implementing key competencies. Collectively, the reports indicated a degree of change fatigue. In Australia's federal system, states had complete responsibility for school education. In 1989, a significant breakthrough had occurred and Commonwealth and State Ministers for Education, sitting as the Australian Education Council (AEC), had agreed to a common set of goals for school education in a document known as the Hobart Declaration (Australian Education

Council, 1989). Leading up to this statement and immediately after it, much work had been put into the development of a nationally consistent curriculum framework, the National Statements and Profiles. It had been widely expected that this would be endorsed at a 1993 meeting of the AEC. However, in October 1992, shortly before the scheduled meeting of the Council, there had been a change of government in Victoria, and the incoming Minister for Education, responding to scathing criticism of some of the science and mathematics standards, withheld approval (Howes, 1997, pp. 110-1). Other states followed and varied the descriptors of the standards. This signalled the end of a period of unusual cooperation between the States and the Commonwealth in education.

The Australasian Curriculum Assessment and Certification Authorities (ACACA) agencies had by this time developed a “minimum position” on the development, assessment and reporting of key competencies (MCEETYA Transition from Schools Taskforce. Working Group on Key Competencies, 1996, pp. 126-127). That position recommended that key competencies should be explicitly embedded in curriculum documents and that students should have opportunities to be assessed on key competencies. However, in their curriculum mapping activities, the ACACA agencies had indicated that the key competencies were already implicit in curriculum documents. The agencies also recommended that assessment should be based on teacher judgments. In order to develop reporting that was consistent across school subjects or schools, it would be necessary to introduce some form of moderation. A third element of their position was that individual State and Territory assessment boards should develop their own guidelines on assessment and reporting of key competencies. Thus, the ACACA position was consistent with the weakened national situation on curriculum cooperation.

Considerable industrial unrest was fermenting during 1995 and 1996. As part of a program of industrial action, bans were placed on work related to implementing key competencies, and in particular on assessing them in Queensland and some other states (Grace & Ludwig, 1997, p. 163). This action was not surprising, because concerns about teacher workload had been raised in relation to the reporting requirements of the National Statements and Profiles initiative. Eltis and Mowbray (1997, pp. 94-5) referred to “horrendous [workload] demands” being made of teachers and to consequent morale problems. The key competencies pilot projects occurred at about the same time as implementation on a state-by-state basis of the National Statements and Profiles. As noted above, both initiatives placed substantially increased demands on teachers.

Full implementation of the key competencies initiative, as envisaged by the Mayer Committee, would have involved reporting of achievement of individual students on each of the key competencies. In order to establish consistency with other curriculum initiatives, individual achievement would be judged against specified outcomes, which would be articulated in curriculum documents. The Mayer Committee's position was that aggregate national data, based on samples of individuals' achievement, would be available. Similar results have been achieved in reporting on other domains of schooling. For example, national benchmarks for literacy and numeracy, for science literacy, and for civics and citizenship have been set (MCEETYA, 2004). It should be noted that literacy and numeracy achievements were assessed through standardised tests. There was no proposal for this form of testing for key competencies. Despite continued advocacy for the inclusion of key competencies and their assessment in schooling (West, 1998), comparable outcomes and benchmarks had not been established for key competencies and this situation has remained. Banks, Chair of the Productivity Commission which has responsibility for reporting on the effectiveness of government service delivery, noted that no progress had been made on MCEETYA reporting against the 'life skills' objectives specified in the Adelaide Declaration (Banks, 2005).

Together, the breakdown in Commonwealth and State cooperation, especially in senior secondary assessment, coupled with workload concerns and the threat of industrial action, stalled full implementation of the key competencies initiative in the schools sector.

The Vocational Education and Training Sector

As for school education, in Australia, vocational education and training is a State responsibility, and each State and Territory had a board that oversaw this sector. However, in this sector, a substantial level of national cooperation had existed since the early 1990s and especially since 1995 when the Australian National Training Authority[5] (ANTA) was established to administer the vocational education and training sector. An important stage in the evolution of the sector was the publication of the Carmichael Report (Employment and Skills Formation Council, 1992). This report recommended the establishment of prescribed levels of training awards under the Australian Qualifications Framework (AQF) and

[5] ANTA was formally established in 1995, but cooperation began before it came into existence. It was disestablished on 30 June 2005 and its functions were transferred to the Commonwealth Department of Education Science and Training (DEST) which, in a subsequent restructure, became part of the Department of Education, Employment and Workplace Relations (DEEWR).

recommended that competency based training should form the basis of recognised education and training.

ANTA was overseen by a Board which included members drawn from industry, unions, and education and training providers. It reported to a council of State and Territory and Commonwealth Ministers for Education and Training. The specification of the content of vocational education and training is the responsibility of ten recently established Industry Skills Councils (ISCs). This oversight was previously the responsibility of national Industry Training Advisory Bodies (ITABs). ISCs made recommendations to ANTA.

Vocational education and training programs are prescribed in training packages. These documents are developed by Industry Skills Councils through an extensive consultation process with relevant industries. Training packages are endorsed for a period of three to five years, after which they are revised and resubmitted for endorsement. Training packages describe the awards that are offered and the levels of those awards within the Australian Qualifications Framework (AQF). Each award recognised within a training package requires candidates to complete specified units of competency. Each unit of competency encapsulates a particular set of competencies (skills) that a person requires in order to carry out certain tasks in an occupation that is covered by the training package. Units of competency may contain optional components, but they must include a unit descriptor that outlines the purpose of the unit. Units also include elements of competency. These are the components that together constitute the competency that is the subject of the unit. Units include a range statement that delimits the contexts in which the competency is to be developed and demonstrated. Units of competency must also include performance criteria for each element of competency. Performance criteria specify what a candidate must demonstrate in order to be judged competent. An evidence guide is provided to assist trainers and assessors in making judgments about a candidate's demonstration of a competency (Australian National Training Authority, 2004, see Chapter 2 pp. 7-19).

Prior to the introduction of training packages, training for vocational awards had been provided under a variety of arrangements. Most of the training had been conducted within state and territory Technical and Further Education (TAFE) colleges, but some progress had been made on nationally recognised curriculum modules in some industries. Training Packages are not curriculum documents. They specify what is to be learned and the performance standards that are to be achieved, but they do not prescribe a process by which the learning outcomes might be reached. This is left to the judgment of the training provider. Thus, the vocational education and training sector had moved in a short period from training delivery based on established curricula to a competency based training regime built around training packages.

Key competencies were introduced into the vocational education and training sector as part of the pilot program that had been recommended by the Mayer Committee and which ran from 1993 to 1996. During this time, the first series of training packages were being introduced into the sector. Some of the early training packages made no reference to key competencies, as those packages had been developed before guidelines on key competencies incorporation had been published. In other training packages, brief tables listing the seven key competencies and the levels that were to be achieved were presented. Indeed, the support materials for developers of training packages did not include advice on how key competencies might be included in training delivery until its April 2001 revision (Australian National Training Authority, 2004).

A consolidated report covering 11 groups of projects conducted in the Department of Technical and Further Education and private vocational education and training providers in South Australia was prepared (Jasinski, 1996). The projects covered many aspects of key competencies and their implementation in vocational education and training programs, and attention is directed here to how key competencies were understood by teachers in the sector, how they were represented in training materials and programs, and how they were assessed. Jasinski (1996, pp. 4.11-4.13) noted that the language used to describe key competencies in the Mayer Report was found by trainers to lack clarity and that key competencies were understood differently by different individuals. The diversity of understandings among providers was a common finding in several research studies (Curtis, 1996; Lawson & Hopkins, 1996; Reynolds & van Eyk, 1996). In her report, Jasinski also drew attention to the lack of explicit reference to key competencies in teaching resources, but noted that they were implicit and that at times, during training needs analyses, a requirement for them was apparent (Jasinski, 1996, see p. 4.14). In relation to assessment and reporting, two important findings emerged. First, the performance levels described by the Mayer Committee were not well understood by practitioners and second, that assessment proved to be quite difficult. Indeed, Jasinski recommended that key competencies should be integrated with the delivery of vocational competencies and therefore that they would be assessed along with the vocational competencies and not separately assessed and reported (1996, pp. 7.33-7.36). Further, the report recommended that:

> Assessment and reporting of the key competencies using the Mayer performance levels not be supported.
>
> National reporting and achievement of the Key Competencies not be supported. (Jasinski, 1996, p. 1.6)

However, others took a different view on assessment. Keeves and Kotte (1996) argued that it was possible to relate the performance levels suggested by the Mayer Committee to the more objective cognitive hierarchy of the SOLO taxonomy (Biggs & Collis, 1982). They then used mathematics achievement data collected from the 1964 International Association for the Evaluation of Educational Achievement (IEA) study and showed that a scale of performance could be constructed for the 'Using mathematical ideas and techniques' key competency. It should be noted that the IEA assessment used standardised testing. Assessment issues and options are discussed in Chapter 4.

More recent research has continued to reveal problems around understandings of key competencies and the meanings of the performance levels suggested by the Mayer Committee. Down (2000) reported that trainers' understandings of key competencies were varied. Many either did not know what was meant by key competencies or confused them with industry-specific skills. She reported that the performance levels of key competencies were confused with AQF levels. A common misconception was that key competencies were a desirable but optional component of training packages. She suggested that some of the difficulties were the result of a rapid shift from state training authority specified training curricula to training packages and to inadequacies in the teaching qualifications available to vocational education and training teachers. Dawe (2002) also reported misunderstandings of performance levels, but found that generic skills were included in training based on most training packages. However, closer inspection revealed a low recognition factor when she asked participants about key competencies. When she switched to using the term 'essential skills,' informants responded more positively. However, in addition to references to some key competencies, such as communication skills, other skills that were not key competencies such as health, safety and security procedures, customer service, professional behaviour, grooming and work ethics, were given as examples. Thus, rather than supporting the contention that key competencies were embedded in training practices, Dawe's work confirmed the lack of understanding reported by Down and other researchers.

In order to investigate key competencies assessment practices in the vocational education and training sector, Clayton, Blom, Meyers and Bateman (2003) studied 12 cases that were thought would provide exemplars of effective practice. They concluded that only two of the cases had paid detailed attention to key competencies assessment and that only one was reporting achievement in accordance with the recommendations of the Mayer Committee. In common with other researchers, they found that key competencies were not well represented in training packages and that they were poorly understood by many practitioners.

The similarities in the contexts into which key competencies were introduced in the schools and vocational education and training sectors are striking. In both sectors, key competencies were introduced following the implementation of other major changes and it is difficult to find evidence of their use in practice. In both sectors, four factors that may explain the low impact of the key competencies initiative on practice are apparent. First, common understandings of key competencies have not emerged, and teachers and trainers remain unclear about what they are required to teach. Second, curriculum documents and training packages have not provided guidance on effective methods for the delivery of these skills. Third, and of particular note, are difficulties that have surrounded key competencies assessment. The performance levels described by the Mayer Committee have not been understood or accepted by teachers and trainers and few models of effective practice are available. Fourth, because of problems in assessing key competencies, reporting achievement either for individuals or at aggregate levels has not occurred.

It appears that the key competencies initiative did not have the sustained impact in the school or VET sectors that employer organisations had anticipated, and they moved to reinvigorate debate on generic skills. This effort was documented in two major national reports, and they are now reviewed.

Training to compete

In 1999, the Australian Industry Group (AIG) commissioned Allen Consulting to report on skills requirements of industry (Allen Consulting Group, 1999). The report *Training to compete* identified factors driving profitability in industry and ways in which enterprises were responding. It paid particular attention to the training that would provide the skills required to meet emerging challenges. Because its focus was much broader than generic skills, it is reviewed briefly. However, it did make several recommendations about desired generic skills and these are outlined. Although the report made many interesting findings about training, personnel recruitment and management practices, attention is drawn to three aspects of the report.

First, this study sought the views of a cross-section of companies through a comprehensive survey, and through focus groups and interviews, many conducted with chief executive officers and senior human resources personnel of over 350 AIG member companies. The main industry sectors represented by the AIG were manufacturing, construction and information technology.

Second, the report highlighted a change in the skill expectations of many companies from a focus on technical skills to an emphasis on generic skills and attributes. This shift was summarised as:

> … an increasing premium is being placed on generic skills, both 'hard' (notably IT skills) and 'soft' (e.g. problem-solving, team skills, willingness and ability to adapt) to be developed prior to recruitment. (Allen Consulting Group, 1999, p. v)

Not only was there an expectation that people would develop generic skills, but that they would demonstrate them prior to recruitment. This change was illustrated by comments made by industry informants during interviews, two of which are reproduced.

> We used to recruit on the basis of knowledge and skills and then fire on attitude, but now we are shifting the balance to take attitude and other personal attributes more seriously early in the process.
>
> In the production areas we look for people with the ability to learn, and with team-work skills. Only rarely do we look for ready-made technical production skills. (Allen Consulting Group, 1999, p. 33)

The expectation that generic skills would be established prior to recruitment placed responsibility for the inculcation of these skills with education and training providers and also suggested that individuals needed to be proactive in developing and demonstrating these skills.

The third notable feature was the structure of the skills taxonomy presented in the report. This is shown in Table 4.

Table 4: The Australian Industry Group skills taxonomy

Generic "core" or basic skills	**Inter-personal or relationship skills**	**Personal attributes**
Literacy Numeracy Information technology capability Understanding of systems relationships	Communication Team working Customer focus Project and personal management	Capacity to learn Willingness to embrace change Independent problem solving and reasoning capability Practicality and a business-orientation

Adapted from Allen Consulting Group (1999, pp. 30-31)

The taxonomy reflected two sets of views. The skills that are shown in Table 4 were those suggested by industry informants. The classification of those skills into the three groups, namely basic skills, interpersonal skills, and personal attributes, appears to reflect a structure developed by the consultant. The basis for the classification was not articulated in the report, but separating these three groups has a logic if it is assumed that basic skills are developed through the compulsory years of schooling and that the assessment of this group of skills was well established. The separation of attributes that had substantial affective and conative dimensions from others that were more narrowly cognitive also seems to have followed a logic dictated by how such characteristics might be developed and assessed. Whether all skills were appropriately placed is open to debate. For example, understanding systems relationships appears to be a high level ability and problem solving and reasoning, although they are influenced by affective and conative factors, are both recognised more commonly as cognitive abilities.

The report *Training to Compete* made a useful contribution to public debate in Australia about generic skills and it came at a time when efforts to implement the key competencies had waned. The skills that were identified reflected the views of industry leaders who drew attention to the need to attend to generic skills. The report suggested a more complex taxonomy than the Mayer Committee had recommended. That structure has implications for developing and assessing these skill categories, and they are addressed in Chapter 4.

Employability skills for the future

The employability skills project was an important step in ongoing efforts to give effect to generic skills in Australian education and training. It was an initiative of two of Australia's peak employer organisations, namely the Business Council of Australia (BCA) and the Australian Chamber of Commerce and Industry (ACCI). They were supported by the Commonwealth Department of Education, Science and Training (DEST) and by the Australian National Training Authority (ANTA). The project consisted of three components, namely a literature review and framework development activity (Curtis & McKenzie, 2002), a survey of senior management and human resources staff from 13 large companies (Field, 2001), and a study of the views and needs of small and medium sized enterprises (McLeish, 2002) were conducted. The literature review resulted in a suggested skills framework and this had been used in the consultations with industry. A report was synthesised by members of the project reference group who were drawn from the four organisations sponsoring the study (ACCI & BCA, 2002).

The report reviewed Australia's position in a global knowledge economy and concluded that education and training programs would be important in providing industry with a workforce capable of responding to current and future demands for skills as new business processes and products were developed, new technologies were implemented, and new opportunities and threats arose.

The term 'employability skills' was chosen at an early stage in the life of the project, the only points of debate being the lack of consistency in other terms being used to describe generic skills and a preference for 'employability' over 'employment-related' skills. The definition of 'employability skills' was:

> Employability skills are defined as 'skills required not only to gain employment, but also to progress within an enterprise to achieve one's potential and contribute successfully to enterprise strategic directions'. (ACCI & BCA, 2002, p. 4)

Despite the focus on employment, it was suggested that the identified employability skills would have wider application: "The skills were as important to effective participation in the community as they were to effective participation in the workforce" (ACCI & BCA, 2002, p. xiv). However, the contextualising annotation to each key employability skill, for example "Communication …that contributes to productive and harmonious relations across employees and customers", reveals a strong emphasis on the work environment. See Table 5 for a list of similar annotations for other key employability skills.

Rather than engage in an extensive analysis of the employability skills framework, differences between this framework and the key competencies initiative are highlighted.

The first point of difference lies in the process by which the two schemes were developed. The Mayer Committee was very broadly representative of industry, government agencies, unions, and teachers and trainers and it consulted widely with the community. The employability skills proposal was developed by a reference group of seven people; the three contributing studies having been developed by four consultants. The Mayer Committee had recommended a review of the key competencies:

> The Key Competencies be reviewed periodically … to ensure that the set appropriately reflects the generic competencies essential for effective participation in the emerging forms of work and work organisation. (Mayer Committee, 1992, p. 9)

The employability skills framework was represented as a continuation of the work of the Mayer Committee. However, the lack of broad consultation has been a factor in generating some opposition to the employability skills framework (Roe, 2003).

A second important difference was an intention to ensure that the employability skills framework should be more universal in its application than the key competencies had been. This is evident in the first of the report's two recommendations.

> That DEST refers the report, Employability Skills for the Future (2002), to relevant agencies including:

> Transition from School Task Force of the Ministerial Council on Education, Employment, Training and Youth Affairs (MCEETYA);
>
> Australian Vice-Chancellors' Committee (AVCC); and
>
> National Training Quality Council (NTQC).
>
> That these agencies be requested to respond to DEST on the following issues arising from this report:
>
> implications for policy development and programs in schools, vocational education and training and higher education; and
>
> strategies and timelines for implementation of the framework in schools, vocational education and training, and higher education. (ACCI & BCA, 2002, p. xviii)

In addition to related developments in the schools and vocational education and training sectors, the report also referred to several initiatives in higher education, namely Candy, Crebert and O'Leary (1994), work done by the Australian Technology Network of universities on generic skills, and to several programs aimed at monitoring the outcomes of university education (ACCI & BCA, 2002, p. 5).

A third important difference between the key competencies specification and the employability skills proposal was apparent in the approach taken in defining these constructs. The Employability Skills Reference Group, which managed the project, was adamant that personal attributes should be included. The Mayer Committee had deliberately excluded values and attitudes because they did not conform to the requirements for key competencies that the Committee had enunciated, and specifically because they could not be taught and assessed. In contrast, the importance accorded to these attributes by industry informants was emphasised in 'Employability skills for the future'. The reference group did acknowledge the assessment difficulties that would attend their inclusion.

> The identification of personal attributes as critical to employability by employers raises a set of issues about how to assess such attributes. Employers are using a range of tools including observation, work placements and references. However, it is essential that the education system now take up the challenge of developing assessment methodologies that can provide advice to the individual. (ACCI & BCA, 2002, p. 28)

Subsequent to the release of the report, representatives of the Australian Chamber of Commerce and Industry had been adamant that the personal attributes were important and that they must be addressed (Australian National Training Authority, 2003). However, in the face of sustained and widespread opposition, ACCI seemed to have resiled from this position to one where they were regarded as important but that there was no expectation that these attributes would be assessed and reported.[6]

The taxonomy suggested in the employability skills report consisted of 13 personal attributes and eight key skills. For each key skill a set of elements or facets were described that elaborated the skill and suggested how it might be applied in industry contexts. It was argued that each substantive skill was generic, but that the relative importance of the elements would vary between job roles and levels of employment (ACCI & BCA, 2002, pp. 29-37). The taxonomy, but without the facets, is summarised in Table 5.

Table 5: Employability skills framework

Personal Attributes		
Loyalty	Personal presentation	Balanced attitude to work and home life
Commitment	Commonsense	Ability to deal with pressure
Honesty and integrity	Positive self-esteem	Motivation
Enthusiasm	Sense of humour	Adaptability
Reliability		
Key Skills		
Communication skills	that contribute to productive and harmonious relations between employees and customers	
Team work skills	that contribute to productive working relationships and outcomes	
Problem-solving skills	that contribute to productive outcomes	
Initiative and enterprise skills	that contribute to innovative outcomes	
Planning and organising skills	that contribute to long-term and short-term strategic planning	
Self-management skills	that contribute to employee satisfaction and growth	
Learning skills	that contribute to ongoing improvement and expansion in employee and company operations and outcomes	
Technology skills	that contribute to effective execution of tasks	

The key skills section of the framework is similar to the seven key competencies that were finally endorsed in that it is a list of required skills. All seven key competencies are included in the employability skills framework. Most of the key competencies are apparent in the labels of the key employability skills but some key competencies are included as elements. Collecting, analysing and organising information is a facet of Planning and organising and Using mathematical ideas and techniques is a facet of both Communication and Problem solving. Self management, Learning, and Initiative and enterprise are new constructs that were not part of the key competencies taxonomy. The concordance between the key competencies and employability skills framework is shown in Table 6.

The decision to include personal attributes extends to some elements of the key skills. Empathising and being assertive are elements of Communication, while being resourceful, taking initiative, having enthusiasm for learning, being open to

[6] Personal communication with Mary Nicholson, ACCI, 19 August 2003

new ideas and being creative are elements of other key employability skills. The key skills component of the employability skills framework shows evidence of a more flexible approach to the definition of these constructs than that taken by the Mayer Committee.

Table 6: Relationship between key competencies and key skills from the employability skills framework

Key competencies	**Key employability skills**
Collecting, analysing and organising information	Communication
Communicating ideas and information	Team work
Planning and organising activities	Problem solving
Working with others and in teams	Self management
Using mathematical ideas and techniques	Planning and organising
Solving problems	Technology
Using technology	Learning
	Initiative and enterprise

In preparing the final employability skills report, the reference group had access to a wide range of resources including the literature review and framework development document (Curtis & McKenzie, 2002) and an earlier review of generic skills schemes (Kearns, 2001). These, and other generic skills schemes, suggested more complex taxonomies in which skills and attributes were clustered, although there was little agreement among the various schemes. The logic driving the Curtis and McKenzie scheme was the separation of skills into those central to compulsory education, those that were predominantly cognitive, and therefore teachable and assessable, and those that were mainly affective or conative, and therefore less readily assessable.

Basic skills are the focus of a substantial body of work being undertaken in the schools sector. The Ministerial Council on Employment, Education, Training and Youth Affairs, through its Performance Monitoring and Reporting Taskforce, is monitoring a range of educational outcomes. In most States and Territories,[7] Basic Skills Tests covering Literacy and Numeracy have been implemented throughout the primary years and performance benchmarks have been identified at several grade levels (MCEETYA, 2004). It is possible that the reference group was aware

[7] The separate State and Territory Basic Skills Tests have been replaced by a National Assessment Program: Literacy and Numeracy (NAPLAN) as of 2008.

of this work and chose not to include basic skills within the employability skills framework.

The definitional problem is revisited in the concluding discussion to this chapter. Before turning attention to the higher education sector in Australia and to approaches taken in other generic skills schemes, a brief review of work done to implement the employability skills framework in the three main education sectors is warranted.

Efforts to implement the employability skills framework

Since the release of the report Employability Skills for the Future, three projects were undertaken to promote the framework. The first of these, 'The acquisition of employment related skills by secondary school students', was proposed in order to ensure that the framework was compatible with secondary school curricula (Erebus Consulting Partners, 2003). The report dealt briefly with the validity of employability skills constructs, saying:

> The Project found broad endorsement across jurisdictions and stakeholders that the attributes and skills in the framework in Employability Skills for the Future are, at face value, ones that young people should have as they make the transition from school. (Erebus Consulting Partners, 2003, p. 10)

The continuation of that paragraph reflected a range of other issues that became the major focus of the report.

> …However, the evidence also indicates that there is a shared belief that a focus on employability skills per se should not be the "driver" of the agenda, but that rather future work should take place in the broader context of the employment related skills and the social outcomes and life skills that young people require as they make the transition from school. (Erebus Consulting Partners, 2003, p. 10)

Much of the report dealt with transition from school and how the employability skills framework might become part of a much broader range of 'employment-related' initiatives designed to support that transition. The change in terminology is significant, as 'employment-related' is meant to convey a concern with very diverse skills rather than the sharper focus of employability which had influenced the development of the employability skills framework. The report evoked the ire of Australia's three peak employer organisations who jointly wrote to the chairperson of the Transition from School Taskforce expressing a view that the purposes of the consultancy had not been fulfilled, that it did not acknowledge the problems that had prevented complete implementation of the key competencies initiative, and that it could derail other cross-sectoral work being undertaken on the employability skills framework. An alternative view was that senior education department and assessment agency officers had taken the opportunity provided by their input to the Erebus review to assert control over an agenda for change. However the report was interpreted, it had the effect of delaying implementation of the employability skills framework in the schools sector.

In the vocational education and training sector, a very wide-ranging review of training packages had been established in 2002 (Schofield & McDonald, 2004). As indicated in previous discussion, training packages are critical documents in the vocational education and training sector because they communicate the required outcomes of training programs and are the basis of funding arrangements in the sector. Key competencies had rarely been prominent in training in the sector and The High Level Review of Training Packages was required to attend to it. Some work had been done on this problem in a separate study (Ratio Pty Ltd & Down, 2003), in which a variety of models for incorporating generic skills into training packages had been explored. Schofield and McDonald affirmed the importance of generic skills.

> We believe that generic skills is the most significant design issue facing the Training Package model, and one which requires immediate attention and urgent resolution. Generic skills are highly valued by industry and employers for their role in facilitating competent workplace performance and in enhancing the capacity of workers to respond, learn and adapt when workplace demands change. (Schofield & McDonald, 2004, p. 19)

Schofield and McDonald also noted that, despite 'widespread support' for generic skills, the employability skills framework had not received formal national endorsement and, therefore, adequate attention to its implementation. They recognised a lack of guidance for the teaching, assessment and reporting generic skills achievement and acknowledged continuing debate about definitional problems, but suggested that:

> …the current work in this area be accelerated and strengthened, and that an agreed national position be reached and implemented as a matter of urgency [and]
> Training Package developers be provided with specific guidance and advice on how to identify employability skills needed within industry and how to incorporate them into Training Packages. (Schofield & McDonald, 2004, p. 19)

Shortly after the publication of the Schofield and McDonald report, the National Training Quality Council did endorse the employability skills framework and did require the Australian National Training Authority to accelerate work on the incorporation of employability skills into training packages.[8]

The third major action designed to give effect to the employability skills framework arose from a working party that included representatives of four

[8] Personal communication, Ms Ann Dunne, Training and Skills Commission, 10 September 2004

education sectors, namely schools, vocational education and training, higher education, and adult and community education.[9] They asked the Department of Education Science and Training to commission a project to explore a uniform approach to the assessment and reporting of employability skills achievement across the four education sectors. The project team, led by the Allen Consulting Group, considered the range of activities that had led to the development of the employability skills framework. One of the documents considered by the project team was a report from a parliamentary committee on education and training which had recommended the establishment of a generic skills qualification.

> The Committee recommends that Commonwealth, state and territory education authorities through ANTA fast track the development of a Certificate I in Generic Workskills for all students to complete by Year 10. (Australian Parliament. House of Representatives Standing Committee on Education and Training, 2004, Paragraph 9.34, p. xxxi)

This recommendation warrants close analysis. First, the establishment of a generic skills qualification had been considered but had enjoyed little support in Australia. A Key Skills Qualification was trialled in the United Kingdom and it was found to be onerous for both training providers and learners and had little traction with employers (Hodgson & Spours, 2000; Powell, Smith, & Reakes, 2003; Pumphrey & Slater, n.d.; Turner, 2002; Washer, 2007). Second, it was to be implemented at about the end of compulsory schooling for most students (Year 10). This was before many students would have had an opportunity to engage in work experience programs or in vocational learning programs, which were offered in Years 11 and 12. Thus, the suggested timing seemed quite premature. Third, the level of the qualification, Level 1 of the Australian Qualifications Framework, while reasonable for a Year 10 program, was unlikely to appeal to key stakeholders when vocational qualifications of any substantive value are at least at AQF Level II, but only AQF Level 3 qualifications and above were regarded as having any substantial positive influence on student outcomes (Stanwick, 2005, 2006). What is significant about the proposed Generic Workskills qualification is that it represents an acknowledgment that certification of generic skills is important and suggests that a requirement for some formal recognition of these skills may be needed to encourage schools and other education and training providers to attend to generic skills constructs in their programs.

---

[9] Normally, the Australian education system is categorised as school education, post-school vocational education and training (VET), and higher education. The Adult and Community Education (ACE) sector is often treated as part of, or perhaps an appendage of, the VET sector, as much of the work of the sector involves assisting adults to re-enter the workforce. It does however, have a broader remit than this. In most of the discussion that follows, reference is made to the three dominant sectors.

The project group released a discussion paper (Allen Consulting Group, 2004), invited submissions, conducted a one-day forum of interested individuals from the four education sectors, and produced a directions paper (Allen Consulting Group & National Centre for Vocational Education Research, 2004). The main objective of the project was to develop a universal recognition and reporting strategy for employability skills achievement. The project group identified four key issues that would need to be addressed by the recommended strategy, namely the definition of employability skills, the development of employability, the collection of evidence and recording of employability skills achievement, and the assessment or verification of achievement (Allen Consulting Group & National Centre for Vocational Education Research, 2004, p. 3).

The discussion on definition of employability skills (pp. 3-5) centred on whether the employability skills framework should be the basis of recording and reporting. This discussion was rhetorical, because it was apparent that the framework would be endorsed, at least within the schools and training sectors where governments and their regulatory agencies could exert direct influence on their adoption. The value of this discussion is doubtful in relation to higher education provision because of the autonomy of universities. Similarly, discussion about the development of employability skills raised the issue of prescription and concluded that it was unnecessary for teaching methods to be prescribed because achievement could be reported, irrespective of how these skills were developed. However, on recording, the report was prescriptive.

> One of the conclusions drawn from information and views provided by stakeholders in the early part of the project was that there appears to be general acceptance amongst stakeholders that the use of individual employability skill portfolios is the most appropriate medium for the recording of employability skills. (Allen Consulting Group & National Centre for Vocational Education Research, 2004, p. 7)

This conclusion is of great concern. The literature review for the project, relegated to an appendix in the report, claimed support for portfolios as a means of assessing student achievement.

> The effective use of portfolios centres the evidence on both evidence [*sic*] of holistic learning experiences and on the achievement of defined competencies to agreed standards. This evidence is collected over time and in multiple ways, thus assuring the validity of the evidence, its reliability and is inherently flexible [*sic*]. There has been a growing acceptance of the usefulness of portfolios across all major education sectors in Australia (Bowden et al 2000, Curtis and McKenzie 2001, Curtis 2004b). This also demonstrates the shift in the nature of skill development as educational and workplace communities become more experienced and mature in their use of a wider range of teaching and learning processes. It is also indicative of a shift from a teacher-centred to a more student focussed (or, more specifically,

> a student-activity focussed) approach to learning and skill development. (Allen Consulting Group & NCVER, 2004, p. 43)

The Bowden at al. paper (2000) canvassed portfolios as a mechanism by which students could maintain a record of their achievement of generic capabilities. However, recommendations were made that these capabilities should be assessed both implicitly and explicitly within university subjects (Bowden et al., 2000, p. 19). The Curtis and McKenzie references to portfolios indicated that portfolios “subsumed other forms of assessment” (Curtis & McKenzie, 2002, p. 19) and therefore they were not, themselves, substantive forms of assessment. The Curtis (2004b) reference did not find that portfolios assured “the validity of the evidence, [and] its reliability”. On the contrary, in that paper Curtis (2004a, pp. 143-144) reported that significant concerns had been expressed about both the validity and reliability of portfolios in assessing and recording achievement.

In addressing assessment, the universal recognition report said: “The preferred option does not prescribe how, or indeed if, skills should be formally verified/assessed.” (Allen Consulting Group & National Centre for Vocational Education Research, 2004, p. 13). Thus, the report avoided any discussion of validity and reliability of what might be contained in and assessed through portfolios.

In summary, the ‘universal recognition’ report offered only one strategy, the portfolio, as a mechanism for ‘recognising and reporting’ achievement, avoided debate about assessment, and therefore did not consider the provenance of the information that might be included in portfolios. The recommended strategy appears to be based on a very flawed understanding of the literature available on portfolios.

The three activities reviewed above, the Erebus Report (Erebus Consulting Partners, 2003), the high level review of training packages (Schofield & McDonald, 2004), and the Allen Consulting Review of assessment methods (Allen Consulting Group & National Centre for Vocational Education Research, 2004), designed to promote the adoption of the employability skills framework in Australia’s four education sectors have had mixed success in engaging their intended audiences. They have, however, served to illustrate five distinct problem areas that must be considered by researchers and by policy makers. Stakeholders groups have not been fully represented in the early stages of this policy development and this has led to some opposition in attempts to implement the employability skills initiative. Some misunderstanding and disagreement about the definition of employability skills have been apparent in the reception of the construct, especially in the schools sector. None of the reports has dealt with issues of pedagogy, and this remains a concern yet to be addressed. Similarly, assessment has been recognised as a key area of concern, and yet it too has been by-passed in an attempt to move quickly to a minimalist approach to assessing, recording and reporting achievement. Because of a failure to attend, seriously, to the issue of assessment, the confidence that consumers of reported achievement, mainly employers, might have in that achievement must be open to question. However, before these matters may be considered more substantially, other developments within Australia and overseas must be reviewed.

Generic skills in Australian higher education

The emergence of generic skills in Australian higher education is treated quite separately from the evolution of these skills in the schools and vocational education sectors because the forces that have driven the generic skills agenda and the mechanisms by which they have operated in higher education are quite different from those influencing the other sectors.

Two factors appear to have contributed to differences in the response to generic skills initiatives of the higher education sector compared with the other two main sectors. First, Australia's universities are relatively autonomous, self-accrediting institutions. Almost all were established by acts of State Parliaments and formally come under state jurisdictions. However, states and territories contribute very little to their funding, much of which is now derived from non-government sources, although approximately 40 per cent of funding is provided through Commonwealth grants. Thus, in contrast with school education and vocational training, the Commonwealth has most influence through higher education policy and funding rather than through direct control. Second, the Mayer Committee's recommendations on key competencies were intended to apply to the schools and vocational education and training sectors. The main interest of universities in the outcomes of the Mayer Committee's deliberations was in whether the achievement of key competencies would be considered in the selection of students for university admission.

The evolution of generic skills in higher education

Despite their detachment from the key competencies initiative, many universities had begun to explore the development of generic skills among their graduates. The principal driver for this exploration was quality assurance of program delivery, and this had two facets, namely a formative function in which the quality of teaching within courses might be enhanced, and an accountability function in which the teaching performance of universities was reported. Three sources for this pressure could be identified, namely government policy and public accountability, industry and professional bodies, and universities themselves.

The Requirement for Public Accountability

The growth in demand for higher education places and in enrolments, referred to as the 'massification' of higher education, coupled with the creation of many new universities led to concerns about the quality of university courses. A group of vice-chancellors of established universities lobbied the government to instigate a quality assurance mechanism. There was some irony in this, as the Australian Vice-Chancellors Committee[10] (AVCC) has been active in circumscribing the scope of the various quality assurance mechanisms that had operated since about 1992.[11] For example, the AVCC guidelines on the use of data from the Graduate Destination Survey and the Course Experience Questionnaire indicated that the data should not be used in making comparisons between universities (Australian Vice Chancellors Committee, 1995). There was further irony in that, in subsequent reviews of university quality, some of the newer universities that had been the subject of AVCC concern were shown to have higher quality teaching programs than some established universities. At that time, the report *Achieving quality* (Higher Education Council & National Board of Employment Education and Training, 1992) was published. It argued that quality in higher education should be measured by the quality of its graduates and that this was best indicated by their achievement of generic skills.

> They [generic skills] include such qualities as: critical thinking, intellectual curiosity, problem solving, logical and independent thought, effective communication and related skills in identifying, accessing and managing information; personal attributes such as intellectual rigour, creativity and imagination; and values such as ethical practice, integrity and tolerance. (p. 22)

They noted that a discipline-based body of knowledge must be developed in order to meet the needs of specific professions and vocations and also to be a vehicle for the development of generic skills, because "it is only through the study of a body of knowledge that they can be acquired" (p. 20). The report went on to argue that generic skills would not be acquired incidentally through higher education, but that they needed to be specific objectives that were deliberately developed through

[10] This organisation is now Universities Australia.

[11] Three phases in quality assurance mechanisms could be identified. Between 1992 and 1996, a series of audits of university teaching, student support and research were conducted. This process was discontinued and between 1997 and 2000, administrative data collections, which used information on demand for places, graduation and subsequent employment rates, and graduate satisfaction, were used to assess quality. Since March 2000, the Australian Universities Quality Agency has operated a series of audits and has used administrative data to assess and report on university quality.

teaching and assessment methods. The last of these assertions was at odds with the Crittenden's (1997) contentions that exposure to a liberal education program, of itself, would lead to the development of the generic skills that he advanced.

Clanchy and Ballard (1995) were very critical of *Achieving quality* on two grounds. Their first claim was that the report's definitions of generic skills were poorly developed. Their second critique was that generic skills had meaning only within a disciplinary context and that the use of generic skills as university wide indicators of quality, being promoted at the time by many universities, lacked substance and that "perhaps a great deal of this is merely display for political purposes and is not intended to be taken too literally or too seriously" (Clanchy & Ballard, 1995, p. 157).

Quality in higher education provision was already on the national agenda. The Dawkins white paper had indicated that accountability would be a priority and that "student satisfaction and completion rates" would be among the performance indicators to be implemented (Dawkins, 1988, p. 85). Ramsden (1991) delivered a report on the Course Experience Questionnaire, an instrument that would assess graduates' perceptions of the quality of their courses. The Course Experience Questionnaire was a 25 item instrument with five subscales, one of which was a five-item indicator of generic skills achievement. One of the items on this scale was: "The course improved my skills in written communication." Other concepts included in this scale were problem solving, teamwork, planning and organising, and analytical skills. It is worth noting the similarity between these constructs and several of the key competencies. After initial trials in the early 1990s, the Course Experience Questionnaire has been administered annually to graduates of Australia's universities since 1994. The logic of using an indicator of student satisfaction was that good teaching led to effective learning and that effective learners were satisfied with their courses (Ramsden, 1998). Thus, student satisfaction (perhaps, more accurately, student perception of course quality) was used as a proxy measure of course quality.

Despite some criticism, the Course Experience Questionnaire remained relatively unchanged since its introduction, at least until about 2003, following a review of the instrument (McInnis, Griffin, James, & Coates, 2001). However, other more direct indicators of quality have been sought. The Graduate Skills Assessment (Australian Council for Educational Research, 2000) was an instrument designed to assess students' skills in problem solving, critical thinking, interpersonal understandings, and report and argument writing. The first three scales were assessed through a series of multiple-choice items while writing was assessed using extended responses to prompts.

Both the Course Experience Questionnaire and the Graduate Skills Assessment were developed at the behest of the various Commonwealth Government Departments that have had responsibility for higher education since 1992.

The Influence of Industry and Professional Bodies

The Business Higher Education Roundtable provided a forum for dialogue between industry and universities. In 1998, that group promoted the development of graduate attributes as one element of quality in the provision of higher education, and proposed a list of generic skills (BHERT, 2003). Later, they commissioned a paper that argued the case for developing generic skills in undergraduate education and that urged their explicit delivery and assessment in higher education programs (Hager et al., 2002). The Business Higher Education Roundtable also published a special issue of their newsletter (B-HERT News) devoted to effective practice examples of generic skills in Australian universities (BHERT, 2003).

Individual professional associations had also sought to encourage universities to pay close attention to the inculcation of generic skills through courses in their disciplines. Speech Pathology Australia supported a project to promote the development of professional competencies, including generic competencies, among students taking speech pathology courses.[12] A particularly compelling case for the development of generic skills was made by the Institution of Engineers[13] in cooperation with the Academy of Technological Sciences and the Australian Council of Engineering Deans, who undertook an extensive review of engineering education in Australian universities. The review found that engineering graduates had highly developed technical abilities, but that a broader range of skills would be required of future professional engineers.

> The Review of Engineering Education is recommending no less than a culture change in engineering education which must be more outward looking with the capability to produce graduates to lead the engineering profession in its involvement with the great social, economic, environmental and cultural challenges of our time. (Institution of Engineers Australia, 1996a, p. 2)

The Review Committee consulted within engineering industries and reported that there was a need for "more generally competent graduates." They recommended that engineering courses should ensure that graduates had, in addition to a sound

[12] Personal communication with Sue McAllister, who undertook this work as a PhD project at the University of Sydney (McAllister, 2005).
[13] This organisation is now called Engineers Australia.

basis in science and engineering fundamentals, communicative ability, problem solving skills, the capacity to work in multidisciplinary and multicultural teams, an understanding of the social impact of engineering, a commitment to ethical practice, and a willingness to engage in lifelong learning (Institution of Engineers Australia, 1996b, p. 30).

Developments within Universities

Universities, individually and collectively, began to explore generic skills as one outcome of their programs that reflected the quality of course provision. A traditional liberal education agenda could be identified in the generic skills literature in higher education. Boyer, Altbach and Whitelaw (1994) articulated a conventional view that universities existed to discharge three principal functions which they described as the 'scholarships' of discovery (research), teaching, and [community] service. The teaching function had been seen as providing a general liberal education and Crittenden (1997) argued that, for an institution to warrant the label 'university', it should be characterised by the following:

> the teaching program should provide a liberal education for all and it should enhance the intellectual virtues of disciplined effort of mind, respect for rational inquiry, intellectual honesty, submission of one's views to public verification, respect for others while criticising their views;
>
> while its education should be broad and liberal, this will have the effect of all sorts of other improvements in the human condition;
>
> teaching and research will have vocational elements, but must be based on the most general theories. (pp. 90-92)

The 'intellectual virtues' that Crittenden raised can be seen as generic capabilities that should emerge through the experience of university learning. The application of these generic attributes was, presumably, the driver of 'other improvements in the human condition.' But, university education had always included strongly vocational elements, with medicine, law and theology having been prominent among early course offerings. By the time Crittenden had published his views on the core characteristics of university education, a more strongly vocational challenge had emerged to his espoused liberal education tradition.

Coaldrake and Stedman (1998) charted the changing demand for university places from the establishment of the first Australian university, Sydney in 1850, to the mid 1990s. After relatively flat demand between 1914 and 1945, demand for higher education had been growing steadily since 1945. Six universities existed in Australia in 1945. This number grew to 14 by 1960, and to 22 by the mid 1970s. However, much of the demand for tertiary places was met through 'second tier' higher education institutions. These included teachers colleges, agricultural colleges and institutes of technology. During the 1970s, many of these institutions were designated ‘colleges of advanced education.’ They expanded their range of programs and began to offer degrees rather than diplomas. New courses were introduced that previously had not been delivered by tertiary education providers, including nursing courses that formerly had been taught within hospitals. In 1988, the then Commonwealth Minister of Education released a key discussion paper (Dawkins, 1988). This recommended the abolition of the ‘binary’ system and the creation of the ‘unified national system’ of higher education. Under these arrangements, some colleges of advanced education and institutes of technology were merged with existing universities and a further 16 universities were created when former colleges and institutes were designated as universities. Thus, particularly from the mid 1970s to the early 1990s, there was great expansion in the tertiary education sector with a much greater range of courses than had been available and with many more students participating.

The Australian Technology Network (ATN), a group of universities that were formerly major institutes of technology, undertook a project to develop frameworks for the definition, development, teaching, learning and assessment of graduate capabilities that would ensure that graduates are “readily employable in a variable job market” (Bowden, Hart, Trigwell, & King, 1998, p. 1).

Graduate attributes were defined as “the qualities, skills and understandings a university community agrees its students should develop during their time with the institution. These attributes include, but go beyond, the disciplinary expertise or technical knowledge that has traditionally formed the core of most university courses.” However, the report restricted consideration to graduate capabilities, being “only a subset of the different attributes …, specifically those concerned with transition to the workplace” (Bowden et al., 2000, pp. 1 and 4). Thus, the scheme that was proposed was vocational, rather than liberal, in its orientation. In discussing reasons for addressing generic capabilities, the project team referred to the desirability of universities differentiating themselves from others in a competitive environment by “adding value” to their graduates.

> By fostering, assessing, and recording judgement of generic capabilities the university demonstrates its commitment to producing potential employees that actually possess the characteristics the university says it values and employers have argued they need. (Bowden et al., 2000, p. 5)

The report did not set out to prescribe a set of generic capabilities. This was left to individual universities and course teams, but a list of common attributes was presented and it included “communication, teamwork, creativity, critical analysis, professional and personal responsibility, leadership, information literacy, IT literacy, international orientation and environmental awareness, among others”

(Bowden et al., 2000, p. 17). However, the report was adamant that, whatever attributes were selected, specific opportunities for students to develop them and to be assessed on them within their discipline in a way that provided a credible record of their achievement were required.

The ATN project report dealt in detail with the issue of assessment. It laid a basis for this through an elaboration of the definition of selected generic capabilities. Four levels of definition were identified, namely scoping, enabling, training and relating. At the scoping level, the purpose for the skill was articulated. At the enabling level, the elements of the skill were elaborated. These components were the focus of training, and that training was specific to the discipline in which the skills were being developed. At the relating level, generic capabilities were contextualised further through an adaptation of the skill to the varied contexts encountered in the practice of the discipline.

The report identified five purposes for assessment of generic capabilities, including a formative function, a summative "authoritative statement of attainment" purpose and a quality assurance role. The purposes were significant because they acknowledged the interests of multiple stakeholders in students' progress through higher education. Both implicit and explicit assessment of generic capabilities was recommended within and across subjects. An important suggestion was the profiling of students' emerging capabilities. This had implications for the way assessment was conducted and reported.

The detailed conceptualisations of generic capabilities and the recommendations on assessment set this project apart from previous approaches to generic skills development in the higher education sector.

A search of Australian university web sites in 2002 revealed that 18 of the 38 universities referred to generic skills in some form. In most cases, these were endorsed at university level but some were faculty or departmental initiatives. Milne (2000) described what were called tertiary literacies at the University of Wollongong which had promoted these skills since at least 1992.

The University of South Australia developed a set of seven 'graduate qualities.' Graduates of the university were expected to:

- operate effectively with and upon a body of knowledge;
- be prepared for lifelong learning;
- be effective problem solvers;
- work both autonomously and collaboratively;
- be committed to ethical action and social responsibility;
- communicate effectively; and

- demonstrate an international perspective. (University of South Australia, 2000)

The University took a strong systemic approach to the development and assessment of graduate qualities. Staff were required to review each course they taught and to show what proportion of the time allocated to the course was spent on the development of these qualities. They were also provided with guidelines on assessment approaches – for example, essays, projects or examinations – and the opportunities for developing and demonstrating graduate qualities that were available using those approaches.

Summary of the evolution of generic skills in Australian education

The discussion in the preceding section on the evolution of generic skills – specifically the key competencies and employability skills initiatives – is summarised in the following two tables. Both tables are organised in an approximate time sequence, but events occurred at different rates in the three main education sectors. Different groups of stakeholders were influential in the three sectors. Senior state and territory education officials were active in relation to the schools sector and their input into some decisions, for example the qualified endorsement of employability skills by MCEETYA, appeared to have occurred because of their influence. In the higher education sector, a desire to preserve university autonomy would appear to have prevented a common initiative across the sector.

Further, several key issues in the success of these innovations are also identified. These issues are elaborated in the next section. Table 7 summarises the development and implementation of the key competencies and Table 8 summarises the events and decisions involving the employability skills framework. Although the sequences of events for the key competencies and employability skills initiatives are shown separately, they reflect a continuing concern with the need to broaden school and vocational education especially to accommodate the growing demand for flexibility in the emerging organisation of work and other social interactions.

Table 7: Summary of key competencies developments in Australia, 1985 to 2001

Milestone	Education sector		
	Schooling	VET	Higher education
Articulation and definition	Identification of the need for higher-level skills (Quality of Education Review Committee, 1985)		
	Proposed key competencies (Australian Education Council, 1991) Key competencies identified and defined; proposals for assessment (Australian Education Council, 1992)		Key competencies were not implemented, although many universities did respond to ‘Achieving quality’ (Higher Education Council & National Board of Employment Education and Training, 1992)
Implementation of key competencies: Enablers and barriers	Key competencies in schools (MCEETYA Transition from Schools Taskforce. Working Group on Key Competencies, 1996) Review of implementation trials (Ryan, 1997) Factors that militated against key competencies implementation (see notes within chapters on statements and profiles in various states Lokan, 1997) (National Industry Education Forum, 2000)	Early versions of training package guide for developers included advice on inclusion of key competencies in unit descriptions (Australian National Training Authority, 2004)	Course Experience Questionnaire (CEQ) was introduced and include a generic skills scale (Ramsden, 1991) Institutional developments beginning early 1990s (See also Bowden et al., 2000; see e.g. Milne, 2000; University of South Australia, 2000) Crittenden (1997) argued that universities had traditionally focused on ‘generic skills’.
Assessment and reporting	Teacher judgment (McCurry & Bryce, 1997, 2000)	Embedded VET in Queensland senior subjects (1994)	External assessment (Australian Council for Educational Research, 2000, 2001a)

The abbreviation KC refers to the key competencies and ES to the employability skills initiatives.

Most of the work that had been undertaken on the implementation of the key competencies had been completed by 1996. It would seem that difficulties in the assessment and reporting of key competencies, especially the workload implications at a time of industrial unrest, led to a failure to implement the initiative effectively. By 2001, peak employer organisations were concerned about this failure and wanted to reinvigorate a generic skills initiative. As a result of a further round of work that built on the proposals for the key competencies, the employability skills framework was developed and further implementation efforts were made. Key events in the development of the employability skills framework are summarised in Table 8.

Table 8: Summary of employability skills developments in Australia, 2002 to 2008

	Education sector		
Milestone	Schooling	VET	Higher education
ES definition and initial validation	Background research including consultations and validation (ACCI & BCA, 2002; Curtis & McKenzie, 2002; Field, 2001; McLeish, 2002)		
Implementation	Recommendation of certificate in generic ES (Australian Parliament. House of Representatives Standing Committee on Education and Training, 2004)		Promotion of generic skills in higher education (Business Higher Education Round Table, 2003; Hager et al., 2002)
	Qualified endorsement of ES and other skills (MCEETYA, 2003)	ES endorsement, following extensive background work (National Training Quality Council, 2005)	
Development, assessment and reporting	The Universal Recognition project (Allen Consulting Group, 2004; Allen Consulting Group & National Centre for Vocational Education Research, 2004)		
	Mapping ES against curriculum and assessment standards documents (Australasian Curriculum Assessment and Certification Authorities, 2003) Relating ES to other initiatives (Erebus Consulting Partners, 2003) Proposals for development, assessment and reporting of ES among senior secondary students (Curtis & Matters, 2008; Matters & Curtis, 2008)	Canvassing assessment options (Australian National Training Authority, 2002, 2003, 2004; Ratio Pty Ltd & Down, 2003) A recommendation on assessment (Allen Consulting Group, 2006a; Cleary, Flynn, & Thomasson, 2006) Development of certificates in employability skills in three industries (ANTA & Ratio Pty Ltd, 2004) Staff development for assessment (Mitchell, 2007)	Learning and teaching performance fund was introduced in 2004 to reward universities for quality teaching and learning achievement. One of the indicators was the generic skills scale of the CEQ (DEST, 2006)

The abbreviation KC refers to the key competencies and ES to the employability skills initiatives.

Critical issues in implementing generic skills

Analysis of the key competencies initiative and its relative lack of success led to the identification of four pairs of issues that are thought to be critical in the success or failure of the venture. They are definition and selection, dissemination and implementation, assessment and reporting, and certification and recognition. They were articulated in the literature review that was undertaken in the initial development of the employability skills framework (Curtis & McKenzie, 2002, pp. 54-61). Although these issues were discussed in the framework development for the employability skills initiative, they were not adequately reflected in the final framework development or in the recommendations for their implementation. In particular, although viable assessment and reporting was identified as a critical success factor, no recommendations were made in the final report on how employability skills should be assessed and reported (ACCI & BCA, 2002).

Given the importance of the 'critical success factors' identified above, attention is paid to definition and to assessment in framing the approach taken to generic skills assessment in this study. Decisions on the other matters, namely dissemination and implementation (teaching) and certification and recognition are not discussed here. Decisions about them flow from decisions about definition and assessment.

A framework for understanding generic skills and for defining problem solving as an assessable construct is developed in Chapter 3. Assessment options for generic skills are discussed in Chapter 4. Below, the argument is developed that failures to define generic skills adequately and failures to develop workable assessment strategies were central to the failure of generic skills as a policy initiative. This discussion summarises the history of the implementation of generic skills in Australia around the issues of their definition and assessment. Work done on other generic skills schemes is brought into this discussion to illustrate where alternative approaches (to those taken in Australia) shed light on the Australian experience. Examples of generic skills schemes developed in other countries, and some discussion of the rationales for them and issues that arose from their implementation, are presented in Appendix 2. Brief references are made to some aspects of them in the discussion that follows. The issues that have attended the two major generic skills schemes in Australia can be identified in relation to the generic skills schemes in the United States of America, the United Kingdom, Canada and in some other countries.

Definition and selection

The problem of defining generic skills occurs at two levels. First, it is necessary to decide what is meant by a generic skill. In all generic skills schemes reviewed, many skills and attributes have been proposed as essential for personal growth and development, social engagement and workforce participation. The Mayer Committee established clear criteria for endorsing skills as being generic (1992, pp. 7-13).

Second, having established criteria accepting a skill as being generic, those skills that are accepted must be defined individually. Providing broadly accepted definitions is difficult. Communication, for example, can be specified from sociological, psychological or information technology perspectives, among many others. Defining the skill from a single disciplinary perspective almost certainly attracts criticism from others. The Mayer Committee, having consulted widely about what skills to admit, provided rather general descriptions of its chosen skills. These were found not to have much traction in the schools and VET sectors and the lack of clarity was one factor among many in the failure to implement the scheme. The Employability Skills framework avoided the disciplinary problem by defining the chosen skills through examples of the skills in practice. These examples were labelled as elements or facets and it was noted “that the mix and priority of these facets would vary from job to job” (ACCI & BCA, 2002, p. 26). For the skill of communication, some of the 13 facets were listening and understanding, negotiating responsively, persuading effectively and being assertive.

Finally, any scheme that endorsed all the candidate skills would become unworkable. It is therefore necessary to establish criteria for judging the relative importance of candidate skills and therefore which are likely to be selected.

While substantial similarities have been identified among the conceptions of employability skills in the countries that have been reviewed in this study, some important differences are also evident. The Mayer Committee had information on the core skills (M. Levy, 1987; National Curriculum Council, 1990) developments in the United Kingdom and the work being done on the SCANS proposal in the United States (The Secretary's Commission on Achieving Necessary Skills, 1991). The Mayer Committee decided to limit the scope of its Key Competencies and precluded values and attitudes (Mayer Committee, 1992, p. 13). By the time the Employability Skills Framework was developed, its promoters had access to extensive work that had been in progress in many countries for more than a decade (ACCI & BCA 2002). It is useful to examine the approaches taken to the definition and selection of generic skills in schemes that were particularly influential on decision making in Australia.

Defining and Validating SCANS Competencies

The SCANS committee in the United States addressed the concern that school leavers who did not go on to tertiary education were being poorly prepared for the workforce. It commissioned two reports on methods for defining and selecting generic skills (Kane, Berryman, Goslin, & Meltzer, 1990; Wise, Chia, & Rudner, 1990). Wise et al. (1990) reviewed previous approaches to defining job-related skills, knowledge and competencies. These approaches included a task analysis approach, a required activities inventory, and the personal characteristics of the worker. Kane et al. (1990) noted that these processes led to the specification of job tasks that depended on the context of the workplace in which they were observed and that were not relevant to skills development in schools. The focus on personal attributes, including intelligence and reaction time tests, were too general and did not address the sorts of skills that were needed in the workplace. The SCANS committee chose to define skills in terms of individual attributes but defined them as general skills and dispositions. A survey of employers was conducted to find the skills and attributes that were required of workers and it yielded a set of 64 characteristics of individuals that were clustered into 10 categories. These became the basis for the SCANS competencies.

A validation exercise was undertaken on the SCANS competencies (Nash & Korte, 1997). The National Job Analysis Study examined the behaviours that were apparent in the tasks that people undertook as part of their jobs. Phase 1 involved the collection of descriptions of job behaviours in order to ascertain those that were common across many jobs. By contrast, the traditional approach to job descriptions was to identify tasks that differentiated jobs. Job descriptions were taken from the Dictionary of Occupational Titles (DOT) (US Department of Labor, 1986). A rating scale of activities was established in which trained observers of job performance were asked to indicate the level of importance and the frequency of job tasks in the DOT listed occupations, both on seven-point scales. An overall index of criticality was obtained from the product of the two ratings. The most important generic skills were crossed with activities undertaken in high performance workplaces. High performance workplaces were characterised as having: a quality focus, customer satisfaction, flexibility, leadership, continuous learning, information sharing, profitability, and productivity.

DeSeCo Key Competencies

In the OECD sponsored Definition and Selection of Competencies (DeSeCo) project, a strongly disciplinary approach was taken to the definition and selection of competencies. In the first phase of the project, the concept of competence was analysed (Weinert, 1999).[14] Following this activity, representatives from six disciplines, psychology, anthropology, politics, philosophy, sociology and economics, were asked to develop a list of 'key competencies' (Canto-Sperber & Dupuy, 1999; Fratczak-Rudnicka & Torney-Purta, 2001; Goody, 1999; Haste, 1999; F. Levy & Murnane, 1999; Perrenoud, 1999). Adopting an anthropological perspective, Goody concluded that it was not possible to develop a list of generic competencies that would apply across societies as such skills were culturally determined. The remaining disciplinary representatives did provide lists of skills

[14] Conceptions of competence and related constructs are discussed in Chapter 3.

with definitions from their perspectives. The lists, however, were quite diverse. In attempting to draw the disparate perspectives together, Rychen and Salganik (2001) were obliged to generate three rather abstract categories, namely acting autonomously, using tools interactively, and joining and functioning in socially heterogeneous groups. Each of these very broad categories was defined further in terms of lower level competencies, but these did not correspond closely to those that had been proposed by the disciplinary experts. The first of these, acting autonomously, included ten sub-skills, many of which could be construed as components of problem solving and of project management. In short, within the DeSeCo project it was not possible to generate interdisciplinary agreement about what skills were important and how they should be defined.

The Employability Skills Framework

The methods used in constructing the Employability Skills Framework involved a review of literature on generic skills schemes and the development of a potential framework based on an analysis of options and constraints (Curtis & McKenzie, 2002). This was followed by consultations with small and medium sized enterprises and employer bodies (McLeish, 2002) and with large enterprises (Field, 2001). The consultations involved focus groups and individual interviews using the options that emerged from the literature review as a basis, but seeking original input from participants. Finally, a panel senior staff from the sponsoring organisations (ACCI and BCA) and the Department of Education, Science and Training considered the information that arose from the literature review and consultations and it developed the final framework. That framework was influenced strongly by the consultations. The issue of establishing criteria for accepting a skill as generic did not arise. The definition of individual skills was managed by specifying facets. Although the final list of key skills was limited to a set of eight, there were over 30 facets and they included many generic skills that were specified in other schemes.

None of the schemes discussed above attended to the specification of criteria for a skill to be considered as generic, to the specific definition of skills endorsed as being generic, and for limiting the set of skills to a manageable number (selection). The Mayer Committee did the first of these tasks well; the DeSeCo project attempted the second and, within disciplinary boundaries, did so effectively.

Summary of definitional issues

In the present study, the second of these activities – the conceptual definition of endorsed generic skills – is regarded as the most problematic. The point is made below that credible assessment of these skills is a critical factor in their success. Assessment, however, is not valid unless there is clarity about what is being assessed, and this clarity can arise only from a conceptually coherent definition. These are not the only issues to be resolved. How generic skills achievement is reported is also important, but reporting depends on having assessment strategies that are capable of supporting the form of reporting that is desired. Once assessment strategies are established, methods for teaching and acquiring generic skills can be discussed.

Assessment, reporting and certification

Assessment appears to be a critical issue in the success or failure of the implementation of generic skills schemes. It requires substantial attention, and the argument is developed below that due attention has not be paid to this matter in any of the generic skills schemes that have been considered in Australia. The issues around assessment are discussed in some detail in a separate chapter (See Chapter 4).

The Mayer Report (1992, pp. 41-56, Assessing and Reporting Achievement of the Key Competencies) dealt extensively with both assessment and reporting issues. It recommended nationally consistent assessment and reporting of individual achievement of the Key Competencies (p. 42). It then moved on to reporting issues and recommended reporting at the individual level through a 'record of performance' using a common format (p. 51), although it said nothing about the format the assessment would take. The Committee also recommended reporting at an aggregate national level that was to be based upon statistical sampling of individual records of achievement and that the performance of equity and target groups should be a specific focus of this approach (p. 55). The suggestion that individual records were to be aggregated would require that the individual assessments must be graded against a common basis. Without a common basis, marks would not be comparable between students and it would make no sense to aggregate the marks. The Committee did recommend that assessment should occur on at least two occasions at each level for each key competency (p. 49). This, they argued was necessary, because the skills were generic and that assessment in one context only might reflect domain-specific knowledge rather than the application of the generic skill. The Mayer Committee would appear to have considered assessment issues in some depth, but it made no recommendations about particular assessment strategies.

The Mayer Committee revealed an awareness of possible roles and purposes of assessment, although it did not make specific recommendations about assessment models. Individual and aggregated reporting have distinct purposes that may require different assessment approaches. Appropriate assessment at the individual level may lead to enhanced individual learning, in part by signalling that what is being assessed is regarded as important enough to assess. It may also reveal individual performance and this achievement may be useful information to individuals to indicate areas of strength and weakness and can lead to feedback and further development. The Committee recognised a formative role for assessment. A summative role is recognisable as potential employers might wish to use the common-format individual reports in selection procedures. Aggregation of individual achievement could be used at the system level to monitor performance.

The Committee recommended that three performance levels should be recognised for each key competency (p. 18). The language used to describe the levels was very general and referred to self management of the application of the skill, the complexity of the task context and individual judgment of the outcome. This might presage a role for self assessment, but this matter was not discussed in the report.

Assessment and SCANS

The SCANS report (The Secretary's Commission on Achieving Necessary Skills, 1991) identified five levels of proficiency for the SCANS competencies. They are shown in Table 9. No basis was articulated for the specification of these levels.

Table 9: SCANS proficiency levels

Level	Example
Preparatory	scheduling oneself
Work-ready	scheduling small work team
Intermediate	scheduling a production line or substantial construction project
Advanced	developing roll-out schedule for new product or production plant
Specialist	develop algorithm for scheduling airline

The original SCANS Report (The Secretary's Commission on Achieving Necessary Skills, 1991) did not make specific recommendations on assessment. In outlining the further work that remained to be done beyond defining ‘workplace know-how’, it did offer numerous general comments on the matter. It suggested that assessment should occur at Years 4, 8 and 12 and that the defined proficiency levels should be applied to assessments from Year 8. The Year 8 assessment was designed to provide information to students on their current levels of proficiency and, therefore, what they would need to do to meet a ‘work-ready’ level by Year 12. They did suggest ongoing assessment through practice, a version of formative assessment, although they did not use this term (p. 17). They also asserted that the assessment system to be developed needed to be fair to all students and provide authentic practice in the application of these skills (p. 25).

The SCANS proposal was not the only generic skills scheme within the United States. Another was the Work Keys system. It had a similar origin to the SCANS competencies, being based on analyses of occupational tasks and worker characteristics. An attempt was made to assess Work Keys (McLarty & Vansickle, 1997). American College Testing developed performance scales for each of the eight Work Keys skills (Reading for information; Applied Mathematics; Listening; Writing; Locating Information; Teamwork; and Observation). Most of these were assessed using a multiple-choice test format, but listening and writing were assessed using a constructed response format and scoring rubrics. Five performance levels were recognised for each scale, the lowest level being regarded as the work-ready level.

Troper and Smith (1997) reported on the use of portfolios for assessing workforce skills. They found that portfolios were not reliable and their assessment could be very time consuming.

Assessment and DeSeCo

In addition to defining and selecting generic competencies (as shown in the preceding section), the DeSeCo project was expected to “develop reference points for the development and understanding of future indicators of competencies and for the validation of education indicators, and to provide a basis for more accurate and appropriate interpretation of empirical results” (OECD & SFSO, 2000, p.2). Salganik (2002) reflected on the progress made in the theoretical work of the DeSeCo project and referred to some success that had occurred in related projects, including the International Adult Literacy Survey (IALS) and in preparatory work for the Adult Literacy and Life Skills (ALLS) survey. However, the DeSeCo project did not establish indicators of the generic skills that it recommended, and as discussed above, the definitions of the key competencies that it did generate were too broad and too abstract to form a basis for the development of indicators. Without such indicators, there was no basis for measuring individual achievement of the competencies.

Assessment of Key Skills in the United Kingdom

A subset of the Key Skills was targeted for assessment leading to the award of a Key Skills Certificate for 16–19-year-olds in England and Wales. The targeted skills were communication, application of number and information technology skills—the basic key skills. Those skills not covered by this qualification were re-labelled ‘wider key skills’. Initially, all 16–19-year-olds in any form of education and training were expected to gain this qualification. The qualification was voluntary, although there were perceptions that it was compulsory and this led to an initial uptake that subsequently waned (Hodgson & Spours, 2000). By 2002, few students who enrolled in academic qualifications, mainly at selective schools, participated in the key skills assessment – presumably because young people enrolled in academic tracks were regarded as competent in these skills and these basic skills were being developed through the academic programs. Uptake of the program was greater in colleges offering vocational qualifications (Powell et al., 2003), and this differentiation would appear to have led to the qualification being seen as remedial rather than as an affirmation of employment-related skills.

Hodgson and Spours (2000, p. 20) summarised the problems of the Key Skills Certificate as having “too many complications, too little currency and [being] too difficult to achieve.” Turner (2002, pp. 15-16), similarly, noted “the unsatisfactory experience of both educators and employers regarding the assessment of the three basic key skills. Testing procedures became too complicated and there is real concern that [the qualification had become] assessment and not learning dominated.” The central problem would appear to have been the complexity of the moderation processes of tests of the basic key skills and portfolio evidence of these skills leading to the award of the qualification. Employers regarded the basic key skills as of lower importance than the wider key skills that were not assessed and were hesitant “to place great store on a ‘certificate’ or even a portfolio of evidence as proof of ‘having these skills’” (Turner, 2002, p. 17). In order to address the lack of breadth, Hodgson and Spours (2000, p. 30) suggested that the wider key skills should become the focus of the qualification.

In summary, the Key Skills qualification was perceived to be too narrow and a test of basic skills only. This added little to what employers and higher education providers would know about candidates based on other school achievement information. The qualification was complex, in that it required an externally set test and a portfolio of activities that was assessed by the school and the two sources of information then had to be moderated. Finally, neither employers nor further education providers placed much weight on the qualification, so there was little point in students doing it or schools offering it. The complexity of the assessment and moderation process is a feasibility problem but the focus on basic skills and the lack of attention to the 'wider key skills' reveals a validity problem with the assessment.

Assessment issues

For generic skills schemes to be implemented successfully, several aspects of their assessment must be established.

Given the range of purposes that have been identified for employability skills assessment, it seems that several approaches to assessment may be required. The main characteristics of assessment approaches are that, collectively, they need to provide:

- a mechanism for communicating the scope of generic employability skills to learners, teachers and employers;
- a means of providing feedback to learners on their acquisition of employability skills;
- a rich source of information about individual achievement, with supportive evidence;
- a summary of the performance of individuals that is readily apparent to employers; and
- a cost-effective means of collecting performance information, individually and at aggregate levels.

Once the priorities of these purposes are established, it may be possible to develop assessment methods (and instruments) to achieve them. However, it is clear that some assessment methods are of limited validity, some are of limited reliability, others are infeasible. Options for assessing generic skills need to be understood in the context of what is known about assessment practices in general. This matter is discussed in some detail in Chapter 4.

Reporting achievement at the individual level is one clear purpose of most generic skills schemes, although the DeSeCo project was part of an international program designed to develop indicators at national education systems levels. This requires assessment at the individual level. Where the proficiency of individuals is reported for high stakes purposes, for example for selection purposes, the assessment methods must support fine-grained judgments. Various methods have been used for reporting generic skills achievement. The key competencies in Australia were to be reported using a standard 'record of achievement.' The intention of the SCANS Committee in the United States was that workplace competencies would be reported as part of an end-of-schooling certificate along with achievement in curriculum subjects. Neither the SCANS workplace competencies nor the key competencies in Australia were implemented. In the United Kingdom, a Key Skills Certificate was developed for England and Wales, although this appears to have failed because of the complexity and cost of the moderation required. It appears that reporting and certification failed because the underlying assessment models were either not adequately specified and were not implemented, or were too complex to be feasible.

Summary of generic skills developments

A number of issues remain to be resolved if a generic skills scheme is to be implemented in Australia. Before effort is expended on this venture, it seems wise to ensure that the conception of employability skills is sufficiently broad to be compatible with international developments and to meet the range of needs of individuals and employers. However, it must also be clearly focused so that a coherent construct, comprising the most important elements of what it means for a skill to be generic, is being assessed. Thus some further work on the definition and selection of generic skills is warranted. This is undertaken both for generic skills as a class of constructs and for problem solving as an example of a particular generic skill in the next chapter.

While considerable effort has been expended on developing definitions of generic skills, there is less evidence of successful assessment models. A variety of assessment approaches have been tried. They need to be grounded in the existing literature on assessment. This is done in Chapter 4.

Chapter 3: Conceptions of Generic Skills and Models of Problem Solving

In Chapter 2, several issues are identified that have prevented the widespread implementation of generic skills in Australia. Specifically, inadequate definitions of generic skills had been developed and inadequate, and even no attention, had been paid to the assessment of these skills. Issues of assessment are discussed in Chapter 4. In Chapter 4, one of the assessment approaches that is canvassed calls for construct-based assessment. That is, assessment needs to be based on standards that are defined in terms of the construct that is the target of assessment. This requires that a coherent conception of the target construct needs to be developed and it is done in this chapter by seeking to define generic skills as a class of psychological constructs. This conception then informs the selection of one of several psychological models of problem solving, which, in turn, contributes to construct definition.

This chapter, therefore, is presented in two main sections. In the first, an attempt is made to address the problem, identified in Chapter 2, of inadequate definitions of generic skills as a class of constructs. This is done through explorations of two major foci of educational-psychological research, namely generic skills as (a) manifestations of intelligence and (b) generic skills as reflective of emergent competence.

In the second major section of the chapter, an attempt is made to arrive at a conception of problem solving that overcomes the definitional deficiencies identified in Chapter 2 and that prepares the groundwork for construct-based assessment. Two main alternative conceptions of problem solving are explored, namely information processing and situative models.

Generic skills as psychological constructs

Three approaches can be identified in the delineation of generic skills. First, skills have been identified by employer organisations through interviews with and focus groups of employer representatives and reviews of other schemes (e.g., Field, 2001; McLeish, 2002). Second, skills have been identified through analyses of the skills enacted by practitioners in workplaces (e.g., Esposto & Meagher, 2007;

Wise et al., 1990). Third, a discipline-based approach has been taken in the DeSeCo Project in which academics from six discipline groups were commissioned to propose lists of generic skills (Rychen & Salganik, 2000). The discipline-based approach appeals because each discipline offers the possibility of providing a coherent theoretically-based construct for each generic skill, and thereby may contribute to a framework for its assessment.

Perhaps not surprisingly, each discipline based group in the DeSeCo project developed a distinct set of generic competencies. For example Haste (1999), writing from the perspective of social psychology, identified technological competence; dealing with ambiguity and diversity; finding and sustaining community links; management of motivation, emotion and desire; and agency and responsibility. Two economists, Levy and Murnane (1999), suggested basic reading and mathematics skills; communicating orally and in writing; the ability to work in groups; the ability to relate well to other people; and familiarity with computers. While there are similarities between these sets of skills, there are also differences. Haste's suggested skills were rather abstract: technological competence included the ability to read, write and do calculations using technologies such as pen-and-paper or a computer. In comparison, Levy and Murnane's suggestions were more concrete, identifying basic reading and mathematical skills. It would appear that the DeSeCo contributors were operating independently and without an agreed metaframework.

It is useful, therefore, to compare generic skills as a class of concepts with some major psychological constructs, specifically conceptions of (a) intelligence and (b) competence and some related concepts.

The problem of definition

A review of terms used to describe generic skills is presented in Chapter 2 and the point is made there that, in most generic skills schemes, no satisfactory definition of 'generic skills' was provided. The terms used to label generic skills in various schemes included skills, competences, competencies, abilities, capabilities and attributes. It is necessary to accept a term and a definition of it in order to establish a sound basis for the construct that is to be assessed.

Terms such as attributes and qualities are rejected. They imply a fixed trait of individuals. Reber (1985) defined attributes as being "relatively constant." There are, no doubt, some relatively fixed traits. For example, certain personality characteristics are taken to be fixed or amenable to change only with considerable effort. The generic skills that are of interest to employers may be relatively fixed traits. If they are, efforts to enhance these skills through instruction or experience are likely to be unsuccessful. It is productive in the interim to operate as though they are amenable to development, and to test empirically the extent to which they can be enhanced through interventions.

The terms ability and capability convey similar meanings (New Shorter Oxford English Dictionary 1997). Ability is the "…power …to perform a particular feat at a particular time. The essence of the term is that the person can perform the task now, no further training is needed" (Reber, 1985). Sternberg (1998) used the terms intelligence and ability interchangeably, and Reber, described intelligence tests as ability tests. The similarity in the meanings ascribed to ability and intelligence creates the possibility that the vast body of work undertaken in the study and measurement of intelligence can be applied to generic skills. This possibility is considered below. (See Generic skills as manifestations of intelligence, page 88).

A third group of terms is used for generic skills, namely competence and competency. Eraut (1994, pp. 170-180) used the term competence to describe occupation-specific skill (that others call competency). Others have described them as generally applicable skills, while Weinert (1999), among others, raised the competence-performance conception (see below). Clanchy and Ballard (1995) drew a distinction between the terms competence and competency. In the Australian context, and especially in the vocational education and training (VET) sector, a competency is narrowly defined as an ability to perform a specific task under specified conditions (National Quality Council, 2007, p. 30). Given that a major reason for promoting generic skills is to prepare individuals for changing work and social environments, such a narrow construct is not appropriate. It may be useful, therefore, to avoid the term competency and to focus on the notion of competence. Apart from legal and clinical definitions, the New Shorter Oxford English Dictionary (1997) did not distinguish between competence and competency, although the Macquarie Dictionary (1981) did refer to linguistic competence – the ability of native speakers to generate utterances that they have not previously heard. This and related definitions are discussed below. The Oxford Dictionary of English (Soanes & Stevenson, 2005) defined competence as "an ability to do something successfully or efficiently" and noted that it was a mass noun – having only a singular form. In the discussion that follows, the plural form is used to denote that, under various generic skills schemes, many 'key competences' are proposed. An individual may be more of less competent – that is display more or less competence, but in proposing components that might be included in instructional programs it is useful to identify potential components separately and to refer to them as competences.

Weinert (1999) explored the concept of competence[15] in detail. While drawing attention to the confusion created by common uses of the term, he provided nine alternative technical definitions of it. These were: "(a) general cognitive ability, (b) specialized cognitive skills, (c) competence-performance model, (d) modified competence-performance model, (e) motivated action tendencies, (f) objective and subjective self-concepts, (g) action competence, (h) key competencies, and (i) meta-competencies" (Weinert, 1999, p. 6). Some, but not all, of these conceptions of competence are examined.

The first way in which Weinert defined the term was to relate it to "psychometric approaches" to intelligence that he said, citing Carroll (1993), reflected "a system of more or less content- and context-free aptitudes and abilities" (Weinert, 1999, p. 6). This conception of competence is close to that implicit in the use of the terms ability and capability as discussed briefly above and in more detail below.

Weinert's second perspective, specialised cognitive competence, reflected the cognitive structures that underlie expert performance. He cited, among other specialised abilities, the playing of chess. In specialised domains, expert performance derives from well structured mental representations. Weinert referred to the form of ability underlying such expert performance as "learned competence" (1999, p. 7). While the development of this form of specialised expertise may be an ultimate goal of much education and training, it is thought to develop through experience long after formal education has been completed. It is therefore not a perspective that can inform general education that seeks to develop generalised competent performance, upon which basis expertise may later develop.

The 'competence-performance' model (Weinert, 1999, p. 7) is based on the notion of linguistic competence developed by Chomsky (see also Eraut, 1994, 1998). This model is interesting in a discussion of generic skills. If the acquisition and deployment of generic skills were consistent with the model of linguistic performance, they would be very powerful skills. It would appear, though, that linguistic ability is innate in that individuals appear to be born with a predisposition for the acquisition of language grammars that they are then able to deploy to generate novel utterances. The observation that linguistic ability is at

[15] Weinert (1999) used the term 'competencies' as the plural form of competence almost exclusively in his discussion: the term 'competences' occurred in only three places in his text. In this chapter, 'competences' is used as the plural to avoid confusion with the narrow concept of competency that is evident in much of the Australian literature.

least partially brain-localised (Sternberg, 1995, pp. 89-93) suggests that linguistic competence is a special case arising from a very strong evolutionary advantage to communicative ability and that this model of competence is unlikely to be a candidate model for a conception of generic skills. Chiappe and MacDonald (2005, p. 26) suggested that language learning, which depends upon an innate domain-specific ‘learning module,’ “is more the exception than the rule in human learning.” However, Chiappe and MacDonald developed the argument further (pp. 27-29), claiming that, in addition to specific learning propensities like that for language, humans must have domain-general learning modules than enable them to recognise stimuli and their contingencies in the environment. Individuals then recognise and focus attention on those stimuli that have important contingencies and therefore focus their learning on specific aspects of the environment. Thus domain-general learning tendencies become attuned to the environment and become domain-specific.

A second aspect of the competence-performance model is its distinction between competence and performance. In the case of linguistic performance, individuals may be capable of generating utterances that conform to the grammar – evidence they have the competence, but often do not produce grammatically correct expressions – that is they do not reveal the competence through performance. Their capability may produce proficient performance but only under certain conditions. Similarly, individuals may be capable of a high level of performance of generic skills, but this may occur only under ideal conditions including those in which individuals are motivated to perform at a high level.

Weinert (1999) described several modifications of the competence-performance model. In these models, competence was seen as having a rule-based conceptual component (as required for linguistic competence) but this component was moderated by general procedural competence, which required access to knowledge about how to operate within domains, and performance competence, which included the metacognitive ability to evaluate the situation and to select appropriate solution strategies. These components were reminiscent of Sternberg’s model of practical intelligence, which is discussed below.

The conception of competence as involving meta-competences invoked a large body of research on metacognition that began with Flavell who defined metacognition as "... one's knowledge concerning one's cognitive processes and products ... [and] refers to the active monitoring and consequential regulation of these processes" (Flavell, 1976, p. 232). Metacognition involves individuals knowing what content-relevant knowledge and skills they already have and what their dispositions and aptitudes are. It also involves an ability to select appropriate strategies for approaching a task, for allocating cognitive resources, and for monitoring, regulating and evaluating performance.

Weinert (1999) described several other conceptions of competence, namely motivational-action, subjective and action competences. These involved aspects of motivation and volition, individuals' beliefs about their abilities and their abilities to contribute to groups. It is acknowledged that such concepts are important in any consideration of individual performance. Dweck and others (Dweck, 1986; Dweck & Henderson, 1989; Elliott & Dweck, 1988; Licht & Dweck, 1984), have demonstrated that motivational dispositions were important influences on performance. However, these conceptions of competence are not discussed further at this point.

Among the various conceptions of competence that Weinert (1999) described, he did not include socially situated models of cognition specifically, although his description of action competence (1999, p. 10) did include some key elements of this conception. This construction of ability was developed by Lave and Wenger (Lave, 1988; Lave & Wenger, 1991). It is not discussed further here, but it is raised below as one potentially useful model for describing problem solving learning and performance (see Situated cognition, p. 111).

Weinert (1999, pp. 11-12) described key competences (key competencies) as a separate conception of competence. While some generic skills have been labelled 'key competencies' (Australian Education Council, 1992; Salganik, Rychen, Moser, & Konstant, 1999), the issue being addressed in this chapter is the search in the psychological literature for a substantive meaning for the term competence in order to locate the underpinning concept of key competences. The use of Weinert's conception of 'key competencies' for this purpose would be a self-referential definition of competence. Nonetheless, Weinert (1999, p. 12) made two particular observations about difficulties that might attend attempts to base interventions on conceptions of generic competencies. First, he noted that general rules or strategies made small contributions to problem solutions, whereas specific rules made larger contributions. Second, in debates about the merits of "systematic versus situated cognitions" he observed that general competencies "have virtually no practical utility" for solving practical problems and that domain-specific knowledge arising from experience was required. On these points his judgment would appear to be rather harsh and this matter is discussed below with respect to the emergence of expertise as an explanatory mechanism for cognitive performance, but see also Chi (2000) and Hunt (1995, Chapter 5). Third, he questioned whether and how generic competencies might be developed through instructional programs. The studies reported in this thesis (see Chapters 7 and 8) are designed to contribute to answers to these questions.

The modified competence-performance model, subsuming conceptual, procedural and performance components, and the meta-competences model can be seen as a composite model of competence rather than as alternative conceptions of it. Together, they capture much of the research attention that has been devoted to the study of developing ability.

Weinert (1999, pp. 15-21) described ten problems arising from the various definitions of competence that he had presented in his paper. Several of these problems arose from attempts to distinguish between domain-general and domain-specific abilities. For example, the third problem that Weinert identified was based on contrasting competence as a set of stable and general primary mental abilities with domain-specific performance that could be enhanced through instruction. He returned to this theme in describing the fifth problem in which he commented that there was no satisfactory way of defining the scope of domains, and Weinert considered labels such as domain-specific and domain-general not to be scientifically useful.

None of the psychological conceptions of the underlying cognitive processes that are applied as generic skills provide indications of which particular skills are generic. Eraut (1994, p. 175), citing Boyatzis (1982), listed as key competences: concern with impact, use of oral presentations, diagnostic use of concepts, managing group processes, efficiency orientation, use of socialised power, proactivity, perceptual objectivity, conceptualisation, self-control, self-confidence, stamina and adaptability. Some of these ‘key competencies’ could be classed as communication skills (concern with impact, and the use of oral presentations), others as interpersonal skills (managing group processes and use of socialised power), and others as personality traits (self-control and stamina). Boyatzis’ list was based on task analysis rather than being based on cognitive requirements of activities and was elicited from managers, and this list might not be applicable to other occupations. Three skills that are almost universal in generic skills schemes (see Chapter 2) are communication, teamwork and problem solving. Other skills are common, but not universal. For example, information literacy has been listed in some schemes but not in others. It is common for generic skills schemes to include from six to ten skills, but some list many more. In part, the variation between generic skill schemes might reflect the lack of specificity about the scope of a domain (Weinert, 1999, pp. 17-18).

Through a brief evaluation of the various terms used for generic skills and Weinert's review of competence, two relatively strong potential approaches to an underlying psychological construct for generic skills emerge. First, generic skills as "general cognitive competences" may be manifestations of intelligence, and many detailed representations of intelligence are available in the psychological literature (see below). A second possibility is that generic skills may be a combination of the conceptual, procedural, performance components of the modified competence-performance model and the meta-competences model of competence. Such a combination is necessary to explain how rule induction and expertise emerge from experience. These two possible conceptions of the competence that underlie generic skills performance are discussed below. Weinert also raised important questions about generic skills, in particular whether they had practical value and whether they could be taught.

In the sections that follow, two of the more productive conceptions that Weinert identified, namely the general cognitive ability (psychometric models of intelligence) and the composite conceptual, procedural, performance and meta-competences models are examined.

Generic skills as manifestations of intelligence

Anderson (1995, p. 1) observed that "the goal of cognitive psychology is to understand the nature of human intelligence and how it works." Conceptions of intelligence have occupied a central place in modern psychology since its inception. Modern psychology has been thought to have commenced with Wundt in 1879 (see, J. R. Anderson, 1995, pp. 7-8; Sternberg, 1995, pp. 45-46). Studies of intelligence have been traced to Galton in 1869, and the first intelligence tests were developed at the beginning of the twentieth century (Pressley & McCormick, 1995, see pp. 531-533).

Psychometric models of intelligence

The 'psychometric approach' to intelligence describes a range of models that are based on tests of ability (Neisser et al., 1996, pp. 78-79). Vigorous debates about the structure of intelligence were joined in the first half of the twentieth century. Pressley and McCormick (1995, pp. 532-533) claimed that Binet was the initiator of intelligence testing and that Spearman was the first intelligence theorist. Spearman had argued for a two-factor model comprising a general intelligence factor, labelled 'g', and a number of tasks-specific factors. Others, for example Thurstone, advocated models with various numbers of basic components of intelligence (Carroll, 1993).

In a major review of the literature on intelligence, Carroll (1993) gathered and re-analysed many of the original data sets that had been used to support the various models of intelligence. Using factor analytic methods, he concluded that intelligence could be represented in the three-stratum hierarchical structure shown in Figure 2. At the lowest level (stratum 1) are so-called 'primary mental abilities.' These are clustered into eight groups at the second stratum, with general intelligence ('g') at the top level (stratum 3).

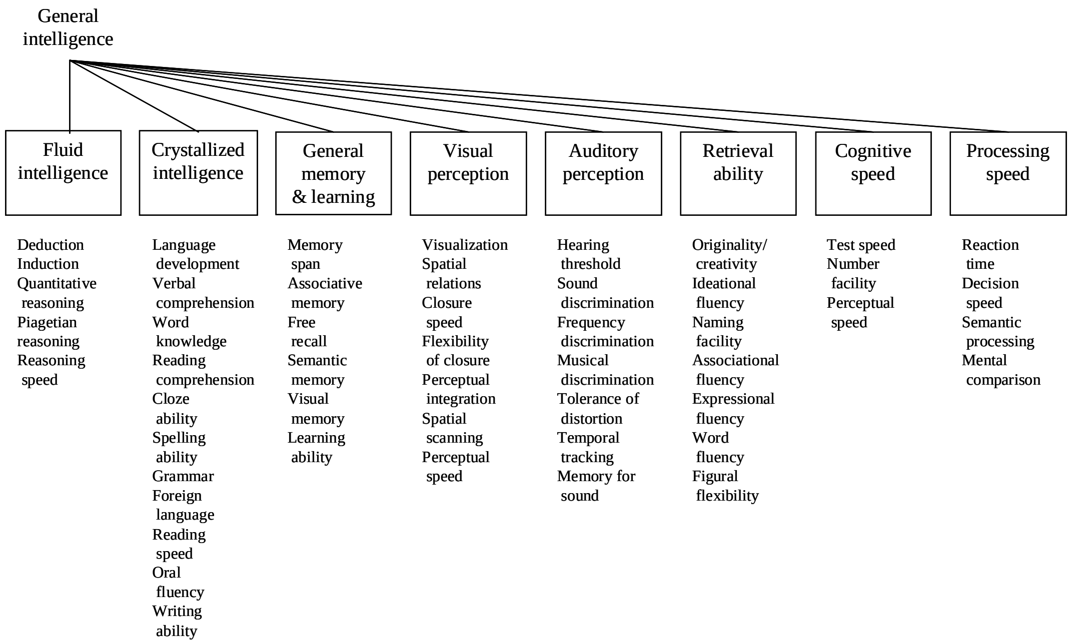


Figure 2: The structure of Carroll's proposed model of intelligence

Gustafsson (1984; 1997) undertook confirmatory factor analyses on data he had collected using three achievement tests and 13 ability tests. He, too, reported that a hierarchical structure was required to represent intelligence adequately, but that in the three-stratum model the loading of Gf on 'g' is unity so the model could be simplified by replacing general intelligence ('g') with fluid intelligence (Gf), a view supported by Jensen (1998, p. 277).

Thus, some agreement would appear to have emerged among psychometric theorists about the structure of intelligence. However, whatever structure is established for intelligence, for this psychometric view of intelligence to be a useful construction of sets of generic skills, it should be possible to map some of the primary mental abilities onto each generic skill and that many, and perhaps most, of them would appear as fundamental processes in an agreed set of generic skills.

The relationship between fluid and crystallised intelligence has been the subject of discussion with respect to the requirements of the workforce. Fluid intelligence is the ability to deal with novelty and to infer relationships, independent of acquired knowledge while crystallised intelligence is the ability to use knowledge and skill acquired through experience (Carroll, 1993). Hunt (1995, Chapter 4) suggested that the ageing of the workforce would be associated with a net increase in crystallised (learned) intelligence, which is dominated by verbal abilities, through the greater participation of older workers, but that there would be a corresponding decline in fluid intelligence, which is dominated by quantitative and reasoning abilities. This argument is premised on the view that fluid intelligence, which includes Piagetian reasoning, develops until about age 25 years and then declines, whereas crystallised intelligence continues to develop into at least middle age (see, e.g., Baltes, Staudinger, & Lindenberger, 1999).

Generic skills may be perceived as general-purpose skills built upon primary mental abilities. Some generic skills, for example communication, would be built upon those primary mental abilities that are acquired through education and training, such as those in the crystallised intelligence group factor. Others, such as solving novel problems, would appear to depend upon abilities in the fluid intelligence cluster (e.g., induction and deduction) and possibly some abilities in the visual perception cluster. In this sense, they may become elements of individuals' crystallised intelligence. The general memory and learning cluster at Stratum II has been the subject of theorising about learning and problem solving performance and this is discussed below in relation to models of problem solving. However, the role of memory as it is represented in theories of intelligence may not relate well to notions of generic abilities. Memory is often tested through recall of words, numbers or symbols and these tests derive from theories of memory such as that proposed by Atkinson and Shiffrin (1968). When learning new material, for example vocabulary in a foreign language, recognition and recall of individual words are important. In the sorts of tasks that are of interest in the application of generic skills, semantic memory performance is more relevant and the processes that may be used to promote encoding and retrieval (Craik & Lockhart, 1972; Stein & Bransford, 1979). Such processes are included in alternative theories of intelligence that are discussed below.

Limitations of the psychometric model of intelligence

The scope of psychometric measures of intelligence

Psychometric measures of intelligence are derived from tests that tap the dimensions of intelligence shown in Figure 2. But, as Neisser et al. (1996) note:

> This [the psychometric] tradition has produced a substantial body of knowledge, though many questions remain unanswered. We know much less about the forms of intelligence that tests do not easily assess: wisdom, creativity, practical knowledge, social skill, and the like. (p. 95)

The most commonly identified generic skills, found in many generic skills schemes (Curtis & McKenzie, 2002, pp. 10-36), are communication, teamwork and problem solving. It would appear likely that the skill of communication depends upon many of the components of crystallised intelligence, specifically verbal comprehension, word knowledge, reading comprehension, Cloze ability, spelling, grammar, reading speed, oral fluency and writing ability. It seems likely also that some elements of cognitive speed and processing speed would be implicated in the skill of communication, including perceptual speed and semantic processing. Similarly, for the skill of problem solving, it would seem likely that some elements of fluid intelligence, namely deduction, induction and reasoning speed would be required and that, for some problems, elements of visual perception including visualisation, spatial relations, perceptual integration, spatial scanning and perceptual speed would be required.

A problem arises, however, when attempting to represent teamwork. Few, if any, of the primary mental abilities identified in Carroll's (1993) model would appear to relate to teamwork or similar conceptions such as interpersonal understanding. Carroll considered this issue in a chapter titled 'Miscellaneous domains of ability and personal characteristics' and suggested that interpersonal skills were not cognitive abilities. He commented:

> Factors discussed in this chapter are not necessarily less important than those dealt with in previous chapters. Some of them concern dimensions of personal characteristics that may not strictly belong under the concept of intelligence or cognitive ability. (Carroll, 1993, p. 542)

The separation of cognitive abilities from the affective and psychomotor reflects Tyler's view that was made prominent in what became known as Bloom's taxonomy (Bloom, Engelhart, Furst, Hill, & Krathwohl, 1956; Krathwohl, Bloom, & Masia, 1964). The cognitive facet of this taxonomy is discussed further in Chapter 4. It may be noted that a complete separation of the cognitive and affective domains has been discarded by many researchers and a prominent role for the affective and conative dimensions of cognitive performance has been recognised (see, e.g., Alexander, 2003b; Bandura, 1997; Dweck, 1986; Dweck & Leggett, 1988).

The decision to omit personal and social skills from conceptions of intelligence has limitations for the skills of communication and problem solving as well as teamwork (see Neisser et al., 1996, especially pp. 77-78 & 95). Communication is an interpersonal activity, whether it is conducted in a face-to-face situation or mediated by a technology such as print. Effective communication requires the originator of a message to be aware of the needs of the intended recipient and, in a face-to-face situation, to monitor the recipient's understanding of the message by observing the recipient's reactions. Similarly, when problem solving is conducted as a group activity, it is desirable that participants share their understanding of the problem, contribute possible solution methods, and share the work involved in the solution. These activities require both communication and interpersonal skills such as negotiation and seeking and offering assistance. Thus, communication and problem solving, as they are described in employability skills schemes (ACCI & BCA, 2002), require some level of interpersonal skill. Further, interpersonal skill is, itself, a valued capability although it is not included in psychometric models of intelligence (Carroll, 1993, p. 88).

The predictive power of psychometric measures of intelligence

For a psychometric model of intelligence to be an acceptable indicator of workplace performance, estimates of it would need to predict performance well. While there is some dispute about its predictive power, it does appear to be limited. Generic skills have been proposed as skills that are necessary for effective performance at work. Thus, for psychometric assessments of intelligence to make a useful contribution to, or even to be proxies for, generic skills substantial predictive power needs to be demonstrated.

Assessments of intelligence and some closely related tests (e.g. the Scholastic Aptitude Test – SAT) have been shown to be good predictors of subsequent academic performance. Pressley and McCormick (1995, p. 523) cited studies in which the correlation between SAT and subsequent course grades was 0.42. Neisser et al. suggested that the correlation between intelligence measures and grade point average was about 0.50 (1996, p. 81). Thus, the predictive power of measures of intelligence for school performance appears to vary from about 17 to 25 per cent.

However, intelligence test scores do not predict workplace performance nearly as well. McClelland (1973) and Spencer and Spencer (1993) showed that intelligence test scores explained between four per cent and 25 per cent of the variance in workplace performance. In part, this could be attributable to the characteristics of intelligence tests, academic learning and workplace requirements. Resnick (1987) identified four key differences between school-learning and workplace learning: school learning is individualistic, abstract, symbolic and generalised; in the workplace it is collaborative, involves tool manipulation, uses contextualised reasoning and develops situation specific competence. Sternberg et al. (2000, pp. 34-35) reported similar differences between academic and practical problems. That is, the content of intelligence tests is more closely related to the demands of school learning than it is to the requirements of the workplace.

As an alternative to intelligence in explaining workplace performance, Spencer and Spencer (1993) proposed six clusters of workplace competence (achievement and action, helping and human service, impact and influence, managerial, cognitive, and personal influence), which, they claimed, included between 80 per cent and 95 per cent of all listed job competencies.

It may be noted that the low correlation between intelligence test scores and workplace performance measures has been disputed. Hunt (1995) and Jensen (1998) claimed that the low correlations between measured intelligence and work performance were substantially an artefact of truncated ranges in the criterion measures used and that the truncation was attributable to testing selected groups. Hunt (1995) argued that people with low ability were unlikely to be included in samples of individuals in the workplaces tested, for example among bank employees making decisions about the credit-worthiness of applicants for loans (Klemp & McClelland, 1986). Jensen (1998, pp. 282-291) used evidence from studies conducted in the armed forces in the United States. He reported that the correlation between measures of 'g' and job selection tests was about 0.4 and that the predictive validity of 'g' for job performance was 0.27. He further argued that the predictive validity of 'g' for job performance increased with increasing cognitive complexity of job tasks.

Jensen also drew attention to the variability in validity coefficients:

> The mean validity coefficient of IQ for educational variables is +.50, but the spread of validity coefficients is considerable, ranging from close to zero up to about .85. (Jensen, 1998, p. 277)

The high variability in these coefficients would appear to suggest that what is measured as general intelligence has differential applicability, even within academic learning contexts. This, in turn, suggests that the psychometric model of intelligence is likely to be of limited value as an underlying mechanism of generic skills. However, other conceptions of intelligence are available and are evaluated below.

Alternative conceptions of intelligence

The lack of any representation of personal or social skills in the psychometric model of intelligence could be overcome through the addition of a component, at

Stratum II of Carroll's model (see Figure 2), reflecting interpersonal skill. This may enhance the suitability of this model of intelligence for generic skills.

Emotional Intelligence

Emotional intelligence has been proposed as an ability (J. D. Mayer, 2001; J. D. Mayer & Salovey, 1993; J. D. Mayer, Salovey, & Caruso, 2000; Salovey, Bedell, Detweiler, & Mayer, 2000). Mayer and Salovey (1993) defended the concept of emotional intelligence against criticisms that it was not an intelligence but rather a set of personality dispositions (see, Carroll, 1993, p. 11). They claimed that emotional intelligence "involves the ability to monitor one's own and others' emotions, to discriminate among them, and to use the information to guide one's thinking and actions" (J. D. Mayer & Salovey, 1993, p. 433). They also asserted that emotional intelligence did include a set of important capabilities. This position was supported by Bar-On (2000) who developed the Emotional Quotient Inventory (EQ-i). He identified ten components of emotional intelligence, namely self-regard, emotional self-awareness, assertiveness, empathy, interpersonal relationship, stress tolerance, impulse control, reality testing, flexibility, and problem solving. While some of these abilities, for example flexibility, would appear to be similar to the primary mental abilities that are shown in Stratum I of Carroll's model, most do not. Impulse control would appear to be a metacognitive activity. Indeed, Mayer's and Salovey's (1993) definition of emotional intelligence – an ability to monitor emotions – would appear to suggest metacognitive activity. The issue of metacognition is discussed in more detail below.

Multiple Intelligences

Gardner (1983; 1993) proposed an alternative to the hierarchical psychometric model of human intelligence. He argued that individuals had seven separate intelligences, namely linguistic, musical, logical-mathematical, spatial, bodily-kinesthetic, and personal (comprising intrapersonal and interpersonal) intelligences. Psychometric models of intelligence include linguistic, logical-mathematical and spatial elements (see Figure 2 above) and they are not discussed further.

In describing intrapersonal intelligence he wrote:

> The core capacity at work here is access to one's own feeling life – one's range of affects and emotions: the capacity instantly to effect discrimination among these feelings and, eventually, to label them, to enmesh them in symbolic codes, to draw upon them as a means of understanding and guiding one's behavior. (Gardner, 1993, p. 239)

Gardner said that interpersonal intelligence was:

> … the ability to notice and make distinctions among other individuals and, in particular, among their moods, temperaments, motivations, and intentions. (Gardner, 1993, p. 239)

Gardner stressed the importance of personal dimensions of intelligence, and clearly disagreed with Carroll's assessment, cited above, that these personal intelligences were not cognitive abilities:

> The capacity to know oneself and to know others is as inalienable a part of the human condition as is the capacity to know objects or sounds, and it deserves to be investigated no less than these other "less charged" forms. (Gardner, 1993, p. 243)

He proposed eight criteria that needed to be satisfied for a capacity to be accepted as a discrete intelligence. Among other criteria, he insisted that these intelligences should be localised in particular areas of the brain, and that this could be demonstrated by examining individuals who had suffered brain injury in which particular capacities were impaired while others remained intact. He also presented evidence of the cases of savants, in which particular capacities were highly developed while other abilities were normal or underdeveloped, to support his contention that these abilities are discrete and do not derive from a notional general intelligence factor.

Gardner's model of multiple intelligences can be criticised on several grounds. Its reliance on exceptional cases (prodigies, savants and brain-injured individuals) may limit its generalisability to the normal population. Allix (2000) was critical of the theory. He argued that Gardner had posited the need for supportive psychometric evidence, but had not provided it. He also argued that there was no need to require specialised brain structures for each intelligence and that a neural network could have produced the same complexity of responses and that such a model should be preferred on the grounds of parsimony. Jensen (1998, Chapter 5) too was critical of the multiple intelligences theory. He argued that four of Gardner's proposed intelligences (linguistic, musical, logical-mathematical and spatial) were highly inter-correlated – violating one of Gardner's criteria. He contended that the bodily-kinaesthetic intelligence was not a mental ability and that neither the intrapersonal nor the interpersonal intelligences had been studied. The latter point must be disputed, given the work by Mayer and Salovey and Bar-On cited above.

Gardner's (1983; 1993) model of multiple intelligences may be attractive to those seeking an underlying mechanism for generic skills. The linguistic, logical-mathematical and spatial intelligences would appear to be closely related to communication and possibly problem solving, at least in some domains. However, this sub-set of Gardner's model of multiple intelligences does not represent an improvement over Carroll's hierarchical model, as it has equivalent components at Stratum II. The inclusion of intrapersonal and interpersonal intelligences in Gardner's model would appear to provide a basis for teamwork. These abilities are explained in Mayer's and Salovey's (1993) model of emotional intelligence. Gardner's other proposed intelligences, musical and bodily-kinaesthetic, do not have equivalent constructs in generic skills schemes. It would appear that, despite its initial attractiveness and because of a lack of supportive evidence and doubts about its generalisability, a model of multiple intelligences does not provide an adequate basis for explaining generic skills.

Practical Intelligence

Sternberg (1985a; 1985b) proposed what he originally called a componential or Triarchic theory of intelligence, which later he developed and labelled a theory of practical intelligence and a theory of successful intelligence (Sternberg, 2003, 2000; Sternberg et al., 2000; Sternberg & Wagner, 1986). The model had three sub-theories, each with a set of component processes. The model is represented in Table 10.

The componential sub-theory of Sternberg's model is similar to other theories of intelligence, such as that proposed by Carroll (see above), in that it posited a number of lower level processes (performance components), for example encoding, inference and mapping. But Sternberg's model differs from others in important ways. First, in Carroll's model, Stratum I processes were recognised as extant processes, for example induction and deduction, whereas in Sternberg's the performance components (e.g. encoding and comparison) had an explanatory function. Second, while some of the processes in the psychometric model identify specific domains of cognitive activity, for example word knowledge and spelling ability, all components in Sternberg's model are domain-general. Third, the meta-components sub-theory also differed from most other theories of intelligence by including high-level (meta-cognitive) components for the management and evaluation of performance and by specifying a set of learning processes (knowledge acquisition components). Sternberg (1983, pp. 3-5) illustrated the application of meta-components and performance components in the solution of an analogy problem. The meta-components were used to recognise the problem, to select a solution method and to allocate attentional resources to it. All the performance components (see Table 10) were invoked in explaining analogical reasoning. Fourth, Sternberg's model included a mechanism for learning through its knowledge-acquisition components. The components did not simply enable the acquisition of domain-specific knowledge but also facilitated domain general learning. Chiappe and MacDonald (2005, pp. 21-22) argued that experience of analogical reasoning tasks developed domain-general mechanisms for problem solving.

Table 10: Sternberg's model of successful intelligence

Componential sub-theory This describes the internal operations of individuals	
Meta-components These processes enable performance to be planned, monitored and evaluated	Recognise that a problem exists Decide the nature of the problem Select processes for solving the problem Select a strategy for approaching the problem Select a representation for the problem Allocate resources to the problem Monitor progress Evaluate effectiveness
Performance components There are many of these processes and each is highly specialised for a narrow range of functions	Encoding Inference Mapping Application Comparison Justification Response
Knowledge acquisition components These processes are used to sort out what is relevant from what is not, then to combine useful information to form a plausible representation, and to relate new information to what is known.	Selective encoding Selective combination Selective comparison
Experiential sub-theory This sub-theory describes how past experience is used to dictate how to deal with completely new tasks in unfamiliar situations, relatively unfamiliar tasks, and completely automated processes.	
Contextual sub-theory This sub-theory seeks to relate what the individual does to the environment in which they operate. The environment may be family, job, sub-culture or culture.	
	Adaptation to the environment Shaping the environment Selecting the environment

Source: After Sternberg et al. (2000, pp. 25-30, 47, 96-97)

Sternberg (1983; 1998) claimed that, under the traditional, psychometric view, intelligence was perceived to be a fixed trait and that this conception of intelligence was a flawed one. A notable exception to the notion of intelligence as a fixed trait can be found in Piaget's theory (Brainerd, 1978). Carroll (1993, p. 11), by including Piagetian development as a component of fluid intelligence, accepted that intelligence had a developmental component. It may be noted that the SOLO taxonomy (Biggs & Collis, 1982) is built upon the developmental ideas of Piaget's theory and it is discussed in Chapter 4. The fixed trait conception was not a necessary consequence of the structure of intelligence as posited in psychometric models. Dweck (1986), without reference to any specific model of intelligence, showed that students who held a fixed trait view of intelligence were less likely to engage deeply and to persist in learning tasks and consequently experienced inferior learning outcomes compared with students who held a malleable view of intelligence.

Under Sternberg's componential model, intelligence was seen as having components that were developed through practice and that what was measured in intelligence tests was emergent expertise (Sternberg, 1998). The conception of practical intelligence invokes the development of tacit knowledge as a component of practical intelligence, and this emerges through experience in a domain (Wagner & Sternberg, 1986). The second major element of practical intelligence, the experiential sub-theory, was intended to explain the emergence of expertise within a domain.

The third major element of Sternberg's model of successful intelligence, the contextual sub-theory, sought to explain interactions between individuals and their environments. This element is considered in relation to situative models of problem solving in the next chapter.

Sternberg's model of successful intelligence raised issues that need to be considered in understanding student performance, especially of generic skills. First, he used the term intelligence and ability interchangeably (Sternberg, 1998), an issue discussed above. Second, Sternberg claimed that what was measured as intelligence was a state in the process of emerging expertise and not a fixed trait (Sternberg, 1998; Sternberg et al., 2000, p. 1). This point was also made by Brown specifically in relation to generic skills (A. Brown, 1998, p. 168). This claim suggests the need to examine the process by which novices transform themselves into experts within domains. If generic skills are perceived to be manifestations of a fixed trait, measuring them may be useful in personnel selection situations, but it would imply that there is little point in trying to develop these desired skills. Third, Sternberg introduced meta-cognitive processes into conceptions of intelligence, and thereby substantially extended the intelligence construct. Many of the processes in his model are metacognitive ones, including allocation of resources, monitoring progress, evaluating effectiveness. However, metacognition was not a new concept. Flavell (1976) had introduced it. But Sternberg incorporated it formally into a theory of intelligence and therefore invoked it as one of a set of processes that are involved in the emergence of expertise.

Limitations of intelligence as a defining concept for generic skills

The psychometric model of intelligence is at best moderately predictive of workplace performance. Analyses of effective and superior job performance showed that six clusters of general competences were implicated in superior job performance (Spencer & Spencer, 1993). It can be noted that these competences have been identified in workers who exhibit superior performance, but measures of these general competences have not been compared with later job performance and therefore their predictive power is unknown (Weinert, 1999, p. 16).

Psychometric models of intelligence do not include personal or social abilities, but these abilities were identified in many generic skills schemes, usually by reference to teamwork, and in everyday conceptions of intelligent behaviour (Ford, 1986; Klemp & McClelland, 1986). The failure to include social or personal abilities in psychometric models of intelligence could be overcome by adding one of several theoretical models that specifically include this dimension, namely emotional intelligence (Bar-On, 2000; J. D. Mayer & Salovey, 1993) or multiple intelligences (Gardner, 1983, 1993).

However, the psychometric model of intelligence, with possible extensions to include a social dimension, does not include any component that would explain how individuals' abilities could be developed. Indeed, the original conceptions of intelligence held that it was a relatively fixed trait of individuals and therefore not amenable to intervention. Sternberg's componential model of intelligence does provide a mechanism by which practical intelligence can be enhanced, but it leads to questions about the nature of the construct and the terminology used to describe it (intelligence, ability or competence). Sternberg's model does include a role for metacognition, and posits a mechanism by which expertise can emerge through experience.

Generic skills as competences

In Weinert's (1999) review, he provided nine technical definitions of competence. The first of them, which he called general cognitive competencies, was substantially synonymous with the psychometric model of intelligence and this is discussed in the preceding section. Among the other technical conceptions of competence, he described a modified competence-performance model, which subsumed conceptual, procedural and performance aspects of competence, and he described meta-competences.

The elements of the modified competence-performance model feature in Sternberg's model of successful intelligence (see Table 10), and they have been invoked in many descriptions of performance. Under this conception, competence is seen as being based in a set of general learning processes that are applied and customised in particular domains. In research on expertise domains are defined in terms of content. But Weinert (1999, pp. 17-18) observed that domain-general and domain-specific notions of competence were not differentiated in "scientific treatments of competence." The problem lies first in the scope of domains and then in whether competence is specific or general in its application. Domains may be defined in terms of content, for example, chess or physics, but even within content areas, the scope of a domain is unclear. A domain could cover all of physics, or be conceived more narrowly as covering problems involving the conservation of momentum or perhaps restricted to problems involving inclined planes. Alternatively, domains could be defined as sets of processes that are deployed in higher level skills such as communication, teamwork or problem solving – constructs that Weinert referred to as "transcurricular competencies," which, he noted, were distinguished from curricular (content-based) competencies on pragmatic rather than theoretical grounds (1999, p. 18). The scope and defining characteristics of domains is an important epistemological problem. It would appear not to have been addressed, despite many references to domain-specific and domain-general knowledge and skills. However, this problem is not discussed further in this study. On the assumption that domains are bounded, either by content or by cognitive process, and possibly by both, attention is focused on the relationship between what are labelled commonly as domain-general and domain-specific processes.

The history of cognitive psychology has witnessed fluctuations in the dominance of domain-general and domain-specific processes. Chi (2000, pp. 161-164) noted that the "dream of psychologists and educators has always been to identify skills or strategies that can be used across domains. The pursuit of domain-general strategies largely characterized the literature in the seventies" (p. 161). The focus on generic skills may be seen as an expansion of the "psychologists' dream" into the wider community. During the 1970s and 1980s, there was considerable focus on domain-general strategies, possibly initiated through the classic work of Newell and Simon (1972) in which a computer analogy was used as a template for human problem solving. Computer programs of that era were designed to read data from various sources, to apply a common programmed algorithm and to produce output reflective of the data that were input. Newell and Simon, using the computer as a metaphor for human problem solving, described "the general problem solver" as a system of general-purpose procedures that would respond to problem situations by enacting general purpose algorithms or heuristics to reach a solution. The general-purpose procedures were various strategies, such as means-ends analysis, analogy and difference reduction. Domain-general and domain-specific strategy instruction was undertaken and shown to be effective in many areas, including reading (Paris, Wixson, & Palincsar, 1986), writing (Bereiter &

Scardamalia, 1987) and mathematics (Charles & Lester, 1984; Polya, 1957). In addition to research in these curriculum areas, other research into strategy use in physics and with the use of puzzles is cited below in the discussion of information processing models of problem solving. During this phase of research, attention was focused on novices and how their performances might be improved.

The dominance of strategy use as a mechanism for improving performance was challenged following comparisons of novices and experts in various domains. Experts were characterised as having extensive knowledge that was semantically organised into cognitive structures called schemas (or scripts for procedural events). Schemas included situation-action rules. Expert performance depended upon recognition of salient features of situations that enabled a representation of the situation to be formed and an appropriate schema to be activated. Experts encoded and represented problems using deep structures of problems, while novices lacked that extensive and well-organised knowledge and represented problems using their surface structures (Chi, Feltovich, & Glaser, 1981; Larkin, 1985). It is interesting to note that playing chess was one of the complex, but well-structured, domains to which the general problem solver model was applied. But it featured prominently in research on expertise. de Groot (1965, 1966, cited in J. R. Anderson, 1995, pp. 292-294) had shown that chess experts were no more intelligent than novices but that they were able to recall accurately the board positions of many more pieces than novices, provided those positions were meaningful in the context of games; they were no better at recalling random positions. This suggested that their superior performance did not depend on greater memory capacity but on their ability to encode meaningful relationships. It may be noted that this observation suggests that the 'general memory and learning' cluster of abilities in Carroll's model of intelligence (see Figure 2) needs to be qualified. Superior memory may be a characteristic of more intelligent individuals, but that memory is not simply an ability that is applied equally to all situations; it appears to be a latent capacity that is influenced by an individual's experience, either to apply general strategies or to activate relevant schema.

The focus of this phase of research would appear to have been on the characteristics of expert performance, rather than on the processes by which expertise emerged.

The shift in focus from general-purpose strategies to schemas was accompanied by a shift in instruction. In mathematics instruction, Sweller and colleagues argued that experts had well-developed knowledge bases and solved problems by the selection of appropriate schemas, that strategy use was inefficient, and therefore that strategy instruction was misguided. They argued that the focus of instruction should be on methods by which novices might develop schemas similar to those held by experts (Owen & Sweller, 1989; Sweller, 1990; Sweller, Mawer, & Howe, 1982). Sweller and colleagues argued that the provision of worked examples enabled novices to induce schema that they would use subsequently in their problem solving. Sweller and colleagues implicitly addressed the question ‘How can instruction most efficiently move novices along the path from their novice status towards expertise?’

Alexander (2003a, p. 3) expressed the view that previous research had failed to provide satisfactory accounts of the processes that led to the emergence of expertise: "Still, after generations of research, there are relatively few models that seek to explain the processes of expertise development or provide means of predicting which individuals will achieve expertise, and which will not."

The emergence of schema-driven expertise as the dominant explanation for enhanced performance was challenged. Bereiter and Scardamalia (1986) raised an important question: “How do novices, who have limited knowledge, transform themselves into experts with extensive and organised knowledge bases?” Their solution to this question was to posit that novices, with a repertoire of general strategies, were able to acquire knowledge and to structure it and thereby to develop expertise. Borkowski, Carr and Pressley (1987) proposed a composite model of expertise development that involved domain-specific knowledge, specific-strategy knowledge and general-strategy knowledge as the basis for observed performance and learning.

Lawson (1990) countered Sweller’s arguments by presenting evidence of the effectiveness of different types of strategy and by students of different ability levels. Lawson distinguished between what he termed task-orientation, executive and domain-specific strategies. Task-orientation strategies addressed the influence of students’ “affective, attitudinal and attributional expectations” in relation to problem tasks; executive strategies included planning, monitoring and reflection; and domain-specific strategies included “heuristics, such as means-ends analysis” (Lawson, 1990, p. 404). Lawson showed that training in strategy use had been effective in a variety of domains. Lawson went on to show that transfer of strategy use was an important factor in the success or failure of strategy instruction. In particular, Lawson reiterated a point made by Ferguson (1954) that learning was a special case of transfer; that is, learning involved the application of a skill in tasks similar to the instructional context. Elsewhere, Lawson (1991) described other dimensions of transfer that extended beyond the maintenance of knowledge or strategy.

The renewed focus on learning, rather than on the transfer of knowledge or strategy to novel tasks (see, e.g., Gick & Holyoak, 1983), was identified by Chi (2000) in her overview of the shift in focus that had occurred between the 1970s and the late 1990s. She pointed to an initial emphasis on general strategies, followed by one based on the structure of knowledge and representations, and then to one in which learning became a central concern.

Alexander (2003b) argued that tracking learning within a domain was not helped by the novice-expert dichotomy and that a continuum of ability must exist between the end-points (p. 10). She referred to stages of acclimation, competence and expertise on this continuum. She advocated a model of learning that involved both domain-knowledge and strategic processing and suggested that:

> ... the journey toward competence or proficiency requires strategic tools for analyzing and responding to the many problems encountered. Students do not come equipped with the cognitive and metacognitive/self-regulatory strategies they need (Winne, 1995). Such strategies must be acquired and practiced in relevant situations that allow students to witness their inherent value (Schonfeld, 1985). (Alexander, 2003b, p. 12)

Dreyfus and Dreyfus (1984) identified five stages of increasing competence, namely novice, advanced beginner, competent, proficient and expert. They did not indicate whether they regarded these stages as representing discrete (discontinuous) levels of performance or whether they were convenient labels for development along a continuum, as Alexander (2003b) has suggested. Whether there is a continuum or not between novice and expert states is not significant in the current study as this study restricts its focus to the novice to competent range. It does appear likely, however, that expert performance is qualitatively different from novice actions. The literature on expertise has identified expert performance as being relatively automatic and reliant on access to a highly conditioned knowledge base. By definition, novices and even competent individuals do not have this level of knowledge and must rely on alternative methods for solving problems. Resolving the issue of continuous or staged progression from novice to expert status may be addressed by measuring the performances of a sample of individuals spanning the novice-expert range on similarly diverse tasks and discovering whether the performances of individuals are clustered or uniformly distributed over that range. The Saltus model (Wilson, 1989b) might be valuable in such an investigation.

Implications for generic skills

Two related trends in conceptions of cognition have been reported. First, there has been a flux in interest initially in domain-general skill, then in domain-specific knowledge, and most recently in an interaction between domain-general strategies and domain-specific knowledge as explanatory concepts in performance (Alexander, 2003b). Second, there has been an associated shift from explanations of performance to explanations of learning (Chi, 2000). Both trends are significant in conceptualising generic skills as a class of constructs. First, if generic skills are conceived as cognitive entities – that is, if they are thought to exist in the same way as other abilities such as mathematical ability or chess playing ability – then the research into these cognitive domains is relevant to consideration of generic skills. Second, the renewed focus on learning rather than performance (Chi, 2000) would suggest that finding ways of developing generic skills was as important as attempting to measure generic skills performance with precision. An implication of the renewed focus on learning is that assessment may be used to evaluate the extent of that learning, an issue that is discussed in Chapter 4. The measurement and development of one generic skill – problem solving – are discussed in Chapters 7 and 8 respectively.

Generic skills, by definition and intent, cannot be based on domain-specific knowledge, although they may use it as it develops. Interest in them has arisen following the realisation that it is not possible for schools, colleges and universities to teach all the knowledge and skill, even in core curriculum areas, that students require during their working lives and that some ‘transcurricular competences’ (Weinert, 1999) are required.

Generic skills, intelligence and competence

The preceding discussion compares two general approaches to the identification of generic skills as psychological constructs. The first conception tested is of generic skills as manifestations of ability, and psychometric notions of intelligence are evaluated. Psychometric models are found wanting on two grounds. First, they do not provide a sufficiently comprehensive account of the elements that are identified in many generic skills schemes and second, they do not predict workplace performance as well as had been anticipated. While the extent of the explanatory power of intelligence for workplace performance has been disputed, it does appear to be the case that some other factors are implicated in successful workplace performance. Generic skills may help fill this gap.

The second conception, of generic skills as competences, appears to provide a basis for understanding and representing the concepts – generic skills – that employer organisations and a variety of other agencies have promoted. Conceptions of competence are varied. Weinert's (1999) review is particularly helpful and his modified competence-performance model appears to be a productive basis upon which to proceed.

It may be worth noting that Weinert (1999) identified the psychometric model of intelligence as one possible representation of competence. The consistency of the competence-performance model with Sternberg's theory of practical intelligence, which may be seen as an evolutionary stage in the understanding of intelligence, appears to suggest that conceptions of intelligence and competence are not discrete alternatives. That is, intelligence and competence may not exist as a dualism, but rather as a duality. There are, however, important and distinctive implications that arise from choosing one or the other representation. While it is not necessary to perceive intelligence as a fixed trait, the origins of psychometric intelligence testing and theorising tend to locate it as a relatively fixed attribute of individuals. Competence, on the other hand, is perceived to be amenable to instruction. But competence, very likely, depends upon an initial level of ability that is built upon by exploiting affordances of the environment, whether they occur by chance or are created by design and instruction. It is in this sense that the linguistic conception of competence is relevant. It appears that individuals are born with a disposition to acquire language. That disposition is not hard-wired, as are some functions, but plastic. Linguistic ability develops according to the language environment of the individual, with local grammars being acquired and the extent of language acquisition depending upon the linguistic richness of the environment. Chiappe and MacDonald (2005) proposed that humans had generalised dispositions to learn and to solve problems that occurred in their environments. That is, generalised learning and problem solving capacities exist, analogous to linguistic potential, using generalised rather than the specialised brain structures that appear to facilitate language learning. If this is so, it suggests that creating an environment in which problems are presented, and by scaffolding problem solving (and substituting generalised cognitive processes for the wiring that facilitates language acquisition), learning and problem solving might be enhanced.

This chapter reports the results of a search for a definition of generic skills as a class of constructs. It finds that that the competence-performance model espoused by Weinert (1999) is a productive basis for defining this class of constructs. A key feature of this notion of competence is its relationship to learning and to the transition along a continuum from novice to expert status. The development of competence depends upon the acquisition of “cognitive and metacognitive/self-regulation strategies” (Alexander, 2003b), strategies that need to be acquired in context.

A second objective of this chapter is a search for a definition of problem solving – the target generic skill of this research. A definition of problem solving must be compatible with the conception of generic skills that is developed above, including a focus on the acquisition of cognitive and metacognitive strategies. Of course, it must also be compatible with at least one substantial conception of problem solving that is recognised in the literature of that field. It is to this second definitional task that attention now turns.

Conceptions of problem solving

Problem solving is one of the very common generic skills and appears, along with communication and teamwork, in almost all generic skills schemes. One of the points made in Chapter 2 is that specific generic skills are not defined or described in any detail. The purpose of this section is to consider various conceptions of problem solving and to select a model that enables problem solving to be operationalised to the extent that it can be described to teachers and learners and can provide a basis for developing an assessment tool that enables problem-solving performance to be measured. The issue of measurement is discussed in Chapter 6 and the development of a problem solving assessment tool is described in Chapter 7.

Mayer defined a problem in the following terms:

> A problem consists of a given state (i.e., a description of the current situation), a goal state (i.e., a description of the desired situation), and a set of operators (i.e., rules or procedures for moving from one state to another). A problem occurs when a situation is in one state, the problem solver wants it to be in another state, and there are obstacles to a smooth transition from one state to the other. (R. E. Mayer, 1995a, p. 4722)

Mayer’s reference to states located his definition within the information processing view of problem solving (discussed below), but definitions that preceded the emergence of the information processing model were similar. Both Polya (1957) and Duncker (1945, cited in R. E. Mayer, 1995a, 1995b) referred to the need to think about ways of moving from the current situation to a desired one with obstacles blocking an easy transition. That is, the main elements of the definition have been constant, despite substantial changes in theoretical conceptions of problem solving.

Mayer (1989; 1995a) elaborated on his definition of problems by indicating ways in which problems might differ. He referred to well- and ill-defined problems. Many puzzles, for example the Tower of Hanoi, were well defined in that the initial state, the desired state and the rules and operators available were well defined. Mayer gave as an example of an ill-defined problem the task of writing a chapter on problem solving because the goal state and the operators available were not clearly specified. Different authors might take quite different approaches to this task and generate quite different, but acceptable, products. Problems varied in complexity. The Tower of Hanoi task was a simple one, as the given and goal states could be described simply and there were few operators and states between the given and goal states. On the other hand, a chess game was much more complex, because although the initial state could be described fully and the number of available moves at any point was limited, there were many solution states and very many possible intermediate states. In principle, all possible intermediate states could be fully described, but there were so many that they would vastly overload human cognitive capacity, unless they were encoded semantically. Mayer also contrasted routine and novel problems. Many text book mathematical problems were routine in that, even if the exact problem had not been encountered previously – perhaps multiplication of two three-digit numbers, very similar problems would have been seen and the algorithm for solving it would have been applied to those examples. It was the capacity for solving ill-defined, complex and novel problems that was valued and it is this capacity that is the target of the current research.

Models of problem solving

Mayer (1992) identified four major problem solving paradigms, which might be labelled associationist, Gestalt, information processing and situated cognition.

Associationist and Gestalt models

The associationist or behaviourist view of problem solving asserted that when problems were encountered, individuals would rely on identifying elements in the problem that were similar to elements they had experienced and that they would generate conditioned responses to the recognised elements (Palumbo & Vargas, 1988). However, it was difficult, except perhaps in controlled experiments, to say what elements individuals might recognise in problems that would enable them to solve those problems. Indeed, Estes (1972) argued that reinforcement theory, the core concept of behaviourism and the mechanism by which responses were learned, had been developed through laboratory studies and that evidence for its generalisation to complex, real-world problems was lacking. Rescorla (1988) posited that early (pre-1960s) conceptions of classical conditioning had been quite limited and that the central concerns was with learning. He posed three questions: "What are the circumstances that produce learning? What is the content of that learning? How does that learning affect the behavior of the organism?" (pp. 151-152). In his answers to each of these questions, Rescorla proposed a model of learning that involved the formation of representations of the learner's (in almost all cases, a rat's) environment. By invoking a theory of a representation, Rescorla departed from the radical behaviourist tradition, which eschewed explanations that involved other than directly observable events. Rescorla's hypothesis reprised a theory that Tolman (1932, cited in Sternberg, 1995, p. 259) had proposed. He had shown that rats were able learn a maze without reinforcement and that they must have formed a mental map of their environment. The allusion to mental representations of the environment, which occurred as early as 1932 and which continued through the period dominated by radical behaviourism (Boden, 1997, p. 56), provided a basis from which information processing models later developed. Rather than attempting to rework behaviourist models in a search for explanations of problem solving, it seems more productive to examine those paradigms that have substantially replaced it.

The Gestalt tradition evolved early in the twentieth century and sought to account for aspects of human perception. In particular, it developed principles to explain the observation that humans tended to perceive complete and organised structures (the Gestalt) in representations of objects (J. R. Anderson, 1995, pp. 44-47). Even when elements of objects were obscured by other components in the foreground or when objects were represented partially, humans tended to perceive complete objects. The principles that were claimed to underlie this perceptual ability were applied to problem representations, with successful solutions requiring that the problem solver was able to perceive a fully-formed solution. The Gestalt model of problem solving posited that problem solving proceeded in four phases, namely preparation, incubation, illumination and verification (R. E. Mayer, 1992, Chapter 3). Preparation involved information gathering and preliminary attempts to solve the problem; incubation involved thinking and reflection; illumination involved a 'flash' or inspiration accompanied by restructuring; and verification involved reflective processes such as checking the reliability of the solution.

Mayer (1992) cited Polya as an example of a Gestalt theorist. Polya (1957, pp. 5-19) suggested that problem solving was a heuristic process that proceeded in four stages, namely understanding the problem, devising a plan, carrying out the plan, and looking back. What is missing from Polya's account is recognition of illumination as an identified stage. Illumination would appear to be central to the

Gestalt model of problem solving. However, Polya did recognise that “seeing the light” (p. 58) could be important in finding a solution, but it did not feature as a major phase in the model of problem solving that he proposed.

This illumination phase is problematic in attempting to build an instructional program or an assessment tool on the Gestalt model. Illumination is regarded as being internal to the problem solver and may be considered to occur at a subconscious level and to be unavailable to intervention or assessment, and that limits the value of the Gestalt model.

One feature ascribed to illumination is restructuring or thinking about the problem in alternative ways. A form of the restructuring concept becomes central to the information processing model of problem solving.

The information processing model

Information processing as a model of problem solving became well known through the work of Newell and Simon (1972). They modelled human problem solving behaviour on computer software and used computer simulations to show how competent human problem solving might proceed. It may be noted that the computer analogy has not been static. When Newell and Simon wrote their seminal text, computer programs were described as belonging to the third-generation of software. The languages were high-level and written as statements using English language words (e.g. If, Then, and Do), and they were procedural. That is, the programmer instructed the computer to follow a sequence of procedures on given data in order to generate output. Recent models of computer programming (object-oriented programs and neural networks) have been based on an emerging understanding of human cognition. Instead of using the computer as a metaphor for human problem solving, recent computer models have been founded on semantically-based human information processing methods using the brain as a metaphor.

The methods of Newell and Simon were applied to many problems, and especially to puzzles such as the Tower of Hanoi, water jug and river crossing problems (Atwood & Polson, 1976; Jeffries, Polson, Razran, & Atwood, 1997; Kotovsky, Hayes, & Simon, 1985; Ruiz & Newell, 1989). The computer software metaphor, in which one of several general-purpose strategies such as means-ends analysis or difference reduction were applied to the initial state, was used to simulate human problem solving behaviour. Under the difference reduction strategy, operators were selected that would minimise the difference between the current state and the goal state. The model was successful in simulating human problem solving. For example, Atwood and Polson (1976) showed that the computer simulation correctly predicted the difficulties of various water jug problems and modelled moves made by competent problem solvers.

The information-processing model was not completely successful. When isomorphs of the Tower of Hanoi puzzle were used, individuals found some of those isomorphs much more difficult than the original towers puzzle (Kotovsky et al., 1985). For example, success on the Tower of Hanoi puzzle did not guarantee success on the monsters and globes problem. This represented a failure of transfer and revealed a limitation of the information processing model. Under that model, the states and the allowable operators were quite similar and the same set of procedures could be used in the problem analogues. In a computer program, different data would be provided to the same program and the same solution path would be followed. Human problem solvers, however, would appear to have been distracted by the different content of the problem isomorphs – disks on pegs and monsters holding globes.

The failure of spontaneous transfer in problem analogues was substantially overcome by prompting problem solvers of the possible relevance of the isomorph. Gick and Holyoak (1983) found that when students were given one prior problem analogue, even with a summary of the principle underlying the problem, they did not demonstrate significant transfer. However, students did demonstrate substantial transfer if they had been exposed to two analogues and had been asked to describe the similarities between them. Gick and Holyoak concluded that exposure to two analogues, with some cognitive effort expended on describing the similarities, did lead to the induction of the analogical principle involved. It would appear that the strategy involved in facilitating the transfer was one that encouraged the induction of an abstract principle.

The successes of the information processing model, especially the demonstration that competent human problem solving could be modelled by computer programs, indicates that some generalised problem solving strategies, such as means-ends analysis, difference reduction, the use of analogy and problem decomposition (the generation of sub-goals) were important methods for problem solving. Their utility might be limited to certain types of problems, for example well-defined problems (R. E. Mayer, 1989, 1995a). The model predicted transfer that did not materialise between problems that could be represented formally as being close analogues (Kotovsky et al., 1985). This failure suggested that the model was deficient in not taking into account the context in which the problem occurs, although Anderson, Reder and Simon (1996, p. 7) cited several examples of successful transfer including one based on the Tower of Hanoi puzzle.

An alternative to the information processing model, situated cognition, does take into account the context in which problems are encountered.

Situated cognition

The situated cognition (or situative, Greeno, 1997, p. 16) model was developed largely through the work of Lave and Wegner (1991), although Bereiter (1999, p. 281) noted that both Tolman (1949) and Woodworth (1958) had described 'place learning' earlier.

Situated cognition has two general elements, namely the way in which competence emerges in contexts and, consequently, what constitutes competence (or expertise) in that context and beyond. The first of these elements, that learning occurs in complex, social situations through legitimate peripheral participation (Greeno, Collins, & Resnick, 1996, p. 23), is not disputed here. Learning does occur through apprenticeship, which has a very long history. It has been described in anthropological studies (e.g., Saxe, 1996) and structured and formalised versions of it have a long tradition in Western crafts and trades (Ray, 2001). It may be noted that the modern, Western apprenticeship model differs from the traditional notions of apprenticeship that are recognised in anthropological studies in that the modern apprenticeship has two components. First, apprentices work and learn on-the-job (in authentic contexts) with experienced tradespersons as mentors, and this is the key element of the traditional anthropological model of apprenticeship. Second, most modern apprenticeships involve off-the-job instruction in trade schools or other educational institutions. The concern in this section is with the characteristics and limitations ascribed to competence under the situated cognition (situative) model.

Several examples of situative learning and problem solving have been described in the literature, including dairy loaders (Scribner, 1984), supermarket shoppers (Lave, 1988) and street vendors (Saxe, 1988, 1996). These cases had two elements in common. First, individuals developed skills in context by interactions with more experienced others. Among street vendors, Saxe (1988) observed young children (aged 10-12 years) buying boxes of candy from suppliers and re-selling it in smaller quantities. Those who were inexperienced depended upon their wholesalers to provide advice on how much to mark-up their goods for resale, while their more experienced peers did not require this support. A second feature of the situative view is that individuals manipulated concrete materials rather than abstract symbolic representations in their operations. This was described by Saxe (1996) with most street vendors counting monetary notes directly, using repeated addition instead of multiplication, when calculations could have been conducted on paper. Similarly, Scribner (1984) noted that the dairy packers used the items they were packing rather than abstract representations to prepare orders. Third, individuals used invented strategies rather than formal operations for their calculations. For example, Lave reported that supermarket shoppers selected a strategy – price difference, ratio or unit cost – that imposed the lowest cognitive load.

Experienced individuals in each domain used concrete rather than abstract representations and used procedures that imposed low cognitive loads. They were fast and accurate in their operations, a characteristic of competent individuals in their domains. The situation of street vendors differed from those of dairy packers and supermarket shoppers. The former had very little formal education (Saxe, 1988) and might not have had sufficient algorithmic skill to apply arithmetic knowledge to the practical problems they faced. The supermarket shoppers and dairy packers, being adults in a developed country, could be assumed to have had enough formal education that they would have known relevant algorithmic procedures. In their cases, the use of a strategy that imposed a low cognitive load would appear to be a matter of choice. In the case of the supermarket shoppers, that choice was associated with a goal – to purchase the best value items. However, the consequence of a poor choice, buying a more expensive product, was low-stakes. This would appear to suggest that the chosen strategy was adequate for its purpose and that there was no requirement to use a more precise formal method.

The apparent adequacy of purpose suggests that transfer from formal school-taught methods is judged not to warrant the cognitive load of forming an abstract representation and manipulating its symbols. However, the selection of a strategy judged fit-for-purpose for that task may limit transfer to other contexts, where different affordances and limitations of the context, and different goals and consequences specific to that context, exist.

Rogoff (1984) summarised situative performance by saying:

> Thinking is intrinsically woven with the context of the problem to be solved... Our ability to control and orchestrate cognitive skills is not an abstract context-free competence which may be easily transferred across widely diverse problem domains but consists rather of cognitive activity tied specifically to context. (cited in R. E. Mayer, 1992, pp. 506-7)

This comment is informative of a key limitation of the situative perspective – a denial of the possibility of substantial transfer. It also implies a limited view of expertise; that expertise resides in proficient routine performance within a domain rather than being adaptive across domains. The issue of transfer is the subject of debate between advocates of situated and non-situated cognition and the two positions are contrasted.

Evaluating alternative models of problem solving

A series of exchanges between proponents of the two major theoretical positions on problem solving, the situative and non-situative[16] perspectives, provides a very useful summary of the differences between them (J. R. Anderson et al., 1996, 1997; Greeno, 1997). Key differences between these positions can be summarised as follows. The non-situative model is concerned with individuals' acquisition of knowledge, the structure of that knowledge, the representations that individuals form of constructs and of problems, and the strategies that individuals use in acquiring, structuring, representing and applying knowledge. The situative model is concerned with individuals' involvement in social processes in complex situations and with their participation in activities that are part of the social exchanges of the groups in which they are located.

Anderson, Reder and Simon (1996) provided a critique of the application of situative models of learning to instructional practices. They attributed four claims about learning to the situative position. The first three of these were based on purported failures of transfer of learning. Their third criticism of the situative position, that the formation of abstract representation was of little use, was central to the debate, as it predicted no possibility of substantial transfer of learning. This was an important claim, because if abstract representations were unhelpful in promoting transfer, then much school-based instruction would be futile. Indeed, Anderson, et al. demonstrated that abstraction was an important attribute of successful transfer. They showed, through reference to many empirical studies, that transfer had varied from being negative in some instances, to being very modest in others, and to being substantial and positive in some cases. What differed, they found, were the instructional conditions that accompanied tests of transfer. Where transfer was anticipated in instruction and the attention of learners

[16] In this discussion, the terms situative and non-situative are used to contrast the two positions. The non-situative position is akin to the information processing model of problem solving. Information processing applies specifically to problem solving, whereas the debate is a more general one about learning. Anderson et al. (1996; 1997) used the term 'cognitive' to differentiate their position from the 'situative' defended by Greeno (1997). Both perspectives, however, are cognitive. Bereiter (1999) used the terms situated and non-situated to distinguish these positions.

drawn to the possibility of transferring that what was learned in one context could be applied in a subsequent one, they showed that transfer was much more likely to occur (see, e.g., Gick & Holyoak, 1983). Miller and Stigler (1991) found that proficient individuals developed one of two representations for abacus use. In the first, labelled ‘conceptual determination,’ individuals’ representations were tied to the calculation functions of the abacus, and led to routine expertise. In the second, labelled ‘conceptual transparency,’ representations were more abstract and led to adaptive expertise. Those whose representation was tied closely to the original task revealed less transfer on a task involving judgments of number similarity than those who had a more abstract representation. The distinction between routine and adaptive expertise would appear to be an important one. Routine expertise is valuable when individuals are required to undertake a limited range of similar tasks and they become fast and accurate in that context. Adaptive expertise is valued when individuals may encounter novel situations for which knowledge and skill gained in related situations can be applied to achieve desired outcomes. In the context of generic skills, it is the latter type of expertise that is valued, and it appears to depend upon the generation of abstract representations.

The fourth criticism offered by Anderson at al. (1996) of the situative position was a critique of the social context of learning. Anderson et al. cited reports in which cooperative learning activities led to ineffective learning, and thus sought to demonstrate that locating learning in social situations was not necessarily an improvement over individual learning contexts. This criticism would appear not to be valid, as cooperative learning – usually learning among novice peers – was not the same as learning in an apprentice-master mentoring relationship in an authentic context. Further, many studies had shown substantial benefits from other forms of social interaction, for example reciprocal teaching (see, e.g., A. L. Brown & Palincsar, 1989).

They also addressed the situative demand for complexity of learning tasks and showed that decomposing complex tasks into smaller elements facilitated initial learning.

Greeno’s (1997) response disputed that the claims Anderson et al. had attributed to the situative position were held by scholars in that tradition. His analysis went beyond simply debating the evidence and included an analysis of the assumptions underlying the two theoretical positions, highlighting the implications of evidence for both theory and classroom practice.

Greeno’s response to claims that, in the situative perspective, knowledge did not transfer between contexts was debated in terms of underlying assumptions about the social compared with the individual nature of learning in the situative rather than the non-situative perspective. Greeno (1997, p. 7) cited several studies in which some transfer of learning had been reported and commented that “These findings are not inconsistent with the situative perspective...” This response indicated that, at least in Greeno’s view, knowledge transfer did occur, and on this matter, he and Anderson et al. agreed.

In responding to the issue of decomposing instruction in complex skills into sub-skills, Greeno introduced a social dimension, and questioned whether skills should be acquired individually or within a community of practice. Unfortunately, this added some confusion to the debate. Under the situative perspective, all learning is social, so perhaps Greeno believed this was a more important facet of learning than the question of complex or sub-skill instruction. The Anderson et al. (1997, p. 18) rejoinder accepted the conjunction of the individual or social and complex or sub-skill issues and noted that the two camps were in agreement and that “instruction need not take place only in complex social situations.”

The Anderson at al. original (1996) critique of the situative perspective had contrasted abstract learning in the non-situative tradition with specific instruction in the situative approach. Greeno took them to task for confusing two separate issues – the balance between abstract and concrete learning and between general and specific instruction. Anderson at al. (1997) accepted this rebuke. However, they asserted their original position that instruction that drew attention to abstract representations could be very effective.

The debate between Anderson et al. (1997; 1997) and Greeno (1997) exposed some key differences between the situative and non-situative positions. Beliefs that learning was fundamentally either individualistic or social were not tested in this debate. What has emerged from it is some agreement that transfer of knowledge can be expected and that instruction that attends to abstract representations can be effective. It appears, though, that the two parties may disagree on the extent of likely transfer and on the degree of abstraction. However, it may be noted that Greeno has worked across the situative non-situative divide, and his views may not be typical of situative theorists who do not have this breadth of experience. In his response to Anderson et al. (1996), he called for continuing work in the two traditions and for dialogue between researchers in these fields. Earlier, he had expressed satisfaction with the constructive dialogue that had occurred between researchers in these traditions (Greeno, 1985). It would appear to be unlikely that protagonists on either bank of the river would change their positions, because the assumptions that underpinned the two positions were not readily comparable.

Bereiter (1999) advanced a critique of the situative position on three grounds. First, he identified artificial or machine intelligence as an archetype of the non-situative position. This, he said, was "profoundly important for understanding human cognition and its situatedness" (p. 287). He examined cases where machine intelligence succeeded and where it failed in modelling human behaviour. He used chess as an example of a game where machine intelligence surpassed all but a few human experts, whereas modelling a group of children playing 'tag' would reveal artificial intelligence robots as clumsy and stupid. The chess example, with a large number of finite states and transition rules – that is a well defined but complex problem, is one of those on which the strength of the information processing model emerges. Playing tag involved actors who had goals that could change quickly, and instead of there being a finite number of quantum states, players' moves were continuous and capricious, and therefore not amenable to the sort of representation required by computers. Most children, on the other hand, were very good at this game. Tag was given as an example of the superiority of the situative model in how it was learned and played.

As Anderson et al. (1996) had done, Bereiter (1999) raised the problem of transfer and he made similar points about it. Under the situative account, the development of expertise involved becoming attuned to the demands of the situation. But, he asserted, that attunement did not transfer to other situations. Bereiter acknowledged Greeno's contribution to a situative model of transfer in which individuals recognised, in a new context, aspects of their relationship with a previous situation. However, Bereiter provided examples where there could have been no similar situation. In such cases, Bereiter claimed that the only viable explanatory mechanism was abstraction, and again this was reminiscent of the debate between Anderson et al. and Greeno. Bereiter's key point in relation to abstraction was that it was not a characteristic of situated models of learning with their focus on specific and concrete experience. Bereiter invoked Popper's notion of 'three worlds' to stress the significance of abstraction. Popper (1978, pp. 156 & 159) drew a distinction between "knowledge in the subjective sense" of thought processes (World 2 knowledge) and "knowledge in the objective sense" that he called thought content (World 3 knowledge). It would appear that the situative view of knowledge is of Popper's subjective or World 2 type but that it is not formalised through language to become shared World 3 knowledge that can exert a causal influence on events in World 1. Bereiter argued that access to World 3 knowledge of theories and models, and its influence on interactions between

World 1 and individuals' experience of it as World 2, would lead to learning that could surpass the constraints of situations. This position suggests a key role for generic skills, at least if they are defined as competences based upon metacognitive skills, such as monitoring and reflection and an ability to perceive the potential utility of solution strategies in new situations that may help individuals to go beyond the initial constraints of their situations.

Bereiter's (1999) third contribution was on the role of goals. He referred to a hierarchy of learning goals – task completion, instructional, and knowledge-building goals. The sequence of goals reflected increasing abstraction. Task completion goals were limited to specific situations, whereas knowledge-building goals required abstract representations to be constructed. In the examples given by the situative researchers cited above (Lave, 1988; Saxe, 1988, 1996; Scribner, 1984), task completion goals could be imputed. Such goals might explain low transfer from formal learning to the contexts that were studied, and would predict low transfer from those contexts, for example dairy packing and supermarket shopping, to other situations. In these cases, the situated actions of the individuals may be seen as an instance of specialised adaptation to the context that reflects a choice of strategy designed to meet immediate task completion goals as efficiently as possible. The use of concrete representations in dairy packing and the selection of low-load computational strategies by shoppers are efficient, in that they involve low inputs to achieve a desired output.

There is strong evidence to support both the situative and information processing models of problem solving. It appears that the information-processing model of problem solving is consistent with the view of generic skills as latent competences, and that a productive way forward is to seek evidence of generally applicable processes that are indicative of an ability to abstract organisational features of problem situations that may subsequently be applied to novel problems.

Problem solving processes

In the search for a conception of generic skills as a class of constructs, Sternberg's theory of practical intelligence (Sternberg et al., 2000) emerges as a bridge between previous theories of intelligence and the competence-performance model identified by Weinert (1999). The theory of practical intelligence and notions of competence require learners to develop a set of cognitive and metacognitive skills and strategies. Alexander (2003b, p. 12) argued that students did not come prefigured with these skills and that they needed to be taught. It is argued above that information processing models of problem solving are superior to situative ones because they involve abstract representations of the problem solving process and therefore facilitate generalisation beyond the constraints of individual situations.

In order to assist the development of problem solving performance among students, it is necessary to focus on the component processes of problem solving. However, Sternberg's components are abstractions of processes that students enact in practical problem solving. A method for representing the higher level components and meta-components is required, and several theorists provide options. Nickerson (1994) reviewed many problem solving schemes and found that they were very diverse and that no single theoretical model could form a basis for them, although he observed that Sternberg's triarchic theory (a precursor to his theory of successful intelligence) was the most useful. He identified three process-based models of problem solving, namely Polya's (1957) four-stage heuristic approach, Bransford's and Stein's (1984) five-stage scheme, and Hayes' (1989) six-stage method (see Table 11) and they are examined here. In addition to the processes identified by these theorists, also includes meta-components and performance components (Sternberg et al., 2000) that are believed to underlie the problem solving processes, and it includes labels for general cognitive and metacognitive processes.

Table 11: Process-based models of problem solving

Cognitive and metacognitive processes	Sternberg's (2000) meta-components	Polya (1957, pp. 5-19)	Bransford and Stein (1984, p. 12)	Hayes (1989, p. 3)
Apprehend	Recognise the problem	Understand the problem	Identify the problem	Finding the problem
Represent	Decide nature of problem Select a representation		Define and represent the problem with precision	Representing the problem
Plan	Select problem solving processes Select a strategy	Devise a plan	Explore possible strategies	Planning the solution
Act	Allocate resources Encode Infer Compare Respond	Carry out the plan	Act on those strategies	Carrying out the plan
Reflect Evaluate	Monitor progress Evaluate effectiveness	Look back	Look back and evaluate the effects of activities	Evaluating the solution Consolidating gains

The shading in Table 11 shows the commonality among the three process-based schemes. The schemes are not competing alternatives; rather they appear to be different representations of the same underlying cognitive and metacognitive processes. They all include locating, defining or understanding problems. These are all early-stage processes in problem solving, although some, for example finding a problem, occur before others such as understanding the problem. The representation phase, which Polya (1957) did not specify, was central to information processing models of problem solving and arose in studies of expertise (e.g., Chi et al., 1981). Chi et al. found that experts represented problems according to abstract characteristics while novices used surface features in their categorisation of problems. However, both novices and experts formed representations. Thus, it appears that this is an important phase in problem solving. Each of the three process-based theories includes a planning stage, and they are quite similar in this respect and in the execution of the chosen plan. Again, there are similarities in the post-solution stage, looking back and evaluating the solution. This phase includes the metacognitive processes of monitoring, reflecting on and evaluating the solution. Hayes included consolidating gains, and this is likely to occur through reflection on and evaluation of the effectiveness and efficiency of the solution. A composite of these processes is selected as the basis of a problem solving assessment tool. Five of these processes are used in the development of that tool, namely representation, planning, execution, monitoring and reflection. Apprehension is not included as the problems are presented to students. The development and testing of this tool are described in Chapter 7.

Summary statements

This chapter reports the results of two searches: one for a conception of generic skills as a class of related constructs and the other for a conception of problem solving that enables learners, especially novices in a domain, to generate possible solution strategies, to apply them and to monitor and reflect on their utility.

Generic skills are competences

It is more useful to perceive generic skills as competences rather than as manifestations of general intelligence (or of specific types of intelligence). As competences, they are regarded as being amenable to development through instruction and experience, whereas if they are viewed as manifestations of a traditional conception of intelligence they may be perceived as relatively fixed attributes. This fixed trait view appears to be mistaken on at least two grounds. First, Piagetian reasoning, a key component of fluid intelligence, develops with age and experience. Second, crystallised intelligence, which includes word knowledge, reading speed and writing ability, all increase with formal education. Sternberg’s theory of successful intelligence provides a link between intelligence and competence and leads to the notion of an intelligence-competence duality.

The set of cognitive and metacognitive processes in Sternberg's model may explain how individuals learn from and adapt to their environments, a requirement of models of competence.

Generic skills are latent traits

If generic skills are competences, they are regarded as latent attributes; that is, like linguistic competence, they may only be observed through performances generated in contexts that elicit their use. Indicators of competences need to be articulated if the state of the emerging competence is to be observed and assessed and so that learners are aware of the behaviours that are being sought.

Effective problem solving, among novices, depends upon the deployment of processes

The most useful conception of problem solving among novices is an information-processing model. The situative model appears to place an emphasis on the immediate context in which performance is elicited and to pay too little attention to general and abstract cognitive processes. The construction of problem solving as a set of processes that can be applied to the representation and solution of problems and the reflection on solution attempts may scaffold novice problem solvers in a domain and may assist them to enhance their initial problem solving, in particular through assessment and feedback. This position suggests an answer to the question "How do novices, who have limited knowledge, transform themselves into experts with extensive and organised knowledge bases?" (Bereiter & Scardamalia, 1986).

The implications of each of these propositions are discussed in Chapter 5. It is to the issues of assessment and feedback that attention is directed in Chapter 4.

Chapter 4: Purposes, Forms and Outcomes of Assessment

Assessment is not the primary focus of the study. The key purpose of the study is to investigate a method for developing and reporting problem solving performance. Assessment is necessarily involved. Because assessment is not the focus of the investigation, but a process required to reveal performance, some propositions – rather than hypotheses – about assessment are developed and used to frame the approach to the development and reporting of problem solving performance in this study.

In this chapter, assessment is defined and two key purposes – summative and formative – are described. Formative assessment is investigated as there is evidence that it can promote learning. Formative assessment has two key elements, namely feedback and self-assessment. Establishing performance standards is identified as a critical issue in assessment. Several options for the specification of performance criteria are explored and the SOLO Taxonomy (Biggs & Collis, 1982) is suggested as a basis for establishing performance levels. Criteria are defined, including validity, reliability, objectivity and feasibility, by which different assessment models can be compared. A variety of assessment methods is reviewed. These methods are grouped into five major categories, and are then compared using the evaluative criteria.

A set of propositions that direct the assessment of problem solving are developed although not all of them are tested in this study. For some of these propositions, there is strong research evidence, while for others the empirical support is modest. The argument supporting them seems sufficiently sound to justify their use in this study.

Assessment defined

Educational sssessment is a process of gathering evidence, making judgments and drawing inferences about student achievement and performance. Pellegrino, Chudowsky and Glaser (2001, p. 42) described assessment in the following terms:

> An assessment is a tool designed to observe students' behavior and produce data that can be used to draw reasonable inferences about what students know.

That description is useful as it draws attention to three key elements common to all assessments: observation, data and inference. The authors went on to describe these elements as 'the assessment triangle with observation, interpretation (of data) and learner cognition (the object of inference) at its vertices. The description leads to several questions. First, what observations are to be made? This question draws attention to the contexts and tasks that are used in order to gather evidence. Second, what data are to be gathered? This element requires that some mechanism be used to rate observed behaviours. This rating may be qualitative only, or qualitative descriptors may be ranked in order to generate scores. Third, what inferences are envisaged? In the current study, inferences are being made about students' current (and likely future) problem solving performance. The inferences depend upon a particular construction of, or cognitions about, problem solving, and that is the subject of Chapter 3. Further, whether the judgments are quantitative or qualitative, they are made about current behaviours in relation to expected standards of performance. Indeed, it seems that the process of recognising standards and of judging students' work in relation to those standards is central to the assessment enterprise.

Each of these questions cascades into further sets of questions. Many constructions of the assessment domain – problem solving – are possible and each leads to different inferences and requires different tasks and contexts. How the tasks are constructed can vary widely, from selection of to construction of responses. Selection occurs in multiple-choice tests while constructed responses may vary from short answers to test questions, to extended responses in essays, to products and to performances. Finally, how the data are used can vary from feedback to learners using qualitative descriptions, to diagnostic investigations and to very detailed psychometric modelling and measurement in order to draw inferences about individuals, class or school groups, or national education systems.

Goals and purposes for assessment

Boud (1995b, p. 38), having summarised a body of research on assessment, concluded that much assessment is of those aspects of learning that are easy to assess, which leads to low-level skills development. He cited Eisner (1993) who listed desirable attributes of assessment as including:

- assessment should be authentic (real-world like);
- assessment should be process rather than results oriented;
- the act of assessment signals the importance of what is being assessed, so assessment is a driver for learning; and

- assessment activities need to be seen by students as worthwhile and interesting activities. (Boud, 1995b, p. 40)

Wiggins (1998, p. 7) asserted, “the aim of assessment is primarily to educate and improve student performance, not merely to audit it.” However, various other authors have identified three broad purposes for assessment, namely to monitor and improve learning, to direct instruction and to monitor system level performance. Airasian (1994a) identified six purposes, namely to:

- diagnose student learning difficulties;
- make judgements about student academic performance;
- provide feedback and incentives to students;
- placement of students;
- planning and conducting instruction; and
- establish and maintain social equilibrium in the classroom.

Airasian (1994a) and Pellegrino, Chudowsky and Glaser (2001) summarised these purposes as:

- promoting learning;
- measuring individual achievement; and
- evaluating programs.

Others have described assessment purposes as assessment *of* and *for* learning (Assessment Reform Group, 1999) and assessment *as* learning (Earl, 2005). Of these approaches, assessment *of* learning relates most closely to evaluating individual achievement. This may be done at the end of a term or school year or at the end of compulsory schooling and may be used to assign grades and to rank students for admission to further stages of education. When appropriately aggregated, using multilevel methods, data on individual achievement can be used to evaluate programs, schools and education systems. Programs such as the OECD’s Programme for International Student Assessment (PISA) and the International Association for the Evaluation of Educational Achievement (IEA) Trends in Mathematics and Science Study (TIMSS) use data in this way to compare national education systems.

Assessment *for* learning is a “process of seeking and interpreting evidence for use by learners and their teachers to decide where the learners are in their learning, where they need to go and how best to get there” (Assessment Reform Group, 2002, p. 2). This approach occurs typically within classrooms during a school term or year, but it is also used, particularly in special education, to diagnose learning difficulties and to prescribe learning programs to remediate those learning difficulties.

Assessment *as* learning exhorts teachers to use assessment activities as opportunities for learning as well as for making judgments about student's understandings to inform teaching processes and to direct individual learning for students. Others have proposed that all classroom activities, including assessment tasks, should be opportunities for student learning. In these cases, assessment tasks become learning tasks and the phrase 'assessment *as* learning' becomes particularly apt.

Assessments have been described as either formative (assessment *as* and *for* learning) or summative (assessment *of* learning) and a clear demarcation has been drawn between them. There are certainly instances where the distinction is clear. In high stakes academic assessments and in professional or trade licensing, summative assessment is the primary and perhaps only interest. Before candidates can be admitted to professional practice, licensing bodies must be assured that candidates have the knowledge and skills that are regarded as necessary. These bodies set assessments to ensure that candidates meet prescribed minimum standards. These assessments are summative only; licensing bodies are responsible primarily for ensuring that practitioners are competent. There are examples of dual-role licensing bodies; they prescribe curriculum and offer courses that prepare candidates for their licensing examinations.

Assessment purposes and potential conflict

Assessment need not be exclusively summative or formative. Stobart (2004) argued that good assessment regimes can be designed both to promote learning and to provide a basis upon which to report achievement – that is, for both formative and summative purposes. Black (2004) cited examples where formative and summative purposes were combined, but in these cases formative assessment was used as students prepared for examinations (having a summative purpose) and these two purposes did not occur at the same time. In a summary of their more extensive review, Black and Wiliam (1998b) commented:

> ...assessing pupils summatively for external purposes is clearly different from the tasks of assessing ongoing work to monitor and improve progress. Some argue that these two roles are so different that they should be kept apart. We do not see how this can be done. (p.13)

In their extensive review paper of formative assessment, Black and Wiliam (1998a, pp. 46-47 & 59), citing studies by Withers (1987) and Butler (1995), recognised that formative and summative assessment goals may be incompatible. As with all studies of assessment, many factors in addition to purpose and including motivation, goal orientation, and feedback, influence learning outcomes and they noted that further research was required into these factors and their interactions.

Biggs' and Moore's (1993) exploration of the implications for assessment of learning approaches (surface, deep and achieving), provided a possible explanation for the conflict between formative and summative assessment purposes:

> ... in only one approach, deep, do students enter an intentional learning mode and become directly engaged with the content of the task itself. In the other two, the institutionalisation of the task, and particularly its assessment with its inevitable consequences of passing or failing, determine the nature of the student's engagement, either in maximising the marks gained (which isn't always the same as engaging the task on its own terms), or settling for an acceptable minimum. (p. 311).

Awarding marks for performances influences students' motives and leads towards an achieving orientation. This may engender better performance among those whose orientation is surface, but it may undermine the performance of students whose characteristic approach is deep. This analysis may contribute to the finding by Black and Wiliam (1998a, pp. 12-13) that formative assessment was particularly helpful for low achieving students. It is likely that low achieving students have either surface or achieving approaches and that the provision of feedback without grades encourages deeper engagement with the assessment tasks.

Black and Wiliam (1998b) concluded their summary paper with:

> …formative assessment is an essential component of classroom work and that its development can raise standards of achievement. We know of no other way of raising standards for which such a strong *prima facie* case can be made. (p.15)

Formative assessment has the potential to enhance learning and performance. In their review paper, Black and Wiliam (1998a) identified two consistently important components of formative assessment, namely feedback and self-assessment and these issues warrant some discussion.

Feedback

Feedback is an essential component of formative assessment. According to Sadler (1989)

> Formative assessment is concerned with how judgments about the quality of student responses (performance, pieces of work) can be used to shape and improve the student's competence by short-circuiting the randomness and inefficiency of trial-and-error learning. (p. 120)

He continued (1989, p. 120) defining feedback as "information about how successfully something has been or is being done." He drew upon Ramprasad's

(1983, p. 4) description of feedback as "…information about the gap between the actual level and the reference level of a system parameter which is used to alter the gap in some way." The notion of a 'gap to be bridged' and the role of feedback in that process were elaborated.

> …the learner has to (a) possess a concept of the *standard* (or goal, or reference level) being aimed for, (b) compare the *actual* (or current) *level of performance* with the standard, and (c) engage in appropriate *action* which leads to some closure of the gap. (Sadler, 1989, p. 121, original emphasis)

Wiggins (1998, pp. 68-9) distinguished feedback (what happened and why) from guidance (what must be done to improve performance), but, Sadler's use of the term 'feedback' encompassed both these facets as well as the idea of a standard. Further, Sadler stressed the direct role of the learner: learners must have access to the standard, who must be able to compare the current performance with that standard and who must take action to close the gap. These issues draw attention to both the description of standards and self-assessment, both of which are discussed below.

Feedback must be clear to the learner and inform performance. Sadler suggested that feedback, coded as letters or numbers, is "too deeply coded" to be useful to students to understand the quality of their own work and cannot lead to improvement. This is consistent with the finding reported by Black and Wiliam (1998a, pp. 12-13 citing Butler, 1988) that feedback as grades decreased performance compared with informative feedback that led to enhanced performance. It might be that the type of feedback interacted with or generated a particular goal orientation. Informative feedback increased a learning motivation while grades as feedback led to an achievement orientation. It might be that grades were not "too deeply coded"; instead they might lead to grade seeking rather than learning goals.

Feedback must be timely and provided frequently. In two related studies, Schunk (1996, cited in Black & Wiliam, 1998a, p. 13) demonstrated that frequent feedback exerted a strong influence on learning, equivalent to having task- rather than ego-oriented goals. However, the issue of frequency is not a simple one, as a learning task is rarely repeated (except perhaps in some mastery learning contexts) so that students can receive feedback specific to that one task. More commonly, students engage in learning and assessment tasks that are sufficiently similar so that the feedback on each task can inform performance on subsequent ones. Wiggins argued that feedback (information about current performance) must be available during assessment so that students had the opportunity to adjust their performance.

> The best feedback is highly specific, directly revealing or highly descriptive of what actually resulted, clear to the performer, and available or offered in terms of specific targets or standards. (Wiggins, 1998, p. 46)

Sadler (1989, p. 123) noted that learning "outcomes are often complex and invariably involve qualitative judgments." These outcomes emerged over time following exposure to many learning and assessment activities. Wiggins (1998, pp. 68-9) concurred with this position suggesting that iterative tasks and longitudinal assessments provided multiple opportunities to meet standards with feedback and guidance after each application of the standards.

Having access to standards is central to effective feedback. The specification of standards is discussed below. Sadler (1989, p. 139) appealed to the argument for assessment as learning in promoting the case for feedback that addressed explicit standards. Knowledge of the desired standard and an ability to make a judgment about current performance in relation to that standard were necessary; that is, students must be able to perceive accurately the gap between current and desired levels of performance. But they must also have a repertoire of productions that could close that gap. He argued "...the possession of evaluative expertise is a necessary (but not sufficient) condition for improvement" (p. 138). Tasks that are more complex needed to have more diverse ways of closing the performance gap, and so it was less likely that simply having information about the gap would lead to success in closing it (p. 139). However, Taras (2003), who used a common assessment sheet for students and tutors, showed that feedback could be effective even when explicit standards were not included in the performance criteria. In her feedback, performance was judged by students and tutors against five performance levels (excellent, very good, quite good, weak and very weak).

Feedback should focus on the gap between the desired and observed levels of performance. In a concern about the use of normative assessment (judging students relative to others), Black and Wiliam (1998a, p. 18) reported a common tendency for teachers to adopt normative assessment practices. They proposed that "...feedback to any pupil should be about the particular qualities of his or her work, with advice on what he or she can do to improve, and should avoid comparisons with other pupils" (1998b, p. 7).

In summary, feedback is central to formative assessment, and that, in turn, has been shown to contribute to enhanced learning. Because students must be able to interpret standards, judge their performances and take corrective action in order to achieve the learning outcomes that are posited for formative assessment, it is apparent that learners must be capable of assessing their own performances. Black and Wiliam (1998b, pp.7&8) note the "link of formative assessment to self-assessment is no accident; indeed, it is inevitable" and "self-assessment ...is ...an essential component of formative assessment." Self-assessment, therefore, demands attention.

Self-assessment

Boud (2002) made a powerful case for the direct engagement of learners in assessment.

> By deliberately keeping assessment out of the hands of learners, we are denying them one of the essential tools – perhaps the essential tool – which enables them to become lifelong learners. (Boud, 2002, p. 43)

Boud went on to observe that, if "assessment is something that is done to learners", even by experts, learners' capacities to become "self-determining lifelong learners" were compromised. What made learners self-determining lifelong learners was the capacity to make judgments about the quality of their own work. This was an ability required by all individuals, in work and other settings, that enabled them to take responsibility for the quality and suitability of their output. Self-assessment in formal learning settings was not about students allocating grades to their work, but about learning to calibrate their judgments about their work against the judgments of their more experienced teachers.

The development of generic skills in individuals contributes to their capacity to lead rewarding individual lives, to engage in their communities and to participate in the labour market. Sadler (1989, p. 139) noted that for formative assessment to fulfil its goals, individuals needed to be able to recognise desired standards, to compare their current performance with those standards and to initiate actions that would close any perceived gap. Self-assessment, therefore, is a necessary element of formative assessment, but it cannot be the only element. Students may need scaffolding to enable them to recognise and close performance gaps.

Definition

Boud (1995b, p. 12) offered a definition of self-assessment. First, he noted that all assessment depended upon establishing appropriate standards and setting criteria for judging whether those standards had been met. Specifically, assessment was a capacity to make those judgments.

Second, he defined self-assessment as "the involvement of students in identifying standards and or criteria to apply to their work and making judgments about the extent to which they have met those criteria and standards."

There is some ambiguity in the role of students in 'identifying' standards and criteria. Boud appeared to suggest that students should have a role in standard setting. Indeed, in his 1995 work, he did argue for students taking a role in standard setting. He acknowledged that some students would lack this capability and that, for them, the application of standards set by others was a minimum requirement (Boud, 1995b, p. 12). It is doubtful that students, by definition novices in the domain in which they are being instructed, would have the capacity to set standards or criteria. Sadler (1989) suggested that students should be able to recognise, rather than set, performance standards (p. 131).

Sadler described the knowledge of standards as 'guild knowledge' (p. 129). That is, there was a knowledge community, defined by the language in which teachers described and negotiated standards, of which teachers were members but learners were not. He went on to say that such ‘guild knowledge’ kept standards inaccessible to learners. Assessment, he argued, should be designed to provide learners with the language that would enable them to describe their performance in the terms that were used by teachers.

Even when standards are communicated to students, there is no guarantee that students understand and adopt the standards intended by the teacher. It seems that a necessary first step is to encourage students to articulate the standards of their own work, so that those standards are available to both the student and the teacher. How this is attempted in the current study is discussed in Chapter 7, in which the development of the problem solving assessment tool is described.

Self-assessment, learning and performance standards

Biggs and Moore (1993) also recognised deep, surface and achieving orientations to learning. They argued that these categories were relational; there was a relationship between students’ own goals for learning and their perception of the context set by the course. If the course context included teacher assessment with the award of grades, and especially if grades were normative, students were likely to adopt an achieving orientation to learning. Boud (1995b, p. 26), however, found no studies of the influence of self-assessment on students’ approaches to learning. He speculated that self-assessment could lead to a greater engagement of students with the subject.

Self-assessment can only be effective if learners understand the expected level of performance, so the expected performance must be expressed in terms that are accessible to them. Wiggins (1998) said that self-assessment was facilitated by the “clear specification of standards and criteria expressed as performance goals.”

Continuing this theme, Sadler (1989, p. 129) contended that standards could be used to specify learning goals, but such goals were only useful when they were owned by learners. That is, self-assessment must involve learners understanding and accepting the standards as goals. He continued by proposing that goals must also be expressed in relation to the learners' current performance: the goals must be challenging and require effort, but achievable. Such standards were necessary to support self-assessment. In turn, this would contribute to student learning through formative assessment.

Most learners complete their schooling without the need to assess themselves. Their experience of assessment is of a process in which they are passive observers. Boud (2002) argued that:

> …most people are ill-equipped to recognise cues which might indicate what is good quality work, or what distinguishes good from not so good task performance, or to work out for themselves whether they need to improve, learn more, ask for help or suggest an innovative change to what they are currently doing. Yet, it is attention to these matters that makes people effective in what they do. It is this understanding of what constitutes effective practice that enables learners to pursue it when they do not have assessors standing by them. (p. 42)

Black and Wiliam (1998b) argued that if self-assessment were to be effective, students needed to be trained in its use.

> ...if formative assessment is to be productive, pupils should be trained in self-assessment so that they can understand the main purposes of their learning and thereby grasp what they need to do to achieve. (Black & Wiliam, 1998b, p. 8)

There is some disagreement on the need for training. Brookhart (2001) reported that successful students made unprompted use of self-assessment, but Taras (2003) cited Wood, Marshall and Hrymak (1988) as having provided six hours of training in self-assessment.

It appears that students are likely to learn self-assessment if they are scaffolded in this enterprise. Boud (1995b) observed that self-assessment did not occur alone. The judgements of others should inform learners' self-assessments and improve learners' knowledge and abilities to assess themselves. This suggests that self-assessment should be accompanied by teacher assessment and that, in addition to feedback about their performance against standards in the domain, students also require feedback about the effectiveness of their own judgments. Taras (2003) noted that students preferred to have tutor feedback to support their self-assessment.

Summary of self-assessment

Self-assessment is an important attribute of lifelong learners. Those who master it are able to take responsibility for their actions and to improve their performance

(Boud, 1995b). Self assessment involves the cognitive processes of reflection and evaluation as shown in the final row of Table 11 (in the preceding chapter).

A capacity for self-assessment, that is for making judgments about one's work, is necessary for effective learning. Biggs and Moore (1993) and Boud (1995b) suggested that a complete reliance on teacher assessment might encourage students to adopt either surface or achieving orientations to learning but that self-assessment could promote deep approaches to learning.

In his concluding remarks, Boud (1995b) saw the possibility that self-assessment could be mandated and become routine and instrumental without being 'creative or emancipatory' and not associated with a learning agenda. He recommended that self-assessment should be introduced in order to promote reflection and metacognition in learning.

There is limited evidence on the effectiveness of self-assessment in promoting learning. Boud (1995b, p. 26) noted this lack of research on aspects of self-assessment. Black and Wiliam (1998a) cited one study (Fontana & Fernandes, 1994) in which self-assessment was compared with teacher assessment only. The self-assessment group achieved approximately twice the learning gain as the control groups. Thus, there was some empirical, and as Black and Wiliam noted (p. 10), ecologically valid, support for self-assessment. There is, however, a need to investigate the influence of self-assessment on learning, and this is done in the second study reported in Chapter 8 of this thesis.

As with all assessment, it is necessary to develop standards against which students' performances can be judged, whether the judgments are made by students or by their teachers. How such standards and criteria are specified is the subject of the next section of this review of assessment.

Frameworks for defining performance standards

Assessment is defined in terms of standards, of criteria that operationalise the standards and of judgments in which performance is compared with those standards. The issue of how standards are defined is an important one.

In particular, the question of how broadly or narrowly standards are defined needs to be addressed. If standards are defined too narrowly and are applicable to a limited range of tasks, the standards are likely to have little value in evaluating students' progress. Feedback against standards that apply to a few tasks may not be useful when students move on to other tasks for which different standards are used. On the other hand, if standards are defined too broadly, and possibly in terms that are quite abstract, it may be difficult to develop meaningful criteria for individual assessment tasks and for students and teachers to gain useful information from the assessment activity.

Norm-referenced assessment

Students may be assessed normatively, and that was common practice in the past and continues to be used when assessment is conducted for selection and sorting purposes. For these purposes, it is effective, as the absolute level of a student's performance is not at issue; rather, the purpose is to identify those students who demonstrate superior performance compared with others.

Black and Wiliam (1998a, p. 17-18, citing studies by Crooks, 1988 and Black, 1993) alleged that the practice had remained common and for other than selection purposes. It could be used in large scale assessments, in which there was reason to believe that the performance of whole cohorts did not change over time. In these circumstances, the top, say, 15 per cent of students might be awarded a distinction grade, the next 15 per cent a credit and so on. In such assessment regimes, it did not matter that an examination in one year might be more difficult than a similar examination in another year.

Norm-referencing was also used when test instruments and questionnaires were 'normed.' This practice is no longer necessary, as it is possible to use item response theory (IRT) to anchor test and questionnaire item parameters so that different groups of individuals can be assessed on a common scale. If selection were the purpose of the assessment, candidates could be ranked by their scores. (IRT and its application to assessment are discussed in Chapter 6).

McGaw (2006) identified a fundamental deficiency with norm-referenced assessment that is particularly salient in the current studies:

> …it cannot readily measure growth or improvement in an individual. So long as the point of reference is the performance of others, the only way in which an individual can be seen to improve is relative to the performance of others, which means effectively at the expense of others. (p. 6)

One of the purposes of the studies reported in this thesis (see Chapter 8) is to evaluate change in student performance over time. For this reason, performance of individuals needs to be measured against a standard that enables change to be detected.

Criterion-referenced assessment

Criterion-referenced assessment has begun to replace normative assessment. It was proposed by Glaser (1963). In criterion-referenced assessment, criteria were specified for each assessed task and students' performances were compared with the criteria. Provided raters applied the criteria consistently, their judgments of student performance should be objective (that it, independent of the student or the rater or the particular set of tasks on which the assessment occurred). This use of the term objective was consistent with its use of it in relation to measurement (Wright & Stone, 1999).

A clear advantage of criterion-referenced over norm-referenced assessment is that students can be informed of the criteria that will be used in advance of undertaking the tasks. Thus informed, students can direct their learning strategies at meeting criteria at the level of their choice. If criteria for acceptable, commendable and outstanding work are specified, students can allocate effort in order to attain their target level. Other, more familiar grade labels such as pass, credit and distinction can be substituted for those just given.

Teachers may set criteria based on their experiences with similar groups of students. In this way, the criteria may have a normative origin, with criteria for acceptable performance being based on the achievements of previous pass students. Even if this is the case, all students can now see the criteria they must meet for a given level, and provided they meet these criteria, they can be confident of their result. Moreover, they can see the criteria for adjacent performance levels and know what more they need to do in order to move to the next level. Thus, criterion-referenced assessment can contribute to student learning, whether the assessment is conducted by a teacher or the learner as self-assessment.

A question about the origins or bases upon which criteria are specified arises. If criteria are specified for a particular assessment activity, while that has the advantages for learners described above, those advantages are limited to a narrow range of tasks for which the criteria are set. If the criteria have a broader base, they may be applicable to a wider range of tasks and the learning that students do on one task may be transferred to others. Two approaches to the development of criteria are apparent in the literature.

Standards-referenced assessment

Standards-referenced assessment was proposed by Sadler (2003; 2005). He noted the movement away from norm-referenced and towards criterion-referenced assessment, but he drew attention to the diversity of criterion based assessment practices, some of which were subjective and failed to inform students of desired performance levels. His solution to the inadequacy of much criterion-referenced assessment was standards-referenced assessment and he defined a standard as:

> A definite level of excellence or attainment, or a definite degree of any quality viewed as a prescribed object of endeavour or as the recognised

> measure of what is adequate for some purpose, so established by authority, custom, or consensus. (Sadler, 2003, p. 10)

According to Sadler, standards-based grading required that a sequence of standards were envisaged by the teacher who wrote descriptors for them and communicated them to students. It was in this process that standards-based grading differed from criterion-based assessment (Sadler, 2003, p. 10).

Sadler's approach to defining standards differed from 'standard setting,' for which there was a long tradition. Much of standard setting occurred after the development of tests in which experts suggested threshold levels that they believed would distinguish between levels of observed performance, particularly between minimally competent individuals and those not yet competent. A variety of 'standard setting' methods were reviewed by Cizek (2001), but these did not deal with *a priori* agreements or frameworks in which standards reflected levels of broad cognitive capability.

Standard setting was used to "refer to the task of deriving levels of performance on educational or professional assessments, by which decisions or classifications of persons (and corresponding inferences) will be made" (Cizek, 2001, p. 1). Later, he resiled somewhat from this strongly psychometric position and wrote it was "much less of a technical challenge and much more of a policy endeavour" (p. 5). Hambleton (1994, p. 5721) observed that all procedures for standard setting were arbitrary and involved judgment. Objective methods for standard setting were desirable.

Sadler's (2003) definition implied that there was a threshold level of a quality that was required in order to meet the standard. Standards continued to be used to describe bands of performance in international and national testing programs such as PISA. In the 2003 PISA study, five levels of literacy, six of mathematics and four of problem solving were recognised (OECD, 2004, pp. 28-31; Thomson, Cresswell, & de Bortoli, 2004, pp. 42-51 & 92-95).

Construct-referenced assessment

Wiliam (1998) proposed construct-referenced assessment. As Sadler had been, Wiliam was critical of norm-referenced assessment. He too, was critical of criterion-referenced assessment, especially for high stakes testing, suggesting that it might lead to 'teaching to the test'. That is, teachers may focus upon the specified criteria and not attend to other aspects of the domain that were important, and because criteria tended to be defined narrowly and specifically in terms of the assessment tasks that were set, criterion based assessment might not yield the learning goals that were desired. Wiliam's solution to the problem was to use expert assessors who had shared understandings of the domain.

Two aspects of Wiliam's proposal are of concern. First, the reliance on expert assessors who share a common understanding of the domain to decide what is 'important' suggests that students, and even their teachers, may not be able to access the 'knowledge community' and may be excluded from an understanding of what is required. Students cannot evaluate their performance against a covert standard and teachers may not be able to provide informative feedback. Second, the suggestion that learning outcomes need not be prescribed appears likely to lead to diverse and potentially inconsistent learning outcomes. Wiliam suggested:

> The innovative feature of such assessment is that no attempt is made to prescribe learning outcomes. In that it is defined at all, it is defined simply as the consensus of the teachers making the assessments. The assessment is not objective, in the sense that there are no objective criteria for a student to satisfy, but the experience in England is that it can be made reliable. (Wiliam, 1998, p. 6)

However, Wiliam did provide a solution to the specification of learning outcomes and suggested a productive way forward:

> …the assessment system relies on the existence of a construct (of what it means to be competent in a particular domain) being shared among a community of practitioners (Lave, 1991), I have proposed elsewhere that such assessments are best described as 'construct-referenced.' (Wiliam, 1998, p. 6)

A key element of this proposal is that there is a recognised construct. In the current study, problem solving is taken to be the construct and the shared understanding is the information-processing model (see Chapter 3). This construction must be communicated to students and teachers so they do join the 'knowledge community.' This construction of problem solving is combined with an approach to standards development (see below) to generate generalised criteria. Making the construct and standards explicit for teachers and learners enables them respectively to generate and benefit from feedback.

It is useful to review Sadler's (1989) requirements for standards.

> In order to make effective judgments, teachers must possess a concept of the quality of the work that they expect and make judgments relative to that conception. For students to improve, they too must hold the same conception of quality that a teacher holds and must be able to use it to make judgments during the production of the work if they are to improve their performance.

The question that arises is 'how can conceptions of competence in the domain be described as standards that are accessible to learners?' This matter is addressed in the present study, first by describing problem solving as a set of processes that emerge from an information-processing model of problem solving and second, by generating standards of performance in relation to indicators that reflect different levels of the application of these problem solving processes. That is, the ‘construct’ that Wiliam (1998) advocated as the basis of a shared understanding of the domain is derived from a theory of problem solving. The key elements of that theory can be communicated to teachers and students. The ‘standards’ that Sadler (2003; Sadler, 2005) advocated emerge from the problem solving construct by generating indicators of problem solving processes and developing descriptors of different levels of performance (standards) that are expected to be observed as students apply their problem solving skills. The information-processing model of problem solving is described in Chapter 3 and the development of the assessment instrument based on this conception is described in Chapter 7.

Despite having a basis for describing problem solving performance, the issue of how objective performance standards may be derived remains unresolved. Most approaches to assessment, of problem solving (see below) and other domains, have defined standards of expected performance in terms that are applicable to the domain but that do not refer to higher order understandings of cognitive performance. A search for a more general model of cognitive performance that may have application across domains is needed. Two systematic approaches to the definition of performance standards are reviewed, namely Bloom’s taxonomy (Bloom et al., 1956) and the SOLO (Structure of the Observed Learning Outcome) taxonomy (Biggs & Collis, 1982).

Bloom’s taxonomy

Bloom’s taxonomy appears to provide a sound basis for creating a set of performance standards that may have application over a wide range of tasks. The taxonomy of cognitive educational objectives was formally published in 1956 (Bloom et al., 1956). The related taxonomy of objectives in the affective domain was published some years later (Krathwohl et al., 1964). The cognitive taxonomy comprised a single dimension and six levels of performance. The taxonomy, summarised in Table 12, included recall of knowledge at the lowest level and the evaluation of information at the highest level. The fact that the taxonomy proposed a single dimension of ordered levels makes it a very attractive basis for measuring performance. The minimal requirements for effective measurement are that performance is judged along a single dimension and that consistent judgements of performance levels are feasible.

The taxonomy resulted in widespread improvements in instruction and assessment, but it was limited. Airasian (1994b) argued, in a review of the taxonomy, that its use had led to an improvement in the specification of educational goals, from a very limited focus on learning outcomes specified in terms of the content being taught to a focus on individuals' learning. Moreover, it included cognitive objectives at a higher level than simple recall. Airasian pointed out, though, that the taxonomy did not suggest any processes that might lead to the desired outcomes.

Table 12: Bloom's taxonomy of educational objectives – cognitive domain

Cognitive level	Description
Knowledge	Recall data or information.
Comprehension	Understand the meaning, translation, interpolation, and interpretation of instructions and problems. State a problem in one's own words.
Application	Use a concept in a new situation or unprompted use of an abstraction. Applies what was learned in the classroom into novel situations in the work place.
Analysis	Separates material or concepts into component parts so that its organizational structure may be understood. Distinguishes between facts and inferences.
Synthesis	Builds a structure or pattern from diverse elements. Put parts together to form a whole, with emphasis on creating a new meaning or structure.
Evaluation	Make judgments about the value of ideas or materials.

Source: After Clark (2002)

The cognitive levels proposed in the taxonomy did not correspond with empirically or epistemologically derived levels. If the taxonomy provides a sound basis for evaluating learning, the empirical difficulty of items, specified according to the levels of the taxonomy, should show an ordered structure reflecting the levels of the taxonomy. Kreitzer and Madaus (1994) reviewed evidence from studies that had attempted to validate the structure of the taxonomy. They found that expert raters, when asked to judge the levels of items using the taxonomy, agreed substantially. However, when items, based on the taxonomy, were administered to students, the difficulty of the items did not reflect the ordering predicted by the taxonomy. Thus, the taxonomy did not provide an empirical basis for measuring performance, because some items that were designed to indicate lower level processes were found, in practice, to be more difficult than others designed to reflect higher level processes. However, Hill and McGaw (1981), who analysed data on over 5,000 students, found that the knowledge level did not conform to a simplex structure and that the application and analysis levels were not separable in the hierarchy. Thus, they found four separable levels in the structure. Furst (1994), through a philosophical analysis, found reversals of the levels of the proposed hierarchy. It appears that these reversals are not simply a result of a mis-specified order of the levels of the taxonomy, but arise because factors other than the proposed levels influence the empirical difficulty of indicators. Consequently, the taxonomy does not capture the complete set of influences on task difficulty.

The levels of the taxonomy can be compared with the problem solving processes identified from the review of models of problem solving (see Chapter 3). The processes that are selected are defining the problem, planning an approach, executing the plan, monitoring progress and reflecting on the result. These processes necessarily imply an order: if a problem cannot be defined, then no effective plan can be devised to solve it. A comparison of the set of problem solving processes and the cognitive activities that may be undertaken in executing them, classified according to Bloom’s cognitive taxonomy, are summarised in Table 13.

Table 13: Comparison of problem solving processes with Bloom’s taxonomic levels

Problem solving processes	Bloom’s taxonomic levels
Define the problem	Comprehend given information
	Recall knowledge
	Analyse information
Plan an approach	Synthesise information
	Generate options (Not included in Bloom’s taxonomy)
	Evaluate options and select most likely
Execute the plan	Apply planned actions
Monitor progress	Analyse
	Evaluate
Reflect on the result	Evaluate
	Synthesise
	Generate alternatives (Not included in Bloom’s taxonomy)

In general, there is a relationship between the set of problem solving processes and the levels of Bloom’s taxonomy with the higher level processes being associated with higher levels of the taxonomy. The relationship, however, is not ideal, with some higher level processes being enacted before lower level ones. Further, the problem solving behaviours envisaged include the generation of options, and this is not included in the taxonomy. This was one of Furst’s criticisms of the taxonomy: it was not comprehensive as it did not include understanding at a level required to support rational action and it did not admit the “continual reconstruction of experience”, one of the requirements of an iterative model of problem solving (Furst, 1994, p. 33).

The taxonomy was not tested empirically in the current study. The criticisms of the original taxonomy (Furst, 1994; Kreitzer & Madaus, 1994; Postlethwaite, 1994; Rohwer & Sloane, 1994) and the lack of a clearly ordered relationship between the problem solving processes and the cognitive levels of the taxonomy suggest that it may not provide a productive way forward in the current study.

The SOLO taxonomy

The SOLO (Structure of the Observed Learning Outcome) taxonomy (Biggs & Collis, 1982), is investigated as a potential framework for defining performance standards. This taxonomy was based upon the cognitive complexity of individuals' responses in applying knowledge in learning and problem situations. It recognised levels of performance from ineffective use of knowledge to very complex and abstract application. Descriptions for each level of the SOLO taxonomy are shown in Table 14. In can be noted that the categories of the taxonomy form a sequence of increasing quality of cognitive performance with performance standards embedded in them. The SOLO taxonomy is, therefore, consistent with Sadler's (1989) requirement for standards.

The structure of the taxonomy can have three dimensions, one for each of the three characteristics of individuals (capacity to recognise and use information, use of relating operations between elements of a problem and consistency and closure). Measurement, in a strict sense of this term, of a construct requires that the construct be unidimensional, or at least, that its dimensions were closely related and functioned in unison, as Bejar (1983) explained:

> Unidimensionality does not imply the performance on items is due to a single psychological process. In fact, a variety of psychological processes are involved in responding to a set of items. However, as long as they function in unison – that is, performance on each item is affected by the same process in the same form – unidimensionality will hold. (p. 31)

This appears to be feasible, as Biggs and Collis (1982) included in their Table 2.1 (reproduced in part as Table 14 below) the Piagetian stages and substages that corresponded to the SOLO levels, suggesting that dimensions of each reflected a common developmental sequence. Whether these dimensions do cohere sufficiently must be determined empirically.

An advantage of the SOLO taxonomy is that its five levels do form a set of ordered responses. These responses are thus amenable to analysis using Item Response Theory (IRT) and may form the basis of interval measurement scales. Scholten, Keeves and Lawson (2002) showed that the SOLO taxonomy could be applied to assessing the quality of students' knowledge within a domain and that it gave rise to precise quantitative measures. This possibility can be tested with data collected using the Problem Solving Assessment instrument (see Chapter 7), but if it is shown that IRT can be used to produce measures of student problem solving performance, a powerful tool for investigating the components of performance becomes available. The SOLO taxonomy together with IRT can provide information on the precision of problem solving ability assessments that in turn indicates the number of performance levels that can be discriminated reliably.

The SOLO taxonomy is selected in this study as the basis for describing performance standards for assessing elements of the problem solving processes. How the levels of the SOLO taxonomy were integrated with the identified problem solving processes is explained in Chapter 7.

Table 14: Performance levels of indicators using the SOLO taxonomy

SOLO Level	Description		
	Capacity	Relating operation	Consistency and closure
Pre-structural	Cue and response confused	Denial, tautology, transduction. Bound to specifics	No felt need for consistency. Closes without seeing the problem.
Uni-structural	Cue + one relevant datum	Can generalise, but only in terms of one aspect	No felt need for consistency, thus closes too quickly; jumps to conclusions on one aspect and so can be very inconsistent
Multi-structural	Cue + isolated relevant data	Can generalise, but only in terms of a few limited and independent aspects	Although has feeling for consistency can be inconsistent because closes too soon on basis of isolated fixation on data, and so can come to different conclusions with same data
Relational	Cue + relevant data + interrelations	Can generalise within given or experienced context using related aspects	No inconsistency within the given system, but since closure is unique, inconsistencies may occur when he [sic] goes outside the system
Extended abstract	Cue + relevant data + interrelations + hypotheses	Deduction and induction. Can generalise to situations not experienced	Inconsistencies resolved. No felt need to give closed decisions – conclusions held open or qualified to allow logically possible alternatives

Source: After Biggs and Collis (1982, Table 2.1)

From standards to measurement

Once standards have been defined, they need to be translated into criteria that lead to measurement of performance and the inference of ability – at least on the measured trait.

Sadler (1989, pp. 124-5) argued that qualitative judgments included multiple criteria that operated as a complex, that some criteria were fuzzy rather than sharp and they might not be decomposed into discrete units. He continued, suggesting that, given the large number of criteria that were involved in judgments of complex tasks, cognitive capacity (of an assessor) restricted the number that could be attended to at a time. This led to two sets of criteria, those that were manifest and consciously attended to, and a latent set that resided in the background. Latent criteria led to tacit standards while manifest ones led to overt standards. Tacit standards yielded holistic or ‘configurational’ judgments while manifest ones led to analytical judgments.

Sadler added (p. 135) that experienced teachers were able to activate latent standards when their salience to student work became apparent and applied them consistently in reaching configurational judgments of student work. This knowledge, he said, could not be codified and communicated directly to students; it had to be done through examples. This is a challenge to the requirements set out by Ramprasad (1983, p. 4) that standards must be communicated to students so they can compare their work against the standard and then engage in actions that seek to close the performance-standard gap.

The conflict between configurational or holistic judgments and analytical ones was raised in relation to self-assessment (Sadler, 1989, pp. 136-7). He suggested that students, like teachers, should make overall judgment first and then identify those criteria that were most relevant before making analytical judgments.

While teachers and students may begin with configurational judgments, the reliable measurement of performance requires that criteria be applied consistently. If a criterion is not relevant, it need not be scored. In an IRT measurement framework, missing data can be regarded as omitted, not wrong, and a missing criterion does not compromise the trait estimate to the extent that it might under a classical test theory model.

Summary

A variety of approaches to the setting of performance standards have been used. They include normative (norm-referenced), criterion-referenced, standards-referenced and construct-referenced assessment. It is argued that a construct-referenced approach, in which standards are developed from the construct, is a preferred basis for formative assessment involving feedback to learners. Reference standards can be established for specific assessment tasks or for classes of tasks. When they are developed for specific tasks, they appear as rubrics that are used to guide markers in order to enhance consistency between markers and can be used by students to guide them in meeting learning goals. In the discussion of formative assessment above, the point is made that formative assessment requires the specification of expected performance standards in terms that teachers and learners can understand and against which they can (a) judge current levels of performance and (b) identify what must be done to raise the performance to a higher level. If students are to learn from a series of assessment activities, the standards must be expressed in ways that are applicable to all tasks in the series. Feedback from one task is likely to feed into a student's approach to the next task. Thus, standards need to be general rather than specific. For this reason, a search of frameworks for standards specifications is required, and Bloom's Taxonomy (Bloom et al., 1956) and the SOLO Taxonomy (Biggs & Collis, 1982) are evaluated. The decision is taken to use the SOLO Taxonomy as a basis for assessing and measuring student performance on problem solving.

Criteria for evaluating assessment approaches

Many attempts have been made to assess and report on students' achievement of generic skills. Before reviewing the various approaches that have been taken, it is useful to examine the criteria by which assessment models can be judged. Typically, validity and reliability are the two main criteria that are used, but others, such as the objectivity and feasibility of the assessment are also worthy of consideration. It is argued that these criteria are not a simple set of discrete bases for judgment; rather, they are interrelated and they include other elements such as fairness and credibility.

Validity and related constructs

The validity of an assessment is now understood as residing in the inferences that arise from the assessment, for individuals and for systems. Traditional conceptions of validity posit that there are several types of validity, namely content, criterion related, consequential and construct validity (see, e.g. Creswell, 2005, pp. 164-166; Zeller, 1997). Messick and three leading research associations in the United States, the American Educational Research Association (AERA), the American Psychological Association (APA), and the National Council on Measurement in Education (NCME) asserted that the validity of an assessment was the extent to which evidence and theory support judgments and interpretations, including the social consequences that followed those judgments and interpretations, that were made from test scores (AERA, APA, & NCME 1999; Messick, 1995). Messick (1995) argued that a range of sources of evidence had to be assembled in order to establish the validity of an assessment. Sources included evidence about the target construct of the assessment, the generalisability of judgments made on the assessment, and the consequences for individuals of the judgments. Messick (1998) concluded his discussion of validity by asserting "validation is an empirical evaluation of the meaning and consequences of measurement. The term empirical evaluation is meant to convey that the validation process is scientific as well as rhetorical and requires both evidence and argument."

Messick (1998) identified two key threats to the validity of judgments: construct under-representation and construct-irrelevant variance. Three implications of these two threats for the current study are identified. First, the target construct, problem solving, must be fully and clearly described. Second, the assessment instrument must address all key aspects of the construct. Third, the assessment instrument should not assess other constructs that are unrelated to the target. These requirements are non-trivial, because problem solving does not exist as an isolated construct. The application of problem solving processes occurs in a domain and that domain has its own knowledge and skill requirements. Attempts to assess problem solving among students who study a variety of disciplines tend to prescribe common tasks that draw upon general knowledge. This is the approach taken in large scale testing such as PISA (OECD, 2003) and the Graduate Skills Assessment (Australian Council for Educational Research, 2000). No argument is advanced against this approach to testing, given the summative purposes for which such tests are constructed. However, the point is made in Chapter 1 that an alternative method of assessment that seeks to promote the development of skills such as problem solving might need to be embedded in the substantive content of students' course-related learning.

The social consequences of judgments for individuals constitute an element of validity and they may be crucial for individuals. Employers, for example, may use problem solving skill assessments for recruitment selection and this use makes the assessment a high-stakes activity. It is critical, therefore, for selection decisions to be supported by the assessment and reporting methods, that the assessments must be informative of an individual's target abilities, free from bias, as well as fulfilling other criteria.

Authenticity has been raised as an attribute of valid assessment. Authenticity, or ecological validity, may be an element of validity in that, for an assessment task to be valid, it needs to have some predictive power for future performances. An assessment task that presages likely future contexts possesses both predictive power and authenticity. For Wiggins (1998), the features that defined an authentic task were that it was realistic, required judgment and innovation, and required action by the learner. Although it is reasonable to expect these features in an assessment task, the assertion of authenticity remains problematic: authenticity is a matter of degree rather than a matter of the mere presence or absence of some characteristic. The extent to which a task reflects a real-world situation can be judged on several dimensions, and the demands of real-world tasks are themselves quite variable, so the requirements for judgment and innovation can differ markedly across tasks. In the current study, authenticity is assumed because the tasks that are set by teachers in order to assess the curriculum objectives of their courses are used as the contexts in which problem solving is assessed. This assumption of authenticity holds to the extent that the tasks routinely set by teachers are reflective of real-world situations that learners are likely to encounter in their work once they have graduated. Classroom learning was contrasted with real world learning by Resnick (1987) who claimed that learning in the classroom was individual, abstract, symbolic and generalised while real world learning was cooperative, practical, contextualised and situation specific. Classroom assessment tasks lack the complexity of real-world problems because, typically, teachers set tasks that focus on specific aspects of the learning they expect students to achieve. However, one of the hypotheses underpinning this study (see Chapter 5) is that problem solving skill emerges in concert with discipline based knowledge and skill. The instrument developed to support judgments of problem solving performance (see Chapter 7) needs to operate in varied contexts but should help also to maintain a focus on the core aspects of problem solving in the domain of instruction and thereby avoid the threats to valid assessment recognised by Messick (1998). Thus, as students move through their courses, the tasks they undertake become more complex and more like real-world ones, and therefore become more authentic over time. The same might be hoped for their problem solving abilities.

Reliability and related constructs

The consistency of scores or grades resulting from an assessment can be estimated using various methods, including test-retest reliability. Comments on the reliability of the various methods for assessing problem solving must be preceded by an examination of potential sources of variability in scores or grades. Fairness is an aspect of reliability because a lack of fairness can be indicated by bias either in the assessment tasks that are set or in the way criteria are applied in judging the work.

The desired source of variance lies in the differences between students' performances on assessable constructs. A variety of confounding sources are also identifiable. Errors in judgment and scoring are obvious sources of variance. It reduces the reliability and precision of measurement but it does not necessarily introduce bias. Other sources of variance including differences in teachers' perceptions of a given piece of work and the opportunities afforded by various tasks and contexts for the development and demonstration of skills, and they have the potential to obscure true variance in ability. These latter sources have the potential to introduce bias and therefore to compromise the fairness of the assessment. Each of these potential sources needs to be identified for each potential method of assessment.

For high-stakes purposes, adequate reliability of a score is not a sufficient criterion. If learners are to be compared by, for example, prospective employers based on their reported achievements on their problem solving skills, the precision of the measure must be adequate to support the granularity of decisions that will subsequently be made. This has implications for the assessment methods, performance levels identified, and reporting formats that are used. The reporting formats must reflect the precision of the measures that are derived from the assessments.

The importance of reliability may have been overstated. Moss (1994) challenged heavy reliance on reliability as a criterion that, she argued, privileged standardised assessment, which tended to have higher reliability than alternative assessment methods. Her assertion was that other methods might be more valid, but by assuming that greater reliability within or between assessments foreshadowed future performances better, assessments that were more reliable were preferred. She argued that the attainment of adequate levels of consistency in teacher judgments has been demonstrated in situations where common tasks were used and teachers were trained to look for particular forms of evidence. Moss suggested that teacher judgments of student performance should include an interpretation of observed performances for future applications. The key issue in balancing the requirements of validity and reliability was establishing the trustworthiness of the score or grade that was reported. McGaw (2006) similarly argued for the ‘fitness for purpose’ of assessment, noting, as did Messick (1998), that the consequences of reported achievement for individuals had to be considered. This also alludes to the fairness of assessments.

Compromises are required between attaining high reliability and high levels of validity in assessment. Some forms of standardised assessment can yield highly reliable and precise estimates of performance, but may not adequately reflect the scope of the target construct. This is the basis of criticisms of standardised tests in some domains. For example, it may be possible, using a multiple-choice test, to assess with great precision a learner’s ability to solve common problems. (See, e.g. ACER 2003). Such tests, however, may not enable the assessment of the candidate’s ability to apply that knowledge in novel contexts. On the other hand, assessing the application of problem solving skills in real contexts may add to the authenticity and validity of the assessment, but at the cost of reliability and precision. Judgments about an appropriate balance between achieving high levels of reliability and precision and high levels of authenticity and validity are required, and the balance may differ when different assessment purposes are invoked.

Objectivity

Objectivity has two similar and related meanings in relation to testing. Specific objectivity is a characteristic of measures and has a technical meaning. It has been shown that the particular form of the logistic function used in Rasch measurement leads to estimates of student ability that are independent of the distribution of the difficulties of the particular set of items used in the test. This matter is discussed in Chapter 6.

A similar, but less technical meaning is ascribed to objectivity in testing. Conventionally, objectivity is assured by creating tests and test items so that candidates' responses can be scored consistently by anyone who marks the test. That is, the subjectivity of judgments by different raters is minimised. This is the sense in which objectivity differs from reliability. Reliability focuses on the consistency with which different sets of items lead to similar judgements of student performance while objectivity deals with the influences of different judges. Objectivity is assured most easily by using structured response (e.g., multiple-choice) formats for test items. Alternatively, constructed responses can be used provided the rules for scoring those responses are clear and comprehensive (Choppin, 1997).

However, Sadler (1986) argued that subjectivity or objectivity was not a property only of the final act of judging or scoring a piece of work. The decision to test certain characteristics, the choice of the assessment method, and the criteria selected are all judgments that precede scoring. That is, in any assessment, a series of judgments is made about what to assess, how to assess it, under what conditions, against what criteria or standards, and, finally, about the quality of the work against agreed standards. In large-scale testing, expert panels are established and processes developed to ensure consistency in the judgments that are made. While all students participate in the chosen assessment methods, students are not necessarily affected equally by those choices. Thus, objectivity may be more ephemeral than is commonly believed and it may be the case that other criteria, such as fairness, need to carry more weight in the selection of assessment methods, especially for complex skills.

Where open-ended responses (e.g. essays) are assessed, variability in judgment by different raters is constrained by the use of scoring schemes, including rubrics.

These are less prescriptive than the rules that may be used to score simpler constructed responses and some post-test checking is required to ensure that objectivity is achieved. This is done by designing marking arrangements so that at least a sample of student responses is marked by multiple assessors. The marks allocated by different markers to different questions on student scripts are then analysed to ascertain whether there are differences in the severity of markers or the difficulty of questions and whether there are interactions between markers and questions (Andrich, 1997a). Where differences are found, marks are calculated taking into account differences in rater severity to ensure that students are treated fairly in the assessment, and feedback to markers can be used to achieve greater consistency in future assessment rounds. This approach is also useful in accommodating the possible effects of rater and question variability so that students who answered different combinations of test items can be compared and graded fairly. In relation to performance assessment, Shavelson, Gao and Baxter (1993) found that, with the support of scoring rubrics, assessor variability was acceptable but that different tasks contributed substantially to score variability. Attention needs to be paid to the objectivity and fairness of those who assess student work and of what tasks are undertaken for assessment.

Feasibility

"Feasibility means capable of being done, with the connotation of convenience and practicability in the doing. While many things are doable, fewer are feasible" (Matters & Curtis, 2008, p. 15). The feasibility of a problem solving assessment requires that all aspects of the assessment meet the criteria of cost- and time-effectiveness and that the benefits flowing from the assessment justify the costs and time required. These aspects include development and validation of assessment instruments, administration of the assessment, gathering students' responses, marking responses, assigning overall grades, recording and analysing student grades, and reporting and maintaining records of student achievement. For standardised assessments, Murray (2003) estimated a practical time limit of 90 minutes for the student testing component. Having students complete a test (or assessment task) is only one component of an assessment regime, but greater time may be justified if multiple purposes are served by an assessment activity; for example, if employability skills assessment are integrated with existing assessment activities.

The interests of a range of stakeholders need to be considered. Institutions (schools, vocational colleges and universities), teachers and students are all stakeholders with direct involvement in the assessment of problem solving skills.

For these individuals and groups, the time and resources committed to the assessment must be perceived to be worthwhile. Other parties, notably parents and potential employers, have an interest in the testing and its outcomes. A net benefit for each stakeholder group needs to be demonstrated, and the benefit clearly must be higher for those assessment and reporting methods that are more resource-intensive. Direct benefits for students and employers include the existence of a report on problem solving skills achievement. An indirect benefit may be that the act of having these skills assessed signals their importance and leads to higher levels of problem solving ability in the student cohort.

The feasibility of assessment and reporting methods may be influenced by a 'backwash' effect – the extent to which the methods influence curriculum content and practices and student responses to the assessment context (Boud, 1995a; Tang, 1994). A positive backwash effect may justify assessment and reporting methods that require greater effort and time than others while a negative effect may militate against those methods.

Summary

Alternative assessment methods may be compared by examining their fitness for purpose. Fitness for purpose consists of judging the validity, reliability, objectivity and feasibility of the assessment. These constructs have subordinate elements, such as fairness and backwash. In the discussion of assessment purposes, both summative and formative goals are identified. Validity is of paramount importance, because if an assessment is not valid – perhaps because certain key elements of the construct are missed or because the assessment is contaminated by unrelated constructs – then what is reported does not reflect the intended construct, no matter how precisely it is measured.

If the main purpose of an assessment is to report individual scores for high stakes purposes, such as selection and sorting, reliability becomes a major criterion. It may be necessary to accept that only part of a construct is assessed in order to ensure that the results reported are comparable. In these assessments, objectivity can also assume high importance, largely to contribute to the perception or warrant of fairness in the assessment process.

In formative assessments, where the development of an ability is desired over a point in time estimate of a candidate's achievement, a lower level of reliability than is required for high stakes summative assessment can be accepted. For this assessment purpose, backwash effects – enhanced through designed feedback –

may be accorded greater significance than they need to be in assessments serving summative purposes.

In the next section of this chapter, several approaches to the assessment of generic skills are examined through examples and they are compared using the criteria of validity, reliability, objectivity and feasibility.

Examples of generic skills assessment

A search for examples of generic skills assessments, with a focus on problem solving, was conducted and selected examples are reviewed below. When the review was undertaken, few examples of the assessment of problem solving as a generic skill were located, although there are many examples of narrower conceptions of problem solving, especially in mathematics education. Since that time, two major international programs have implemented assessment of problem solving (OECD, 2004; Reeff, Zabal, & Klieme, 2005). Although the current study is restricted to the assessment of problem solving, it is desirable that any method that is adopted needs to be adaptable to other generic skills and the review presented below is of assessment methods in general.

The theoretical basis of the constructs being assessed is considered in reviewing assessment methods and examples of them. Some examples were based on a clearly articulated theoretical account of the construct, for example the Graduate Skills Assessment, while others, for example the Employability Skills Profiler, were not. The point is made in Chapter 2 that one of the reasons for the lack of success in implementing generic skills schemes was the lack of a clear definition of the target skills. In order for assessment to meet the requirements of validity, it is necessary to articulate the target construct and to ensure that assessment tasks adequately reflect it (Messick, 1995).

The various approaches to generic skills assessment are categorised as:

- standardised assessment;
- common assessment tasks;
- performance assessments;
- teacher judgment; and
- portfolio construction.

Each of these approaches is defined and examples of them are cited and described briefly here. The examples are described in more detail in Appendix 3.

Standardised assessment

Standardised tests comprise items for which students select responses from prescribed options (typically multiple-choice items) or for which students provide limited constructed responses. These items are developed according to item specifications that include the particular constructs to be tested and the scope and range of abilities that are being assessed. The broad scope of a test is subdivided into small units of knowledge or skill and each element is assessed by a number of discrete items. Test items are evaluated by expert panels then trialled and are accepted, rejected or modified before final testing.

Where constructed response items are included, e.g. items eliciting short answers, raters are trained to recognise key features in responses and multiple raters are used, at least on a subset of scripts, to check that inter-rater reliability reaches an acceptable standard.

Although many judgments are made in delimiting the scope of the test and in prescribing the scope and range of items, variability in grading responses is either eliminated (for multiple-choice items) or minimised (in grading constructed responses).

The results of trial and final tests are analysed statistically to ensure that acceptable criteria are achieved, usually in relation to reliability, but validation studies are also conducted to check that test scores correlate with constructs that are thought to be related and to other criterion performance measures.

Standardised assessment, often in the form of multiple-choice tests, has many critics. A common criticism is that the selection of responses tests much lower level cognitive skills than does the generation of a response. Calfee (1994, pp. 345-6) contrasted externally (to a school) and internally mandated assessment approaches. His comparison is reflected in Table 15.

Table 15: Calfee's comparison of externally and internally mandated assessment

Assessment facet	Externally mandated	Internally mandated
Origin	Development and validation of assessment methods by a central agency responsible [accountable] to a top level policymaker.	Development and validation of methods by a professional community of teachers directly responsible to themselves and their clientele.
Methods	Adherence to standardised procedures and routinised administration; professional judgment is neither needed nor allowed.	Reliance on procedures springing from a shared understanding of curriculum and instruction, procedures that are adapted according to situational context.
Outcomes	Cost-effective (i.e. cheap) methods yielding simple numbers that either pass or fail a set criterion.	Case-effective (i.e. expensive) methods requiring informed judgment and yielding complex "portraits."

Source: Based on Calfee (1994, pp. 345-6)

He contended that multiple-choice testing reflected a behaviorist orientation to learning, that this orientation was incompatible with the sorts of learning that were valued, and that authentic performance assessments were compatible with cognitive models of learning and were more conducive to promoting such learning (p. 346). He addressed the need for assessment for learning, but did not acknowledge the need for summative assessment, especially at the system level, and did not acknowledge examples of effective practice in standardised assessment.

Examples: The Graduate Skills Assessment (ACER 2000; 2001a) was developed by ACER and administered on a trial basis over several years. In the development of this assessment, attention was paid to the characteristics of the constructs. For example, the developers used work of Polya (1957) and of Bransford and Stein (1993) in planning the assessment of problem solving (Hambur, Rowe, & Luc, 2002, pp. 16-18). The items used in the assessment of problem solving presented prompts (often scenarios) followed by one or more questions based on the information provided in that prompt. Example items provided for the test include scheduling tasks in which information about bus departures or individuals' timetables must be read an understood and plans developed to enable efficient travel or meetings to be scheduled. Successful performance on these tasks requires the various problem-solving processes to be enacted. Where there are several questions for a single prompt, they can be graded so that the easiest questions require information encoding (problem representation) while the more difficult questions demand the generation and application of strategies and possibly their evaluation (Australian Council for Educational Research, 2003). As students select responses from a list of five options, it is not possible to observe the application of problem solving processes. The application of problem solving processes is inferred from students' scores on items.

The Programme for International Student Assessment (PISA) project provided an example of the processes that surround standardised testing. Sample test items were available and reports based on the testing were published (see, for example, Thomson, Wernert, Underwood, & Nicholas, 2008). As is the case with the GSA, problem solving in conceived as the application of problem solving processes (OECD, 2004, pp. 26-31). Problem solving performance is inferred from scores on sets of test items. In reporting problem solving performance, three levels are identified (with an additional level below performance level 1). The highest level is labelled the 'reflective, communicative problem solver' (OECD, 2004, p. 29). Invoking communication skills as an element of performance is problematic. Communication is identified as one of the processes involved in problem solving (see Figure 2.1, OECD, 2004, p. 29), and in 'real-world' problem solving, there would be a requirement to communicate or in some other way act on a problem solution. However, in this form of assessment, it introduces another generic skill into the assessment of problem solving and inferring communication from selected-response items would appear to require a substantial extrapolation.

An alternative approach to generic skills testing was implemented in Australia in the Employability Skills Profiler. This was an unusual approach to generic skills assessment because it was based on identifying basic skills and dispositions that were thought to be implicit in the Employability Skills Framework (ACCI & BCA, 2002) then locating and adapting existing standardised assessment instruments for each of these components. The final example was developed in England, namely the Key Skills Certificate. Strictly, this was not a standardised test: it included a standardised test and a school-based component. The latter was moderated and the results of both components combined to generate a common score (Hodgson & Spours, 2000; Powell et al., 2003; Pumphrey & Slater, n.d.).

Common assessment tasks

Common assessment tasks, because they are tasks not tests, typically allow for the assessment of a wider range of skills and practices than standardised tests, in a range of formats and settings that more closely approximate how people function in the wider world outside the classroom. Standardised tests are closely controlled in their development, implementation and marking. Common assessment tasks are designed or selected so that they provide opportunities for students to demonstrate (and possibly develop) the constructs that the tasks are intended to assess. The standards by which student performances are judged are also established when the suite of tasks is selected.

Whereas the marking of standardised tests is often automated, the assessment of common assessment tasks requires expert judgment, typically by the teachers involved in the implementation of the task. The credibility of the grades awarded to students across (and even within) sites depends upon teacher–assessors having a shared understanding of the standards. A moderation process is required to ensure fairness and comparability of the assessment across schools. The process also helps to create common understandings among teachers from different locations. The moderation provides a basis for the fairness and credibility of the reported results. In high stakes testing, scoring is conducted by raters trained in the use of the rubrics and normally a sample of responses is assessed by at least two raters so that inter-rater reliability can be checked.

Example: Problem solving assessment tasks were developed for the International Life Skills Survey (Herl et al., 1999). In this example, two tasks were trialled. Common assessment tasks have also been developed and trialled in Queensland schools as part of the New Basics program (Queensland Department of Education, 2003) and the Core Skills Test (Pitman, Matters, & Nuyen, 1998).

Performance assessment

"Performance-based assessment is a type of testing that calls for demonstration of understanding and skill in applied, procedural, or open-ended settings" (E. L. Baker, O'Neil, & Linn, 1993, p. 1210). Performance assessment has been used in judging activities that do not normally leave an artefact that could be evaluated later. Examples have included gymnastics and dance, but the act of making an object was also a performance that could be evaluated separately from any assessment of the object that was produced. Performance assessment has been extended to a wider range of activities, including science laboratory classes, medical diagnoses and building brick walls. In these cases, products existed that were evaluated, but generic skills are more likely to be observed in their execution during construction of the artefact than in the artefact that is produced. It is the act of producing, rather than the product, that is being judged.

Performance assessment was characterised by the use of open-ended tasks with responses constructed by the student rather than being selected from defined options and by a focus on complex skills in a curriculum context, and were judged in terms of domain specific criteria rather than against a generalised trait (E. L. Baker et al., 1993, p. 1211). The last of these claims is a challenge to the generalisability of any inferences that may be drawn from a performance assessment. One goal of assessment is to judge and provide feedback about a specific activity, but if that feedback cannot be used in subsequent performances, it is of little value to the learner.

Mumford, Baughman, Supinski and Anderson (1998) disagreed with the evaluation given by Baker et al. They pointed out that:

> Performance assessment exercises are often designed to assess several skills and putatively provide a more comprehensive description of the individual's performance capabilities than do traditional methods. When these characteristics of performance assessment are considered in the light of the demonstrated predictive validity of at least some performance assessment systems, they appear to provide a compelling argument for the use of this approach. (p. 79)

Their contention called into question those characteristics that distinguished the more effective performance assessment systems. They specified four criteria for effective performance assessment. First, developing viable measures of complex performance skills required defining the nature of a specific skills and the way that skill was applied in performance. Second, assessing these skills might be accomplished using simplified performance tasks expressly intended to elicit expressions of the target skills. Third, performance on the simplified, low fidelity simulations should be structured to elicit the crucial component characteristics of skilled performance. Fourth, scoring systems should be designed to capture these component characteristics of skilled performance (Mumford et al., 1998, p. 82). The first and last of these criteria suggest the need for a construct based approach to the definition of the skills that are being assessed.

Performance assessment differs from holistic teacher judgment (see below), since performance assessment co-locates the act of judgment of skills with a specific performance by a student. It differs from the Common Assessment Tasks model in that the choice of tasks is made by the teacher, or possibly the learner, rather than using prescribed tasks. Moreover, performance assessment is argued to be more authentic than pencil and paper alternatives, although the claim for authenticity depends upon the setting in which the performance is observed. The validity of the assessment depends upon the context of the performance. In workplace assessment, the workplace practice is what is being performed and assessed, so the validity is high, provided the criteria used to judge the performance do indeed reflect the desired characteristics of the intended practice. In school-based performance assessments, the validity can be judged by comparing the tasks and contexts with a criterion domain of practice.

A potential disadvantage of performance assessment is its reliability. If the performance is judged by a lone rater, no information is available about the reliability of the score. If multiple raters are used, their ratings can be compared, and if there is close agreement the score can be accepted as reliable. For high-stakes assessments, such as Olympic contests, multiple trained judges, strict criteria and scoring protocols are used. The complexity and cost of ensuring this level of reliability across schools and jurisdictions is likely to be prohibitive and thus is not feasible. The reliability of performance judgments can be improved by the development of descriptive standards and consensus judgment by small panels within schools.

Shavelson et al. (1993) argued that the assessment of achievement on a particular task reflected the ability of the student, but it might be compromised by the severity of the person making the judgment and by the affordance of the task. The use of rubrics or other assessment tools may help to minimise the rater severity variability. The variability associated with the characteristics of the task may be minimised by requiring judgment across a range of tasks or by the use of prescribed tasks. (See the section on common assessment tasks.)

Herl et al. (1999) provided an example of performance assessment of problem solving. They used two tasks, each involving three sub-tasks. Their data suggested that males and females responded differently to the two tasks, thus drawing attention to the need to ensure that tasks are fair to all participants.

Examples: An example was provided by the validation of problem solving measures for the International Life Skills Survey (Herl et al., 1999). A second example was provided in the assessment of creative problem solving (Mumford et al., 1998). The assessment of Core Skills in Queensland (Matters, 2005; Pitman et al., 1998) provided a third example of the performance assessment model.

Teacher judgment

Teacher judgment has featured prominently in the literature on assessment and has been shown to be central to almost all forms of assessment. An obvious exception is the multiple-choice format, although in that format, judgments are made when the tasks and response alternatives are chosen. An issue raised in the literature reviewed above is a restriction of the 'guild knowledge' of standards and their application to a community of assessment practitioners (Sadler, 1989, p. 129). This is shown to have implications for feedback and for a role for students in self-assessment. If individual teachers are not privy to this information, they are constrained in the quality of feedback they can provide.

In many situations (other than multiple-choice assessments), teacher judgment could be scaffolded, for example by scoring rubrics, which might increase the consistency of judgments between teachers and increase the reliability and fairness of the assessment. In other contexts, teachers' judgments could be reviewed through moderation processes, a method used commonly in school-based assessment (see, e.g., International Baccalaureate Organization, 2004, pp. 36-37; SACE Board, 2009).

The assessment formats and examples considered above involve the assessment of particular tasks. The form of assessment considered in the approach being considered here is a global summary judgment made after observations of students on many varied tasks and made by groups of teachers.

Teachers may make individual judgments based on either the explicit or the inferred assessment of generic skills in existing courses. These judgments may be based on students' achievement of content goals, the attainment of which are thought to depend on students having the generic competencies. This is referred to as the inferred assessment of generic skills (Ratio Pty Ltd & Down, 2003, pp. 42-43). Teachers' judgments may also be based on observation of the application of generic skills. For example, an assignment may require a presentation to the class. In addition to assessing the content of the presentation, teachers may assess aspects of the student's communication skills.

Groups of teachers may consider student performance and seek to reach a consensus view of student achievement. In this case, teachers meet and consider the employability skills of individual students whom they have taught or otherwise interacted with in co-curricular activities during a school year. Teachers consider each employability skill in turn, and describe the evidence they have been able to gather that illustrates each student's achievement of that skill. This process may be supported by a rubric describing the behaviours likely to be observed in students performing at different levels of the targeted employability skill. Judgments by teacher groups that seek to achieve consensus has led to consistency in reported grades (McCurry & Bryce, 1997).

The teacher-group judgment method is subject to several limitations. First, it assesses student performance in school contexts and depends upon teachers having regular contact with students in those contexts. Students may develop and demonstrate skills in other contexts, for example through part-time paid work or through community involvement independent of their school activities and these contexts are generally unavailable to teachers. Second, the standards against which teachers make their judgments may differ between schools. To overcome this limitation, some form of moderation is required, but this is difficult as it is unlikely that there is any trace of the activities that teachers in a school use to form their views and that the teachers can share with moderators. Third, the method (as described by McCurry & Bryce, 1997) occurs some time after students have been observed and so any feedback is likely to occur long after the observed performance and therefore to be of limited value to students. Fourth, although the method has been shown to yield consistency between teacher judges within schools, it appears unlikely to work in tertiary education contexts where teachers have very little contact with students out of scheduled classes and where some classes are quite large and teachers have little opportunity to observe individual students.

Example: An example of consensus judgment by groups of teachers is provided by Victorian Board of Studies Key Competencies levels assessment trial (McCurry & Bryce, 1997, 2000).

Portfolio construction

Portfolio construction is distinguished from portfolio assessment. The distinction is an important one in the Australian context as the construction of portfolios, but not their explicit assessment, was recommended for the 'recognition and reporting' of employability skills in schools, vocational education and training, higher education, and workplaces (Allen Consulting Group & National Centre for Vocational Education Research, 2004, p. 1). A line of argument pursued in this chapter is that assessment is a key driver for learning, both directly and indirectly. If portfolios are constructed, but not assessed, there appears to be little motivation for students to engage with the content expected in their portfolios. Further, in the absence of assessment, there is no need to articulate standards and certainly no opportunity to apply them, by either teachers or learners, and no motivation to provide feedback and to act on it to enhance performance.

The construction of a portfolio is the selection and aggregation by individuals of evidence of their own achievement of particular skills (which may include employability skills). Portfolios may be paper-based or electronic. Electronic versions are popular because of the ease with which they can be updated. Templates can be provided, and electronic portfolio systems may have facilities, such as filters, that enable students to select and present views of the contents in print or electronic form for specific purposes and audiences.

Having constructed a portfolio, a student may submit it for assessment, although portfolios are often used simply as devices for recording supportive evidence that may be used for job applications.

Two approaches to portfolio assessment were described by Troper and Smith (1997). The first was the Michigan Work Readiness Portfolio. This was a framework for evaluating students' analyses of how their activities revealed the skills student claimed to have developed. The framework comprised three categories, namely academic skills, personal management skills, and teamwork skills. Raters were trained to judge items in the portfolio as either credible (having been produced by a third party rather than the student) or not, and to recognise whether evidence for a skill was present (present or not). Dichotomous judgments were used to limit the amount of time taken in preparing portfolios and in making judgments about them, but this limited the level of evidence that was presented.

A second scheme was developed by the Center for Research in Evaluation, Standards and Student Testing (CRESST). The scheme consisted of a rating scale comprising five main elements (general impression; communication skills; personal management skills; interpersonal and team skills; and thinking, problem solving and technical skills). The instrument included 16 items and for each a seven point rating scale (poor to superb) was used (Troper & Smith, 1997, p. 362).

Troper and Smith (1997) expressed some reservations about the validity and reliability of portfolio assessments. One of the problems was that if the portfolio were assessed as an object along with its content, then a construct other than the target one was brought into the judgment, threatening the validity of the assessment. An argument advanced in Chapter 2 and 3 is that there has been a failure to define generic skills as overt constructs and much of the discussion in the second part of Chapter 3 seeks to build a psychologically valid conception of problem solving. However, even the CRESST assessment framework described by Troper and Smith is rather general having 16 items to cover five broad concepts and appears not to provide a basis for the valid assessment of the three generic skills that were identified (communication, teamwork and problem solving).

A further problem occurs if different templates provide different levels of support to students. Some templates may be very comprehensive and require students merely to ‘fill in the blanks.’ This approach engenders low levels of student engagement, and this must limit the learning that may occur in developing a portfolio. Assessing portfolios in ways that encourage student learning is time-consuming, and if they are to be useful resources to students, they need to be assessed repeatedly with feedback provided to guide improvement.

Example: The Conference Board of Canada developed a template to assist students in developing portfolios of evidence of their employability skills (Conference Board of Canada, 2000b).

Summary evaluation of potential assessment methods

The methods that are described above are evaluated against criteria established to enable the selection of assessment methods that are fit for purpose. The outcome of the evaluation is summarised in Table 16. The evaluation is general, in that it compares the alternative assessment methods for a range of domains and is not restricted to the assessment of problem solving skill, although where they could be found, examples of the method used for problem-solving assessment are cited. It is argued that effective assessment methods are general and that such practices can be applied to the assessment of problem solving.

If the purpose is summative assessment followed by reporting at a system-wide level, standardised testing is the method of choice, albeit with some limitations on the aspects of problem solving that can be assessed. If summative reporting of performance at an individual level is required, a similar conclusion may be reached, although attention needs to be given to the standard errors of the assessment to ensure that they are within acceptable bounds, given the purposes to which the results may be put. For high stakes purposes, such as selection, the precision of measures derived from assessment needs to be high. If the report is intended to assert that a criterion level of performance has been reached, the assessment tasks can be targeted to that level and provide a sound basis for reporting.

If the purpose is formative and the focus is on the development of the skills, then it is more likely that common assessment tasks or performance assessment can be pursued. These tend to have high validity, although their reliability does not meet the standards required for large scale, high stakes assessments. Both methods, especially performance assessment conducted in the context of students' courses, provide opportunities for feedback and both can be adapted for use in self-assessment.

Neither teacher group judgment, which is a summative assessment method that does not provide reliability and objectivity between sites, nor portfolio construction meet the criteria of supporting timely feedback. Portfolio construction can incorporate self-assessment, but there is potential for the assessment to focus upon the quality of the portfolio as an artefact rather than on the substantive set of skills that are meant to be substantiated by the portfolio.

The potential of methods for the assessment of problem solving

A key concern raised in the thesis is the need to ensure that each generic skill should be defined as a construct. A failure to do this is implicated in a lower impact of the key competencies initiative than employer organisations had anticipated. Further, it is shown that validity demands that the assessment should focus upon all attributes of the target construct. For this reason, it is necessary to define the construct in order to develop assessment tasks and methods for it.

In the examples cited above, standardised assessment, the use of common assessment tasks and performance assessment methods can incorporate a clearly articulated construction of problem solving. This has not been shown in relation to teacher judgment or portfolio construction, although it appears that both methods could incorporate a clear definition of problem solving.

Table 16: Summary of the application of evaluation criteria to prospective assessment methods

Criteria for evaluating assessment methods	Assessment methods Standardised assessment	Common assessment tasks	Performance assessment	Teacher group judgment	Portfolio construction
Validity	The validity of standardised assessment is questioned. Tasks are largely pencil-and-paper ones that may not enable all aspects of the construct to be assessed. This assessment format does not permit contextual assessment of problem solving, as tasks common to all students, irrespective of their courses, need to be provided.	The careful selection of tasks so that they provide opportunities for students to apply and develop their problem solving skills can enhance the validity of this assessment method. Tasks that are common across a range of courses would compromise the authenticity	Provided tasks are selected or constructed to provide opportunities to develop and apply the target skill, the method needs to have high validity. The method gives rise to authentic assessment, as the target construct is being developed in the context of the students' disciplinary or vocational domain.	If teachers are made aware through professional learning activities of the scope of the target constructs, the validity can be moderate to high.	The validity is called into question because of the likely attention paid to the portfolio as a product rather than directed to the target construct.
Reliability	This assessment method normally has the highest level of reliability. Careful attention to item design followed by piloting normally leads to highly reliable assessment.	Assessment is shown to have low reliability at the individual level unless many tasks are provided. However, as the number of tasks is increased, the costs of administering and rating student performance rises sharply.	The method is unlikely to have the reliability of the numerous small scale tasks typically administered in standardised tests. There is likely to be some variation in the affordances of the tasks for the target construct (problem solving) and this reduces the reliability.	Within schools, the consensus judgement forum is likely to lead to high levels of reliability. However, this cannot be guaranteed across sites, so the net reliability is likely to be low.	Reliability is likely to be quite low. This problem can be addressed be developing rubrics for aspects of the portfolio.
Objectivity	Such tests are highly objective. No student is either advantaged or disadvantaged by the test, and objectivity can be verified through pilot testing in which any systematic bias in items can be detected and rectified.	With well-developed scoring rubrics accompanied by the use of multiple raters, at least for samples of scripts, can reduce any lack of objectivity. Methods are available to control for systematic differences between raters.	Likely to be medium to high provided well-developed standards are established.	The assessment may not be objective. Halo effects may be apparent in teacher judgments. This can be overcome to some extent through the in-school moderation of the consensus judgment.	This assessment model is unlikely to yield objective assessment as each student produces a unique assemblage of evidence to be judged.

Table continued over page…

Table continued

Feasibility	On a large scale, this assessment approach is feasible. Establishment costs are high, as the infrastructure for the testing needs to be established. The design and development costs of assessments are high, but once these costs have been met, large scale assessment can be conducted at a low unit cost.	The method is feasible. If the intention is to sample population performance, this method is feasible. If the number of tasks has to be increased to provide more reliable estimates of individuals' performances, the costs rise sharply and the method becomes more difficult logistically.	Review of existing tasks is necessary. If they do not provide opportunities for the application of problem solving skills, they do not enable their development. New tasks may need to be developed.	The method is feasible and imposes a low to moderate load on teachers.	Simply developing a portfolio as a vehicle to assemble and present evidence imposes a low load on schools. If the portfolios are to be judged, the assessment load is likely to be quite high.
Other considerations	This assessment method does not provide opportunities for feedback, nor for self-assessment. A low level of learner engagement in these tasks is anticipated. No backwash effect is expected to arise from this method of assessment.	The method can provide feedback to learners as scripts are returned. The feedback is unlikely to be immediate. Self-assessment can be used in that students can be asked to assess their work using the same criteria as raters prior to submitting their work for assessment. There is some doubt that any backwash effect can influence the development of the target skill in the students' primary learning domain.	The use of established tasks enhances the authenticity of the assessment. They are amenable to both self-assessment and frequent feedback. Backwash effects may be substantial.	The method is unlikely to work in settings other than schools. It is a summative assessment model and does not provide scope for feedback. Backwash can be established if schools provided opportunities for students to engage in activities that are likely to lead to favourable judgments and informed students of these opportunities and of the judgment process.	Self-assessment is implicit in this model because students need to make judgments about what to include and how to relate what is presented to target constructs. The method is likely to have very little impact on programs offered to enhance learning of the target constructs.

Implications for assessing problem solving in the current study

The preceding discussion focuses on assessment purposes, identifying two dominant approaches to assessment, formative and summative, that subsume several related sets of purposes. Summative assessment provides a basis for reporting student achievement at the conclusion of a period of learning. It is important when assessment is used to rank students for selection. This occurs when students apply for entry into higher-level courses, for example when completing secondary schooling and applying for admission to higher education courses, and it is used in employment selection and recruitment decisions. The results of summative assessments of individuals can be aggregated to provide information on system performance.

Formative assessment, in which students receive feedback on their performance, has been shown to be effective in promoting learning (see especially, Black & Wiliam, 1998a). Several elements contribute to the effectiveness of formative assessment in promoting learning, namely feedback and self-assessment. Feedback is most effective when it occurs frequently and when the standards against which performance is judged are overt and readily accessible to learners. That is, learners are likely to gain from feedback when they appreciate the gap between their performance and the desired goal. Self-assessment has received little empirical attention, although there is some evidence for its effectiveness.

Performance standards must be established, for without them, there is no basis for consistent and reliable assessment. How performance standards are established has not received due attention. Rubrics are often specified for specific assessment tasks. These are very useful for teachers and other assessors and for students. Explicit rubrics specified in terms of expected learning outcomes for particular tasks are likely to have immediate salience for learners. Learners need to be able to set goals and they need to know whether they have achieved the intended outcomes. However, if the focus is on the learning of complex cognitive skills, it is very likely that students need to apply their skills over many tasks. The specific feedback on one task may not generate information that learners can use to help them with subsequent tasks. This raises a dilemma. Specific criteria (specified in rubrics) are helpful on individual tasks, but the feedback that is generated from comparisons of current performance with specific goals may not generalise. General standards that may apply to a range of tasks may not be immediately useful to learners. A solution is to establish a principled basis for the specification of standards. This requires two elements. First, the complex cognitive skill –

problem solving in this study – must be defined as a construct whose components are clearly identified. Second, a generalised basis for evaluating cognitive performance must be established. Two candidates for the generalised framework, Bloom's Taxonomy and the SOLO Taxonomy are evaluated. It appears that the SOLO Taxonomy is likely to be the most productive framework for establishing generally applicable standards.

In the preceding sections, criteria for the evaluation of assessment methods are reviewed, including validity, reliability, objectivity and feasibility. Elements of these criteria include authenticity, fairness and backwash effects. These criteria are applied in evaluating assessment methods. Many approaches to assessment are examined and categorised as: standardised assessment; common assessment tasks; performance assessments; teacher judgment; and portfolio construction. Examples of each category are presented in Appendix 3. The categories are not neatly separated. For example, there is potential for overlap between common assessment tasks and performance assessment and several types of assessment lead to products that may be included in a portfolio.

Each of these methods can be applied to the assessment of problem solving, and indeed most have been (Australian Council for Educational Research, 2001b; Conference Board of Canada, 2000b; Herl et al., 1999; McCurry & Bryce, 1997; Mumford et al., 1998; OECD, 2004; Pitman et al., 1998). However, as argued above, there have been some limitations to their implementation.

Summary statements on assessment

Together, consideration of purposes for assessment, criteria for judging assessment methods and existing examples of assessment can lead to a set of propositions about assessment that can guide the current research study. However, the statements below are no more than propositions: all inform the study, but only two are the subject of specific investigation in this study. Those that are not tested are important claims that emerge from the literature on assessment, and they do warrant empirical investigation. For some of the propositions, evidence does exist, but often not in relation to the assessment of generic skills. It is hoped that other researchers may see these propositions as worthy of detailed investigation.

A key goal of problem solving assessment is to enhance students' performance

Assessment may be undertaken for formative or summative purposes, and in some circumstances, perhaps both. Summative assessment of generic skills has been undertaken and examples are given below. In the current study, the focus is on formative assessment and on testing the proposition that formative assessment of problem solving does lead to enhanced performance. While summative assessment is important, and in programs such as PISA, provides valuable information on the relative performance of education systems in different countries, there is also a case for attempting to increase the stock of these skills rather than simply reporting current levels of them.

The literature on lifelong learning and on generic skills indicates that increasing the stock of human capital is important for the competitive performance of advanced economies. (For a recent discussion of the role of cognitive skills in economic performance, see e.g., Hanushek & Wößmann, 2007). Skills that are specific to particular occupations are important, but change is a feature of dynamic economies and enhancing the generic capacities of the future labour force is an important goal.

At the individual level, ensuring that learners develop high levels of generic skills is extremely important. Methods for encouraging students to develop generic skills – and in this study, their problem solving skills – need to be advanced. Generic skills can be taught and learned, but they are likely to be acquired in the contexts in which students learn the substantive content of their disciplines and vocations. Formative assessment is part of the solution to the issue of encouraging the development of generic skills.

Assessment is a driver for learning and teaching

In research in the vocational education and training sector, students indicate that if skills are not assessed, they may not attend to them (Callan, 2002).

Assessment can be an indirect or direct driver of enhanced performance. If teachers are aware that students may be assessed on problem solving, perhaps using an external test like the Graduate Skills Assessment, they may pay greater attention to problem solving in classroom instruction. This approach uses assessment as an indirect driver in that teachers have no immediate role in the assessment. If teachers become involved in the assessment, it seems likely that their classroom practices may address more closely the constructs that the teachers assess. But students, too, need to become involved in the assessment of their own skills. That is, assessment may drive both teacher and student behaviour.

Formative assessment leads to learning gains

Compelling evidence that formative assessment can lead to substantial learning gains is presented above (see Black & Wiliam, 1998a, 1998b). Formative assessment necessarily involves feedback to learners about their developing skills, and for feedback to make an effective contribution to learning, students need to engage sequentially in learning tasks in which the feedback from one task feeds into their performance for the next task.

Black and Wiliam (1998a, pp. 9 & 13) argued that, for assessment to be ecologically valid, it must be conducted by the students' regular teachers and it needed to address learning tasks that were part of the normal curriculum. This argument applied also to the issue of authenticity raised by Wiggins (1989). Thus, for formative assessment to be effective it must be integrated into routine classroom instruction and learning practices.

It appears likely that, for formative assessment to deliver the gains that are claimed for it, and that are found in some instances, formative assessment procedures including task selection, setting performance standards and providing feedback need to attend to key aspects of the target construct. In relation to problem solving, the feedback that is critical to the success of formative assessment should focus on identified problem-solving processes (e.g., those identified by Bransford & Stein, 1993).

Self-assessment leads to enhanced learning

Boud (1995b; 2002) argued a sound case for the greater use of self-assessment and Black and Wiliam (1998a) presented some empirical evidence of its effectiveness. Self-assessment should encourage students to engage with the standards that have been established to describe possible levels of performance. It is anticipated that repeated engagement with self-assessment of a sequence of formative learning tasks contributes to the development of the skills that students are asked to assess.

Feedback on the substantive content assessed and through students' self-assessment leads to improved performance

In addition to feedback on the substantive content of their courses, feedback through students' judgments of their performance against standards may help students to perceive any gaps between their performance and the desired standards. An assessment mechanism that includes this feedback appears likely to contribute to enhanced performance over time.

Self assessment of problem-solving performance needs to focus on identified problem solving processes and clearly articulated performance standards. Feedback on the substantive content of the assessment tasks and on students' judgments of their performance on problem-solving processes against standards is expected to lead to improved problem-solving performance.

A generalised standards framework needs to be applied across tasks

In order to establish consistent standards across a diversity of tasks, a general framework is required. The SOLO Taxonomy is believed to provide a fruitful framework. That framework is applied jointly with an information-processing model of problem solving to provide construct based assessment standards.

Performance assessment integrated with the substantive course content leads to improved generic skills performance

Performance assessment is seen to be a method for integrating the formative assessment of generic skills into the discipline or vocational context set in the design of students' courses. Other assessment methods are rejected as not complying with one or more of the requirements for skills development through assessment. Standardised assessment, while meeting desirable and necessary criteria for large scale summative testing does not provide a basis for formative assessment with feedback. The construction of portfolios, which may be assessed, is regarded as excessively time consuming for students, and if they are assessed adequately, time consuming for assessors. Although judgments by teachers is fundamental to almost all assessment methods, and is essential in providing informative feedback to students, the 'teacher group judgment' method is found to have limitations. It is likely to provide a sound basis for school reports, but not to be applicable beyond schooling. Common assessment task and performance assessment can provide contexts in which the target construct – problem solving-can be developed and assessed. If common tasks are set outside the context of a course, it appears likely that these tasks may favour those students whose courses most closely match the content of the task. If the tasks are common to students within courses, and if students' problem solving performances on those tasks are a focus for assessment, the requirements of ecological validity and authenticity can be met. However, it does create a question about the comparability of grades of student performance across courses. Within-course performance assessment does not deliver objective grades for generic skills achievement that has the same meaning across different courses and disciplines. Indeed, this appears to be a critical point in comparing assessment methods.

If the main purpose of assessment is to generate comparable results between students, schools and countries on a common scale, then standardised assessment methods such as those used in the PISA assessments appear to be necessary. If, on the other hand, the main purpose is to foster the development of generic skills, it appears that performance assessment within course contexts is the preferred option.

The main purpose identified in this study is contextual development of problem solving skills, so cross-course comparability is not seen to be a core requirement, although it is desirable.

Chapter 5: Summary of Research Themes and a Strategy for Their Investigation

In this chapter, key issues raised in Chapters 2, 3 and 4 are reviewed. They key issues identified in Chapter 2 are the lack of a definition of generic skills as a class of constructs and of each proposed generic skill, and too little attention being paid to their assessment. It is necessary to examine these issues together, as effective assessment depends upon appropriate definition. The issues of definition are addressed in Chapter 3, with a focus on problem solving, and of assessment in Chapter 4. Following this summary, a set of propositions is advanced and these frame the inquiry reported in this thesis. In this chapter, these issues are summarised and a plan for investigating them is outlined.

Critical issues in implementing generic skills

The two major Australian generic skills initiatives are the Key Competencies (Mayer Committee, 1992) proposal and the Employability Skills scheme (ACCI & BCA, 2002). In Chapter 2, these schemes are reviewed and four pairs of issues are identified as being crucial to the effective implementation of initiatives such as these, the issues being their:

- definition and selection;
- dissemination and implementation;
- assessment and reporting; and
- certification and recognition.

Of these issues, definition and assessment are argued to be the most important because reporting, certification and recognition depend upon having a sound approach to assessment, and assessment depends critically on the clarity of the construct being assessed. In Chapter 3, the meaning of generic skills as a class of constructs and the definition of problem solving are discussed. In Chapter 4, possible purposes and approaches to assessment are canvassed.

Defining constructs

It is argued in Chapter 2 that a failure to define generic skills, at two levels, was a factor in the very limited success in the implementation of both major generic skills initiatives in Australia, namely the key competencies (Mayer Committee, 1992) and the employability skills (ACCI & BCA, 2002). First, clarity is required about which skills qualify as being generic and which do not; that is, a definition of what makes a skill generic is required. Second, having defined generic skills, each candidate skill requires definition, for without such clarity, instructors may generate quite different conceptions of what is intended and may work towards diverse, and possibly incompatible, goals. In summarising the discussion of Chapter 3, three propositions were advanced, namely:

- generic skills are competences
- generic skills are latent traits
- effective problem solving, among novices, depends upon the deployment of generalisable processes

The implications for the study of problem solving as a generic skill are now discussed.

Generic skills are competences

One of the difficulties with generic skills is that they are widely applicable and it would appear to be relatively easy to gain acceptance for skills such as communication, teamwork and problem solving; their importance would appear to be self-evident. There is considerable agreement about the main skills that are endorsed as being generic across many schemes and this suggests that there is tacit agreement about the meaning of 'generic.' However, a common tacit understanding of what skills should be included is not an adequate basis for developing assessment standards. Thus, Chapter 3 includes considerable discussion of possible foundation conceptions of generic skills. Models of intelligence are examined as they include sets of primary mental abilities that may underlie competent performance. Notions of competence are also examined, and a view of competence as the capability to learn and adapt to an environment is considered the most useful conception of generic skills. This view does not construe intelligence and competence as incompatible alternatives, but rather as a duality. Competence may be related most closely to the cluster of abilities that Carroll (1993) labelled crystallised intelligence, although the emergence of this set of abilities may depend upon the cluster of fluid intelligence abilities, which appear to have particular salience for problem solving, and the opportunity to exercise them. However, a more useful model of intelligence emerges from Sternberg's theory of successful intelligence (see, e.g., Sternberg et al., 2000). This view is considered more useful in instructional contexts because, unlike the psychometric view that identifies a set of abilities that exist in varying amounts in individuals, the theory of successful intelligence posits a set of cognitive processes that may explain observed performance and that may provide a basis for interventions designed to enhance cognitive performance. It is argued in Chapter 3 that these component processes constitute the propensity to learn and adapt that are identified as competent behaviour. Sternberg's theory appears to provide the internal structure that links the two facets of the intelligence-competence duality.

Competences are believed to depend upon basic cognitive processes (see Table 10 in Chapter 3) that are activated according to the demands of the context. In order to encourage the development of these processes, it is necessary to ensure that individuals are exposed to contexts that provide an opportunity for (or perhaps demand) the deployment of these processes. This is not a particularly challenging requirement, for if the skills are as generic as claimed, many learning and most work contexts should require their use. In the present research, existing assessment activities are used as the contexts in which students demonstrate (Chapter 7) and develop (Chapter 8) their problem solving skills.

Generic skills are latent traits

The cognitive processes that underlie generic skills are not directly observable. Their existence may be inferred through observations of performance in situations in which these processes are expected to be deployed.

Indicators of each of the problem solving processes are used as a method for gathering evidence of students' application of the cognitive processes that underlie effective performance. These indicators are described in Chapter 7.

However, these processes may not be activated even though an individual may have demonstrated them in similar contexts previously. Their activation appears to depend upon motivational and conative factors.

An attempt to assess motivational and dispositional components of students' engagement in problem solving processes (Heppner & Petersen, 1982) is described in Chapter 7.

Effective problem solving, among novices, depends upon the deployment of processes

Problem solving is commonly recognised as a generic skill. Generic skills are construed as forms of competence, and competence is a function of cognitive and metacognitive skills, so a model of problem solving that is consistent with current psychological views of problem solving and that is consistent with the notion of competence is developed. Two current conceptions of problem solving are apparent in the psychological literature, namely an information processing model and a situative model. While strong evidence for both approaches exists in the literature reviewed in Chapter 3, the information processing model is perceived to have significant advantages. First, because it is based on a set of processes for which indicators can be found, it relates closely to cognitive conceptions of competence. Second, the information processing model provides a theoretical basis for anticipating transfer to novel contexts. The situative model implies the specialisation of skill to contexts and does not readily allow for the generalisation of skills through abstraction.

Three descriptions of the processes involved in problem solving are reviewed (Bransford & Stein, 1984; Hayes, 1989; Polya, 1957, see Chapter 3). They are shown to be very similar to each other and to the set of cognitive and metacognitive processes listed in Sternberg's theory of practical intelligence (Sternberg et al., 2000). Consequently, this conception of intelligence is shown to be consistent with the model of competence most closely aligned with an ability to learn as new situations arise (Chiappe & MacDonald, 2005).

It is shown in Chapter 2 that the there has been a lack of a clear definition of generic skills and of problem solving. Having defined problem solving as a psychological construct and operationalised it as a set of processes, it would appear to be necessary to communicate this definition to teachers and learners. This is done through a problem solving skills assessment tool, whose development and testing is described in Chapter 7. A range of support material was developed for teachers and students and this is described in Chapter 7 and shown in Appendix 5.

An assessment model

Simply defining generic skills is not regarded as an adequate solution to their dissemination and development. Their assessment is regarded as an essential element of their dissemination and implementation.

Assessment is shown to be multi-faceted (see Chapter 4). It has several purposes, namely promoting learning (formative assessment), measuring individual achievement (summative assessment) and program evaluation (Airasian, 1994a; Pellegrino et al., 2001). In the discussion in Chapter 4, seven propositions about assessment are advanced, namely

- assessment is a driver for learning;
- a key goal of assessment is to enhance students' performance;
- formative assessment leads to learning gains;

- feedback by teachers on the substantive content assessed and on students' self-assessment leads to improved performance;
- self-assessment leads to enhanced learning;
- a generalised standards framework needs to be applied across tasks; and
- performance assessment integrated with the substantive course content leads to improved generic skills performance.

The implications of these propositions for the study of problem solving are now discussed.

Assessment and learning

The first three of the above propositions link assessment and learning. The first summarises the observation that unless concepts such as generic skills are assessed, neither students nor teachers are likely to attend to them. The second proposition asserts that while assessment can have several purposes, an important goal is the development of students' skills. This may occur indirectly in that summative assessment may lead to greater attention being paid by teachers to instruction in generic skills, but that assessment deigned specifically to enhance learning through a formative approach is more likely to generate desired improvement, and this is the point of the third proposition.

These propositions are taken as axioms in the current study. They are not investigated as hypotheses but are supported by a body of literature reviewed in Chapter 4. These propositions influence the present study as assessment is the focus for the development of generic skills.

Enhancing returns from assessment

An essential element of formative assessment is the provision of feedback to learners on their performance (Proposition 4). Assessment either without feedback or with feedback provided only as grades has been shown to be ineffective and even counterproductive of student learning (see Biggs & Moore, 1993; Black & Wiliam, 1998a; Sadler, 1989; Wiggins, 1998, cited in Chapter 4). Proposition 5, advocating student self-assessment, is seen as a way of enhancing student engagement in the assessment activity and therefore with the target construct.

These propositions influence the current study as the assessment instrument is constructed to require both student self-assessment and assessment by, and feedback from, their teachers.

Assessment standards

Expected standards of performance need to be communicated to learners (Sadler, 2005). In the current study, diverse assessment tasks are used (see the comment on integrated assessment below) so a broadly applicable framework of performance standards is required.

For the present study, after evaluating (a) the development of rubrics for each assessment task, (b) the use of Bloom's taxonomy and (c) the SOLO taxonomy, the decision was taken to use the SOLO taxonomy as a basis for defining performance standards. These standards are applied to the set of processes that operationalise the definition of problem solving.

Integrated assessment of generic skills

The final proposition, which arises from an evaluation of a variety of approaches to the assessment of generic skills, is that problem solving can be assessed using tasks that constitute existing requirements of courses. That is, the assessment of generic skills can be integrated with the assessment of substantive course content. The implementation of this assessment approach suggests the need for a common standards framework. The strategy entails some risk, as it can be shown that diverse assessment tasks have different affordances for the skills being assessed. This risk can be accommodated through the use of measurement models that take into account item difficulty (see Chapter 6).

In the current study, teachers are asked to nominate, from among the tasks they set in their courses, a set of assessment activities that they believe can provide students with the opportunity to demonstrate and develop their problem solving skills.

Towards an assessment tool

The preceding discussion points to the need to develop an assessment tool that (a) is based upon a clearly defined construct for each generic skill; (b) operationalises the defined construct as a set of cognitive and metacognitive processes; and (c) embodies a set of standards to describe observable performances.

One of the purposes of this study is to develop such a general-purpose problem solving assessment tool. This objective constrains the options for assessing problem solving performance.

Gay (2001) noted that most research on metacognition in problem solving had used self-report data, derived from either concurrent or retrospective 'think-aloud' protocols, written self-reports or self-evaluations, and he commented that "each of these methods suffers from validity problems." This would appear to be a harsh judgment. Verbal protocols have yielded very useful information about problem solving processes, although care must be taken with the instructions given to participants (Ericsson & Simon, 1980). In the current investigation, the assessment of problem solving uses the results of the many studies that have relied upon verbal reports and that have contributed to knowledge of the processes that problem solvers use. In the current study, interest lies in quantifying the use of those processes. However, rather than relying on self-reports as the primary data sources, performance standards, which are based upon past research that used self-report data, are employed. Students are invited to compare their work against the described standards. Students' judgments are validated by their teachers.

The development and testing of a general-purpose problem solving assessment tool is described in detail in Chapter 7.

Research questions

The research project has been guided by several major research questions and they are now posed and discussed.

Research Question 1: What are the components of generic skills?

This question is raised because there are several bases – individual, civic, and economic – that could be used to frame responses. Much of the debate contained in the documentation of generic skills schemes reflects an economic orientation. It seems, however, that these three perspectives depend upon a common set of competencies. From the argument developed in Chapter 2 it is apparent that generic skills schemes represent useful tacit theories of human ability and that these schemes may be a valuable starting point in an exploration of this question.

Psychological theories of human ability also suggest a range of components that contribute to an overall conception of competence. These theories suggest a framework in which the tacit theories embodied in generic skills schemes can be evaluated and formalised.

As a result of investigating this question, the key issues, of definition and assessment of generic skills, arise and they are discussed in detail in Chapters 3 and 4. This first question is introduced in Chapter 2.

Research Question 2: Can a foundation for generic skills be found in psychology?

A subsidiary question (to Research Question 1) frames much of the discussion in Chapter 3, namely:

- To what extent can the elements of the generic skills schemes be represented by psychological constructs? That is, is there a valid basis in psychological theory for these elements?

This question, which follows from the debates about the characteristics and potential bases for generic skills, is addressed in Chapter 3. In that chapter, various theories of intelligence are examined as potential bases for understanding and representing the skills that have featured prominently in many generic skills schemes. Although Carroll's (1993) summary of the field proved to be an extremely useful account of many basic cognitive abilities, it is found wanting in several respects. The concept of competence appears to be a useful representation for generic skills and Sternberg's theory of successful intelligence provides a link between conventional views o intelligence and the notion of competence.

Research Questions 1 and 2 are addressed through analysis rather than empirical investigation.

Research Question 3: Can valid and reliable indicators of performance in generic competences be developed?

This question is, to some extent rhetorical, but not entirely so. It was expected, at the commencement of this investigation, that generic skills could be assessed, provided they could be defined as coherent constructs from a sound disciplinary perspective. The question is given added impetus following the review of generic skills schemes reported in Chapter 2, and the identification of assessment as a key issue gives rise to the discussion in Chapter 4.

Following the discussion in Chapter 4, the question has been refined and elaborated. The focus is narrowed initially to problem solving, rather than all generic skills, and modified to the assessment of a common conception of problem solving across a diversity of tasks. The focus has also been restricted to relative novices, or at least non-experts, as expert problem solving is qualitatively different from that of non-experts. This question is addressed in the study reported in Chapter 7. It is hypothesised that an assessment tool that is based upon a psychologically valid construct for problem solving and that employs a sound basis for the assessment of cognitive performance leads to the valid and reliable measurement of problem solving achievement.

A subsidiary question is also developed. In Chapter 4, the proposition is advanced that formative assessment with feedback can lead to enhanced learning. A further research question therefore arises.

Research Question 4: Can problem solving performance be enhanced by repeated assessment and feedback cycles?

This question is addressed in the study reported in Chapter 8. The assessment tool, whose development is described in Chapter 7, is used in a series of assessment tasks over an academic year. In each cycle, students assess their own performance against specified standards; their performance is assessed by their teachers; and they receive feedback on the substantive content of the assessment task, on their problem solving performance, and on their self-assessed performance judgment. It is hypothesised that repeated self-assessment and feedback leads to enhanced problem-solving performance.

Answering the research questions

As noted above, the first of the research questions has been answered substantially in Chapter 2. Its response leads to the other questions. Question 2 is addressed in Chapter 3 where a basis for the tacit theories of human ability that are embodied in generic skills schemes is found in psychological theory. This, in turn, provides a basis upon which to pursue Questions 3 and 4, answers to which depend upon constructing and testing an assessment tool (Chapter 7) and evaluating the influence of its application in existing learning contexts (Chapter 8).

In order to gather and evaluate evidence for the questions that have guided this research, several analytical methods are required. They include item response theory (IRT) and methods for the analysis of longitudinal (repeated measures) data. In addition, in order to maximise the use of data that are gathered, methods of missing data imputation are required. These methods are described in Chapter 6.

Chapter 6: Analytical Methods

The questions that drive the current research project require the analysis of data on student achievement. Over more than a century, a variety of approaches have been taken to the assessment and reporting of student ability, achievement and performance. One objective of this research is to measure student problem solving performance on an interval scale and the issue of measurement frames much of the discussion in this chapter. The second major issue discussed in this chapter is the evaluation of change in student performance over time. In the second of the two studies reported in this thesis (Chapter 8), change in student performance over time is evaluated. Approaches to studying change that were common in the past are contrasted with contemporary thinking about the study of change. Finally, to include cases for which there are some missing data, approaches to missing data imputation are discussed.

Measurement

A brief history of measurement in the social sciences

Reliable measurement is essential to the smooth operation of commerce and to science. The need for standard measures was recognised in Article 25 of the Magna Carta:

> There shall be one measure of wine throughout our whole realm, and one measure of ale and one measure of corn--namely, the London quart;--and one width of dyed and russet and hauberk cloths--namely, two ells below the selvage. And with weights, moreover, it shall be as with measures. (King John, 1215)

In the physical sciences, standards were established for units of length, mass, time and temperature. The measurement of quantities in the physical sciences has evolved, with more precise measures being required for current purposes. The standard metre, adopted in 1793, was defined as one ten-millionth of the earth's quadrant passing through Paris. Recent definitions of the standard metre are based on the distance travelled by a helium-neon laser beam in a given amount of time. This requires accurate measurements of time, and this is now accomplished using radiation emitted from Caesium 133 atoms. Of particular interest is that, at the atomic level, these emissions arise stochastically, so physical measures that are thought to be deterministic have a random basis. In this sense, measurement in the physical sciences may not be as different from measurement in the social sciences as some have argued (Bond & Fox, 2001, p. 7). What is apparent in the physical sciences is the attention paid to measurement and the continuing refinement of measures.

Measurement in the social sciences has a more complex recent history. Michell (2003, p. 300) posited that all continuous quantitative attributes must satisfy seven conditions. The first four of these, which include being associative and commutative, are necessary for attributes to be additive; that is, for any magnitudes a and b of an attribute, there must be a magnitude $c = a + b$. An important consequence of this property is that c is composed only of a and b and does not include any other component. In other words, the attribute must be unidimensional for it to be measurable. The final three of Michell's propositions require that quantitative attributes are both continuous and unbounded. Michell noted that Holder had identified these requirements for measurement in 1901.

The requirements for measurement were also recognised in the social sciences in the early years of last century. R. M. Thorndike (1999) noted that both E.L. Thorndike and Thurstone had attempted to base approaches to social science measurement on the principle of additivity. He added that the adoption of Likert scaling (see below), from which numbers assigned to ordered responses are taken as measures, curtailed further development of a rigorous approach to measurement. Andrich (1997c), in a review of Thurstone's contributions to social science measurement during the 1920s, noted Thurstone's attention to the principle of linearity that coincides with Michell's propositions for additivity and with those advanced by Holder in 1901 and Campbell in 1917. Andrich (p. 821) quoted from two of Thurstone's publications that reveal Thurstone's recognition of the possibility of both item independent measures and person independent scales. In 1926, Thurstone observed:

> It should be possible to omit several test questions at different levels of the scale without affecting the individual score. (Andrich, 1997c, p. 821)

Thurstone, in a series of papers in the 1920s, extended psychophysics – the application of scientific principles to explore the relationships between stimuli and psychological responses – to the measurement of achievement, attitude and opinion (Bezruczko, 2000). Thurstone perceived the necessity of additivity (which he called linearity) and he described absolute scaling, which would now be called objective measurement. Thurstone did not include a person ability parameter in his procedures.

Keats (1998) summarised the debate around social science measurement between about 1900 and 1940. He observed that Campbell, in 1917, had declared that for an attribute to be measurable in the physical sciences, two requirements must be met. First, objects must be comparable and transitive on the attribute; for example, if A was judged heavier than B, and B heavier than C, then A had to be judged heavier than C. That is, for an attribute to be measurable, objects had to reveal consistency in their ordering on that attribute. Second, the attribute had to be shown to be additive, as Michell (2003, p. 300) posited.

Thurstone's and others' proposals for measurement led to concerns by physical scientists about the use of the term measurement in the social sciences. These concerns led to the establishment of a committee of review into measurement in the social sciences by the British Association for the Advancement of Science (BAAS). The BAAS committee reported in 1938 and it found that measurement in the social sciences complied with Campbell's first requirement, but not the second, and concluded that measurement in the social sciences was of a different kind than measurement in the physical sciences (Keats, 1997, p. 713). Apparently, the BAAS committee did not accept Thurstone's proposals for absolute scaling.

In a retrograde step in the evolution of measurement in the socials sciences, Likert, in 1932, proposed a method for the 'measurement' of attitude (Dunn-Rankin & Zhang, 1997). This involved generating sets of statements reflecting an attitude and providing a set of ordered response categories, such as strongly disagree, disagree, agree and strongly agree. Each response category is scored using successive integers. Likert claimed that the response scores for a set of related statements could be summed. This summation of scores ignored Thurstone's contention that raw scores were not measures. He had argued, in relation to spelling tests, that each prompt represented a different amount of the trait being assessed and that counts of correct answers were not additive and were not estimates of the trait of interest. However, the simplicity of Likert's procedure, and the complexity of Thurstone's, led to the widespread use of raw scores derived from Likert scales.

Lawley (1943) formalised the basis for checking additivity. He recognised the need for person ability and task difficulty to be considered. His key insights were that ability was a student trait, that difficulty was an item trait, that the difference between a student's ability and the difficulty of a test item (β-δ) was a single parameter and that the parameter should be normally distributed. Unfortunately, computations involving the normal function were complex and, without the power of computers, proved to be largely intractable at that time.

In what became a further major retrograde step in the evolution of measurement, Stevens (1951) proposed four ‘measurement’ categories for attributes, namely nominal, ordinal, interval and ratio. Michell (1997) found that the definition of measurement that appears in many psychology texts, and which is attributed to Stevens, reads something like “Measurement is the assignment of numerals to objects or events according to rules.” Michell defended Stevens, as the quote attributed to Stevens is only a part of what he said about measurement, omitting some important qualifications. Stevens’ statement was:

> But measurement is a relative matter. It varies in kind and degree, in type and precision. In its broadest sense measurement is the assignment of numerals to objects or events according to rules. And the fact that numerals can be assigned under different rules leads to different kinds of scales and different kinds of measurement. The rules themselves relate in part to the concrete empirical operations of our experimental procedures which, by their sundry degrees of precision, help to determine how snug is the fit between the mathematical model and what it stands for. (Stevens, 1951, p. 1)

Michell (1997) was critical of what has passed for measurement in psychology. It would seem that the simplicity of Likert scales and the limited interpretation of Stevens’ definition have led to counts of events and numbered ordinal responses being taken as measures, implicitly assuming that the requirement of additivity is met.

A significant development in the theory of measurement occurred with the formulation of simultaneous conjoint measurement (Luce & Tukey, 1964). Wright (1997) noted that conjoint measurement is a formal representation of Campbell’s requirement for additivity.

Until about 1960, although all the requirements for true measurement were known in the social sciences, the requirements were not brought together in a single method and measurement was not routinely undertaken in the social sciences. That began to change when Rasch, who had been working on a variety of intelligence and achievement tests, developed a method for converting raw scores to interval additive measures (Rasch, 1960, 1980). Rasch brought together a series of insights into the problem of estimating person ability independent of the particular set of test items and of the population on which the items had been calibrated. He referred to this property as specific objectivity. This property arose from his choice of the Poisson distribution to model responses to items. This distribution has the property that, if the outcomes of two tests are compared for a person, the person parameter is factored out of the ratio of the Poisson functions (Wright, 1998, p. 18 citing Rasch, 1977). This property represented a development of the one parameter (β-δ) model because relative values of these traits (person ability and item difficulty) could be separately estimated. When applied to items limited to one trial, the Poisson function simplifies to the logistic model shown in Equation (1) below.

Rasch's focus was on measuring individual achievement, whereas previous work in psychology had operated on the distributions of ability in groups and reporting individual achievement relative to group norms (Wright, 1998). Before describing the family of methods named after him, it is useful to consider classical test theory.

Classical test theory

Under classical test theory (CTT), a linear relationship is presumed to exist "between a person's observed number-correct test score and the error-free score that it estimates" (Weiss & Yoes, 1991, p. 70).

Items that purport to reflect an attribute must be shown to be unidimensional. This is often asserted through an item reliability index, although unidimensionality is usually demonstrated through a form of factor analysis. In the past, this was commonly done through principal components analysis but is now achieved through confirmatory factor analysis (Weiss & Yoes, 1991). Unidimensionality does not imply that responses to all items reflect a singular underlying process (Bejar, 1983, p. 31). Demonstrating this in practice requires that the set of test items are represented by a single dominant factor (Hambleton, 1989, p. 150).

Numerous criticisms have been made against CTT. First, the item parameters generated in CTT – item difficulty and discrimination – depend upon the sample of individuals who take the test (Hambleton & Swaminathan, 1985; Weiss & Yoes, 1991; Wright, 1988). Second, estimates of candidate ability based on the number-correct score depend upon the difficulty of items (Weiss & Yoes, 1991). Administering a test of easy items would yield higher scores than would a test composed of more difficult items. Third, since estimates of individual ability are item dependent, it is not possible to treat students with omitted items, whether by design or accident, in the same way as students who complete all items (Wright, 1997, p. 2). Fourth, test reliability indices, such as Cronbach's alpha, are based on the variance of the test scores of the sample of individuals tested and are therefore sample dependent (Keats, 1997, p. 718). Fifth, number-correct scores are non-linear with respect to the trait of interest. Wright (1997, pp. 3-4) illustrated this by example, showing that a difference of ten percentage points near the extremities of a test reflect about five times the difference in the trait score compared with a similar difference in number-correct scores at about the mean test score. The non-linear relation between raw scores and traits has implications for secondary analyses that researchers might want to undertake. If the trait being measured is what the researcher wishes to model, the number-correct score, being non-linearly related to the trait, will lead to truncated parameter estimates in the models. Sixth, under CTT, errors are assumed to be constant across the measurement range. In practice, this does not hold (Keats, 1997). This has implications for the validity of test scores. "Validity refers to the degree to which evidence and theory support the interpretations of test scores entailed by the proposed uses of tests" (American Educational Research Association et al., 1999, p.9). If the errors are too large in any part of the score range, the test may not be fit for its intended purpose for that part of the range.

In his concluding remarks on classical test theory, Keats (1997) observed:

> At the practical level, an atheoretical approach to testing will undoubtedly continue (with all its inefficiencies) as will weak true score theory with its emphasis on item selection in terms of difficulty and discriminating power and on reliability coefficients which are often meaningless figures. As far strong true score theory is concerned, it will be noted that this was formulated in the early 1960s at a time when the theory of conjoint measurement was also developing. (p. 718)

Despite its shortcomings and its vocal critics, CTT continues to be used and has its adherents. Lord (1980) argued that CTT and IRT are not alternatives; rather they are mutually supportive. Certainly, CTT is useful in initial screening of data, but the scores generated for individuals under CTT are not measures (Wright, 1997). Where it is demonstrated that items cohere to represent a single construct and there are no missing data, the number correct score for items and persons are sufficient statistics for the measures that are generated through measurement approaches.

The Rasch measurement model

There is a family of Rasch measurement models including the dichotomous model, the rating scale model and the partial credit model as well as others (Wilson, 1989a, 1999; Wright & Mok, 2004). These Rasch models may be considered a subset of a more extensive set of models under the umbrella term item response theory (IRT), although some regard the Rasch model as being fundamentally different from IRT models and not a special case of them (Andrich, 2004). What is common to IRT models is that they posit functions that link the probability of success on a trial (e.g. getting a correct answer to a test item) to a characteristic of the person attempting the item and a characteristic of the item. The IRT models differ in the number of parameters they use to represent the relationship between the probability of success and the item and person characteristics. The dichotomous Rasch model, sometimes referred to as a one-parameter IRT model, uses the difference between a person ability parameter (β) and an item difficulty parameter (δ), the single modelled parameter being (β-δ). The two-parameter IRT model adds a term to represent item discrimination and the three-parameter model adds a further term to model guessing behaviours.

Equations for dichotomous one-, two- and three parameter models (after Bond & Fox, 2001, p. 201; and Weiss & Yoes, 1991, pp.78-79) are[17]:

$$P_{ni}(x=1\,|\,\beta_n,\delta_i)=\frac{e^{(\beta_n-\delta_i)}}{1+e^{(\beta_n-\delta_i)}} \tag{1}$$

$$P_{ni}(x=1\,|\,\beta_n,\delta_i)=\frac{e^{a(\beta_n-\delta_i)}}{1+e^{a(\beta_n-\delta_i)}} \tag{2}$$

$$P_{ni}(x=1\,|\,\beta_n,\delta_i)=c_i+(1-c_i).\frac{e^{(\beta_n-\delta_i)}}{1+e^{(\beta_n-\delta_i)}} \tag{3}$$

The first equation, for the Rasch or one-parameter model, relates the probability (*P*) of a person (*n*) attempting an item (*i*) scoring a correct answer, given the person's ability (β) and the item's difficulty (δ). In the second equation for the two-parameter model, a parameter (*a*) is added to model the discrimination of items. A third parameter (c_i) appears in the third equation, for the three-parameter model, to accommodate a tendency for individuals to guess correct answers to items. This parameter, with the *i* subscript, is taken to be a characteristic of items (Stocking, 1997, p. 836), but others (e.g. Choppin, 1997) have shown that guessing varies between students.

The Rasch formulation does not include item discrimination or guessing parameters: if items in a test have varying discriminations or if there is substantial guessing on some items, these items should reveal poor fit and typically would be removed from the analysis. The two- and three-parameter models might provide better fit to the data when item discriminations vary or when guessing is common. However, Wright (1997, pp. 12-13) showed that the two- and three-parameter models do not comply with the additivity requirement of true measurement. He also showed that the introduction of the additional parameters of those models do not provide reliable estimates of person ability unless constraints are imposed on the range of permitted values. Further, the discrimination parameter is sample dependent and so does not permit the objective calibration of item difficulties.

---

[17] Weiss and Yoes (1991) showed a scaling factor of 1.7 in the exponent terms of their equations, but others, including Bond and Fox (2001) and Wright and Mok (2004) omitted this. Andrich (2004, p. 154) pointed out that this scaling factor is used to ensure that the logistic probability closely matches that of the normal function used by Birnbaum in his IRT model.

In many statistical procedures, e.g. regression, the process involves collecting data, proposing a model that is thought to account for the patterns of variation in the data, and checking the model against the data. If the fit is adequate, the model is accepted, albeit tentatively, but if the fit is poor, the model is rejected. In these approaches, the data have primacy and the model is negotiable. In measurement, certain requirements are non-negotiable (e.g., see Michell, 2003, p. 300). When data are collected for the purpose of calibrating a measurement instrument or to generate measures for individuals, one task is to confirm that the data comply with the requirements of measurement. If they do, the process of measurement can continue; if not, the data are considered as not complying with the requirements of measurement. There is some scope to reconsider the particular measurement model adopted. Data that do not comply well under the rating scale model might comply better under the partial credit model. (See below for a description of these variants of the Rasch measurement model). Wright (1997) pointed out that the introduction of discrimination and guessing parameters into IRT models takes those models out of the realm of measurement and into the domain of data modelling. In the measurement analyses undertaken in the studies reported in this thesis, the Rasch measurement model is used. In the sections below, variants of the Rasch model are described.

The dichotomous Model

The dichotomous Rasch model is represented in Equation (1) above and is described briefly. It has been described in numerous texts and papers (Allerup, 1997; Andrich, 1988; F. B. Baker, 2001; Bond & Fox, 2001; Embretson, Schmidt, & McCollam, 2000; Harwell & Gatti, 2001; Keeves & Alagumalai, 1999; Rasch, 1960, 1980; E. V. Smith & Smith, 2004; Weiss & Yoes, 1991; Wilson, 2005; Wright, 1988, 1999a, 1999b; Wright & Mok, 2004; Wright & Stone, 1999; Zhu, 1996).

The logistic function of the dichotomous Rasch model is a special case of a more general model, based on the Poisson distribution that Rasch developed in analysing students' scores on reading tests. The model conforms to the requirements of measurement, as defined by Campbell and later by Thurstone (cited in Wright, 1997). The general model, the dichotomous special case and its extensions (see below) have the property that differences in item difficulties are independent of the sample of persons used to compare the items, and that differences in the abilities of two individuals are independent of the particular set of items used to compare them (Wright & Stone, 1999, p. 2). This property is not shared by IRT models that include discrimination and guessing parameters.

Polytomous models

The basic Rasch model has been applied to the analysis of test results, especially from multiple-choice tests, where answers to questions are coded as either right or wrong. The model has been extended to many other applications. Andrich (1978; 1997b; 1999) extended the model to the analysis of ordered response data. Such data arise from the administration of, e.g. attitude survey instruments in which items have a common response format. The requirement is that the responses to each item must reflect an ordered sequence. Examples of complying formats include levels of agreement with propositions (e.g. 'strongly disagree', 'disagree', 'agree' and 'strongly agree') or frequency of events (e.g. 'almost always', 'often', 'sometimes' and 'rarely or never'). Responses such as these are coded 0, 1, 2 and 3 to reflect the ordering of levels of agreement or frequency. Importantly, these numerical codes are ordinal, and as Andrich (1997a, p. 878) pointed out, "the integer score on each rating is not a measure."

In the rating scale model, difficulty (or location) parameters are estimated for each item, and a common set of threshold parameters are estimated for alternative responses to the items. That is, the threshold parameters are common across the set of items. In the examples given above with four response options, in addition to a difficulty parameter for each item, three threshold parameters are estimated for the set of items. The threshold parameters correspond to the trait levels at which an individual is equally likely to choose either response *k* or response *k-1* (where *k = 1..m* for the *m* response categories available) and are the thresholds that mark the boundaries between the trait regions where one or other of response *k* or *k-1* is more probable. The rating scale model parameters are shown in equation (4) below, which is taken from Bond and Fox (2001, p. 204).

$$P_{nik}(x = k \mid \beta_n, \delta_i, \tau_k) = \frac{e^{(\beta_n - \delta_i - \tau_k)}}{1 + e^{(\beta_n - \delta_i - \tau_k)}} \tag{4}$$

In an instrument with *L* items, each with *m* response options, *L+(m-1)* parameters must be estimated.

Masters (1982; 1997; 1999; Wright & Masters, 1982) extended the basic model to situations in which items have multiple response options, but in which not all items need have the same number of response categories. In cases where items have a common response format, either the rating scale model or the partial credit model can be used. The main difference between the rating scale and partial credit models is in the threshold parameters. In the rating scale model, the distances between the thresholds are common to all items. In the partial credit model, the distances are unique to each item. For an instrument with *L* items and *m* response categories, under the partial credit model, *L*(m-1)* parameters must be estimated. The partial credit model parameters are shown in equation (5) below, which is adapted from Bond and Fox (2001, p. 205). The Delta parameters in that equation have two subscripts, one for the item and the second representing the threshold between categories.

$$P_{nik}(x = k \mid \beta_n, \delta_{ik}) = \frac{e^{(\beta_n - \delta_{ik})}}{1 + e^{(\beta_n - \delta_{ik})}} \tag{5}$$

Thresholds

The value of the trait at which an individual is equally likely to choose either of two adjacent categories is the threshold between those categories. In the dichotomous case, there is a single threshold per item, and this is usually referred to as the item difficulty. In polytomous model, there are (m-1) thresholds for m response categories. In the rating scale model these thresholds are represented by the τ_k estimates (see equation 4), while in the partial credit model, they are represented by the δ_{ik} estimates (see equation 5).

While it is common to estimate thresholds at the point where a candidate has an equal probability (0.5) of failing or succeeding on a dichotomous item or failing to achieve or achieving a particular category in a polytomous item, it is possible to use an alternative probability in estimating thresholds. In studies requiring mastery performance, the probability of success could be set at 0.8. In the analyses presented in Chapters 7 and 8, the default probability of 0.5 is selected. Rather than demanding mastery, the models being developed seek to locate students' current learning level for formative diagnostic and developmental purposes (see, Griffin, 2001, p. 8).

The program Quest (Adams & Khoo, 1999) generates estimates for the Tau and Delta parameters. These thresholds are functionally identical. The Tau thresholds are reported relative to item difficulties (locations) while the Delta thresholds are reported relative to the scale mean, which is usually the zero point on the scale unless it has been anchored at some other value. From equations 4 and 5, it can be seen that:

$$\delta_{ik} = \delta_i + \tau_k \tag{6}$$

where δ_{ik} is the Masters (Delta) threshold, δ_i is the item difficulty, and τ_k is the Andrich (Tau) threshold.

The Delta and Tau thresholds represent the trait level at which a person is likely to succeed at one level, conditional on having succeeded on the lower level (Adams & Khoo, 1993, p. 86).

Quest reports a third threshold, referred to as the Thurstone threshold. This is the point on the scale at which a person has a 50 per cent chance of falling into any of the performance levels below that category or of achieving that level or any above it (Griffin, 2001, p. 8). While the Delta and Tau thresholds need not be ordered, the Thurstone thresholds must be ordered (Adams & Khoo, 1993, p. 86).

Delta or Tau threshold disordering indicates potential problems with items. In a survey using the Likert response format, if a reversed item is not recoded, the apparently more favourable response categories should have lower thresholds than the less favourable ones. However, reversals are seen in the Delta and Tau thresholds when some response categories attract low frequency responses in a particular sample. This is illustrated in an item taken from an analysis of data from the Multidimensional School Anger Inventory (MSAI, Boman, Curtis, Furlong, & Smith, 2006). The category probability plots, generated using RUMM (Sheridan, Andrich, & Luo, 1997), for two items from this scale are shown in Figure 3 (Item 2) and Figure 4 (Item 9). These plots show the likelihood of individuals at varying trait levels choosing one of the available response categories. Item 2 has a regular sequence of categories, with successive categories becoming more likely as the trait level (school anger) increases. The thresholds, the points at which the category probability curves cross, are ordered, the Delta thresholds being -2.52, -1.18 and -0.02.

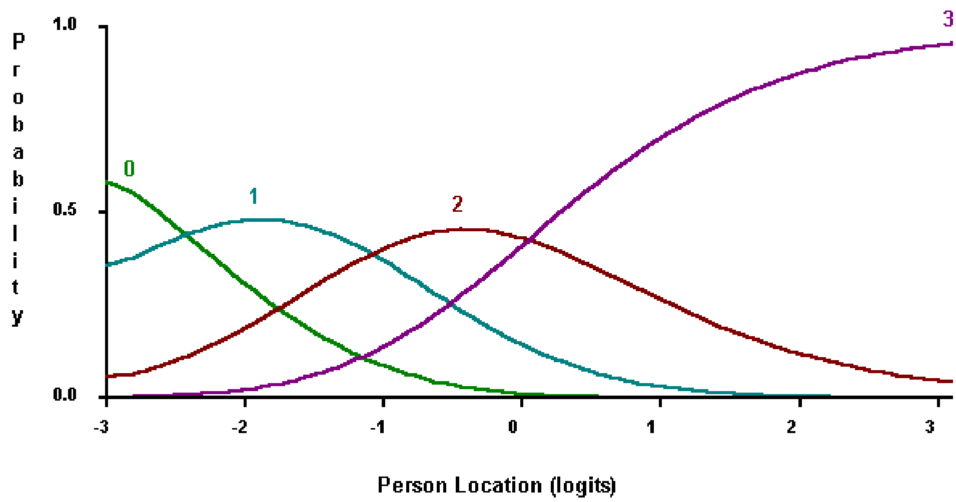


Figure 3: Category probability curves for MSAI Item 2

Item 9 has disordered Delta and Tau thresholds, the Delta thresholds being +1.58, +1.16 and +0.65; that is, the thresholds are in reverse of the expected order. It might be noted that there is no region of the trait in which either of these response categories is the most likely. This item appears to be operating as a dichotomous one. Andrich (1997b, p.879) observed that such disordering indicates there is a problem with the item, although it does not suggest what the cause of the problem might be. In the case of Item 9 in the MSAI, it was found to be highly skewed, with very low frequency responses (<5%) to each of the top two categories (2 and 3).

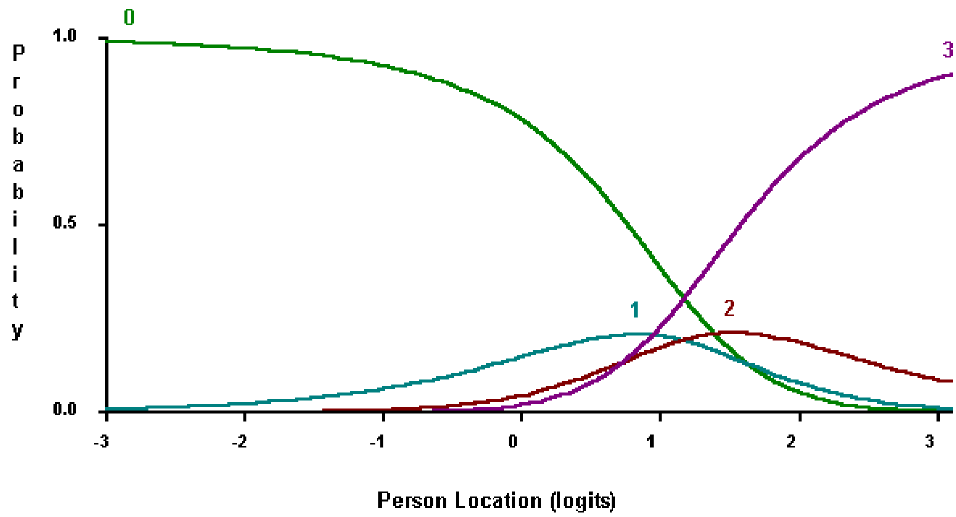


Figure 4: Category probability curves for MSAI Item 9

Because Delta and Tau thresholds can be disordered, it is useful to examine them for reversals, as any disordering will indicate potential problems with items. It is useful also to examine the Thurstone thresholds. Although they cannot be disordered, the degree of separation among these thresholds can provide information about the effective trait ranges of response options. Well separated thresholds, relative to the standard errors of their estimates, indicate that the response categories cover a useful trait range.

Detecting measurement model misfit

Weiss and Yoes (1991) stated four requirements of measurement which may be paraphrased as (a) individuals respond honestly to item prompts; (b) items are indicators of a uni-dimensional latent trait; (c) items are locally independent; and (d) item responses can be modelled using a monotonic function. (The last of these applies to polytomous items and is discussed below.) Threats to each of these requirements exist. For example, not responding honestly may include individuals attempting to guess items they do not know (perhaps in a high stakes test) or responding carelessly to items (in a low stakes test). In these cases, the data may not conform to the requirements of measurement. Departures from the requirements of the measurement model are detected through indices of item fit.

In order to check that data conform to the requirements of the Rasch measurement model, fit indices based on residuals are calculated. Bond and Fox (2001, pp. 176-186) provided an account of the assessment of measurement model fit. Wright and Stone (1999, pp. 47-55) provided a more technical account. Bond and Fox began with the data matrix of individual responses to items. In the dichotomous case, the matrix elements are 0s and 1s corresponding to incorrect and correct responses. Using the Rasch model, ability estimates are generated for the respondents and difficulty estimates for the items. Using these estimated parameters, the expected probability of a correct response can be generated for each person to each item. These data form a second matrix of expected response probabilities (P_{ni}). A third matrix of differences between elements of the first two matrices is the set of residuals between observed and expected responses.

Because responses are coded as integers (0 or 1 in the dichotomous case) and probabilities of correct responses must lie between 0 and 1, there will always be a residual. When the person ability is equal to the item difficulty, the probability of a correct response will be 0.5, so the residual, whether the answer was right or wrong, will be 0.5. A correct answer by a high ability student, relative to item difficulty, will yield a very small residual, but an incorrect response will produce a high residual. Wright and Stone (1999, p. 50) argued that the expected variance of the residuals is $Q_{ni}=P_{ni}(1-P_{ni})$. Using this variance, observed residuals can be standardised. The mean squared standardised residual over all persons generates an index of item misfit, and over all items yields an index of person misfit.

The mean squared standardised residual is influenced equally by all cases, including those at the extremities of the distribution of abilities. However, the information function for the Rasch model shows that the maximum information occurs when the person ability matches the item difficulty. Adjusting the unweighted mean square residual using the information function generates an information weighted misfit index. The unadjusted indicators are referred to as ‘outfit mean square’ statistics “…because they are heavily influenced by outlying, off-target, unexpected responses” (Wright & Stone, 1999, p. 53). The information weighted mean square index is referred to as the ‘infit mean square’ fit statistic.

Bond and Fox (2001, p. 179), citing (Linacre & Wright, 1994; R. M. Smith, 1994a, 1994b, 2000), suggested ranges of acceptable fit, depending upon the purposes of the test or survey. For high stakes tests, they recommended that infit and outfit statistics should fall within the range 0.8 to 1.2, while for 'run of the mill' tests, they suggested the range 0.7 to 1.3.

Indices of scale measurement efficiency

In CTT, the Cronbach alpha (α) is often taken as an indicator of scale reliability (R. L. Thorndike & Thorndike, 1997). In effect, it compares the within-item variance with the variance of test scores.

The Quest program computes two reliability indices, one for items and one for cases. The item reliability index reported in Quest is labelled the 'reliability of item separation' by Wright and Masters (1982, p. 92) and is based on the measurement range of the instrument compared with the precision of the item and case estimates. This index (after Wright & Masters, 1982, pp. 91-92) is shown as:

$$R_i = \frac{SA_i^2}{SD_i^2} \qquad (7)$$

where

$$SA_i^2 = SD_i^2 - MSE_i \qquad (8)$$

and

$$MSE_i = \frac{\sum_{i=1}^{L} s_i^2}{L} \qquad (9)$$

The reliability of item separation is an index of the measurement effectiveness of the instrument. The variance of the item difficulties (SD_i^2) is adjusted (SA_i^2) for the mean standard error (MSE_i) of those difficulty estimates. An effective measurement instrument is one that covers a broad range of the trait, and therefore has a large variance in item difficulties (SD_i^2), and for which the estimates of the item difficulties have high precision, i.e. have a low MSE_i. Because the reliability of item separation uses the standard error of item estimates, it is sensitive to sample size, with large samples generating larger values for a given instrument.

A similar index is computed for cases. In this case the spread of cases and the standard error of their estimates are used to calculate a reliability of person separation index. Andrich (1982) has shown that this is equivalent to the Cronbach alpha statistic. This index is sensitive to the item sample size for a given sample of persons.

Using the variance of item difficulty estimates and the standard error of their estimates, Wright and Masters (Wright & Masters, 1982, pp. 90-92) calculated an item separation index and, from it, showed that the number of separable item strata (performance bands) can be calculated. They wrote:

> …if we define statistically distinct levels of item difficulty as difficulty strata with centers three calibration errors apart, then this separation index G_i can be translated into the number of *item strata* defined by the test. (Wright & Masters, 1982, p. 92)

The reliability of item and person separation indices and the number of strata are useful diagnostics in evaluating the measurement properties of instruments.

Standard error of estimates

The indices of measurement effectiveness described above are based on estimates of standard errors of item and person parameters. An assumption of CTT is that errors are assumed to be constant across the measurement range. This is not so under Rasch measurement. This is illustrated (see Figure 5) using data from the 13-item Anger Intensity scale of the MSAI (Boman et al., 2006). The figure shows the non-linear relationship between raw scores and interval scaled scores (as shown by Wright, 1997, pp. 3-4) and the non-uniform error of estimates across the trait range. The standard error is at a minimum in the mid-range of the trait and increases at the extremities. The precision of a measure in the mid-range of the trait may be adequate for the intended purpose, but this may not be so at the extremities.

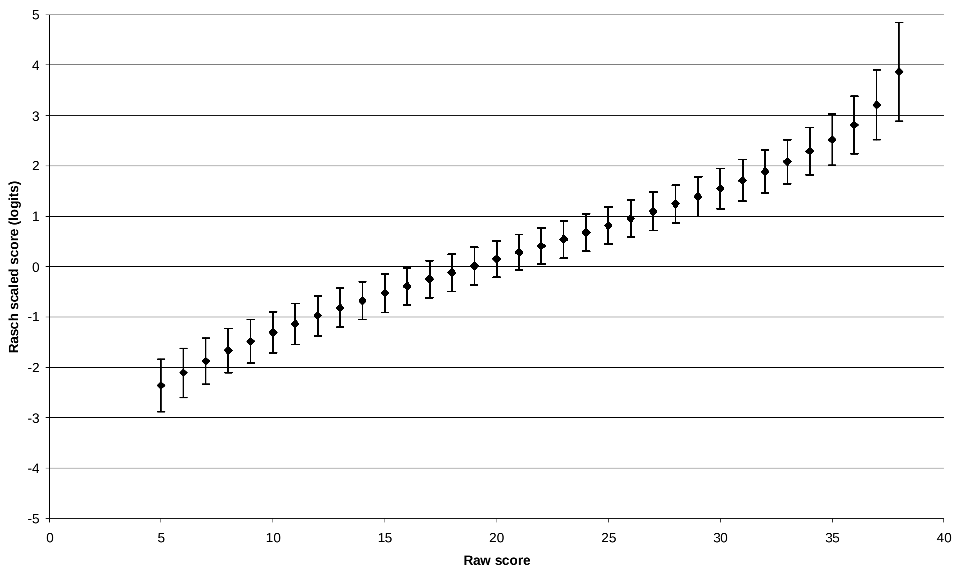


Figure 5: Relationship between raw scores, Rasch scaled scores and the standard error of person estimates for the MSAI Anger Intensity scale

Summary of Rasch measurement

A review of the history of measurement in the social sciences has revealed that the requirements of true measurement have been known in those fields since at least 1930. Several efforts have been made to establish measurement in the social sciences, but there have been retrograde steps as simple approaches to scaling became popular. The development of the Rasch model for measurement, accompanied by the availability of computers, have made measures, that comply with the axioms of measurement (Michell, 2003, p. 300), accessible to social scientists. The Rasch measurement model produces item calibrations that are independent of the particular sample of respondents and measures of individuals that are independent of the particular sample of items selected to reflect the trait being measured. Variants of the Rasch measurement model enable dichotomous and polytomous test and survey data to be analysed to calibrate instruments and to generate measures for individuals.

Software programs that implement the Rasch model, e.g. Quest, generate fit statistics that show whether the data conform to the requirements of measurement. These software programs also produce indices of item and person separation reliability that show the extent to which instruments are capable of generating effective measures. Along with estimates of thresholds, they generate standard errors for those estimates.

In the calibration of instruments, during which estimates of threshold locations are provided, the precision of those estimates and the relative positions and separations of thresholds can be used as diagnostic tools in the evaluation of instruments and of their constituent items. When an instrument has been calibrated and found to have acceptable measurement properties, it can be used to provide measures of individuals on the trait being assessed using that instrument.

Constructing Measures

The Rasch model provides a technology for measurement. However, in order to gather data that afford measurement, it is necessary to conceive of a construct and from it to develop a variable that may exist in greater or lesser amounts within individuals and tasks. Individuals are expected to vary in the extent to which they demonstrate this variable, and a range of tasks or situations should exist that vary in their demand for this variable. Wright and Masters (1982) demonstrated the creation of measures of various constructs including liking for science and fear of crime. Michell (1997; 2003) argued that the requirements for measurement go beyond showing that data derived from the administration of an instrument conform to a mathematical model. He argued that it is necessary that the object of measurement, the construct, must be shown scientifically to have a quantitative structure. The basis of this argument can be derived from the structure of the construct and from the structure of performance levels for it.

In the present study, the concept that is defined is problem solving. It is shown (see Chapter 3) that problem solving can be conceived as a set of observable processes based on latent cognitive processes. These observable processes are expected to vary in the ease with which students apply them; students are expected to use some processes readily while other processes, particularly monitoring and evaluation, are likely to be apparent only among more proficient problem solvers. For each process, a set of indicators is developed, and for each indicator, a set of performance levels is proposed. (The development of the instrument is described in Chapter 7). Levels of performance described in the instrument are based on the SOLO taxonomy (Biggs & Collis, 1982). In Chapter 4, two alternative principle-based methods for specifying performance levels are canvassed, namely the SOLO taxonomy and Bloom's taxonomy of cognitive objectives (Bloom et al., 1956). A preference is expressed for the SOLO taxonomy.

The specification of performance levels for a construct leads to the possibility that individuals' performance over time may reveal discrete stages. Such stages are fundamental in Piagetian conceptions of children's cognitive development (Brainerd, 1978). Wilson (1997, pp. 908-909) observed that several staged models of cognitive development have been used, including Piagetian models and the SOLO taxonomy, and that staged models exist in other domains such as moral development. Wilson also noted that such levels need not produce discrete performance strata. He wrote:

> The changes in the individual towards maturity may be fundamentally continuous, in which case the levels will constitute no more than convenient labels for parts of the continuum, or the progression may be fundamentally discrete, in which case the levels must be conceptualised as a systematic part of the changes that occur. (Wilson, 1997, p. 908)

Whether the variable of interest, performance on problem solving, is continuous or discrete can be decided through the analysis of performance data, as Wilson has demonstrated (1989a; 1989b; 1997). Whether problem solving develops as a continuum or through a series of discrete and staged levels is an interesting problem. It is likely that the emergence of problem-solving ability is very closely related to the development of expertise within a domain. Such an investigation lies outside the scope of the present research. However, examining student performance to see whether it changes as a result of the assessment model and process that is developed in the study is a central concern of this research. The investigation of change requires more than the measurement of performance. It requires the a particular method for the analysis of measured performance, and the evaluation of change is now considered.

Evaluating change over time

Change has been inferred from cross-sectional studies (von Eye & Spiel, 1997), but this requires strong assumptions and when the results of cross-sectional and longitudinal analyses of similar data have been compared, they have not agreed well. There is agreement that the study of change requires longitudinal designs with observations on at least two occasions. However, a critique of two-wave designs is presented below.

Keeves (1997a) identified several types of longitudinal approaches in the study of stability (or constancy) and change in individuals. One dimension of these approaches includes monitoring the influences of biology, environment and interventions on stability and change over time. A second, design, dimension includes cross-sectional, trend, time-series and intervention studies. The present study is an intervention approach using a time-series design.

In the past, it was common to assess change using a pre-test of some criterion, a treatment and a post-test. Growth was taken to be the difference between the pre-test and post-test scores (Singer & Willett, 2003; von Eye & Spiel, 1997; Willett, 1997). This method appeals because of its simplicity. The dependent variable in analyses is the difference score between the pre- and post-tests. Typically, this method involves a paired-sample t-test of the pre- and post-test scores. A limitation of this approach is that it cannot distinguish measurement and sampling error in the pre- and post tests. Measurement error is conflated with change in the test scores. Numerous criticisms have been made of two-wave or pre- and post-test designs. Willett (1997, pp. 328-330) defended the approach suggesting that many of the criticisms have been poorly founded. For example, correlations, both negative and positive, have been reported between initial the score and the change score. These correlations have been interpreted as indicating poor reliability of the change score. Willett noted that the correlations described are understandable and expected, depending upon the research context as he demonstrated (pp. 330-331). These methods may yield information on relative change within individuals. Despite repudiating many of the criticisms of pre- and post-test studies of change, Willett concluded that modern methods for the study of change involving multiple waves of data analysed using multilevel models are superior to pre- and post-test designs. Assessment of true growth requires assessment on at least three occasions (Bryk & Raudenbush, 1992, pp. 130-131; Singer & Willett, 2003, pp. 9-10).

An alternative to the pre- and post-test approach is to use multiple assessment occasions and to use a repeated measures analysis of variance on individuals' scores over the assessment occasions. This method partitions variance between occasions and individuals and is able to show whether there are significant differences between individuals' scores on the multiple test occasions and may provide information about absolute change in performance. *Post-hoc* tests are used to indicate where the differences occur. The method is limited, however. Tabachnick and Fidell (2007, p. 44) noted that tests of significance in within-subjects (repeated measures) designs are typically conservative and this makes it difficult to reject the null hypothesis of no difference between occasions. Similarly, Bryk and Raudenbush (1992, p. 83) drew attention to misestimation of standard errors and heterogeneity among participants as problems not addressed in conventional analyses that do not take into account the hierarchical structure of the data. Analysis of variance does provide an indication of differences in aggregate performance over test occasions, but it is not informative about individual growth trajectories. Now, multilevel methods are available for the study of change.

Bryk and Raudenbush (1992, pp. 133-134) compared the use of multivariate repeated measures (MRM) methods with hierarchical linear modelling (HLM) and gave five advantages of HLM over MRM approaches. First, they noted that HLM models growth directly, whereas MRM methods model individual change as an interaction between individuals and occasions. Second, HLM is more flexible as it is able to accommodate time varying data collection points, while MRM methods require the same sequence of time-spaced observations for all cases. Third, and unlike MRM, HLM permits modelling of influences on the structure of change – specifically the intercept and slope parameters for individual growth. Fourth, when MRM data restrictions apply, HLM methods provide the same estimates of the fixed parameters of the MRM methods. Fifth, the specification of individual growth as a two-level model, with within-individual repeated observations at Level-1 and between-individual variables at Level-2, can be extended to higher level models. For example, individual change could be compared between institutions, which would enter the model at a third level.

Multilevel modelling

Multilevel models, in which change within individuals is modelled at one level and differences between individuals at a separate level, are given a variety of names. Bryk and Raudenbush (1992, p. 3) identified the names multilevel linear models, mixed-effects and random-effects models, random-coefficients regression models and variance components models. What the models have in common is a data structure in which observations are nested. In organisational research, e.g. in schools, where student achievement is of interest, students are nested within classes, and classes within schools. In this instance, there are three levels of clustering. There are many examples of this type of clustering (see e.g., Bryk & Raudenbush, 1992; Hox, 1995; Singer & Willett, 2003). In the study of change over time, several observations are made of individual achievement and these observations are nested within students, while there may be characteristics of students, at a between-student level, that influence change in the performance of students over time.

If data that are clustered are analysed using conventional one-level models, seriously biased parameter estimates can be reported. Such bias is called aggregation bias. For example, in an analysis of student achievement, when students are sampled within schools, if school type (government, Catholic or independent) is disaggregated to students so that all students in a school are assigned their school type, the parameter for this variable may be found to be highly significant. However, since school type is a characteristic of the school and not of the student, if it is modelled at the school level, it is much less likely to be reported as significant. Conversely, if student-level variables, e.g. gender, are aggregated to the school level and used in analyses, similarly biased parameter estimates may occur.

Bryk and Raudenbush (1992, p. 131) noted that the evaluation of change has been beset with measurement problems because instruments, designed to measure achievement on one occasion, were standardised to have a constant variance. They described this as "fatal to the study of change". Keeves (1997a) noted the availability of several methods for equating measures across time, including item response theory methods. The discussion of the Rasch measurement model above shows that such measurement problems have been overcome. Bryk and Raudenbush added:

> The development of hierarchical linear models, however, now offers a powerful set of techniques for research on individual change. When applied with valid measurements from a multiple-time-point design, they afford an integrated approach for studying the structure and predictors of individual growth. (p. 131)

Individual growth trajectories can be modelled by regressing performance scores on time, provided there is a "sensible metric" for time (Singer & Willett, 2003, pp. 10-12), and growth is inferred from model parameters. The simplest model is a linear one, and that is used in the present study as there are only three test occasions per candidate (see Chapter 8). Additional test occasions allow more complex growth trajectories to be modelled.

The Representation of Multilevel Models

Where data are hierarchically structured, either in organisational units such as students within classes or in time series studies where repeated observations are nested within students, hierarchical models are required for their analysis. Two approaches are common in representing multilevel models. In the program HLM (Bryk, Raudenbush, & Congdon, 2005), separate equations are used to represent relationships at each level of the model. In MLwiN (Rasbash, Browne, Healy, Cameron, & Charlton, 2005) a single equation, in which the level-2 equations are substituted into the level-1 equation, is used. General forms of these equations (after Bryk & Raudenbush, 1992, pp. 9-14) are shown below.

Equation 10 represents level-1 of a general hierarchical linear model. An outcome, Y_{ij}, is measured for an individual (*i*) in organisational unit (*j*) (or at time *i* for individual *j* in a time series study). This outcome is modelled as a linear function that includes an intercept term (β_{0j}) and an explanatory variable (X_{ij}) with a parameter (β_{1j}). In a more general model, there may be many (N) explanatory variables at level-1 and the set of variables would be written as $\Sigma\beta_{nij}.X_{nij}$, where n=1..N. There is a residual term (r_{ij}), which in some representations is described as an error term and written as e_{ij} (Rasbash et al., 2002). The variance of the residual term (r_{ij}) is σ^2, a parameter to be estimated in the model.

$$Y_{ij} = \beta_{0j} + \beta_{1j}.(X_{ij} - X_{.j}) + r_{ij} \tag{10}$$

At level-2 of the model, each level-1 parameter is modelled as a function of level-2 explanatory variables. Thus, there are as many level-2 equations as there are parameters in the level-1 part of the model, excluding the residual term. Equation 11 models the level-1 intercept term and relates it to an overall intercept (γ_{00}) for all level-2 units with variation about that overall intercept explained by level-2 explanatory variables (Wj) with their parameters (γ_{01}). The residual term for this level-2 function is μ_{0j}. The variance of this residual term is τ_{00}.

The second level-2 equation (equation 12) models the parameter for the first explanatory level-1 variable. The parameter is described as varying about the overall mean for that parameter (γ_{10}) for all level-2 units, explained by a level-2 variable (W*j*) and its parameter (γ_{11}) with a residual term (μ_{1j}), whose variance is τ_{01}.

$$\beta_{0j} = \gamma_{00} + \gamma_{01}.Wj + \mu_{0j} \tag{11}$$

$$\beta_{1j} = \gamma_{10} + \gamma_{11}.Wj + \mu_{1j} \quad (12)$$

The set of equations (equations 10, 11 and 12) can be combined into a single equation. This is the representation used in MLwiN (Rasbash et al., 2005).

$$Y_{ij}=\gamma_{00}+\gamma_{01}.Wj+\gamma_{10}.(X_{ij}–X_{.j})+\gamma_{11}.Wj.(X_{ij}–X_{.j})+\mu_{0j}+\mu_{1j}.(X_{ij}–X_{.j})+r_{ij} \quad (13)$$

Bryk and Raudenbush (1992, p. 15) pointed out that this equation is not a typical single-level regression model. "Efficient estimation and accurate hypothesis testing based on OLS require that the random errors are independent, normally distributed, and have a constant variance." This is not the case in hierarchical data structures as the random error term ($\mu_{0j} + \mu_{1j}.(X_{ij} – X_{.j}) + r_{ij}$) includes elements ($\mu_{0j}$ and μ_{1j}) that are common to all level-1 units within each level-2 unit. Further, they depend on X_{ij}, which varies across level-1 and level-2 units. Thus, the error terms are not independent. Because the structure of hierarchical models violates the requirements of independent error terms and consistent variance across level-2 units, ordinary least squares (OLS) regression provides unsatisfactory estimates of model parameters and maximum likelihood methods are used (Raudenbush & Bryk, 1997, p. 551). In the analyses reported in Chapter 8, for which MLwiN (Rasbash et al., 2005) is used, the iterative generalised least squares (IGLS) estimation method is employed.

In addition to differences in the conventions used by different authors to represent models, Bryk and Raudenbush (1992, pp. 130-154) changed their conventions in considering the study of individual change over time. Instead of using β and γ as symbols for their level-1 and level-2 parameters, they used π and β respectively. The reason given was that models of individual change could be extended to three level models of change in performance over time of students nested in classes. The student and class level parameters would then have the same notation as in organisational models.

The Null Model

The null model, or random effects model (Goldstein, 1999, p. 4; Rasbash et al., 2002, p. 28), is a special case of the more general model presented above. In this model, all explanatory variables are removed and only variance parameters are estimated. Because the only parameter at level-1 is the intercept term (see equation 14), there is only one level-2 equation (15). The random effects model is represented as:

$$Y_{ij} = \beta_{0j} + r_{ij} \quad (14)$$

$$\beta_{0j} = \gamma_{00} + \mu_{0j} \quad (15)$$

This model can be shown as a combined equation.

$$Y_{ij} = \gamma_{00} + \mu_{0j} + r_{ij} \tag{16}$$

The variances of r_{ij} and μ_{0j} are σ^2 and τ_{00} respectively and together represent the total variance in the model, distributed between its two levels. Estimating these parameters enables ρ, the intraclass correlation (Bryk & Raudenbush, 1992), to be calculated. This coefficient is also called the 'variance partition coefficient' (Rasbash et al., 2002) or the 'intra-level-2-unit correlation'[18] (Goldstein, 1999, p. 5). It is the ratio of the level-2 variance to the total variance.

$$\rho = \frac{\tau_{00}}{(\sigma_r^2 + \tau_{00})} \tag{17}$$

It is informative as it indicates the proportion of the variance at each level. If there were almost no variance at level-2, perhaps because there had been random assignment to level-2 units, the problem could reasonably be treated as a one level model and analysed using conventional regression methods. When the variance at level-2 is significantly different from zero, the problem does need to be approached using a multilevel model. As explanatory variables are added at each level, the proportion of variance explained at that level can be calculated, and the refinement at either level of the model can be terminated when further improvement in the variance explained is unlikely given available variables.

Summary of evaluating change over time

The evaluation of change over time requires observations on at least three occasions. Because the repeated observations are clustered within individuals, multilevel models are required to account for change. Such models are potentially powerful because they are able to provide robust estimates of individual change and they are able to model change parameters (initial status and growth) as functions of individual characteristics. The analysis and evaluation of multilevel methods require the use of specialised software such as HLM (Raudenbush, Bryk, & Congdon, 2005) or WLwiN (Rasbash et al., 2005). These programs implement maximum likelihood and generalised least squares methods for the estimation of parameters, as the OLS method used in single-level models is inadequate for multilevel models. Several conventions, namely separate equations at each level of the model and combined equations, are used to represent multilevel models. Each has its advantages.

---

[18] Goldstein preferred not to use 'intraclass correlation' as he wished to avoid confusion when the statistical 'class' of units he was investigating was the school. Keeves (1997b, pp. 589-591) argued in favour of the term 'intraclass correlation.'

Missing value imputation

In surveys, it is common for data to be missing (Beaton, 1997). The default treatment for missing data in many statistical packages is listwise deletion. Beaton (1997) noted:

> A technical advantage of this method is that it produces a positive definite or semidefinite cross-products matrix, whether or not the data are missing at random, and estimated or predicted values and their residuals for the observations with complete data can be produced. A serious disadvantage is that this method ignores the potential bias introduced when scores are not missing at random; the resultant estimates are, therefore, biased to an unknown degree. (Beaton, 1997, p. 763)

An alternative to listwise deletion is pairwise deletion. Acock (2005) and Beaton (1997) both commented that this method is a poor choice because it distorts the matrices generated from the data and may produce misleading parameter estimates.

A range of missing data substitution methods is available. Mean value substitution is implemented in some statistical routines in standard statistical packages. Acock (2005) and Beaton (1997) noted that this method reduces the variance and covariance. Acock described it as a poor choice.

Beaton (1997) described the imputation of values estimated using correlated variables, and a variant to this in which a random error term is added to the estimated value. Beaton commented that the use of an estimate is inadequate, but that the use of an estimate with an error is a better method.

Hot-decking, in which a missing value is replaced with the value from the most recent record in the data file with the same characteristics, is used in large scale surveys such as the British Household Panel Survey (Acock, 2005; Starick, 2006, p. 3). Acock (2005) described a single imputation method in which the EM algorithm is used to generate values for the missing observation. This method preserves the variance and covariance matrices of the data set, so avoids the problems noted by Beaton (1997) and described above.

Acock (2005) and Schafer (2000) described multiple imputation methods. The EM algorithm is used to generate values for the missing observations. Each imputation is generated through an iterative approach in which the covariance matrix is preserved, but the imputation process is conducted, typically, five to ten times, so that five to ten versions of the data file are generated. Tabachnick and Fidell (2007, p. 69) reported that multiple imputation can be used with longitudinal data sets and that they retain their sampling variability. They also commented that five, and sometimes as few as three, imputed data sets suffice for subsequent analyses. Beaton (1997, p. 765) reported that he had used five sets of plausible values in the National Assessment of Education Progress surveys. Multiple imputation requires analyses to be repeated and the results of the set of analyses pooled to generate a final set of parameter estimates. Acock noted that software, such as Mplus (Muthen & Muthen, 2006) can accept multiple data files in its procedures, conduct the specified analyses on each file, and then report pooled results.

Acock (2005) concluded (a) that steps should be taken to avoid missing data, (b) that substitution methods, especially mean substitution, are poor choices, and (c) that multiple imputation methods should be used for generating values to be used in analyses.

Summary of missing value methods

A range of methods for handling missing data, including deletions, substitution, single value imputation and multiple imputation have been identified. It would appear that multiple imputation methods, in which multiple versions of data files with imputed values are generated, are the best solution when it is not possible to avoid the problem of missing data.

Chapter 7: Development and Testing of the Problem Solving Assessment Instrument

The project involved two major stages, namely the development of a problem solving assessment tool and an investigation of the growth in problem solving performance over time. In this chapter a study to develop, test and refine a problem solving assessment tool is described.

Purposes of the study

The main purpose of this specific study is to develop, refine and evaluate the measurement properties of a problem solving assessment instrument. The instrument is developed based on the conception of problem solving outlined in Chapter 3. The conception of problem solving is deliberately cognitive. That is, it does not seek to assess affective or conative aspects of problem solving. It is recognised, however, that attitudes and dispositions influence approaches to problem solving and problem solving performance. For this reason, a separate instrument to address those dimensions of problem solving was located and adapted. Its adaptation for the current study is described.

The research was conducted at the Torrens Valley Institute of Technical and Further Education (TVI) in Adelaide, South Australia at the invitation of the director of that Institute. Students and staff in two programs at TVI volunteered to participate in the study. Approval for the study was conveyed in letters from the Department of Education, Training and Employment (see Appendix 1). A copy of the consent form signed by participating students is shown in Appendix 10. The main part of the study was conducted in the School of Electronics and Information Technology with a related study undertaken in the School of Business with students enrolled in the Certificate IV program in Assessment and Workplace Training.

In this chapter, the recruitment and characteristics of participants are described. This is followed by a description of the Problem-Solving Inventory (PSI) and its modification for the current study. The development of the Problem Solving Assessment (PSA) instrument is then described. The results of the administration of the two instruments, the PSI and PSA, are presented. The methods used for analysing the data arising from the administration of the instruments include both classical approaches to scale and item analysis and the Rasch measurement model. The results section focuses on the calibration of the PSA and establishing its measurement properties. Rasch scaled scores for participants from both course groups are used to establish the measurement equivalence of the instrument across groups. The chapter concludes with a discussion of the results.

Recruitment and characteristics of participants

The main project was conducted in the Electronics and Information Technology (E&IT) school of the Torrens Valley Institute of Technical and Further Education (TAFE). The school offered courses ranging from Certificate II to Advanced Diploma levels of the Australian Qualifications Framework (AQF). Students were quite diverse in age and in education and work backgrounds. Some students were full-time, but most were part-time and had work commitments ranging from casual and part-time to full-time jobs.

All students undertaking E&IT programs were informed of the project through the processes outlined in the previous section and were invited to participate. In addition to these promotional efforts, staff noted students' progress through their modules, and when students were about to undertake a module that included one of the assignment tasks that had been identified as having potential for the demonstration of problem-solving ability, an individual email was sent to them advising them of the opportunity available. However, participation was voluntary and no pressure was exerted on students to participate.

Thirty-three students participated and completed research consent forms. Some other students also submitted technical assignment tasks for problem-solving assessment, but did not complete research consent forms, so their data are not included in this report.

In addition to the main study, a subsequent validation study was undertaken, also at Torrens Valley Institute of TAFE, but involving 48 learners who were enrolled in the Certificate IV in Assessment and Workplace and Training.

Selection and development of instruments

Two instruments are used within this research project. The Problem-Solving Assessment instrument is the principal one and is a modification of a prototype developed to assess problem-solving skills. In order to assess participants' attitudes towards problem-solving the Problem-Solving Inventory (Heppner & Petersen, 1982) is used. The development and administration of these instruments is now described.

The Problem-Solving Inventory

The Problem-Solving Inventory (PSI) was developed by Heppner and Peterson (1982). In its original form it had 35 items, three of which were so-called 'test' items. Six response categories were used. These categories were neither labelled nor described in the paper. The items invited respondents to indicate the frequency with which they used the behaviour described in each item. Its developers reported the results of a principal components analysis. They retained three factors and accepted items with loadings above 0.3. Heppner and Petersen labelled the three components Problem-Solving Confidence (PSC, 11 items); Approach-Avoidance Style (AAS, 16 items); and Personal Control (PC, 5 items). Cronbach alphas for the 3 subscales were 0.85 (PSC), 0.84 (AAS), and 0.72 (PC), and 0.90 for the whole instrument. The instrument was shown not to assess the same thing as either intelligence or social desirability scales. It showed concurrent validity with other measures of problem-solving and locus of control.

The instrument was reviewed favourably by Camp (1992) who found that, although primarily a research instrument, it could also be used for “contrasting problem-solving appraisals with more objective measures of actual abilities” (p. 699). The PSI was used by Haught, Hill, Nardi and Walls (2000) to assess the influences of age and education on practical problem solving tasks. These authors did not check the properties of the scale, but did find a positive relationship between the problem solving confidence scale and performance on their practical problem solving tasks (p. 100). Given the favourable reviews of the PSI and the desire to assess conative aspects of problem solving, the decision was taken to use it in the present study. Cognitive aspects of problem solving are assessed using the Problem Solving Assessment (PSA) instrument. This instrument is described below.

In order to avoid response set effects, the original form of the PSI instrument included 15 reverse-scored items. Some of the items included negative statements and one had multiple negative expressions.

> When a solution to a problem was *unsuccessful*, I do *not* examine why it *didn’t* work. [Emphasis added]

Osterlind (1998) suggested that approximately 10 to 15 per cent of items should be reverse scored in order to detect response set effects. The almost 50 per cent of reversed items, including eight with negative statements, in the instrument was thought to add to the cognitive load in interpreting items, so four negative items were reworded. In the revised instrument 11 reverse scored items remained, but only five of these included negative statements.

It is believed that the six unlabelled response categories were unclear. Attempts to develop well differentiated labels for the six categories of behaviour frequency were unsuccessful. Four well-differentiated category descriptors are chosen and are labelled ‘Almost always’, ‘Often’, ‘Seldom’ and ‘Almost never.’ The three test items (items 9, 22 and 29 and for which data were not reported by Heppner and Peterson (1982)) were excluded leaving the revised instrument with 32 items, each with four response categories. The revised version of the PSI is shown in Appendix 4.

Development of the Problem-Solving Assessment instrument

As outlined in the review of problem solving assessment (Chapter 4), most past efforts to assess problem-solving have focused on individuals' performance on common problem-solving tasks (see, e.g., ACER, 2001a). Common tasks avoid the problem of task variability identified by Herl et al. (1999), but the chosen tasks may function differently across groups. It is possible, for example, that some of the differences in performance between discipline-based groups of students result from the salience of the tasks to those groups rather than or in addition to differences in their underlying ability. For example, in the Graduate Skills Assessment trials, engineering students outperformed nursing students on the problem solving scale, but nursing students scored higher than engineering students on the interpersonal understandings scale (ACER, 2001b).

When assessment is based on a set of common tasks, the tasks chosen have been shown to contribute a substantial component of performance variability and therefore to mask the individual contribution to performance variability (M. D. Miller & Linn, 2000; Shavelson et al., 1993). Since the purpose of the problem-solving assessment is to identify individual ability, approaches in which this is contaminated by other factors are compromised. Rater variability is also a source of variation in performance judgments (Andrich, 1999; Linacre, 1997). Where a limited range of tasks is used, the variation in task difficulty can be modelled using programs such as ConQuest (Wu, Adams, & Wilson, 1998). Similarly, rater effects can be modelled. In the current study, these effects are not modelled. It was not feasible, because of the assignment of teaching staff to particular modules each with their own assessment requirements, to construct a design in which raters and tasks were crossed. The implications of this are discussed with the results.

In past efforts to assess problem-solving in a componential, rather than holistic, way separate scoring rubrics were developed for each task (Herl et al., 1999; Shavelson et al., 1993). If this approach were to be taken in the VET context within Australia, the load on assessors would be excessive. Each training package has many units of competency and each unit has many tasks. The process of developing separate rubrics for this number of tasks and then of providing professional development to ensure that they were used consistently would be onerous at system and provider levels and for individual assessors. A similar situation would apply in schools and higher education, although the load may be less obvious as nationally consistent approaches to assessment are not required in those two sectors. Thus, in this project the intention was to develop either a single instrument or a very small number of generally applicable instruments.

The Problem-Solving Assessment (PSA) instrument was designed to assess the use of problem-solving processes directly, as these processes are thought to be important in the emergence of expertise within a domain and also to be transferable between tasks within and possibly between domains. The stages in the development of the PSA instrument are now described.

Converting a conception of problem solving into a measure

In this section, the development of a measurement instrument for problem solving performance, depicted in Figure 6, is described. At the top of the figure, the process of taking a construct (problem solving) and developing a conception of it is shown. In this case, the construct is conceived of as a set of cognitive processes, namely representation, planning, execution, monitoring and reflection. The basis for choosing this set of processes is outlined in Chapter 3. These processes are considered to be cognitive activities and as such, they are not directly observable. Therefore, for each of these processes, a set of observable indicators is developed. In Figure 6, the set of indicators chosen for the planning process is shown. For each of the other processes, a similar set of indicators is also developed. For simplicity, they are not shown in the figure, but they are shown in full in the Problem Solving Assessment instrument presented in Appendix 5. The next stage in the process is the specification of performance standards for each indicator. Again for simplicity, only the performance levels for the 'Plan an approach' indicator are shown. These performance levels are derived from the SOLO taxonomy (Biggs & Collis, 1982). The justification for using the SOLO taxonomy for defining performance standards is presented in Chapter 4.

The set of processes, indicators and performance levels, along with guidelines for the application of these elements, formed the Problem Solving Assessment (PSA) instrument. The instrument is used to assess students' problem solving performance on tasks that students must complete as part of the normal assessment requirements of their course. The results of this problem solving assessment are analysed to test whether they conform to the requirements of measurement.

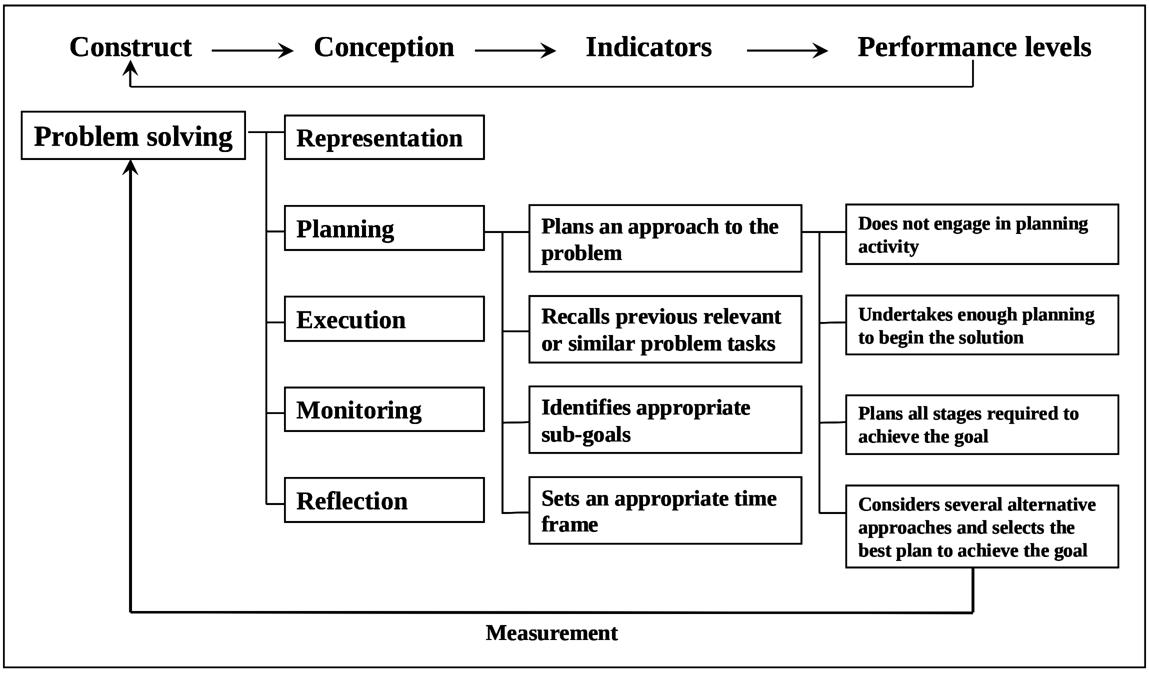


Figure 6: The process of developing a measurement scale for problem solving performance

Problem solving processes

In Chapter 3, several major approaches to describing problem-solving are canvassed. For the reasons elaborated in that chapter, an approach to problem solving assessment based on a set of general processes is preferred. The applicability to the assessment of problem solving within domains for non-experts is the major factor in selecting this approach. Expert problem solving depends upon a well-developed body of domain-specific knowledge that is highly conditioned to the situations in which it is applicable (Alexander, 2003b; Bereiter & Scardamalia, 1986; Chi et al., 1981; Chi et al., 1982; Yekovich, Thompson, & Walker, 1991). Novises, by definition, lack this extensive knowledge base and must proceed using more general strategies. It is through the application of strategies that the applicability of knowledge to situations becomes apparent (Hatano & Oura, 2003; Lajoie, 2003). Having opted for a description of problem solving as a set of general processes, considerable agreement is found among several authors on the processes that are deployed in problem solving (Bransford, Sherwood, Vye, & Rieser, 1986; Bransford & Stein, 1984; Hayes, 1989; Polya, 1957). These processes were:

- apprehending, identifying and defining the problem;
- planning an approach to the problem including selecting strategies;

- carrying out the chosen plan;
- monitoring progress towards the goal; and
- reflecting on the effectiveness of the solution attempt.

In the Problem-Solving Assessment instrument (see Appendix 5), these processes are labelled representation, planning, execution, monitoring and reflection.

Indicators of problems-solving processes

For each of these five major processes, a set of indicators is sought. The indicators used in the PSA are established by answering the question: 'What would a competent person do in showing that he or she is able to apply the component process in a real situation?'

Indicators of performance are the basic elements of measurement – they are the items that form the hypothesised scales of the problem-solving construct being measured. In the case of the PSA instrument an overall scale of problem-solving ability is hypothesised; it is assumed that there is such an ability and that all the indicators form a single factor that reflects that construct. Further, this single factor does have components, each of which is internally coherent and each of which contributes to the overall factor. In measurement terms, it is hypothesised that there are a set of sub-scales that together contribute to a coherent overall scale (Bejar, 1983, p. 31).

In order to avoid a major threat to validity – that of construct under-representation (Messick, 1998) – each scale must reflect fully the content of the construct, so that each component process must be represented in the overall scale (content validity). However, to be practically useful, the scale must have a limited number of items – probably between 15 and 25 – so that each component process is limited to between three and six indicators. The indicators that are finally selected are shown in the PSA instrument (see Appendix 5). In order to maintain compatibility with the Mayer Committee's recommendations on key competencies assessment,[19] one indicator, Application of strategies, is added to the Execution process. This has three performance levels reflecting the three levels suggested by the Mayer Committee (AEC, 1992).

Performance levels for problem-solving indicators

Several bases for the establishment of performance levels are available. They are canvassed in some detail in Chapter 4 (Purpose, Forms and Outcomes of

[19] In this chapter, reference is made to 'key competencies.' When this component of the study was undertaken, the Mayer key competencies had been endorsed in the VET sector. The employability skills framework was under consideration, but had not been adopted.

Assessment) and are summarised here. The Mayer Committee's suggested that performance levels could be used. The three levels are: (a) apply a given procedure, (b) select a procedure from several alternatives, and (c) adapt or create a procedure. They were designed to be pragmatic and no theoretical basis was presented for them in the Committee's report, although it is possible that, respectively, the 'apply', 'evaluate' and 'synthesise' or 'create' levels of Bloom's taxonomy informed the Committee's deliberations. The Mayer Committee's recommended performance levels have been used in making holistic judgements of performance, but finer grained judgements have been shown to be reliable (McCurry & Bryce, 1997). The main concern with the levels suggested by the Mayer Committee is that they are applied to all key competencies and appear to reflect increasing autonomy rather than key aspects of the target construct. Their validity as measures of problem solving is, therefore, suspect.

A second alternative is the revised version of Bloom's Taxonomy of Cognitive Objectives (L. W. Anderson & Krathwohl, 2001). This provided six levels of cognitive skill: remember, understand, apply, analyse, evaluate and create. These cognitive skill descriptions represented a hierarchical sequence of increasing cognitive skill, albeit with some doubt about the relative positions of three highest levels. They were designed to operate in conjunction with the levels of the knowledge dimension of the revised taxonomy, and do not appear to be sufficiently consistent with the major problem-solving processes that were derived from the literature on problem solving.

A third alternative, the SOLO (Structure of the Observed Learning Outcome) taxonomy (Biggs & Collis, 1982), is selected. This taxonomy was based upon the cognitive complexity of individuals' responses to the application of knowledge in learning and problem situations. It recognised levels of performance from ineffective use of knowledge to very complex and abstract application. Adapted descriptions for each level of the SOLO taxonomy are shown in Table 17. The simplified descriptors are used in the current study to inform staff and students of the basis for the assessment of problem solving performance.

Table 17: Performance levels of indicators using the SOLO taxonomy

SOLO level	Description	Score
Pre-structural	No knowledge, inaccurate recall, or does not use relevant knowledge	0
Uni-structural	Uses relevant knowledge/skill elements in isolation	1
Multi-structural	Uses relevant knowledge/skill elements in combination	2
Relational	Can generalise using knowledge within the problem situation	3
Extended abstract	Can extend what has been found through the current problem to other situations	4

Note: The descriptions are adapted and simplified versions of the original descriptions provided by Biggs and Collis (1982). See Table 14 in Chapter 4 for the original descriptors.

An advantage of the SOLO taxonomy is that its five levels form a set of ordered responses. These responses are thus amenable to analysis using Item Response Theory (IRT) and may form the basis of an interval scale of problem solving performance. This possibility is tested with data collected using the PSA. IRT modelling provides information on the precision of problem-solving ability assessments and indicates the number of performance levels that can be discriminated reliably.

The levels of the SOLO taxonomy are applied to descriptions of performance on each indicator and provide a means of scoring student performance. Not all SOLO levels are applied to all indicators. In order to make the instrument as easy to use as possible, the number of SOLO levels selected for each indicator is based upon the anticipated ability of assessors to make reliable judgements of student performance. Thus, for some indicators, for example 'Sets a realistic goal', only two levels are suggested, while for others, for example 'Plans an approach to the problem,' four levels are used.

Refinement of the Problem Solving Assessment Instrument

The teaching staff of the School of Electronics and Information Technology formed a reference group for this phase of the project. A draft version of the PSA with a rating scale layout was developed, see Appendix 5, and this was presented to staff. The focus group discussion was framed around the useability of the instrument. In its draft form, some staff were concerned about the format of the tool, expressing a preference for a more familiar rating scale design in which all items had the full range of response options. This would have meant applying a common set of performance levels to all indicators but, as argued above, it is apparent that some indicators afford only low level responses. Offering a full range of response options may have led to some staff using only the lowest two categories to discriminate between non-performance and performance while other staff may use the extreme categories. This would compromise the meanings imputed to the various response categories and therefore the measurement properties of the instrument.

A second focus of discussion was the useability of the instrument by students, as self assessment by students was included by design in the assessment model. (The assessment model is discussed below). The decision was taken to integrate the criteria for each performance level for each indicator in the text of the instrument. While this requires more information to appear on the assessment form, it does mean that the criteria are immediately apparent to students and to teachers and it avoids the need for extensive separate information about the criteria. Additional information about the criteria is provided on the reverse side of the form, but the instrument is limited to a single sheet. The revised version of the PSA is shown in Appendix 5 and this is the version that is used in this study.

Administration of the Problem Solving Assessment instrument

A set of procedures is established for the application of the PSA. These include gathering background information on student participants, the selection of tasks for assessment, self assessment by students, and validation by their teachers. These procedures are now described.

The project was promoted extensively to students. Posters were displayed in the laboratories and common areas where students gathered and the home page when students logged on to computers on campus displayed information about key competencies. A lunchtime meeting was organised at which pizza and soft drinks were provided and representatives from the electronics industry addressed students on the skills they sought when employing graduates. The industry representatives, all employers, had been asked to draw attention to generic as well as the technical skills they sought in making hiring decisions.[20] Students were invited to participate in the assessment project, and those who volunteered were asked to complete a Consent Form, a Personal Details Form, and the Problem-Solving Inventory. The completion of these forms took approximately 20 minutes.

Task Selection

In order to ensure that the assessment of generic skills occurs in context, only existing assessment tasks are used for the problem solving assessment. Deliberately, no new tasks were constructed. In preparation for the problem solving assessment, each teacher was asked to make a list of all assessment tasks in all modules of work. From the list, teachers were asked to nominate those that provided scope for students to apply, develop and demonstrate their problem solving skills and to restrict this list of designated tasks to three or four in each module. The tasks recommended for problem solving assessment are shown in Table 18. The lists of tasks that were proposed for assessment were discussed at a staff meeting so that all teachers were aware of the tasks that their colleagues had chosen. Other key competencies were also being assessed, and it was important to ensure that the total assessment load should not become onerous for teachers or students. By having teachers nominate a limited number of assessment tasks that were identified as being suitable for problem solving development and

[20] These activities were facilitated by Rob Denton, whose very enthusiastic assistance in this research is greatly appreciated.

assessment, students were guided in their selection of tasks to submit for this assessment. Students were free, however, to choose any tasks for this assessment.

Table 18: Course modules and recommended problem solving assessment tasks

Module	Assessment activity
AC Principles	Practical Test
Applied Electricity 1	Practical Activity 2.3 – Electrical Circuits
Digital Electronics 1	Practical Activity 5.1 – Digital Displays
Digital Subsystems 1	Practical Activity 2.2 – Switch Debouncing
Electrical Fundamentals	Practical Test 1
Electrical Fundamentals	Distinction Activity (extension to Prac test 1)
Embedded Systems	Practical Activity 2.3 – Interfacing a Microcontroller Development Board to a PC
Hardcopy	Practical Activity 5.1.1 – Practical Application
Intro to Electricity & Electronics	Practical Activity 2.5 – Basic Circuit Measurements
Intro to Programming	Final Programming Project – Game Program (5-in-a-row)
Microprocessor Fundamentals	Practical Activity 5.3 – Using a RAM Chip
PC Systems Support	Practical Activity 6.1 – Application/Hardware Installation
Power Supply Principles	Topic 11 – Faultfinding Techniques
Single User Operating Systems	Topic 4.1 – Configuring Operating System
Single User Operating Systems	Topic 4.1 Credit Activity (extension)
Soldering – Basic	Project kit

Note: An expanded version of this task list is presented as Appendix 7.

Self Assessment

The assessment of problem solving in this project is designed to be an iterative process. It is hypothesised that students learn about problem-solving by being involved in self-assessment and that they improve their performance in successive assessments. Thus the assessment procedure is intended to be an overt process of both learning and assessment.

Students and staff use the same form of the PSA instrument. This is done in order to ensure that the process is an open one in which all assessment criteria are clearly laid out for students. They are aware of exactly how their lecturer conducts the validation of the assessment (see below) and what evidence their lecturer will seek in the validation. Indeed, the reverse side of the PSA form includes brief scoring instructions and sets of questions that relate to each indicator used in the assessment. The questions are designed to focus attention on the evidence that is expected to support the levels of performance that are suggested for that indicator.

The process is an iterative one, in which students undertake an initial assessment. Through this, they are expected to learn about the processes of problem solving and how these processes are applied in authentic situations. Students are also expected to become consciously aware of the processes they apply and to identify evidence of the application of these processes in practice. Thus, in subsequent assessment activities, students should become more proficient users of these processes and be better able to identify and present evidence of them.

Intended outcomes of this process are that students become both more proficient in their use of problem solving processes by gaining explicit knowledge of them and are better able to describe what problem solving is in the context of their work.

Lecturer Validation

The term 'lecturer validation' is adopted within the project as it describes the process by which students' judgments of their own work are reviewed be lecturers and feedback is provided to students. Lecturer validation comprises two elements. First, students are required to present their self-assessment forms and to present or refer to evidence in their assessment reports to support their self-assessed result. The second element of the process is the judgment, by the lecturer, of the problem solving performance based upon the evidence presented by the student.

When the lecturer makes a judgement, the assessment is discussed by the lecturer and the student. The purpose of this discussion is to draw attention to aspects of the assessment in which the student has either not presented evidence or has misinterpreted either the evidence or the criteria. The purpose of this discussion is clearly instructional with the aim of enhancing the student's understanding of both problem solving and the process of assembling and judging evidence against specified criteria.

The performance level that is recorded in the student record system is the lecturer's judgment. Since the assessment of Key Competencies at Torrens Valley Institute (TVI) results in the award of a Statement of Attainment,[21] accountability for the recorded level of attainment lies with the institution and its staff. It is essential that the processes that underlie a Statement of Attainment are credible and robust and that they withstand external review such as quality audits.

The assessment of problem solving

The assessment of problem solving occurs through the use of an assessment tool (the Problem Solving Assessment instrument) and a particular assessment process. Together, the instrument and the process by which it is administered constitute an assessment approach that seeks to assess problem solving as a construct that is defined according to a particular cognitive view (see Chapter 3). Second, that particular construction of problem solving is used to define a "progress variable"

[21] Statements of Attainment are no longer used for certifying generic skills achievement. They were a means of certifying learning when a student had completed a component of a module, but had not completed that module. They were issued under the authority of a Registered Training Organisation (RTO).

(R. E. Mayer, 2000, pp. 524-526; Wilson & Sloane, 2000, p. 184) against which students' acquisition of problem solving competence is revealed. Third, that progress variable, instantiated in the PSA, is common to all tasks on which students' assessments of problem solving are undertaken (Wilson & Sloane, 2000, pp. 191-192). Fourth, the assessment of problem solving (a co-curricular generic skill) is embedded in the substantive curriculum that students experience in their courses. Fifth, the use of self-assessment followed by lecturer validation and feedback is designed to close the learning and assessment loop so that assessment contributes to learning. The current study, designed to evaluate the Problem Solving Assessment instrument, does not test the possible relationships between assessment, feedback and learning. This matter is addressed in a follow-up study (see Chapter 8).

Data collection and analysis

Data arise from four main sources. First, when students were recruited into the project they completed a brief Personal Details form through which demographic data were acquired. At that time, students also completed the Problem-Solving Inventory which provided information on students' approaches to problems, their confidence in problem-solving, and their sense of control over problem-solving processes. The third, and principal, source of data for the project was the lecturers' assessments of student performance using the Problem Solving Assessment instrument. The fourth data source was qualitative and includes comments made by students about the assessment process. Comments are sought from both students who participated in the study and those who had chosen not to take part.

Quantitative data collection and analysis

The data that are recorded on printed versions of the instruments were entered into data files for subsequent analysis using a range of software packages including SPSS (SPSS Inc., 1995), Quest (Adams & Khoo, 1999), and RUMM (Sheridan et al., 1997). Several approaches are taken to the analysis of quantitative data. In order to inform those who are familiar with more traditional approaches to data analysis, factor analyses and scale reliabilities analyses are conducted. However, one of the purposes for conducting the project, and for the chosen design of the instruments, is to establish the validity of measurements of the identified constructs. In the discussion of measurement in the Chapter 6, the point is made that simply assigning numbers as a result of a set of scoring rules does not constitute measurement. Measurement requires that (a) scores are reported on an interval scale, (b) the measurement is independent of persons and tasks, (c) estimates have a known precision, and (d) the items that contribute to the measurement are unidimensional. A powerful tool for the conversion of ordered responses to interval scores of known precision is available through the Rasch measurement model. This is described in Chapter 6, Research Methods.

Qualitative data collection and analysis

Although the study is designed primarily as a quantitative one, some qualitative information is gathered as a result of an online evaluation survey in which students are asked to comment on various aspects of the new approach to the assessment of problem solving that is being trialled. These comments are recorded in the student records system and are reviewed and analysed.

Results

In this section, the results of analyses of data collected through the study are presented. Data were collected from two sample groups, both at Torrens Valley Institute of Technical and Further Education. The first sample consisted of students enrolled in Electronics and Information Technology (E&IT) courses and the second of students undertaking a module of the Assessment and Workplace Training (AWT) course. The E&IT results are discussed first, followed by the AWT data. The characteristics of participants are summarised. Then the results of analyses of the Problem-Solving Inventory (PSI) are presented in summary form, followed by the results of more detailed analyses of the Problem Solving Assessment instrument. In both cases, the outcomes of factor and scale analyses are shown before the results of the Rasch method. Relationships among attitude, problem-solving ability and educational achievement are presented. This chapter concludes with an analysis of students' evaluations of the processes of assessing Key Competencies.

The main study involves 43 assessments undertaken by 33 students enrolled in Electronics and Information Technology (E&IT) courses. A subsequent replication study, conducted within the Certificate IV in the Assessment and Workplace Training course, involves 48 participants, each of whom submitted one assessment. A further eight assessments were subsequently submitted by E&IT participants. In all, 99 assessments were completed using the PSA.

Participants

Of the 33 students who participated in the main study, 25 students submitted one piece of work for problem-solving assessment, six submitted two, and two individuals submitted three assignments, for a total of 43 problem-solving assessments. Each assignment consisted of a single task.

Thirty of the students who participated are males and three females. The ages of participants range from 17 to 50 with a mean of 29.5 years.

The previous educational attainment of participants varied. Three had completed Year 10, six had completed Year 11, 22 had completed Year 12, and two did not respond to this question. The post-school educational attainment also varied. Thirteen individuals had no post school education other than the E&IT course that they were undertaking, six had undertaken other VET sector courses, seven had completed other VET sector awards, three had undertaken study at university, and three did not provide this information.

The distribution of work experience among participants is bimodal. Nine students report having no work experience, six have less than a year in the workforce, two have from two to five years experience, and 22 people have more than five years experience in the workforce. One student did not respond to this item. Among the students who submitted work for a problem-solving assessment, there appears to be a bias towards older individuals with considerable experience of work and this may influence aspects of the results.

The Problem-Solving Inventory

The Problem-Solving Inventory (PSI) is not a principal focus of the current study but is used to provide some information on participants' attitudes to problem solving. For this reason, although the results obtained from its administration are subject to extensive analyses, only summaries of these analyses are presented here.

Factor Analysis and Classical Item Analysis

A factor analysis is conducted on the data gathered from the administration of the PSI to 33 participants from the E&IT program. The analysis employed SPSS (SPSS Inc., 1995) and, following the analysis presented by Heppner and Petersen (1982), a three factor model is specified. The Eigen values for the three factors are 8.01, 4.62 and 2.76. Six components have Eigen values greater than 1.0, but inspection of the scree plot suggested that the three factor solution is adequate. The three factors together account for 48.1% of observed variance. The component loadings are shown in Table 19 and the correlations among the components in Table 20. The loadings indicate that the scales are poorly differentiated in the current data set. Given the modest size of this data set, it is considered inappropriate to revise the instrument structure, and the scales proposed by the developers are retained for subsequent analyses.

The Personal Control scale reflects confidence to proceed with problems, and not always with a perceived need for planning.

Classical item analysis is conducted using SPSS. The Cronbach alpha values for the three sub-scales are AAS 0.85 (16 items), PSC 0.86 (11 items), and PC 0.79 (5 items). Rather than explore these further with classical analysis, the instrument is re-analysed using the Rasch measurement model with Quest (Adams & Khoo, 1999).

Rasch Analysis

Rasch analyses are undertaken for two purposes. First, the Rasch model is able to detect deviations from expected patterns of responses of both items and respondents. Items that fail to fit the measurement model are not measuring the intended construct and contaminate the measurement. Second, once a set of fitting items has been established, estimates of respondents' attitudes are generated on interval scales for later comparison with PSA scores.

The partial credit model (Masters, 1982) is used in examining the PSI. While the rating scale model is applicable to the scale, since all items have a common set of response options, the partial credit model can detect differences in threshold patterns between items and is therefore a useful diagnostic tool. This analysis indicates that three of the 32 items reveal poor fit to their intended scales, and in stages, these items are removed from the analysis. The main basis for removing items is that they show a weighted MS of >1.35, although item discrimination is also examined. Generally, items with high Infit MS values also show low (<0.4) item discriminations. The Cronbach alpha values of the revised scale and sub-scales are: AAS 0.83 (16 items); PSC 0.87 (9 items); and PC 0.77 (4 items).

Table 19: Results of principal components analysis of PSI responses

Item No.	Scale	Item text (abbreviated)	Components 1	2	3
1	AAS	When a solution to a problem was unsuccessful…			0.475
2	AAS	When I am confronted with a complex problem…	0.496		0.443
4r	AAS	After I have solved a problem, I do not analyse…			-0.597
6	AAS	After I have tried to solve a problem with a…	0.512		
7	AAS	When I have a problem, I think up as many…	0.532		
8	AAS	When confronted with a problem, I consistently…			0.686
13r	AAS	When confronted with a problem, I tend to do…	-0.773		
15	AAS	When deciding on an idea or a possible solution…	0.532		
16	AAS	When confronted with a problem, I stop and…			0.751
17r	AAS	I generally go with the first good idea that comes…	-0.603		
18	AAS	When making a decision, I weigh the…	0.570		
20	AAS	I try to predict the overall result of carrying out a…			0.627
21r	AAS	When I try to think up possible solutions to a…		0.707	
28	AAS	I have a systematic method for comparing a…	0.708		
30	AAS	When confronted with a problem, I usually…		0.474	0.484
31	AAS	When I am confused by a problem, one of the…	0.495		0.715
3r	PC	When my first efforts to solve a problem fail…		0.579	
14r	PC	Sometimes I do not stop and take time to deal…		0.737	
25r	PC	Even though I work on a problem, sometimes I…		0.762	
26r	PC	I make snap judgements and later regret them…		0.449	
32r	PC	Sometimes I get so charged up emotionally that I…		0.853	
5	PSC	I am usually able to think up creative and…	0.653		
10	PSC	I have the ability to solve most problems even…	0.524		0.510
11r	PSC	Many problems I face are too complex for me to…	-0.741		
12	PSC	I make decisions and am happy with them later.			0.617
19	PSC	When I make plans to solve a problem, I am…	0.408		
23	PSC	Given enough time and effort, I believe I can…	0.479		0.580
24	PSC	When faced with a novel situation, I have…	0.725	-0.439	
27	PSC	I trust my ability to solve new and difficult…	0.699		
33	PSC	After making a decision, the outcome I expected…	0.566		
34r	PSC	When confronted with a problem, I am unsure of…		0.735	
35	PSC	When I become aware of a problem, one of the…	0.609		

Notes: The complete item text is presented in Appendix 4. Item numbers are from the original form of the PSI (Heppner & Petersen, 1982). N=33 respondents, Promax rotation. Loadings <0.4 are suppressed. AAS= Approach avoidance style; PSC= Problem solving confidence; PC=Personal control; r=reversed item.

Table 20: Correlations among PSI components following Promax rotation

Component	1	2	3
1	1.000		
2	-0.142	1.000	
3	0.305	0.137	1.000

Note: Promax rotation is used as it is expected that the components are correlated.

A final set of 29 items is retained and estimates of individuals' scores on these scales are generated in preparation for later analyses.

The Problem Solving Assessment

The Problem Solving Assessment (PSA) instrument is a newly developed assessment tool, and requires validation. Conventional approaches (factor and scale reliability analyses) and the Rasch measurement model are used in the validation of the instrument and the results of these analyses are reported.

The decision was taken early in the design of the current study to use existing assessment tasks and to have problem solving performance assessed by the teachers who are responsible for the routine technical assessment of tasks. A consequence of this decision is that a balanced block design with assessors crossed with both tasks and candidates, which would be a feature of an experimental approach to the assessment, has not been possible. The naturalistic design used in the Electronics and Information Technology (E&IT) course is presented in Table 21, which shows the number of student assessments conducted on the available tasks by participating teaching staff. A much more balanced representation of the seven assessors and 18 tasks was anticipated. It is apparent that one assessor accounted for about two-thirds of all assessments and a single task (STET12) accounted for one-third of all assessments. It is not feasible, therefore, to model task and rater effects. A consequence of this is that measurement error is conflated with rater and task effects with expected randomness in the judgment of student work against criteria for the indicators in the instrument. In addition to the E&IT assessments a further 48 assessments were undertaken in the Assessment and Workplace Training (AWT) course, all on a single task and graded by the same assessor.

The form of the PSA that is used in this study is presented in Appendix 5 (Version 2 of the PSA). In the tables that follow, abbreviations are used for each of the indicators in that instrument. The numbered indicators and their abbreviations are shown in Table 22. For convenience, indicator labels are used in subsequent tables and indicator numbers in subsequent figures.

Table 21: Assessors who participated and tasks used in problem solving assessment

	Assessor ID							
Task ID	STEA01	STEA02	STEA03	STEA04	STEA05	STEA06	STEA07	Total
STET02	2	0	1	0	0	0	0	3
STET03	1	0	0	0	0	0	0	1
STET04	3	0	0	1	0	0	0	4
STET05	0	0	0	0	3	0	0	3
STET08	1	0	0	0	0	0	0	1
STET09	0	1	0	0	0	0	0	1
STET12	13	0	0	0	0	0	0	13
STET13	1	0	0	0	0	0	0	1
STET15	2	0	0	0	0	0	0	2
STET17	1	0	0	0	0	2	0	3
STET18	4	0	0	0	0	0	0	4
STET19	1	0	0	0	0	0	0	1
STET20	1	0	0	0	0	0	0	1
STET21	1	0	0	0	0	0	0	1
STET22	1	0	0	0	0	0	0	1
STET23	1	0	0	0	0	0	0	1
STET24	1	0	0	0	0	0	0	1
STET25	0	0	0	0	0	0	1	1
	34	1	1	1	3	2	1	43

Table 22: PSA Major Processes, indicators and abbreviations

Indicator		Indicator text
No.	Label	
		Defining the problem
1	Rep01	Forms a correct understanding of the problem
2	Rep02	Recognises relevant given information
3	Rep03	Identifies the need for additional information
4	Rep04	Recalls relevant information
5	Rep05	Sets a realistic goal
		Planning an approach
6	Pln01	Plans an approach to the problem
7	Pln02	Recalls previous relevant or similar problem tasks
8	Pln03	Identifies appropriate sub-goals
9	Pln04	Checks that required equipment is available
10	Pln05	Sets an appropriate time frame
		Carrying out the plan
11	Exe01	Begins to follow the set plan
12	Exe02	Activates relevant knowledge
13	Exe03	Uses relevant skills
14	Exe04	Application of strategies
		Monitoring progress
15	Mon01	Checks that set plan leads toward problem goal
16	Mon02	Response to deviations from expected progress
17	Mon03	Reviews original plan
18	Mon04	Checks problem representation
		Reflecting on the result
18	Ref01	Reviews efficacy of problem approach
20	Ref02	Compares current problem with previously encountered ones
21	Ref03	Anticipates situations in which current problem approach might be useful

Response Frequencies

Before analysing the results, it is useful to examine the frequencies with which performance levels are endorsed for the indicators used in the PSA. The frequencies at which these performance levels were endorsed in both the E&IT and AWT assessments are shown in Table 23. The lowest performance level for each indicator is, in general terms, 'the student did not engage in this activity.' With few exceptions, very few students were judged not to have engaged in indicator activities at all. This is perhaps not surprising as students were asked to use the PSA in judging their own performance and the instrument would have served as a prompt in their approaches to the problem solving activities.

Table 23: Frequencies of Problem Solving Assessment indicator performance levels

Indicator	Response category					Total
	0	1	2	3	Missing	
Rep01	1	22	75		1	99
Rep02	0	44	55		0	99
Rep03	4	90			5	99
Rep04	1	45	49		4	99
Rep05	3	96			0	99
Pln01	0	24	32	41	2	99
Pln02	8	88			3	99
Pln03	11	86			2	99
Pln04	2	96			1	99
Pln05	14	83			2	99
Exe01	5	92			2	99
Exe02	3	28	67		1	99
Exe03	0	23	74		2	99
Exe04*	23	29	43		4	99
Mon01	5	93			1	99
Mon02	2	20	18	52	7	99
Mon03	8	34	47		10	99
Mon04	10	76			13	99
Ref01	2	47	45		5	99
Ref02	13	41	32		13	99
Ref03	4	90			5	99

Note: *EXE04 is the Mayer key competency performance level indicator.
Blank cells represent performance levels that were not available on the assessment tool.

One of the reflection activities (REF02) and two of the monitoring activities (MON03 and MON04) are exceptions in that more students did not demonstrate these activities. Three of the planning activities reveal greater variation in performance levels than most items. The highest available performance level was endorsed in over half of all assessment for 12 of the 21 indicators. This skewed response pattern may reflect the voluntary nature of the activity with more able students being more likely to seek this additional form of assessment. Performance levels were not assigned in cases where the lecturer believed that the indicator was not relevant to the activity. Significant numbers of missing performance levels occurred for one reflection and two monitoring indicators. In calibrating the instrument, these missing responses were coded as being at the pre-structural stage; that is, at the lowest performance level as lecturers reported omitting criteria for which no evidence was presented.

Factor analysis

An exploratory principal components analysis is undertaken on the E&IT data (n=43) using SPSS (SPSS Inc., 1995). A scree plot suggests that three components can be identified, although 5 have Eigen values greater than 1.0. However, some caution must be exercised in interpreting the structures because items that are skewed tend to have suppressed correlations with comparable items that are not skewed. This leads to the recognition of additional factors in order to account for the pattern of correlations. Certainly, there are some skewed indicators, and possible reasons for this are discussed later. In addition, some indicators have zero variances (see Table 24) and these items are removed from the analysis. Interpretation of the component structure therefore is not completely clear. There appears to be a separation between the Representation, Planning and Execution indicators on one hand and the metacognitive ones of Monitoring and Reflection on the other. The most readily interpretable structure involves only two components. Only two indicators have substantial loadings (>0.4) on the third component, and both have stronger loadings on the first component.

Table 24: Rotated factor solution for the Problem Solving Assessment

	Factor 1	Factor 2	Factor 3
Rep01	.418	.769	
Rep02	.683		
Rep03	.586		
Rep04	.571		
Pln01	.663		.594
Pln02	.801	.439	
Pln05	.484	.578	
Exe02		.671	
Exe03	-.532	.508	
Exe04	.782	.405	.459
Mon02		.801	
Mon03		.664	
Mon04		.712	
Ref01		.753	
Ref02		.679	
Ref03	-.515	.593	

Principal components extraction with Oblimin rotation. Factors loadings <0.4 are suppressed.

The lack of clear component separations is not surprising. Although problem solving is construed as comprising identifiable processes, these processes are believed to be applied iteratively. It is believed that successful problem solvers apply these processes at high levels while less successful individuals may not apply them all or may apply them to a limited extent. The apparent component separation may reflect a tendency for some individuals to operate in a lower cognitive register only and for others to exercise the full range of cognitive, including meta-cognitive, activity.

The constraints on the variables (low variance and skewed responses) along with the lack of clear component separation suggest that the optimum solution may have a single factor. This proposition requires confirmatory factor analysis for which a greater number of cases are required. The single factor solution is supported by the Rasch analysis (reported below).

Classical item analysis

The results of a classical item analysis, conducted using the SPSS Reliabilities command, are shown in Table 25. Some items have zero variance. This is likely to reflect some bias in the candidates who volunteered for this assessment. It seems likely that more confident and more able students might have volunteered and therefore to have performed uniformly well on some of the indicators. Cronbach alpha for the scale is 0.80.

Table 25: Results of reliabilities analysis for the complete PSA scale

Item	Scale Mean if Item Deleted	Scale Variance if Item Deleted	Corrected Item-Total Correlation	Alpha if Item Deleted
Rep01	24.375	19.050	0.535	0.778
Rep02	24.563	19.729	0.251	0.792
Rep03	25.313	19.829	0.378	0.787
Rep04	24.563	18.929	0.441	0.781
Rep05	25.188	21.096	0.000	0.797
Pln01	24.125	15.850	0.664	0.760
Pln02	25.438	18.929	0.505	0.779
Pln03	25.188	21.096	0.000	0.797
Pln04	25.188	21.096	0.000	0.797
Pln05	25.563	19.063	0.409	0.783
Exe01	25.188	21.096	0.000	0.797
Exe02	24.313	20.363	0.200	0.794
Exe03	24.313	20.763	0.070	0.799
Exe04	25.188	14.429	0.781	0.747
Mon01	25.188	21.096	0.000	0.797
Mon02	23.500	19.733	0.187	0.798
Mon03	25.188	18.563	0.318	0.792
Mon04	25.688	18.363	0.557	0.774
Ref01	24.813	18.429	0.563	0.774
Ref02	25.625	17.450	0.513	0.775
Ref03	25.250	21.000	0.015	0.799
Scale alpha				0.795

Problem solving is conceived as having five major processes. These processes, however, are not expected to form discrete scales. They are considered to be processes that are enacted iteratively during problem solving. That is, if it becomes apparent during monitoring that the solution is not as expected, the candidate is likely to revise the representation of the problem or to consider alternative plans. Thus, separate subscale analyses are not presented.

The results of the classical item analysis indicate that the PSA instrument has satisfactory scale properties despite the lack of variance of some items. However, more detail is available from Rasch analyses.

Rasch analysis: Calibration

Analyses are conducted in two stages. In the first stage, the instrument is evaluated and calibrated by examining the properties of (a) the scale as a whole, (b) the indicators, and (c) the performance level thresholds. In the second phase, Rasch scaled scores are estimated for individuals and indicator locations are compared between the two course groups. Data from both the Electronics and Information Technology (E&IT) and the Assessment and Workplace Training (AWT) courses are used for these analyses. The indicator for the Mayer performance level was removed from the analyses as this characteristic is not one of the identified problem solving processes. A judgment of the Mayer performance level does reflect an aspect of task difficulty, but this is a separate matter from individual performance. The scale as it is analysed, therefore, includes 20 indicators with data from 99 participants.

Rasch partial credit analyses are undertaken on the PSA using Quest (Adams & Khoo, 1999). The partial credit model is required as the indicators have different numbers of response categories. The purposes of these analyses are to evaluate (a) the measurement properties of the scale, (b) the fit of individual items to the scale, and (c) the separation of thresholds between performance levels within indicators. The analysis is also used to generate interval estimates of students' abilities. Results from the Quest analyses are reported in this section.

Scale Coherence

The instrument has adequate scale properties. The Item Separation Reliability estimate, a measure of the separation of indicator difficulties and performance level thresholds, is 0.78. The Person Separation Reliability index is 0.71. This is an estimate of the extent to which the indicators of the scale are able to discriminate among participants of different interval scaled abilities and is approximately equivalent to the Cronbach alpha (0.74), which is based on raw (non-interval) scores. These two indices reflect both the range separation of item locations on the scale formed by the set of indicators and the standard errors of these location estimates. Wright and Masters (1982, pp. 91-92) introduced these indices and showed they could be used to estimate the number of performance bands that could be recognised. Using their formulae, three distinct performance bands can be recognised for problem solving performance. Thus, while up to five potential SOLO performance levels were available for individual indicators, three overall performance levels appear to be separable for problem solving assessed using the PSA.

The standard errors of estimate of indicator and performance level thresholds normally are taken to indicate measurement error. However, it should be noted that rater and task facets are not modelled, so the standard errors of indicator difficulty estimates include those facets. If these were able to be modelled separately, it seems likely that the true measurement error would be somewhat lower than is indicated in the current analyses. If this were the case, more favourable Item and Person Reliability indices might be found.

In addition to the indices reported above, it is useful to compare the range of performances captured by the indicators and their performance levels with the range of problem solving performances shown by individuals in the sample. The set of locations of indicators and of performance level thresholds effectively mark out meaningful levels on the scale of problem solving performance. By default, the zero point on Rasch scaled measures is set at the mean item difficulty – in this case the mean indicator difficulty. If the range of performance level thresholds is truncated, the measurement scale will not provide satisfactory estimates for individuals whose performance lies substantially beyond the calibrations of the scale. The lowest performance level Thurstone threshold was -1.83 logits and the highest was 2.71 logits, a range of 4.54 logits. The mean person measure is 2.02 logits, ranging from -1.34 to +5.10 logits. In order to examine the adequacy of the range of thresholds for the sample of individuals being assessed, it is useful to examine plots of the frequencies of individual scores and performance level threshold locations. This is done in Figure 7. It is feasible to estimate person performances slightly above and below the range of performance level thresholds because of the stochastic nature of scoring, but it is apparent from the mismatch between the distributions of person competence levels and indicator performance levels that the instrument is not well targeted to this sample. It should be noted, however, that these samples are of volunteers and are likely to include more high-achieving students rather than being representative of all students. The specification of additional higher-level performance standards is likely to improve the match between the abilities demonstrated by this sample and the indicators.

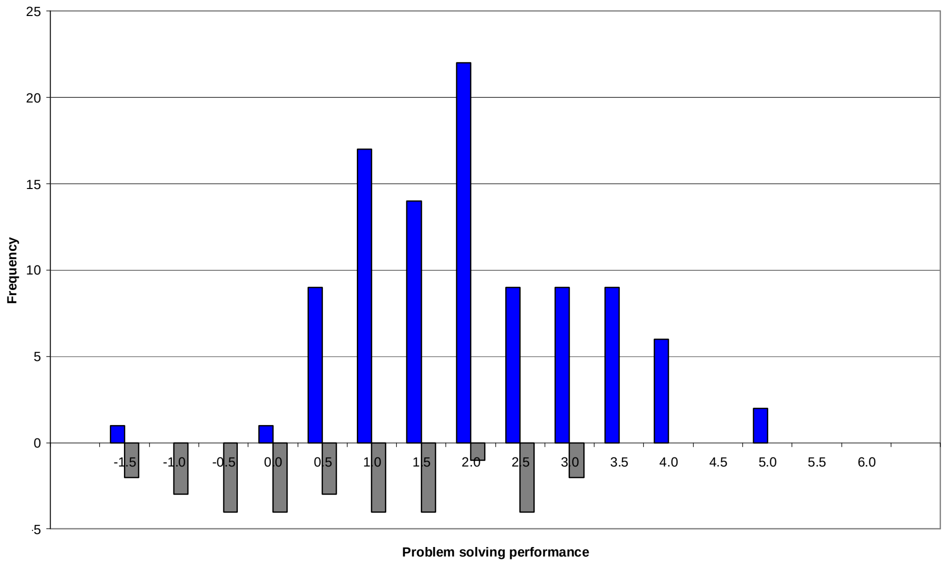


Figure 7: Frequency distributions of individual competences (above the horizontal axis) and performance level thresholds (below the axis) along the problem solving performance scale

Indicator Fit

Item fit is assessed in Quest using the information weighted mean square (IMS) deviation from expected, Rasch modelled, responses. A diagram showing the fit parameters of the PSA indicators is presented in Figure 8. In that figure a vertical line at an IMS value of 1.0 indicates the expected amount of deviation from perfectly predicted responses. Two other vertical bars are shown at IMS values of 0.71 and 1.30. This is the range recommended for “run of the mill multiple choice tests” (Bond & Fox, 2001, Table 12.6, p. 179). The information weighted mean square fit indices for indicators lie within the 0.7 to 1.3 range. This is taken to indicate acceptable fit for these items in this study. For high stakes testing, more restricted fit criteria, perhaps in the range 0.8 to 1.2, would be used, while for rating scales and survey instruments more lenient fit criteria (e.g. 0.6-1.4) were recommended.

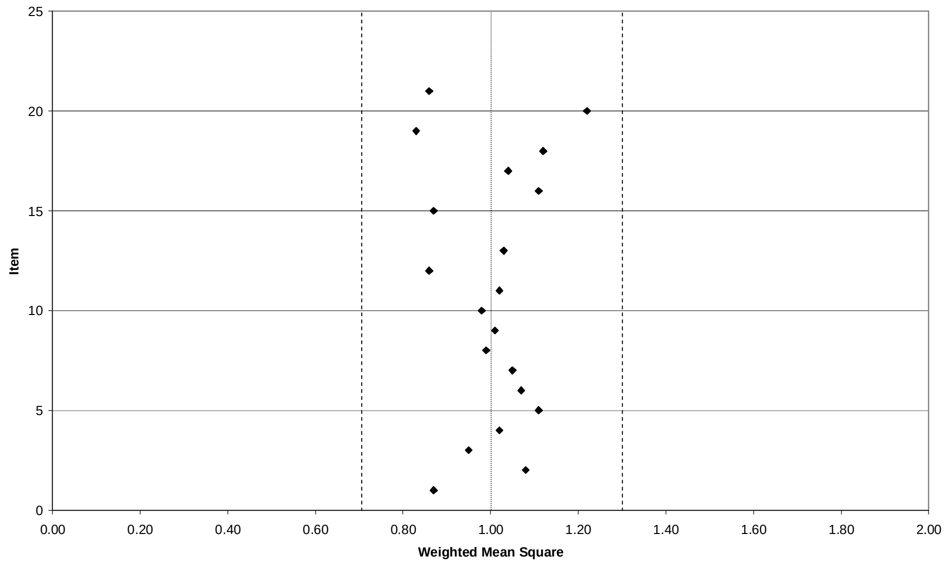


Figure 8: Fit parameters (Infit MS) of the Problem Solving Assessment

Indicator Difficulty and Precision

The positions of indicators, on the measurement scale formed by the set of indicators, are estimated as are performance level threshold locations. In items with multiple responses, the ability level at which a person is likely to move from one performance level to the next, for example from level 1 to level 2, is the threshold between those performance levels. Dichotomous indicators have a single threshold (equal to the item location or difficulty) while items with four performance levels (the maximum number used in the PSA) have three thresholds. The indicator locations and the performance level Thurstone thresholds and their standard errors are shown in Table 26.

The distribution and separation of item locations and thresholds contribute to effective measurement. The indicator locations are reasonably well spread along the scale, with difficulty values ranging from -1.83 to +1.68, a range of more than 3.5 logits. The precision of these difficulty estimates is quite variable, with REP02 having a standard error of 0.23 and EXE03 having a standard error of 0.94. The latter is quite an imprecise estimate and is 27% of the range of indicator difficulties. It is the lack of precision of indicators such as this one that leads to the lower than desired item separation reliability index for the scale.

Table 26: Estimates of PSA indicator locations and performance level thresholds

		Indicator location		Performance level (Thurstone) thresholds					
Item no.	Indicator label	Estimate	se	Th1	se	Th2	se	Th3	se
1	Rep01	-1.47	0.91	-1.47	0.91	0.68	0.57		
2	Rep02	1.68	0.23	1.68	0.23				
3	Rep03	-0.63	0.37	-0.63	0.37				
4	Rep04	-0.94	0.66	-0.94	0.66	1.96	0.44		
5	Rep05	-1.83	0.60	-1.83	0.60				
6	Pln01	-1.56	0.88	-1.56	0.88	0.98	0.44	2.27	0.38
7	Pln02	-0.40	0.34	-0.40	0.34				
8	Pln03	-0.19	0.32	-0.19	0.32				
9	Pln04	-1.83	0.60	-1.83	0.60				
10	Pln05	0.07	0.29	0.07	0.29				
11	Exe01	-0.92	0.41	-0.92	0.41				
12	Exe02	-0.94	0.72	-0.94	0.72	1.12	0.49		
13	Exe03	-1.50	0.94	-1.50	0.94	0.73	0.55		
14	Exe04								
15	Mon01	-1.09	0.44	-1.09	0.44				
16	Mon02	0.00	0.50	0.00	0.50	1.19	0.41	1.80	0.40
17	Mon03	0.50	0.44	0.50	0.44	2.02	0.40		
18	Mon04	0.57	0.26	0.57	0.26				
19	Ref01	-0.63	0.59	-0.63	0.59	2.14	0.43		
20	Ref02	0.91	0.41	0.91	0.41	2.71	0.41		
21	Ref03	-0.63	0.37	-0.63	0.37				

Notes: Indicator 14 (Exe04, Mayer performance level) was removed from the analyses as it is not an indicator of a fundamental problem solving process.
The Rep02 threshold can be considered a series 2 (uni-structural to multi-structural) threshold, as there were no pre-structural performance ratings.

Thurstone[22] thresholds are necessarily ordered, and it is useful also to examine the Delta or Tau thresholds. These latter thresholds need not be ordered, but any disordering warrants investigation. The Delta thresholds are shown in Figure 9. The vertical bars in the figure are the standard errors of the threshold location estimates. For most indicators with relatively large standard errors for the threshold estimate, for example indicators REP01, PLN01 and EXE03, the lowest response categories have very low frequencies so the estimate is based on few cases and, therefore, is estimated with low precision.

The indicator REP02 (Recognises relevant given information) requires comment. It has three performance levels, but the lowest level was not used at all. The indicator has become, by default, dichotomous and the threshold shown in Figure 9 as a ‘Delta 1’ threshold is in reality a ‘Delta 2’ threshold. The assignment as a Delta 1 threshold is a consequence of the software assigning the lowest estimable threshold to the first set of thresholds.

Of the polytomous indicators, indicator 16 (Mon02, Responds to unexpected problems along the way), which has four performance levels, shows a reversal of the top two Delta thresholds. It can be noted that the confidence intervals for the

[22] Indicator locations and performance level thresholds are described in Chapter 6.

two thresholds overlap, so the reversal is not statistically significant. The poor separation of these thresholds (circled in Figure 9) is not attributable to low frequency use of adjacent performance levels. It would appear that, for this indicator, there is inconsistent use of the higher performance levels. There may be a case for revising the indicator or the number and descriptions of its performance levels or even removing the indicator.

For comparison with the Delta thresholds, the Thurstone thresholds are plotted in Figure 10. Attention is again drawn to indicator 16, which is highlighted. The Thurstone thresholds in question, although necessarily ordered, are not well separated.

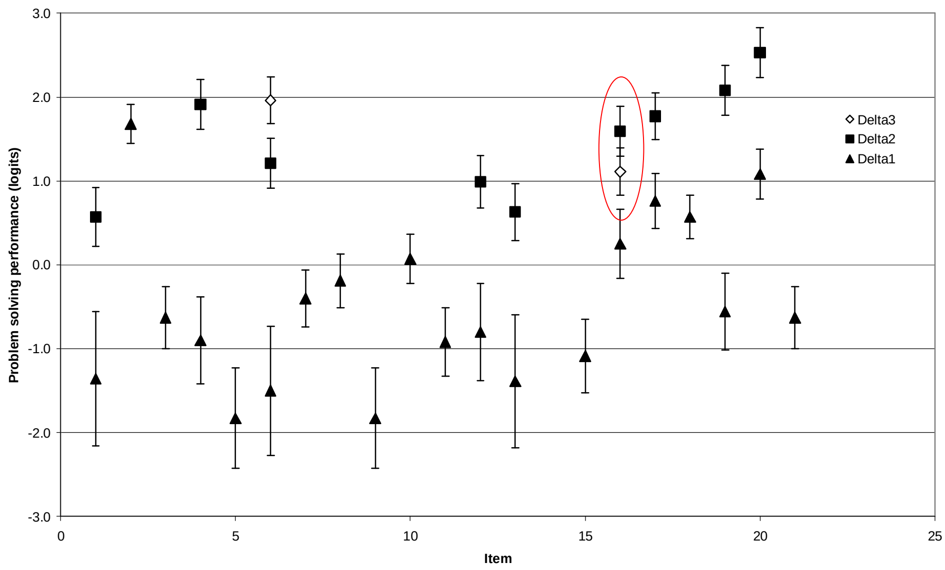


Figure 9: Delta thresholds for PSA indicator performance levels

Reflection on the estimated thresholds and SOLO performance levels is warranted. The performance level descriptors for indicators are based on the SOLO taxonomy. While some variation in the threshold locations of different indicators is anticipated, it is expected that the lowest set of thresholds, separating the pre-structural and uni-structural bands, should cluster in a range that is clearly below the second set of thresholds, which separate the uni-structural and multi-structural performance levels. The expected separation of the performance levels is not as marked as was anticipated. For items 1 to 13, the lowest set of thresholds lie in the range -2.0 to 0.0 logits; the second set of thresholds lie between 0.0 and 2.0 logits; the only two thresholds of the third series are approximately 2.0 logits. There appears to be a difference in the pattern of threshold locations between the representation, planning and execution indicators on one hand and the monitoring and reflection indicators on the other. The sets of thresholds for indicators 1 to 13 are somewhat higher than the corresponding thresholds for the indicators 15 to 21. These two groups of indicators could be classified as cognitive and metacognitive respectively. Two possible explanations for the observed lack of performance level separation arise. First, it is possible that the performance level descriptors do not adequately reflect the levels of the SOLO taxonomy or, second, that the performance levels have not been applied consistently in judging student performance. If the latter explanation is more accurate, it will be necessary to revise the performance level descriptors and to provide better information to teaching staff so that the levels are used more consistently.

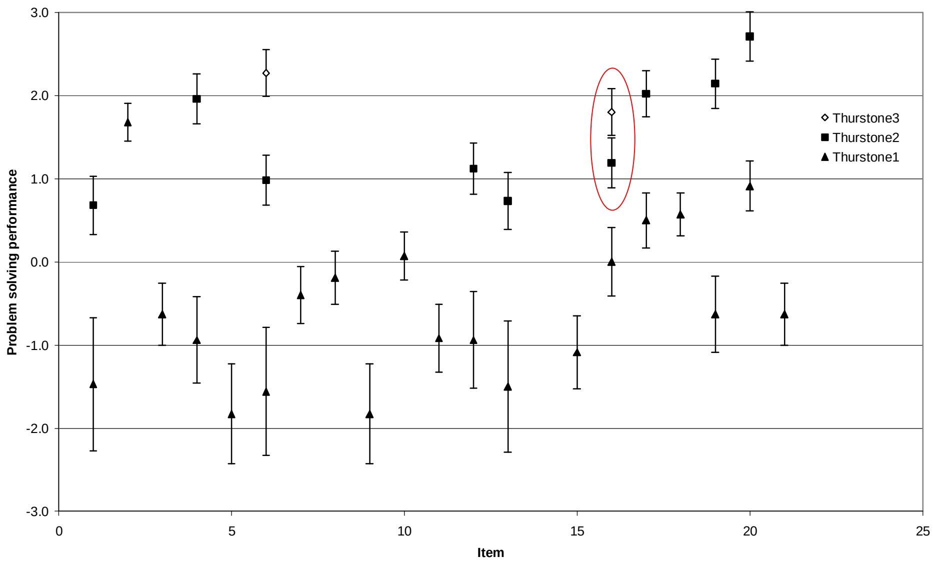


Figure 10: Thurstone thresholds for PSA indicator performance levels

Reporting student problem-solving ability

The problem solving abilities shown in the above tables and figures are expressed in the natural unit of Rasch measurement – the logit. The threshold range is approximately 4.5 logits and the person range is approximately 6.5 logits. Performance, measured using logits, includes decimals and negative numbers that are inconvenient, and a metric for problem solving ability is required in units that may be more attractive to practitioners, students and employers. Rasch scaling generates interval values that can be transformed linearly. A scale must be found that will communicate ability clearly and simply and not lead to misunderstanding. This suggests that measurements on the scale should be expressed as positive integers. However, if values are expressed in units for which common values are below 100, the reported ability may be interpreted as being on a percentage scale.

In many other scales, it has become common practice to transform the measures found from Rasch analyses into scales with means of 500 and standard deviations of 100. The Scholastic Aptitude Test (SAT), for example has a mean of 500 units and a standard deviation of 115 units. Similar parameters are used in reporting achievement test results in Programme for International Student Achievement (PISA) studies. This practice is suggested for the PSA and has been implemented in this study. The transformed scale is labelled the Problem-Solving 500 (PS500) scale. The distribution of problem solving scores on this scale is shown in Figure 11, with separate distributions shown for students enrolled in Electronics and Information Technology (E&IT) and Assessment and Workplace Training (AWT) courses.

The indicator locations and performance level thresholds (not shown diagrammatically) are also transformed to the PS500 metric. The average indicator location, 0 on the logit scale, lies at 302 units on the PS500 scale. The performance level threshold range, from -1.83 and +2.71 on the logit scale, lies between 137 and 546 on the PS500 scale.

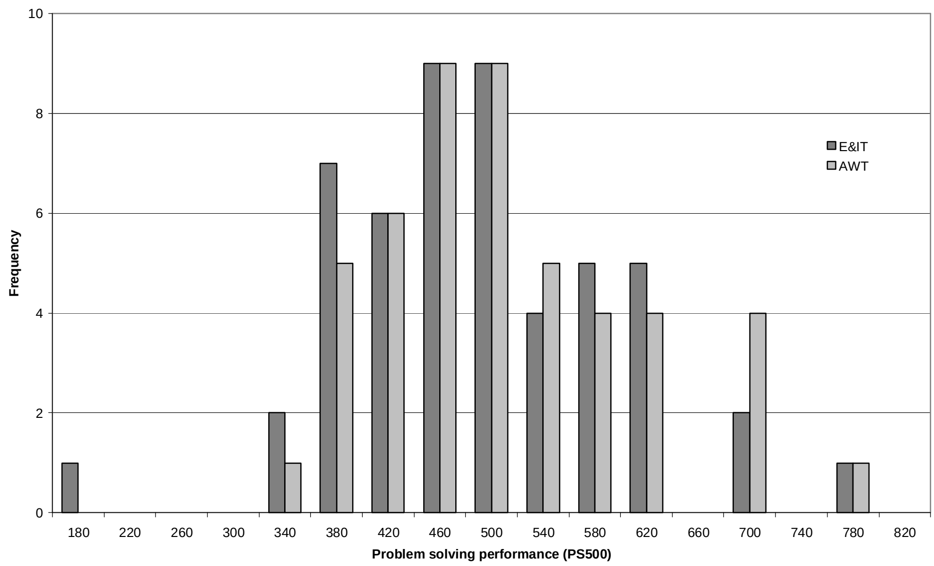


Figure 11: Distribution of problem solving performance measured on the PS500 scale

An advantage of using the Rasch measurement model is that it produces estimates of known precision – a precision indicated by the standard errors of the estimated problem-solving ability. Figure 12 depicts individuals' PS500 scores, sorted in ascending order, showing the standard errors of the estimate. For estimates within the range of indicator thresholds, standard errors are at a minimum and the precision is at its highest, but for abilities near the extremities of the distribution and especially at higher performance levels where few indicators are located, the measurement is less precise. This is a consequence of the mis-targeting of the PSA instrument for these samples of respondents, but, as noted above, the samples are substantially self-selected and appear to over-represent higher achieving students.

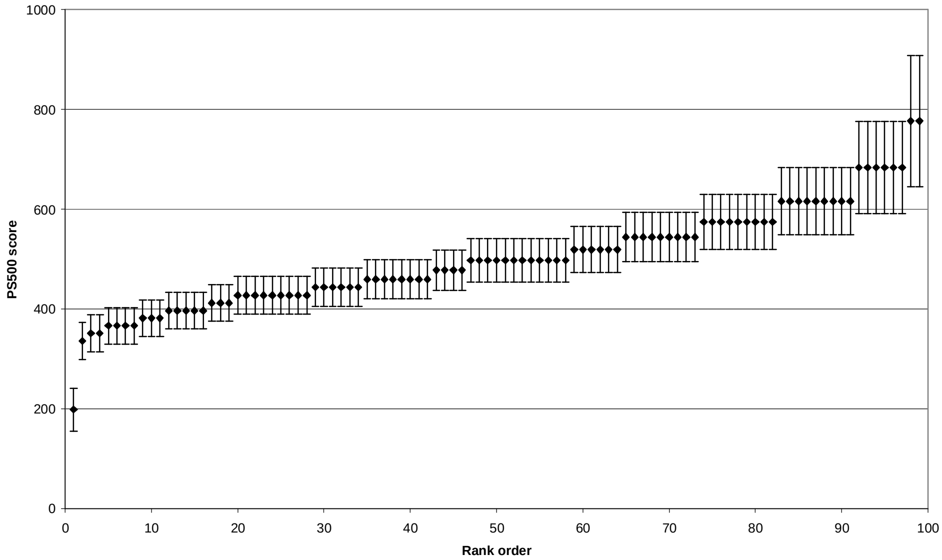


Figure 12: Individual's problem solving scores in rank order, showing standard errors of estimates

The PS500 scale may be a useful metric for reporting problem-solving ability. For this metric to be accepted by policy makers, it would be necessary for other generic skills to be assessed in a way that yielded performance measures that could be reported on a similar scale. Individual performances measured on metrics like the PS500 scale, for example those reported from PISA studies, are useful for secondary analysis.

Performance bands are of greater practical use for reporting achievement. This is done, for example, in reporting adult literacy scores arising from the Survey of Aspects of Literacy[23] (ABS, 1997). Five literacy performance bands were recognised, the lowest two (Bands 1 and 2) being regarded as below the level required for people to participate effectively in Australian society. In the analyses of the problem solving scale, the precision of the indicators and thresholds leads to the suggestion that three performance bands can be discriminated. Better targeting of the current instrument and modifications to some indicators and performance levels could lead to more precise measurement and therefore additional performance bands. However, there are in the current study no external criteria

---

[23] The Australian component of the International Adult Literacy Survey (IALS) is referred to in Australia as the Survey of Aspects of Literacy.

against which assessed problem solving performance can be evaluated to identify transition points between bands.

The current study is conducted in the Australian VET sector. For that sector, employer organisations, which have been active in promoting generic skills assessment and reporting, may have views on the levels of problem solving performance required for various occupations. Different criteria are likely to apply to the schools sector, since, at the end of formal schooling, students pursue a variety of pathways, including direct entry to the workforce and further study in the VET and higher education sectors. Graduates of higher education programs might be expected to develop higher levels of generic skills than school or VET graduates as university courses prepare individuals for professional and managerial roles. It seems, therefore, that effective problem solving assessment will need to be flexible enough to generate performance bands discriminating achievement covering a wide range of performances. The number of performance bands required and, especially, the thresholds between them need to be decided.

Validation study

The current study began with the development of the Problem Solving Assessment (PSA) instrument in the School of Electronics and Information Technology (E&IT). It was later extended to include students enrolled in the Assessment and Workplace Training (AWT) course. In the factor analysis and classical item analysis reported above, data from the E&IT sample were used, but in the Rasch calibration of the instrument, data from both samples were used. The calibrated estimates for indicator difficulty and performance level threshold locations were used to estimate individuals' problem solving performance scores, and they are shown combined in Figure 7 and separately in Figure 11. The purpose in extending the study to include the AWT program was to see whether the instrument could be used to assess problem solving in a different course context.

In order to test whether the instrument and its indicators work effectively in both course contexts, the instrument needs to be calibrated separately for both samples. Rather than undertaking the comparison using entirely separate analyses, the Quest program offers the option of contrasting the calibrations of items between groups. If measurements of problem solving on both samples using a common instrument are commensurate, relative indicator difficulties should be similar between the samples. The *compare* command produces a plot of standardised indicator location differences between the comparison groups, and this is shown in Figure 13. The two vertical lines in the figure mark differences whose *t* values that lie between -2 and +2, that is differences that are not statistically significant. Three indicators lie outside this range. Indicator 2 (Recognises the significance of given information) is easier for E&IT than for AWT students. Two indicators, 17 (Reviews original plan) and 20 (Compares current problem with previously encountered ones) are easier for AWT students. These differences may be explained by the content of courses and the characteristics of students taking them. AWT students are required to plan and present training sessions. Those processes require students to adapt their training plans to the needs of particular audiences and situations, so monitoring and refection processes are likely to have been raised specifically in their instruction. AWT students are preparing to take on training and assessment roles within the VET sector and the course they are doing is a requirement for those roles. These students are generally older and more experienced than most students undertaking the electronics program. Their experience may provide them with a basis for their use of monitoring and reflection processes.

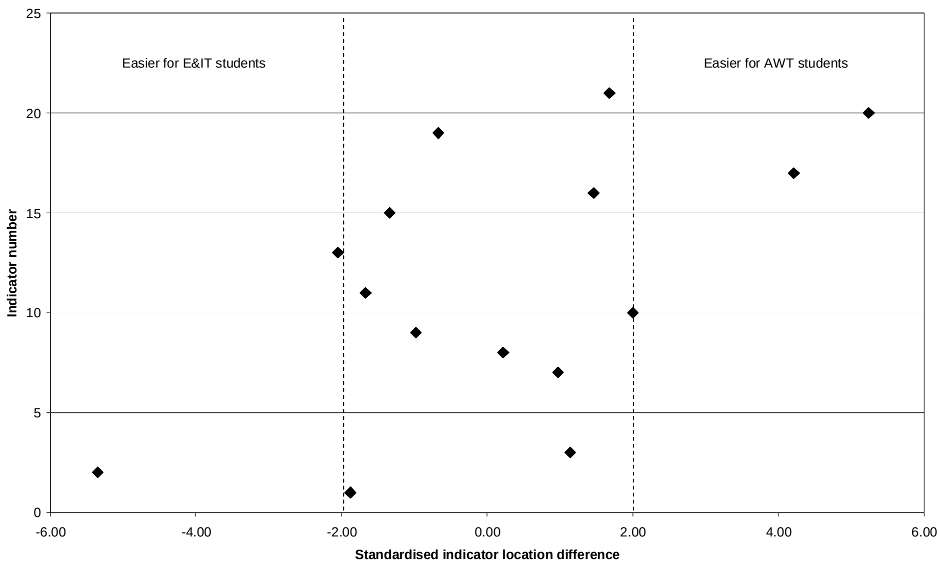


Figure 13: Standardised indicator location differences between E&IT and AWT students

Some caveats are warranted in interpreting this analysis. First, and as noted above, the standard errors of some indicator estimates are rather high. These large standard errors and the smaller sample sizes of the separate groups reduce the power of the *t*-tests used to compare indicator difficulty. It is worth noting that the three indicators for which significant differences are found have relatively small standard errors (see Table 26). Second, only 15 of the 20 indicators are evaluated in the comparison of indicator difficulty. The five indicators that were not estimated in the comparison had very little variance in either the E&IT or the AWT sample and could not be reliably estimated.

A major advantage of the Rasch measurement model is that it produces sample-independent estimates of item difficulties provided responses to items conform to the requirements of the measurement model. Comparing the two groups of learners was a test of this assumption. While most indicators revealed consistency in the estimates of item parameters between the two groups, the fact that some indicators did not fit raises concerns about the responses to those items. It is possible that the performance level descriptors were interpreted differently between the two groups.

Student evaluation of the Problem Solving Assessment

A student evaluation of the problem solving assessment and its processes was undertaken in order to identify any difficulties that students might have experienced and that might be informative in revising the instrument and its processes.[24]

The views of students who participated in the problem solving assessment and of those who declined to participate were sought. An online survey form was developed. It included a form that has been reproduced as Table 27. In the table, responses to the online form are summarised. In addition to the form, students were invited to provide comments. The comments are reported in Appendix 6. Both the survey form and the comments were volunteered anonymously.

Feedback from the Online Survey Form

Of the 29 responses to the form, 22 were from students who had completed a problem solving assessment and seven were from students who had not participated.

The students who had not participated in the study reported that they were adequately informed about generic skills and their assessment. They were aware of the information that had been made available and knew who to ask if they required more information (questions 1 and 3). Most indicated that they had not given much thought to undertaking these optional assessments (question 2), but indicated they were likely to consider taking part in future (question 4).

Students who did participate in the study reported that they were adequately informed about Key Competencies and their assessment (questions 1 and 2). In

[24] This survey was presented to students through the Electronics and Information Technology student web site. This was arranged by Rob Denton to whom the author is grateful for this assistance.

this respect, their views were similar to those students who chose not to participate. This suggests that lack of information does not explain the non-participation of those students who elected not to take part.

Seven reasons for deciding to have Key Competencies assessed were suggested in the survey form (question 4) and all were strongly endorsed by participants. These questions focused on the development and recognition of these generic skills – perhaps intrinsic reasons for participation, although one reason offered focused on employment prospects and this could be classified as an extrinsic motivation.

Students' positive views of the assessment are evident in that most would recommend them to other students (question 5).

Participants are somewhat ambivalent about the clarity of the assessment process (question 6). While students' views of this aspect of the assessment are mostly positive, they are much less positive about this issue. This is, therefore, an area in which the instrument and the processes in which it is embedded could be improved.

Question 8 sought students' opinions on the extent to which the assessment process had helped them improve their skills. This question is similar to one of the reasons offered for undertaking this assessment, but responses are slightly less favourable. This suggests that students perceived some value in the assessment process, but they were slightly less happy with the improvement that occurred. The process of undertaking an assessment, gathering evidence of the generic skill, presenting that evidence and receiving informative feedback appears to have led to favourable, but not completely satisfying, perceptions by students of their emerging abilities.

The final question (9) and its components asked students to compare a previous assessment form that used only the three levels recommended by the Mayer Committee with the new PSA form. Students were generally happier with the new form as it gave them more information than the previous one, although it is unclear how many students had used both forms as compliance under the previous assessment arrangements had been quite low.

Feedback from Comments

Twenty-two comments were received from students through the online feedback form. Numbered comments are presented in Appendix 6 and the bracketed numbers in the text below refer to the numbering in that appendix. Students' comments reflect some of the issues raised in the survey. For example, several comments refer to an ability to present evidence of generic skills in job interviews [1] and that it is a way of having these skills recognised [5, 20 and 22].

Some comments [7, 12 and 16] refer to difficulties in the processes in which the assessment tool is embedded. It is apparent from these comments that there was some misunderstanding about the processes. Some students believed they needed to inform their teachers of their intention to seek a key competencies assessment before commencing that assignment. This was not required, but the comments could reflect earlier procedures.

Several comments relate to difficulties in understanding performance levels. Some refer to previous assessment approaches in which the Mayer performance levels were used [3]. Difficulties in understanding these levels are evident in the literature (Jasinski, 1996; Ryan, 1997). Other comments refer to difficulties in comprehending the text in the PSA that described performance level criteria [15 and 17]. Students in many VET courses have very varied prior educational attainment and achievement backgrounds. Some students are very able and are undertaking VET rather than similar higher education courses by deliberate choice. Some others enrol in VET courses because they do not meet the entrance requirements of higher education courses and have limited academic skills. Still others enrol after periods of unemployment and some of these students have poorly developed basic skills. To meet the diverse needs of this varied student cohort, it appears that the guidelines provided with the PSA will need to be simplified and targeted at a lower reading level.

Several comments [18 and 19] refer to the scaffolding provided by the PSA in helping students recognise the component processes of problem solving. However, in comment [20], there is evidence that more needs to be done to show students how these process skills can be enhanced. This comment refers specifically to the need for assistance in improving current performance before moving on to the next assignment.

One comment [4] is very strongly negative. It seems that this student received feedback that suggested his or her performance was not at the desired level and did not find the feedback helpful in improving that performance.

Despite some comments about difficulty in comprehending the text of the PSA, the feedback from respondents was positive, indicating they believed the PSA helped them to develop and demonstrate their problem-solving skills. Most would undertake a similar process for the assessment of other generic skills.

Table 27: Questions and summary of responses to an online student evaluation survey

	Yes			No
Have you completed any Key Competencies Assessments?	22			7
	Yes	**Largely**	**Partly**	**No**
IF NO …				
1. Were you given enough information to help you make an 'informed decision' about doing it or not?	5	1	1	0
2. Have you given it much thought?		2	5	0
3. Do you know how to get more information or who to ask about Key Competencies Assessment?	6	0	1	0
4. Do you think you might consider Key Competencies Assessment in the future?	5	1	1	0
IF YES …				
1. Do you feel well informed about the Key Competencies Assessment process?	10	8	2	0
2. Is there enough information available to inform people about Key Competencies Assessment?	12	6	3	0
3. Do you think you will apply for more Key Competencies Assessments in the future?	13	5	3	0
4. Why did you decide to have your Key Competencies assessed?				
To get these skills formally recognised	18	0	0	1
To help give you an edge when applying for a job	16	0	0	3
To help you prove and explain your skills at a job interview	14	0	0	3
To get recognition for the extra skills you use in this Flexible Learning program	14	0	0	3
To better understand your skills in this area	13	0	0	4
To improve your skills in this area	15	0	0	3
It was suggested that I do it	1	0	0	
Other	0	0	0	0
5. Would you recommend it to other students?	15	5	2	
6. Is the assessment process clear and easy to follow?	7	9	5	1
7. Have you received adequate assistance from facilitators?	17	2	2	
8. Have these assessments helped you understand & improve your skills?	10	6	5	1
9. If you have used the NEW Problem Solving Assessment …				
9a. Do you think it is:				
informative?	10	5	0	0
easy to use?	5	9	1	1
effective in helping you understand your skills?	7	6	3	0
better than the other assessment forms?	4	3	1	0
9b. Do you think Key Competencies Assessments will help you prove and explain your skills at a job interview?	14	5	3	0

Note: Question numbers did not appear in the online form and have been added for reference in discussion of the results.

Discussion of results

Participants

The key question in relation to those people who volunteered to participate in this study relates to the generality of any conclusions that can be drawn from the current project.

First, the study was conducted in the Electronics and Information Technology (E&IT) program at Torrens Valley Institute of TAFE with a replication study in the Assessment and Workplace Training program of the same TAFE institute. The E&IT program has a history of innovation in course delivery and implements a flexible, student-centred delivery strategy. Key Competencies have been an element of course delivery since its inception and they have been assessed specifically since 1998, although relatively few students had volunteered for the previous form of Key Competencies assessment. This raises the question: "Would similar results be achieved in similar courses offered by other providers who have pursued more conventional course delivery strategies?"

The E&IT program is a technical one, and many of its graduates can be expected to work in fault finding and repair jobs. The course modules may therefore place greater emphasis than other courses on problem solving. This leads to the question: "Would similar results be achieved in other courses, either in technical ones in different areas, for example automotive, or in non-technical courses, for example in human services courses?"

The students who took part in the study reflect a range of ages and backgrounds. However, it is possible that a greater proportion of more able students submitted work for problem-solving assessment. The age and experience distributions of participants are bimodal. Some participants were relatively young recent school leavers with limited work experience while others were somewhat older people having considerable experience in the workforce. There is no significant relationship between age and problem-solving performance (r=0.07, p=0.70). However, the possible bias towards high ability participants raises the question: "Would similar results be achieved for students of all ability levels?"

The issue of possible ability bias in the sample of participants has implications for the generality of the results of the study. If there is no ability bias, the PSA instrument would appear to be targeted at too low a level of problem-solving ability. Alternatively, having access to the PSA instrument with its prompts for problem solving approaches may have encouraged students to use processes that they may not otherwise attend to. Conversely, if there is an ability bias, the instrument may be adequately targeted. The possible existence of an ability bias must therefore be established before a conclusion can be reached on the targeting of the instrument, and therefore before any decisions are made about revisions to the instrument.

The Problem-Solving Inventory

The Problem-Solving Inventory (PSI) is used in order to assist in developing an understanding about the role of attitudes to problem solving in problem solving performance. The instrument was reported to have reasonable psychometric properties, although the methods used to establish this, principal components analysis and classical item analysis, are not able to provide the penetrating analyses available using the Rasch method.

As administered in the current study, the PSI does not show the same degree of separation of its three subscales – Approach/Avoidance Style, Problem-Solving Confidence, and Personal Control – as originally reported (Heppner & Petersen, 1982). However, the current study used a modified form of the instrument and was trialled on only 33 participants. For these reasons, some caution is warranted in an interpretation of its structure. The Personal Control factor is apparent, but the other two seem to form an undifferentiated factor. A substantial body of theory exists around motivation, goal orientation, self-efficacy and attributions for performance within domains (Bandura, 1989; Dweck, 1986; Dweck & Leggett, 1988). Some of these concepts had been described before the development of the PSI, but considerable work has been done since that time. There appears to be a case for reconsidering the PSI given the more recent work on the issues that the PSI sought to address. Scores on the PSI subscales and PSA were used in a path model. This model yielded unexpected relationships. Negative path coefficients were found between the PSI Approach/Avoidance and Problem Solving Confidence scales and both educational attainment and problem solving ability (Curtis & Denton, 2003, pp. 61-62). These findings are at odds with other findings reported for the PSI (Haught et al., 2000). The PSI is not used further in the current study as it appears to warrant further investigation.

The Problem Solving Assessment Tool

The Problem Solving Assessment (PSA) tool is the central focus of the current study, and was developed to test the hypothesis that an instrument designed to assess individuals' application of a set of generally accepted key problem solving processes would provide both a valid and a reliable basis for measuring problem solving ability.

The tool is developed following wide ranging reviews of theories of problem solving and is thought to reflect high levels of both construct and content validity.

Test sample

Using data from the E&IT group, conventional analyses, using exploratory principal components analysis and classical item analysis, suggest that the scale formed by the items of the PSA is a coherent measure of problem-solving ability. However, there are some caveats. Some dichotomous items showed little variance and this is thought to reflect a generally high ability sample of participants. However, it may be necessary to revisit these items and to provide additional performance levels for them.

Analyses based on the Rasch measurement model also provide good support for the instrument. Reliability indices for both items and persons of approximately 0.80 to 0.85 suggest that the instrument is robust. Analyses of response patterns to individual items suggest that most items are quite sound. Some departures from expected category response patterns were found. These may reflect the modest sample size of this study, but it may be necessary to revise the number of performance levels for some items and the descriptions of those performance levels. Such revisions, however, should await trials of the instrument with larger and more diverse samples of respondents.

Validation sample

The overall PSA scale indicators found for the validation study sample (AWT) are quite similar to those computed from the original E&IT test sample. This is encouraging and does suggest scale coherence. However, variations in the precision of item locations within scales do suggest inconsistency in the interpretation of at least one indicator. Variation in indicator locations between the test and validation samples suggest that there are systematic differences in interpretations of indicator performance or differences in the context of the assessment that are not invariant between groups. This is an undesirable feature of the instrument and suggests the need to revise the performance criteria specified for each indicator to improve the consistency of their application.

The several analytical approaches taken in examining the properties of the instrument have provided a wealth of information than can assist in the revision of indicators.

Student evaluation

An evaluation by students of the Key Competencies assessment process employed in the current study shows that students were quite clear about the benefits to them of participating in Key Competencies assessments. They understood that through these assessments their skill levels will receive formal recognition and that this will assist them in seeking employment. In addition, they understood that through the processes used in which they self-assessed they would enhance both their Key Competencies and their abilities to describe them.

There was some support for the more detailed approach taken in the PSA instrument compared with previous tools. However, there is a need to review the terminology used in the instrument to make it more accessible to students.

The concerns that students expressed about performance levels suggests the need for further work to be done on this. Performance levels have been a difficulty from their initial articulation in the Mayer report. In the consultations undertaken by the Mayer Committee, this issue was identified as a problem. Since that time several authors have commented upon confusion surrounding these levels (Jasinski, 1996; Ryan, 1997). In the current study, the issue that taxed students was the clarity with which the different levels were distinguished. The support expressed by students for the more detailed indicators of the PSA compared with previous assessment tools suggests that this might be an effective approach to the resolution of this matter.

Summary

The primary purpose of the study reported in this chapter was the development of a tool to assess problem solving in the contexts of routine assessments undertaken in students' courses.

The instrument, the PSA, was developed using a particular conception of problem solving – one based on a set of cognitive and metacognitive processes. The application of these processes is assessed using a set of indicators of these problem solving processes. Student performance on these indicators is assessed against standards based on the SOLO taxonomy.

The PSA was used in an assessment procedure in which students were provided with the PSA, including its processes, indicators and performance standards. They were invited to assess their own problem solving performance on routine assessment tasks and to submit their responses to those tasks along with their self-assessments of problem solving. They received feedback on their responses to set tasks and on their problem-solving performance.

While there is scope to improve the PSA, it does appear to have reasonable measurement properties and it is capable of being used to assess students' problem-solving performance.

What remains to be shown, given the proposition that self-assessment followed by feedback can lead to enhanced learning (see Chapter 4), is whether the use of the PSA can lead to improvement in students' problem solving skills over time. The PSA can yield a measure of problem-solving performance, and that measure can be used in analyses of change over time. That is the issue addressed in the second study reported in the next chapter.

Chapter 8: The Growth in Students' Problem Solving Performance over Time

The study reported in this chapter is built upon the Problem Solving Assessment (PSA), the development and initial testing of which is described in Chapter 7. A modified version of the PSA, the Problem Solving Skills Assessment Tool (PSSAT, see Appendix 8) is used to assess problem solving performance over time on a range of current assessment tasks. The instrument was used by students in a self assessment mode and by their teachers as a means of providing feedback to students on their emerging problem solving ability. Repeated measures of problem solving performance over an academic year enabled the development of problem solving ability to be evaluated.

In Chapter 3, alternative conceptualisations of problem solving are elaborated. These include information processing and situated cognition models. It is argued that information processing models are likely to be more productive conceptualisations of the emergence of problem solving ability in non-expert learners. It is also argued that novice problem solving is of a different kind than expert problem solving; it is not simply a less efficient version of expert performance. During the early stages of the emergence of problem solving ability, it is argued that novices require scaffolding of problem solving processes and that they are likely to apply these processes consciously rather than automatically. The PSSAT is designed to scaffold and evaluate the application of problem solving processes during solution attempts of problems that occur routinely within the learning domain.

One of the issues raised in Chapter 2 is that failures in the implementation of generic skills schemes have several origins, including a failure to assess performance against theoretically sound conceptions of these skills and a failure to provide feedback to students following that assessment. In Chapter 4, which deals with assessment, evidence is presented that shows (a) the importance of feedback in formative assessment and (b) the potential value of self assessment, with feedback, as a means of engaging students in the constructs being assessed.

The approach taken in the current study to the assessment of problem solving differs from other approaches that are used. In the Programme for International Student Assessment (PISA) tests and those used in the Graduate Skills Assessment (GSA), problems are posed that are common to all candidates undertaking the assessments, irrespective of their field of study, and therefore that are not related to the disciplines or vocations the candidates are studying (ACER, 2000; OECD, 2003). The approaches taken in those tests are legitimate for their purposes. In the case of PISA for example, the purpose is the comparative assessment of general problem solving ability across populations.

The purpose of the study reported in this chapter is to investigate an approach to the enhancement of problem solving ability through an iterative formative process of instruction, assessment and feedback. The PISA and GSA assessments are summative and designed to measure the stock of skills in the target populations. The PSSAT is one component of this iterative process. Other components are described below.

Organisation of the chapter

In this chapter, the procedures followed in selecting and recruiting institutions, teachers and students are outlined. The information provided to participants is described and the documentation provided to them is presented in several appendices to this chapter (see Appendix 9 for the teacher information package and Appendix 10 for the student information package).

Students' prior generic skills achievement was assessed using a shortened form of the Tertiary Skills Assessment (TSA). The Tertiary Skills Assessment is based on the GSA, but uses only the interpersonal understandings, critical thinking and problem solving scales of the GSA (ACER 2000). These are assessed using multiple-choice tasks. The two writing tasks of the GSA are omitted. Analyses of the TSA results are presented.

It is noted in the previous chapter that the PSA, while providing an adequate measure of problem solving performance, could be improved. Changes were made to the PSA to produce the PSSAT that is used in the study reported in this chapter. Those changes are described. The procedures followed in students' self-assessment and in teacher assessments of problem solving are outlined and the methods used to analyse and scale the PSSAT results are described.

The main purpose of the study is to evaluate the growth in students' problem solving performance over time. The case is made that multilevel modelling methods are required to model that change. Some exploratory analyses are presented, but the focus of this chapter is on the multilevel analysis of students' self assessed problem solving performance.

The chapter concludes with a summary of the main findings of the study and a brief discussion of the implications of those results.

Selection of research sites and students

The Problem Solving Assessment (PSA) was shown to yield a coherent measure of problem solving performance in an electronics course and an 'Assessment and Training' course in a Technical and Further Education (TAFE) college (see Chapter 7). The study reported in this chapter builds upon that work by investigating whether a revised version of the instrument would work in other courses and whether it could be used to promote and monitor growth in problem solving performance over time.

In order to get greater diversity in the problem solving assessment context, approaches were made to senior staff in two university courses and four TAFE institutes in South Australia seeking their participation in the proposed research. A research plan was provided to potential participant institutions. A presentation was made to staff of those institutions that expressed interest in the research. The presentation included information on:

- work that had been done on the delivery, assessment and reporting of generic and employability skills development in post-secondary education;
- the findings of the project to measure problem solving skills; and
- the proposed research plan and what it would involve for participants.

Two of the TAFE colleges declined to be involved. Staff responsible for an information technology course in one TAFE college and staff from a human services course in the fourth TAFE college agreed to participate. Staff responsible for a first year engineering course and a third year teacher education course at a South Australian university agreed to participate. At each site, a presentation was made to students in the courses for which staff had agreed to participate in the research. The student presentation focused on the importance of developing and demonstrating generic employability skills and on what would be required of student participants. This information was provided in some detail in the information kits provided to students (see below). In particular, voluntary participation and the right to withdraw at any time were emphasised.

Staff who agreed to participate in the research project were offered a modest payment based on an allowance of 15 minutes per assignment to assess and provide feedback to students on their problem solving performance. The estimate of 15 minutes per assessment was based on feedback from staff involved in the first study (see Chapter 7). The hourly payment rate was based on a senior tutor casual payment rate.

In order to encourage students to participate, gift vouchers were sourced from two clothing store chains, two cinema groups, three fast food chains, a music retailer and two book stores. Most gift vouchers were donated and some were purchased by the researcher. The value of the donated and purchased vouchers was $6,420. All participating students were given fast food vouchers with the first problem solving assessment. Cinema passes were allocated randomly with the second assessment, while vouchers of greater value for DVDs, books and clothing were allocated, also in a random draw, after the final assessment.

Participation and attrition

Staff from the TAFE information technology program withdrew from the study. In the TAFE institute, there was a change in senior management during the project. The original institute director had been a strong supporter of the need to develop generic skills and was enthusiastic about research in the VET sector. His departure led to other management changes and to a decline in interest by staff. The VET sector has been characterised by more than a decade of very substantial change, and a degree of change fatigue may have contributed to the loss of interest by staff. An enrolment of 60 students had been expected in the course, but only 30 students enrolled. Of them, 24 completed the Tertiary Skills Assessment (see below) and of that group, 12 students submitted a problem solving self assessment for their first assignment. After this, staff withdrew from the study and no further problem solving assessments were undertaken.

In the teacher education course, there was a change of staff, with tutorials being managed by casual staff, some of whom were recruited after the project had commenced. They were fully occupied with the routine requirements of the course and their marking and few encouraged students to participate in the problem solving assessment. Although some students submitted self assessments, too few students participated to warrant the inclusion of their data in the study. Of the 60 teacher education students who undertook an initial test activity, only ten completed three problem solving assessments. This low response and high attrition rate would make results from this group unrepresentative, so the data that were gathered from this group are not considered in this chapter.

Eleven students were enrolled in the human services course at a regional TAFE institute. Eight students completed the TSA and three problem solving assessments. Although the attrition was low, there were too few cases for the forms of analysis reported in this chapter.

Participation in the engineering course was strong. There were 64 enrolments in the first semester course. Of them, 54 participated in the first activity, the Tertiary Skills Assessment (TSA) and of that group, 42 students completed three problem solving self- and lecturer assessments over two semesters. The analysis of problem solving development is based on the data gathered in the engineering course. In this course, students undertook four problem solving assessments; two in each semester. An administrative error on one assessment occasion resulted in scripts being returned to students before data could be entered, so data from only three assessment occasions were available for analysis.

Information for participants

The PSSAT was developed as one component of a process designed to enhance the problem solving skills of students. Information about the PSSAT and the assessment process was provided to students and teachers.

Information for teachers

Information about the development approach was made available to teachers and to students. A booklet (see Appendix 9) was prepared for teachers. Teachers were also provided with copies of all material prepared for students. The booklet explained the origins of the approach being taken to problem solving and outlined what was expected of them if they chose to participate in the research project.

In particular, the booklet describes the assessment model of self-assessment and the presentation of evidence, validation by the lecturer, followed by feedback to the student. The purpose of this assessment model is to encourage the development of problem solving and to assess and report on problem solving performance. The assessment procedure is outlined. Existing assessment tasks are selected; students undertake the substantive content assessment through those tasks; they identify evidence for problem solving processes through the set tasks; and they submit their self assessment including evidence when they submit their substantive course assignments.

The booklet describes and depicts the model of problem solving used in the current study. The relationship between the model of problem solving and the structure of the PSSAT is shown. The problem solving indicators and their performance levels are described.

Information for students

An information kit was provided to students as a folder of resource materials (see Appendix 10). The folder contained the following information leaflets:

Leaflet	Contents
What's in this package?	This is a description of the package contents, including a brief summary of the leaflets in it and advice on using the package
What's this project about?	This describes the project, emphasising likely employment advantages of developing and demonstrating generic skills.
What's in it for me?	This leaflet emphasises the development of generic skills and their employment implications. It also includes information on the TSA and on the student incentives.
Participation consent form	This is the consent form as approved by the Research Ethics Committee.
Personal details	This form seeks basic demographic information from participants.
Other employability skills	This leaflet provides information on other employability skills. In particular, it emphasises communication and teamwork as they and problem solving are 'the big three.'
How do I get an assessment?	This leaflet describes the problem solving process. In particular, it shows how the PSSAT is used, it describes the indicators and performance criteria, it describes self assessment and lecturer validation, and it presents the problem solving model diagrammatically.
The Problem Solving Skills Assessment Tool	This is the assessment instrument that students use in assessing their problem solving performance and having it validated by their lecturer.

Students were encouraged to retain the folder and its contents and to add a record of each problem solving assessment to the folder as evidence of their developing problem solving ability for use in employment interviews.

Revisions to the Problem Solving Assessment tool

A revised version of the Problem Solving Assessment (PSA) instrument is used in the current study. Version 2 of the PSA was used in the first phase of the research and is shown in Appendix 5. The revised version is referred to as the Problem Solving Skills Assessment Tool (PSSAT) and appears in Appendix 8. It was also provided to teachers and students and appears in their information packages.

Several changes were made to the PSA following the analyses presented in Chapter 7 and based on feedback from staff. One of the changes was to the format of the instrument. In the PSSAT, space was made available so that students could refer to the evidence they believed supported the performance level they had selected. The way this is used is shown in Figure 14, taken from 'How do I get an assessment?' leaflet in the student information package.

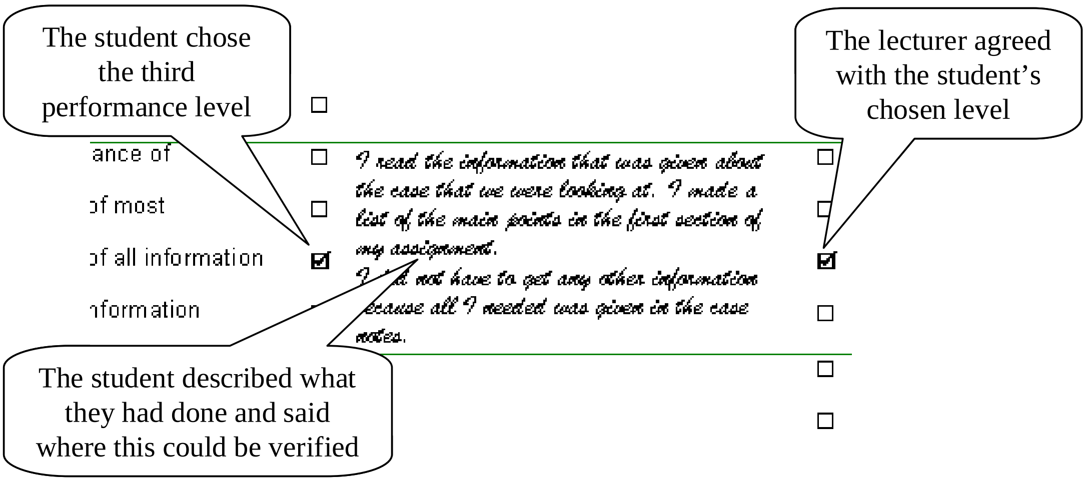


Figure 14: Using the PSA to record and judge evidence of the student's selected performance level

The number of indicators was reduced from 21 to 18. This was accomplished by removing an indicator from the planning process. The indicator in the PSA was 'Checks equipment availability'. This indicator had a very low threshold so it did little to differentiate most students and the indicator was believed to have limited applicability beyond courses where equipment use was required.

One indicator, 'Uses relevant skills,' was removed from the execution process. This indicator was believed to have limited applicability beyond very practical tasks and the 'Activates relevant knowledge' indicator was believed to reflect a similar and perhaps more important problem solving activity.

Two representation process indicators were combined. In the PSA, there were separate indicators for recognising the significance of given information and identifying the need for additional information. These were combined into a composite indicator, as recognising the need for additional information was regarded as a higher level activity. The former indicator is thought to be uni-structural in the SOLO hierarchy and the latter at least multi-structural.

The 'Sets a realistic goal' indicator was simplified to be dichotomous. The former three levels conflated given goals with those set by the student. Similarly, the 'Sets an appropriate time frame' indicator was simplified by removing the third performance level. This level had been little used in the first phase of the research.

The number of performance levels of the fourth monitoring indicator, 'Checks original understanding and definition of problem,' was increased from two to three. The metacognitive monitoring and reflection indicators had higher difficulty thresholds than other, cognitive, indicators and an additional threshold was believed likely to improve the discrimination of the instrument among higher performing students.

Finally, the indicator 'Application of strategies' was retained despite the fact that it was not used in the analyses presented in Chapter 7. At the time the instrument was revised, the anticipated decision to replace the Mayer key competencies with the key skills of the employability skills framework had not been taken. Because the project was being undertaken in TAFE institutes as well as in two university courses, the 'Application of strategies' indicator would provide a basis for reporting that was consistent with the requirements of training packages that dictated elements of TAFE course reporting.

The indicators used in the revised instrument are shown in Table 28. Indicators are referred to by their numbers in subsequent graphs and by their labels in tables.

Table 28: Indicator numbers, labels and text in the PSSAT

Indicator No.	Indicator Label	Indicator text
		Defining the Problem
1	Rep1	Forms a correct understanding of the problem
2	Rep2	Recognises significance of given information or the need for new information
3	Rep3	Recalls relevant information
4	Rep4	Sets a realistic goal
		Planning an Approach
5	Pln1	Plans an approach to the problem
6	Pln2	Recalls previous relevant or similar problem tasks
7	Pln3	Identifies appropriate sub-goals
8	Pln4	Sets an appropriate time frame
		Carrying out the Plan
9	Exe1	Begins to follow the set plan
10	Exe2	Activates relevant knowledge
11	Exe3	Application of strategies
		Monitoring Progress
12	Mon1	Checks progress towards goal
13	Mon2	Responds to unexpected problems along the way
14	Mon3	Reviews original plan
15	Mon4	Checks original understanding and definition of problem
		Reflecting on the Result
16	Ref1	Reviews efficiency and effectiveness of problem approach
17	Ref2	Compares current problem with previously encountered ones
18	Ref3	Anticipates situations in which current approach might be useful

Assessment and data collection procedures

At each research site, the following procedures were used.

After completing the research consent and student details forms, students undertook the Tertiary Skills Assessment (TSA) test. Students enrolled in the various courses had diverse educational backgrounds and bases for course admission, so there was no common measure of prior educational achievement. The TSA was used to provide an aggregate measure of achievement. It also had three identified subscales that might be related to problem solving performance. In particular, it includes a problem solving scale and a positive relationship might be expected between scores on this scale and the assessment of problem solving using the PSSAT. The TSA was administered during a scheduled class period and took one hour to complete. Data were entered, checked and analysed and students were provided with a report of their performance on the test. As this is a secure test, all test forms were returned to ACER.

A meeting was held with all staff who were scheduled to be involved in the assessment of students' problem solving. At this meeting, the purpose and use of the Problem Solving Skills Assessment Tool (PSSAT) was explained. In particular, attention was drawn to the performance level descriptions for each indicator on the form.

An information session was conducted for students, during a scheduled class session, on the use of the PSSAT as a self assessment tool. The purpose of this demonstration was to model the use of the PSSAT as a self assessment activity. A key point made was that the purpose of the form was to provide an indication of students' current problem solving performance and that no grade would be attached to each form. This was to discourage students from seeing the exercise as an opportunity to maximise their course grade. An example problem was selected by the lecturer who proceeded to work through the problem as a demonstration to the class referring to the PSSAT and selecting performance levels. When the performance level was selected, the lecturer made a note of the evidence that supported that level. The researcher provided commentary on the performance levels and the evidence provided to explain them.

Students were provided with a copy of the PSSAT with each course assessment activity. In each course, except for the human services program, it had been planned that four assessment exercises would be undertaken by students. In the human services course, three assignments were planned. Students were asked to undertake the assessment task and then to complete the PSSAT and to submit it with their assignment. Lecturers graded the assessment task as they normally would then examined the PSSAT and the evidence that students had provided for their chosen performance level. The lecturers then selected the performance level that they believed reflected the student's performance on that indicator and made a comment where their judgment differed from that of the student. The lecturers made copies of completed PSSAT forms and provided them to the researcher. The original PSSAT forms were returned to students with their marked scripts.

At the end of the semester, the lecturers provided a final mark and grade for participating students.

The above processes yielded three main data elements. First, a score for each sub-scale on the TSA is available. Second, self assessment and lecturer scores are available for each of the three problem solving assessments completed. Third, there is an end of semester and end of year mark and grade for each participating student.

Analytical methods

This section contains brief descriptions of the data preparation, including the treatment of missing values, and analytical procedures used in the analyses. A more detailed discussion of each of these methods is presented in Chapter 6.

Data from students test forms were entered directly into SPSS and saved as standard SPSS data files. Data were scanned for obvious data entry errors using basic descriptive statistics.

Tertiary Skills Assessment

The purpose of the study reported in this chapter is to assess the emergence of problem solving ability over time. Problem solving is assessed 'in context' – that is, in existing courses using the assignments that are set as part of those courses to assess the knowledge and skills that form the curriculum for the courses.

In addition to evaluating growth in problem solving ability, there is interest in investigating whether problem solving ability reflects prior achievement and whether it is related to subsequent performance in the courses being taught. Students are admitted to tertiary programs on a variety of bases. In South Australia, the conventional criterion for entrance to university degree programs is a student's end-of-secondary school examination achievement. However, not all students enter post-secondary programs on this basis. Some students come from other states or countries, and for these students the end-of-school achievement is not readily available or comparable. Some students are admitted as mature age entrants, and most sit mature-age entry tests, while some individuals are admitted based on other work or study experiences. For these reasons, a separate test of prior achievement is required.

The Australian Council for Educational Research (ACER) was approached to see if one of their many tests could be used as a measure of initial achievement. The Graduate Skills Assessment (GSA) was identified because it was a test of generic skills, including problem solving. The components of the GSA had been shown to measure discrete attributes and the performances of students in different degree programs suggested that the test was tapping into a general ability construct (ACER, 2000; ACER, 2001a). At the time the test was being considered for use in the current study, it had four components. It was expected that the test would subsequently be broadened by the addition of further sub-scales. The four sub-scales were critical thinking, problem solving, interpersonal understandings, and report writing and argument development. Each of the first three sub-scales had a multiple choice format, while report writing and argument development were assessed using two extended writing activities. The test took three hours to administer, which was impractical for the purposes of the current study, as the testing had to fit into the one-hour timetabled sessions for the courses.

A new test, the Tertiary Skills Assessment (TSA) based on the GSA, was developed by ACER.[25] This test included items similar to those used in the multiple choice components of the GSA. The number of items in the TSA was reduced so that the test could be administered in one hour. All TSA items were trialled by the author with a small sample of nine undergraduate students and recent graduates. The purpose of this trial was to assess the time taken for each of the available items and to generate approximate difficulty estimates for these items to select items covering a wide range of difficulties. A set of 55 items was selected for the shortened form of the TSA comprising 16 critical thinking items, 19 problem solving and 20 interpersonal understanding items. This is referred to below as the Tertiary Skills Assessment -Short Form (TSA). The GSA and TSA were secure tests so items were not available for publication. However, a set of sample items was prepared by ACER. These items were similar to, but not the same as, the ones used in the shortened TSA and are available in the publication 'GSA Sample Questions' (ACER 2003).

The TSA-SF was administered in scheduled class sessions at each of the four research sites. Each administration was supervised by the author. Test booklets were distributed to participants. Instructions for participants were included on the cover sheet of each booklet and these instructions were read aloud to participants by the author. The test was invigilated by the author and regular teaching staff. At the conclusion of the test sessions, booklets were collected by the author. The short form of the TSA was administered to 149 volunteer candidates at the four research sites. All test booklets, used and unused, were returned to ACER as soon as the scoring and data entry had been completed.

Analysis of the Tertiary Skills Assessment

For each candidate, a scaled score on each of the three sub-scale components of the test was required. This was achieved in a two-stage process. First, it was

[25] The TSA was used by permission of Professor Geoff Masters, Chief Executive Officer of ACER. The author is grateful for permission to use the secure test in this research.

necessary to verify the measurement properties of the instrument and to calibrate it by estimating difficulty parameters for all items. Second, scores on each of the scales were estimated for participants. Verification of measurement properties, calibration and scaling were achieved through the application of the Rasch measurement model (Rasch, 1960, 1980).

Calibration and scaling

In order to check the measurement properties of the instrument and to calibrate it, items were scored dichotomously. Some candidates did not complete all test items in the allocated time. Isolated items within the test that were skipped were coded as wrong. Where two or more items at the end of the test were missed, they were coded as 'not-reached'. The not-reached items were coded as missing during calibration. Failure to do this would result in biased estimates of item difficulty for items at the end of the test. However, when generating ability estimates for individuals on each of the scales, the not-reached items were scored as wrong. In the calibration analysis, item parameters were exported to a file to be used as anchored values in generating Rasch scaled estimates for individuals.

Item difficulties, their standard errors and fit indices for items in each of the three scales are shown in the following tables; Table 29, interpersonal understanding; Table 30, critical thinking; and Table 31 problem solving. The scales have useful ranges of item difficulties and item fit, as indicated by the weighted fit index (IMS) for all items lies within an acceptable range (0.7 to 1.3, Bond & Fox, 2001, p. 179).

Table 29: Item parameters for the TSA interpersonal understanding scale

Item no.	Difficulty	Std err	IMS	OMS	Infit *t*	Outfit *t*
1	-0.19	0.26	1.14	1.28	0.88	1.01
2	0.11	0.24	1.03	0.99	0.30	0.03
13	-0.48	0.29	0.93	0.76	-0.33	-0.67
14	-0.41	0.28	0.96	1.11	-0.16	0.44
15	0.80	0.21	1.13	1.11	1.42	0.69
16	-0.13	0.26	1.08	1.19	0.55	0.75
17	0.84	0.21	1.05	1.08	0.61	0.51
18	-0.95	0.34	0.91	0.58	-0.26	-1.00
19	0.01	0.25	1.01	1.09	0.09	0.41
23	-0.41	0.28	1.03	1.12	0.24	0.45
24	-0.75	0.32	0.91	0.62	-0.34	-1.00
25	-0.56	0.30	1.04	1.21	0.29	0.67
29	-0.26	0.27	1.08	1.27	0.51	0.94
30	0.66	0.22	0.97	0.99	-0.29	0.00
37	0.65	0.22	1.10	1.24	0.95	1.27
47	-0.24	0.28	0.93	0.77	-0.34	-0.72
48	0.25	0.25	0.97	0.92	-0.16	-0.28
49	0.13	0.25	0.88	0.72	-0.80	-1.14
50	0.40	0.24	0.88	0.82	-0.95	-0.80
51	0.53	0.24	0.86	0.74	-1.27	-1.33

Note IMS=Infit mean square; OMS=Outfit mean square

Table 30: Item parameters for the TSA critical thinking scale

Item	Difficulty	Std err	IMS	OMS	Infit *t*	Outfit *t*
8	-0.39	0.25	1.11	1.17	0.87	0.68
9	-1.75	0.37	1.04	1.88	0.22	1.42
10	-0.39	0.25	1.05	0.97	0.42	-0.02
11	0.57	0.21	0.96	0.96	-0.40	-0.16
12	0.44	0.22	0.84	0.74	-1.75	-1.52
22	0.99	0.21	1.25	1.70	2.63	3.61
26	-0.50	0.25	0.96	0.87	-0.24	-0.38
27	-1.13	0.30	0.90	0.78	-0.46	-0.44
28	0.53	0.21	1.27	1.48	2.61	2.43
38	-0.04	0.23	0.89	0.79	-0.99	-0.94
39	-0.64	0.26	1.00	1.04	0.06	0.23
42	0.04	0.23	0.88	0.84	-1.02	-0.66
43	-0.06	0.24	0.91	0.83	-0.78	-0.68
44	0.58	0.22	0.95	0.91	-0.50	-0.46
45	0.93	0.21	0.87	0.80	-1.48	-1.20
46	0.81	0.22	1.03	1.05	0.33	0.36

Note IMS=Infit mean square; OMS=Outfit mean square

Table 31: Item parameters for the TSA problem solving scale

Item	Difficulty	Std err	IMS	OMS	Infit *t*	Outfit *t*
3	-2.37	0.41	0.86	0.32	-0.38	-1.10
4	-1.19	0.28	0.79	0.55	-1.16	-1.33
5	-1.04	0.27	1.05	0.93	0.37	-0.10
6	0.59	0.20	0.87	0.91	-1.46	-0.49
7	1.40	0.20	1.32	1.45	3.91	2.04
20	-1.17	0.29	1.05	0.96	0.32	0.01
21	-0.68	0.25	1.05	1.06	0.40	0.31
31	-2.14	0.39	0.87	0.40	-0.40	-1.06
32	-0.55	0.24	0.88	0.87	-0.80	-0.40
33	-0.33	0.23	0.90	0.97	-0.79	-0.05
34	1.11	0.20	1.00	1.05	0.05	0.32
35	-0.55	0.24	1.03	0.94	0.26	-0.15
36	1.11	0.20	1.01	1.01	0.21	0.14
40	-0.15	0.23	0.94	0.87	-0.47	-0.56
41	-0.06	0.23	0.84	0.67	-1.35	-1.64
52	-0.29	0.25	1.21	1.57	1.42	1.94
53	1.99	0.23	1.00	1.17	0.09	0.66
54	2.15	0.23	0.80	0.76	-2.22	-0.88
55	2.18	0.23	1.23	1.92	2.20	3.06

Note IMS=Infit mean square; OMS=Outfit mean square

Having calibrated the three TSA scales and verified that items fitted their scales adequately, the analysis was run again. In this scaling analysis, all item parameters were imported from the file of anchored values created in the calibration analysis. In the scaling analysis, item parameters were not re-estimated, all missed items were coded as wrong, and individuals' scores and their standard errors were estimated.

There is no intention to undertake a diagnostic and refinement exercise on the TSA, but in order to establish the utility of the instrument and its scales for the current study, it is useful to examine summary statistics for the scales (see Table 32). The scales have acceptable indices of reliability, with the Cronbach alpha values being 0.81, 0.84 and 0.83 for the interpersonal understanding, critical thinking and problem solving scales respectively. Rasch item reliability indices for these scales are 0.75, 0.90 and 0.96.

The scales all have useful measurement ranges. The range of the interpersonal scale, at 1.79, is somewhat limited, but these ranges need to be considered in relation to the distribution of person abilities. The mean person score on the interpersonal understanding scale, at 1.76, is rather high, suggesting that the scale is not well targeted to the current sample. Mean scores within 0.5 logits of the item mean are regarded as being very well targeted, those between 0.5 and 1.0 logits are adequately targeted, while those greater than 1.5 logits from the items mean are poorly targeted (Curtis, 2004b). Similarly, the mean person score on critical thinking, at 1.34, also indicates some mistargeting. That scale, however, has a reasonably wide measurement range, so most individuals can be measured adequately. The problem solving scale is adequately targeted, with a person mean of 0.90 and a wide range of item thresholds.

A further indication of the utility of a scale is the typical standard error of the person estimate. The standard errors range from approximately 0.5 to 1.0 logits, their mean values being shown in the final column of Table 32. The mean standard errors are less than 10 per cent of the range of person scores and about one half of the standard deviation of person scores, suggesting that individuals are well differentiated on the three scales.

Finally, person fit indices and differential item function (DIF) are examined for all individuals on the three scales. No cases reveal poor person fit on either the interpersonal understanding or the critical thinking scales. Two individuals show marginal misfit on the problem solving scale with weighted person fit indices greater than 1.8 (Curtis, 2004b). No items on the interpersonal understanding or critical thinking scales show DIF, but one item (Item 21) on the problem solving scale is marginally easier for males than females (t=-2.31). The effect is regarded as small and the item retained.

It appears that the three subscales of the TSA are adequate for their intended purposes in the current study. The targeting of the interpersonal understanding scale is not ideal, but it does lead to scores that have an acceptable precision.

Table 32: Summary statistics for the three TSA scales

Scale	Scale indices		Item difficulty (logits)			Person measure (logits)		
	Cronbach alpha	Item reliability index	Minimum	Maximum	Range	Mean	Std dev.	Mean std err
Interpersonal understanding	0.81	0.75	-0.95	0.84	1.79	1.76	1.36	0.74
Critical thinking	0.84	0.90	-1.75	0.99	2.74	1.34	1.58	0.77
Problem solving	0.83	0.96	-2.37	2.18	4.55	0.90	1.43	0.64

Note: N=144 cases

PSSAT calibration and scaling

Because the problem solving assessment instrument was revised, it is necessary to check that the revised assessment instrument conforms to the requirements of measurement and to calibrate it.

The PSSAT was used by students in assessing their own problem solving performance and by their lecturers in examining the evidence students presented and making judgments about each student's problem solving achievement. The instrument was used by students and their lecturer on three assessment tasks. It was expected that student problem solving performance would improve over the three assessment occasions. In order to compare individuals' scores on all occasions on a common metric, one occasion was selected as a baseline for calibration and students' scores on this and other occasions were scaled using anchored item parameters from the chosen assessment occasion. The second of the three assessment occasions was chosen, although any of the three could be used. The analysis of PSSAT data proceeded in two stages. In the first, the instrument was calibrated; that is, item parameters were estimated and these values were exported to a file. In the second stage, individuals' problem solving performances were estimated using the anchored item parameters.

Indicator thresholds (Deltas) and their standard errors and weighted indicator fit statistics (IMS) are shown in Table 33. The parameters for indicator 11 – the Mayer key competency performance level – is shown. The weighted fit index for this indicator is recorded as 1.47 and reveals poor fit. This indicator was removed in estimating individuals' scores.

For clarity, the thresholds shown in Table 33 are plotted in Figure 15. Indicator 11, the Mayer key competencies performance level indicator, has been removed from the analysis. Indicator 13, a monitoring item, shows relatively poor separation of the top two thresholds. This is indicator 16 in the PSA (see Figure 4 in Chapter 7) where it also has poorly separated upper thresholds.

Table 33: PSSAT indicator thresholds (Deltas) and indicator fit statistics

Indicator		Thresholds (Deltas) and standard errors						Fit
No.	Label	Delta 1	se	Delta 2	se	Delta 3	se	IMS
1	Rep1	-2.93	0.36	-0.44	0.23			0.91
2	Rep2	-3.52	0.43	-1.05	0.23	1.82	0.23	1.20
3	Rep3	-2.43	0.31	0.06	0.22			1.17
4	Rep4	-1.61	0.28	-0.11	0.22			0.97
5	Pln1	-2.38	0.31	-0.38	0.22	2.27	0.26	0.91
6	Pln2	-0.51	0.17					1.11
7	Pln3	-1.30	0.19					0.80
8	Pln4	1.26	0.22	2.54	0.32			0.95
9	Exe1	-2.06	0.31	-0.52	0.23			0.96
10	Exe2	-2.88	0.34	0.30	0.22			0.81
11*	Exe3	-0.97	0.24	1.42	0.23			1.47
12	Mon1	0.51	0.17					0.78
13	Mon2	-0.40	0.22	0.79	0.21	1.20	0.23	1.20
14	Mon3	0.08	0.21	2.03	0.26			0.82
15	Mon4	-0.07	0.22	1.89	0.25			1.00
16	Ref1	-0.14	0.22	2.38	0.27			0.86
17	Ref2	0.58	0.22	2.87	0.33			1.06
18	Ref3	1.21	0.17					0.89

Note: *Indicator 11 (Exe3) is the Mayer key competency performance level. It is not used in estimating individual problem solving scale scores.

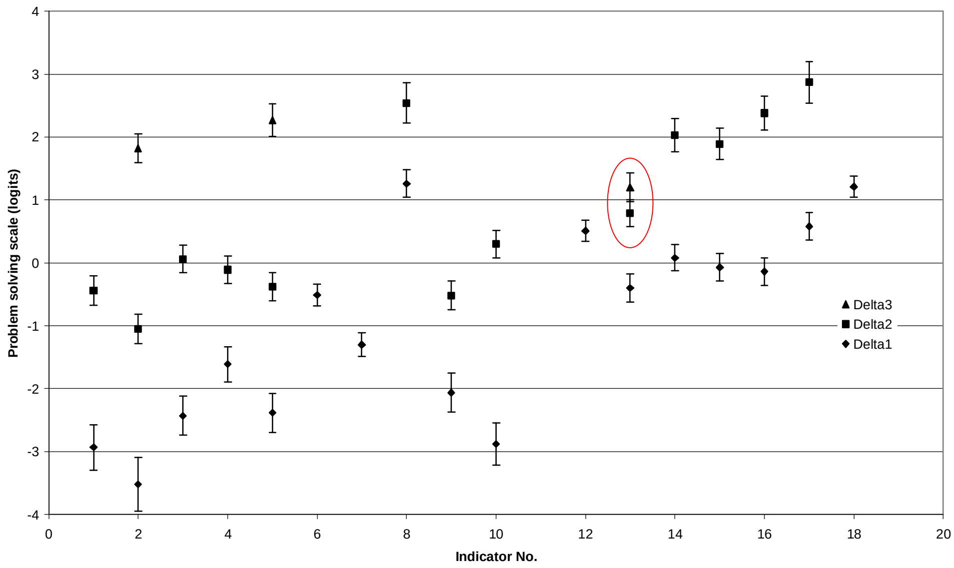


Figure 15: PSSAT indicator thresholds (Deltas) and standard errors

Indicator 10 (sets an appropriate time frame) has relatively high thresholds. In the PSA, this was a dichotomous indicator, but the context for the assessment is now different. Students in the current study were required to undertake the assignments in a set time frame and had little choice in this matter. This indicator might not be as relevant for this course as it had been in the first study. One of the observations made in the analysis of the PSA instrument (Chapter 7) was that performance level thresholds did not show the consistency between indicators that might be expected given the use of the SOLO taxonomy as a basis for describing these levels. In that analysis, it was observed that the cognitive indicators (representation, planning and execution) had lower difficulties than the metacognitive ones (monitoring and reflection). The same pattern is apparent in the current data set. In the current study, the lowest thresholds are between -3 and -1 logits for the cognitive indicators and between 0 and +1 logits for the metacognitive indicators. For the polytomous indicators, similar differences are seen in the ranges of the second set of thresholds. This implies that the metacognitive indicators are more difficult to achieve than the cognitive ones and that the SOLO levels are not equally applicable to the two sets of indicators.

The calibrated indicator parameters were exported to a file. These values are used to generate problem solving scores for each assessment occasion for students' self assessed performance and for the teacher assessed performance. These scores are then imported into the original SPSS data file in which demographic information and students TSA-SF results are recorded.

Comparing self assessment and lecturer validation

In order to test whether self assessment and lecturer validation were measuring an intended common trait, a differential item function (DIF) analysis was undertaken. If students and lecturers interpret indicators and their performance standards differently, their assessments would not be comparable, and this is expected to be revealed through DIF. For this analysis, indicator difficulty parameters estimated from self-assessments were compared with those estimated from the lecturer validations over the three assessment occasions. In Figure 16, standardised differences in indicator difficulty parameters are shown on the horizontal axis and indictors are shown on the vertical axis. A dashed vertical line indicates no difference between the difficulty estimates and solid vertical lines indicate differences corresponding to *t* statistics of -2 and +2. Indicators located within that range show no significant DIF, while those outside the range do show DIF.

Of particular statistical interest in the differential indicator function (DIF) plot is the pattern of differences of the cognitive and metacognitive indicators. Lecturers found it easier than students to endorse favourably four of the cognitive indicators while students found it easier to endorse four of the metacognitive ones. In particular, students were more likely than their teachers to endorse Indicator 12 (Checks progress towards goal). It is likely that students did this, but might not have presented evidence of this activity. The pattern of differences between students and lecturers in assessing cognitive and metacognitive indicators suggests that students may not be adequately aware of what is involved in metacognitive activity. Some instruction in this area may be beneficial for students.

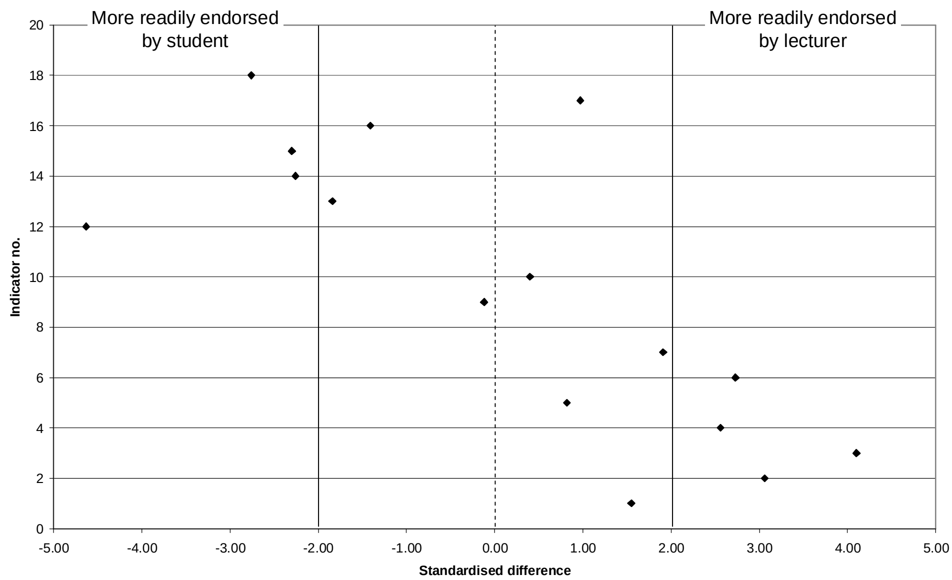


Figure 16: Differences in indicator difficulty estimates between self assessments and lecturer validation

Changes in problem solving performance over time

Of the 62 students who enrolled in the engineering course, data are available for 42 on three problem solving assessments. The students who did not complete three assessments included some who chose not to continue in the project and some who left the course.

Eight students who participated in the problem solving assessments were absent for the TSA testing. Because the test is a secure one, it was not possible to arrange to test these students. In addition, four students who had undertaken self assessments did not have teacher assessments recorded on the final assessment task. In order to use the greatest number of available observations, missing data imputation was used. Alternative methods for missing data imputation are discussed in Chapter 6. Multiple imputation, employing the program Norm (Schafer, 2000), was used in the current study. This program preserves the covariance matrix of the data and imputes values for each missing observation, creating multiple data files for the several imputed values. In this study, the recommended five imputations and therefore five data files were created.

In order to investigate ability bias in the attrition, scores on the TSA sub-scales can be compared for those who continued with those who dropped out. On each sub-scale, the students who dropped out have slightly lower average scores than those who continued (see Table 34), but none of the differences is significant. The *t* statistics for differences on the three TSA sub-scales (interpersonal understanding, critical thinking and problem solving) are -0.86, -0.06 and -1.33 (corresponding p values are 0.395, 0.951 and 0.190).

Table 34: Comparison of TSA sub-scale scores of those who continued in the study with those who dropped out

Sub-scale	Status	N	Mean score	Std. dev	Std. error mean
TSA IP	Dropped out	20	47.17	22.91	5.12
	Continued*	42	52.14	20.62	3.18
TSA CT	Dropped out	20	49.61	22.40	5.01
	Continued*	42	49.94	18.12	2.80
TSA PS	Dropped out	20	46.58	24.32	5.44
	Continued*	42	53.52	16.42	2.53

Note: *Includes cases with imputed values.

In designs that depend on paired comparisons, sample attrition can lead to biased estimates of effects. In this case, there appears to be little ability bias in the attrition of individuals. In longitudinal designs, individuals act as their own controls, since change within individuals is modelled, so any effect of attrition bias is minimised.

Approaches to assessing growth

The case for using multilevel modelling, rather than difference scores between initial and final achievement or repeated measures analysis of variance, for evaluating change over time is made in Chapter 6. There, the points are made that the evaluation of growth requires assessment on at least three occasions (Singer & Willett, 2003, pp. 9-10) and that multilevel modelling is required to evaluate both growth within individuals (their intercepts and slopes with respect to time) and differences in the growth trajectories between individuals.

Before describing the growth modelling that is undertaken, it is necessary to establish a common metric on which student problem solving performance can be compared between students and from one assessment occasion to the next for individual students. This is done using the Rasch measurement model.

Exploratory analyses of problem solving development

For each student on each assessment occasion, two scores are recorded, namely the student's self assessed score and the teacher score. These sets of scores are treated separately. In the design of this study, the intention was to have students assessed on each occasion by the same lecturer to avoid differences in rater severity influencing the result. Unfortunately, for the final assessment activity, the lecturer's workload increased and the marking of that assignment was undertaken by a colleague. All lecturer and student information was provided and an induction session was conducted with the colleague. However, as is shown below, the second marker appears to have been more severe than the course lecturer. For this reason, the final teacher assessment is not readily comparable with previous teacher assessments. Consequently, emphasis is given to the change in self assessed scores, although analyses on the teacher assessed results are also reported.

Summary descriptive statistics are presented along with results of *t* tests and simple analyses of variance to indicate the findings that might arise from past approaches to the evaluation of change. Most attention is given to results of the multilevel analyses.

For the analyses presented in this chapter, Rasch scaled scores, derived from the PSSAT instrument, were rescaled to a mean of 50 and standard deviation of 15.

Summary statistics for problem solving assessments

Summary descriptive statistics for self- and teacher-assessments on the three assessment occasions are shown in Table 35. There appear to be modest increases in the self assessed scores over the three assessment occasions. The teacher assessed scores are low on the first occasion, much higher on the second, but are lower on the third assessment occasion. It should be noted that the third assignment was marked by a colleague and not the lecturer responsible for the course.

Table 35: Descriptive statistics for problem solving assessment by occasion

Assessor		Assessment occasion		
		Occasion 1	Occasion 2	Occasion 3
Self assessment	Mean	50.75	55.88	61.13
	St dev	13.72	15.74	13.58
Teacher assessment	Mean	14.87	49.29	36.23
	St dev	6.04	18.40	20.28

Note: N=42.

Had the first two assessment occasions been used in a two test design, the differences between the test results would show a significant difference, the difference in the means being 5.13, t=2.42 (p=0.02). Similarly, had the second and third test occasions been compared in this way, the difference would again be significant, with the difference of the means being 5.25, t=2.52 (p=0.02). Given that there are three test occasions, a repeated measures analysis of variance was run on self assessed problem solving performance on the three occasions. It showed a significant effect for occasion ($F_{82,2}$=11.85, p<0.001).

Similar analyses were conducted on the teacher assessments of problem solving. The results are not shown, but given the data on teacher assessment shown in Table 35, it is not surprising that they reveal differences by occasion, but this effect might be attributable in part to the final assessment being undertaken by a second assessor.

One of the questions of interest in this study is whether feedback from teachers enhances students' abilities to make reasonable judgments of their own performance. If students do learn to calibrate their judgments against those made by their teachers, correlations between teachers' and students' assessments should increase over time. The correlations between students' and teachers' scores on the three assessment occasions are -0.16 (p=0.30), +0.71 (p<0.001), and +0.54 (p<0.001). The third of these correlations is between students and a different assessor than is the case for the first two assessments, but on the basis of the increase in the correlation between self and teacher assessment between the first and second occasions, it might be argued that students did improve their self assessment skills.

While the above results, based on their self assessments, suggest that students' problem solving performance has improved over time, such analyses do not provide the opportunity to investigate factors that may influence observed change.

In subsequent analyses of problem solving performance in this chapter, the procedures outlined in Singer and Willett (2003, pp. 16-41) have been followed. The key point that Singer and Willett, and many others, make about the study of change is that it requires a multilevel approach. They suggested, however, that it is instructive to begin a study of change by exploring relationships using standard statistical packages such as SPSS. In this section of the chapter, the results of exploratory analyses are presented. The original data file had a single record for each student with assessment results (self and teacher assessments) for each of the three occasions recorded in separate variables. Singer and Willett (2003, p. 19) described this as a person-level data set. The file was restructured so that a separate record was created for each person and each assessment occasion – a person-period data structure. The person-period file is used for the exploratory analyses summarised below.

Self assessment results

The study of learning requires an examination of student performance over time. As learning occurs, student performance increases on successive observations. Researchers are often interested in change within individual students and in how this change differs between students. Thus, there are two related problems. The first is to examine the change that occurs within individuals (their starting points or intercepts and the rate of change or slope). The second is to examine relationships between features of individuals' learning trajectories and the student characteristics that might explain differences in their intercepts and their slopes. Multilevel modelling simultaneously evaluates these two models: the changes within individuals and the relationships between learner characteristics and their rates of learning.

Singer and Willet (2003) recommend undertaking exploratory analyses before running multilevel models in order for researchers to visualise the relationships that they are seeking to model. Trellis plots enable researchers to discern patterns in the slopes and facilitate model conception. It is for these reasons that trellis plots are presented for the data being analyses in the investigation of change in learner problem solving performance.

The trellis plots presented in Figure 17 shows students self assessed problem solving scores, derived from their performance level ratings on the indicators of the PSSAT instrument, over the three assessment occasions. For each student, a linear regression plot of performance over time is shown.

For some students, for example cases 12, 25 and 57, the anticipated pattern of improvement over time is observed. For others, for example cases 2, 20 and 39, no change is apparent. For some students, for example cases 16, 50 and 61, a decline in performance is seen. The distance between the regression line and the plotted observations in Figure 17 gives a visual indication of the magnitude of residuals and therefore the fit of the linear regression model to the observations. The intercepts for these plots appears to cluster around 50 units, but ranges from about 30 to 70 units.

A composite of the trellis plots using a curve for each case is shown in Figure 18. Because there are data for only three assessment occasions, it is necessary to use a linear growth model. The curves shown in Figure 18 reveal departures from the modelled linear relationship. This plot shows a similar variety in the patterns of change in self assessed problem solving performance over time. There are some approximately linear increases in performance, there are also some slightly non-linear patterns, including some that curve either up or down initially before returning to a value on average slightly larger than the performance level observed on the first occasion. There are also some highly curved patterns, with one case increasing sharply between the first and second assessments then declining on the third and two showing a sharp decline between the first two assessments then recovering in the third. The composite plot provides a visual indication of the likely fit of a linear model of change to the data. The highly variable patterns of change and the several non-linear trajectories over time appear to suggest an overall weak relationship between performance and occasion, however this is investigated in the multilevel modelling described below.

Teacher assessment

The trellis plots for teacher assessed problem solving performance in Figure 19 reveal similar variation in the regression line slopes, although there appear to be more upward trending lines. Cases 10, 54 and 57 show substantial growth trajectories over time. Cases 2, 38 and 52 show little change over time, while cases 22 and 58 indicate a decline in performance over the three test occasions. The intercepts of the teacher assessments show less variation than the self assessed regression curves, with most clustering close to 15 units. Of concern, though, is the apparent poor fit of the regression lines, with cases 22, 38 and 48 showing very high residuals. Indeed, there are relatively few cases, for example 9, 25 and 61, where the fit appears to be good.

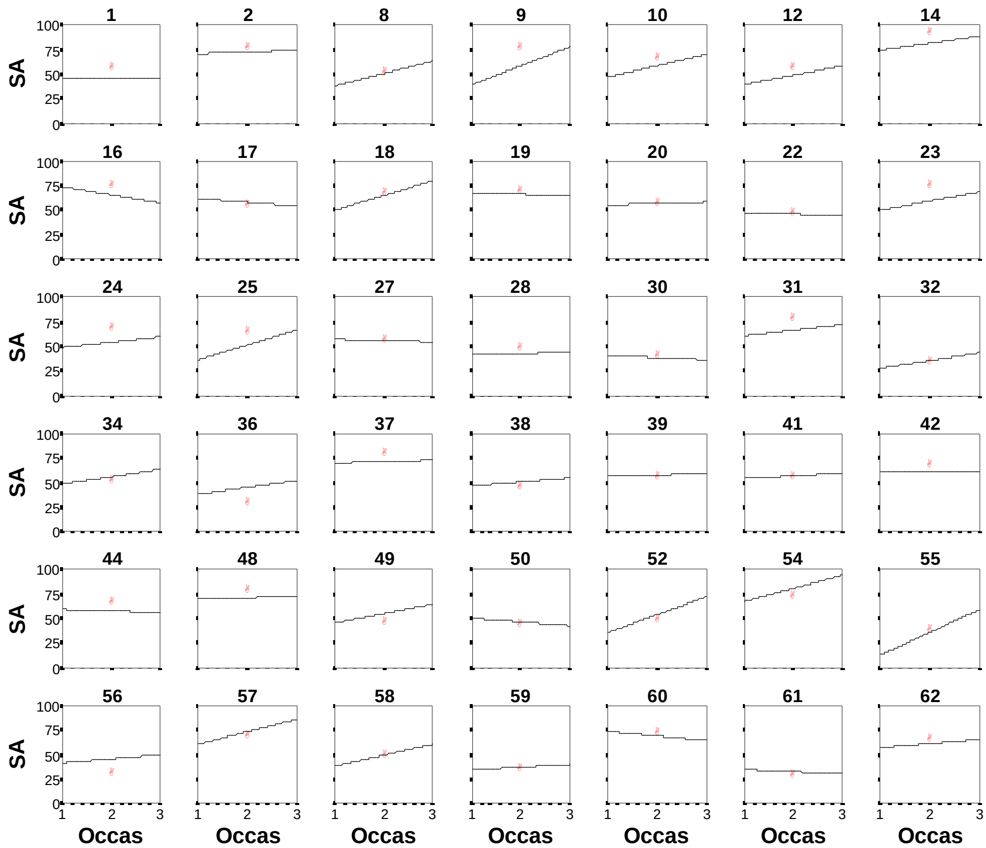


Figure 17: Trellis plots for students' self assessed problem solving scores over three occasions (N=42)

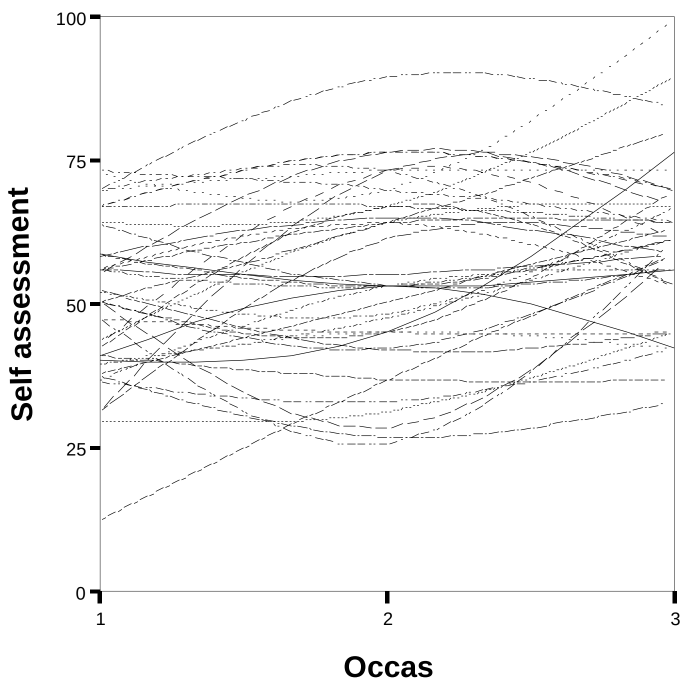


Figure 18: Composite plot of students' problem solving self assessments by occasion showing an interpolated 'best fit' curve (N=42)

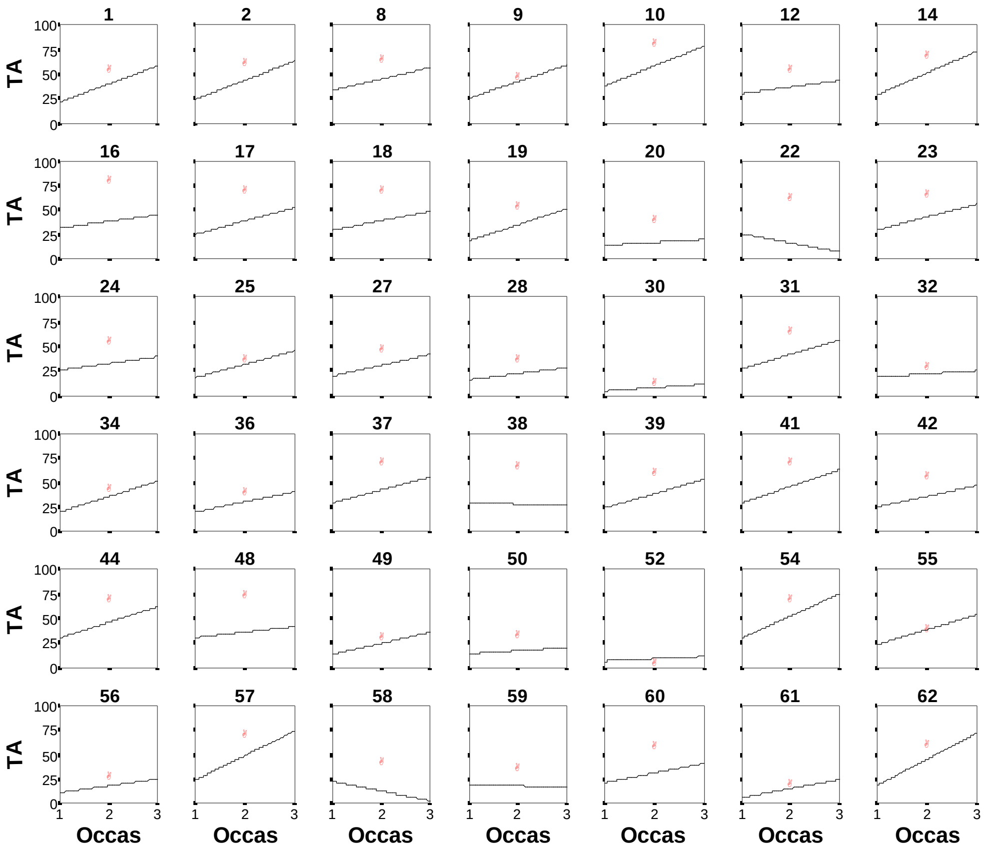


Figure 19: Trellis plots for teacher assessed problem solving scores over three occasions (N=42)

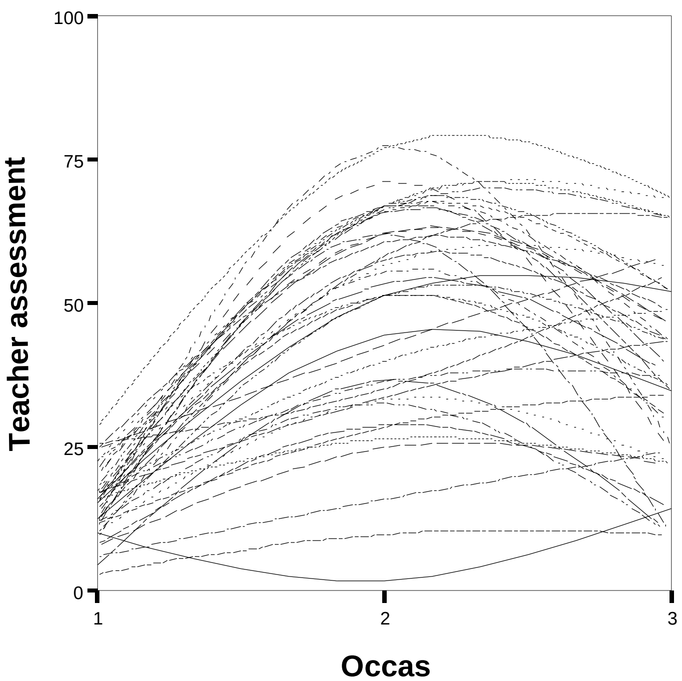


Figure 20: Composite plot of problem solving teacher assessments by occasion showing an interpolated 'best fit' curve (N=42)

The composite plot of the teacher assessed problem solving performance in Figure 20 suggests a reason for the likely poor fit of the linear regression model. Almost all the splined plots show a very strong curve with a low initial score, a higher and more variable score on the second occasion, followed by a lower and variable score on the third occasion. As suspected from the trellis plots, few of the relationships appear to be linear.

The preceding exploratory analyses suggest that the self assessment results might be appropriately modelled using linear regression. Some cases are clearly non-linear, but most appear to sufficiently close to enable linear modelling. The effect of non-linearity will be revealed as poor model fit statistics. It does appear to be unlikely that the teacher assessed scores can be modelled well by similar linear regression functions, but with three observations, this is the only growth model available. In the next section, results of the multilevel modelling of these data are presented.

Multilevel models for problem solving development

The representation and development of a sequence of multilevel models are described in this section.

Representation of multilevel models

Two representations of multilevel relationships are common; in one, the various levels are shown separately by sets of equations while the other uses a single composite equation to characterise the relationships. The former approach is embodied in the HLM program (Raudenbush et al., 2005) and the latter in MLwiN (Rasbash et al., 2005). The two approaches are illustrated in the following equations that represent a simple linear two-level model for growth, with one explanatory variable for the intercept parameter and one for the growth parameter.

The HLM representation uses separate sets of equations for the two levels. For the occasion (within individual) level (Level-1) the equation is:

$$\mathrm{Y}_{ij} = \beta_{0j} + \beta_{1j}.\mathrm{Occ}_i + e_{ij}$$

At the inter-individual level (Level 2), there are two equations. One explains variation in the individual intercept term (β_{0j}), and the second explains variation in the slope parameter (β_{1j}), of the Level 1 equation.

$$\beta_{0j} = \beta_{00} + \beta_{01}.\mathrm{A}j + \mu_{0j}$$

$$\beta_{1j} = \beta_{10} + \beta_{11}.\mathrm{B}j + \mu_{1j}$$

The Level-1 equation can be read as: 'The outcome measure (Y) for individual *j* on occasion *i* is equal to that individual's intercept, plus an effect (β_{1j}) due to the flux of time (occasion), plus a random component (e_{ij}) to capture variation about the modelled outcome.' This equation has two parameters (β_{0j} and β_{1j}) that vary across individuals. Each of these parameters is modelled using an equation at Level-2 of the model.

The Level-2 equations can be read as: 'The intercept for any individual j (β_{0j}) is equal to the overall intercept for all individuals (β_{00}), plus some systematic variation about that value explained by that individual's score on variable *A* (with parameter β_{01}), plus a random residual component (μ_{0j}).' Similarly, the second Level-2 equation may be read as: 'The growth trajectory (slope, β_{1j}) for an individual j is equal to the mean slope for all individuals (β_{10}), plus some variation about that attributable to the individual's score on variable *B* (with parameter β_{11}), plus a random residual variation (μ_{1j}).'

The models can become substantially more complex by the addition of further explanatory variables for intercepts and slopes and by considering non-linear, perhaps quadratic or higher order, models of change over time.

The composite equation, used commonly in representations of analyses conducted with MLwiN is found by substituting the Level-2 expressions for β_{0j} and $\beta_{1\,j}$ into the Level-1 equation. The composite equation thus becomes:

$$Y_{ij} = \beta_{00} + \beta_{01}.Aj + \mu_{0j} + (\beta_{10} + \beta_{11}.Bj + \mu_{1j}).Occ_i + e_{ij}$$

Terms in the equation are usually re-arranged to separate the fixed and random components so the equation appears as:

$$Y_{ij} = \{\beta_{00} + \beta_{01}.Aj + \beta_{10}\,.Occ_i + \beta_{11}.Bj.Occ_i\} + \{\mu_{0j} + \mu_{1j}.Occ_i + e_{ij}\}$$

The fixed terms are included in the first set of braces and the random terms in the second. In this representation, the variation in the slope parameter over individuals ($\beta_{11}.Bj.Occ_i$) is seen to be a cross-level interaction of occasion at Level-1 with a moderating explanatory variable B at Level-2.

Singer and Willett (2003, pp. 74-84) used both representations, as each illustrated particular aspects of multilevel models. The separate equations make the relationships at each level clearer while the combined equation shows the cross level interaction between an individual characteristic (B_j) at Level-2 and assessment occasion (Occ_i) at Level-1. Although MLwiN is used for the analyses presented below,[26] the HLM representation makes some of the relationships more apparent, and it is used in describing the models that are developed.

The case was advanced above for using multiple imputation of missing data. Five imputed data files were generated using the program Norm (Schafer, 2000). In the analyses reported below, each of the models to be tested was run using each imputed data set. The mean parameter estimates from the repeated runs are presented in the tables within this chapter. The results of the separate analyses for each model and each data set are included in Appendix 11.

Problem solving growth: Self assessment

Problem solving performance is assessed on three occasions and, because there are only three occasions, regression models of change over time are limited to linear ones (see Figure 18). Where time (assessment occasion) is used in the models, the intercept is centred on the first assessment occasion and this represents the modelled initial problem solving ability of students. The slope parameter represents the change in performance over assessment occasions. Additional variance parameters are estimated when intercepts and slopes are modelled as random variates across individuals.

A sequence of models is developed, each subsequent one adding additional explanatory elements over its precursor. Two criterion variables are measured, namely student self-assessed and teacher assessed problem solving performance. Similar models are developed for both measures, so it is expedient to describe the models for both criterion measures in general terms first, and then to present and discuss the results separately for the two measures. The self-assessed problem solving achievement variable (Self) is used in the descriptions and equations of the models, but the models apply equally to teacher-assessed performance.

Problem solving is assessed on three occasions for each of the 42 students for whom adequate data are available. Data on the performance of a group of students, assessed on multiple occasions, has a two-level structure and two subscripts are attached to the variable self; subscript i reflects the occasion on which the performance is assessed, and subscript j represents the student for whom the assessment is reported.

Because the models are nested — that is, each model is an extension of an earlier one — the models are evaluated using a likelihood ratio test. For each model, the

[26] The analyses of the self-assessment data were repeated using HLM (Bryk et al., 2005). This program used the restricted maximum likelihood estimation method, compared to the iterative generalised least squared method employed in MLwiN. The results were almost identical, and the MLwiN results are presented in this thesis.

deviance of the observed data from what is estimated to occur if the proposed model completely accounts for variability in the population is computed. The deviance is computed using a likelihood function (the logarithm of the likelihood). When two models are nested, that is where one model is a simplified version of the other, improved fit is indicated by a reduction in the value of the log-likelihood function. Models are compared by considering the ratio of their likelihoods, and since the logarithm of the function is taken, the likelihood ratio is the difference between the two log values. The ratio of likelihoods has a Chi-square distribution and the number of degrees of freedom of the distribution is the number of addition parameters in the more complex model. Consequently, the expected value of the difference between the estimated log-likelihood values divided by the difference in the number of degrees of freedom is unity (1) and can be tested for statistical significance with the Chi-square test.

A sequence of models of problem solving performance over time is developed. The first model (Model 1) is the simplest possible representation of observed problem solving performance and the later models (Models 3A, 4A and 4B) add explanatory variables in an attempt to explain change in problem solving performance. Each model in the sequence is described below. The relationships among the models are represented in Figure 21.

In the discussion of the models presented below, reference is made to parameters for self-assessed problem solving performance. These can be found in Table 38. The deviance statistics for the models of self-assessed problem solving performance are in Table 39. The results for the teacher assessed problem solving model parameters are shown in Table 40.

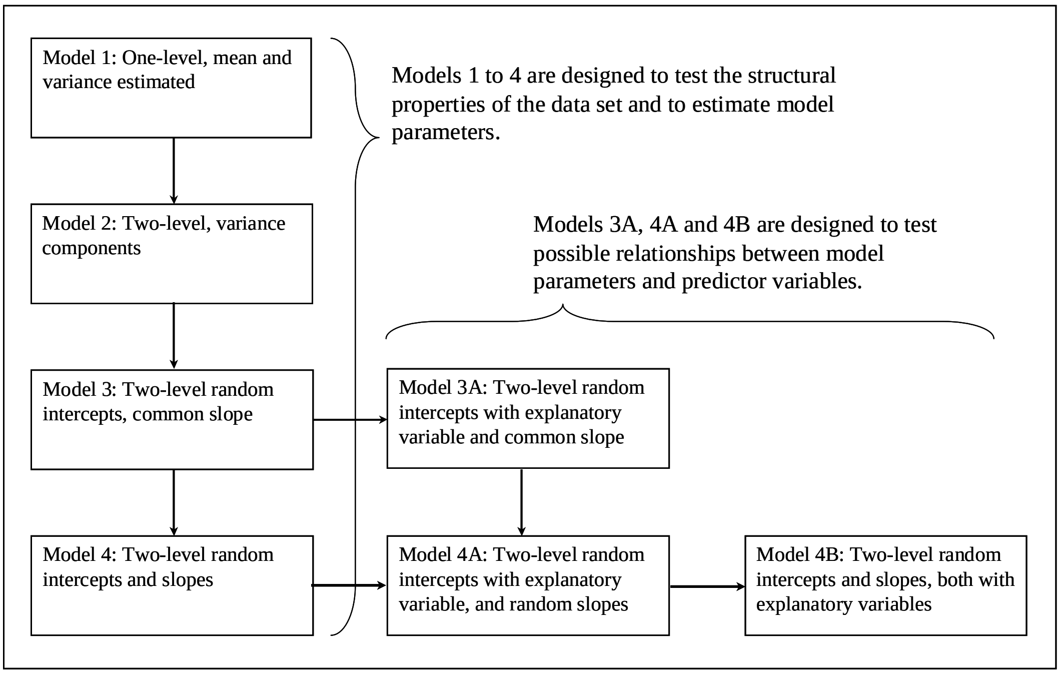


Figure 21: Representation of relationships between models in the sequence of models developed to explore growth in problem solving performance

Model 1: A one-level mean and variance model

The simplest model asserts that there is a common intercept parameter for all students, and that self-assessment scores deviate minimally from that value across individuals and occasions. Since there are no time-related variables in the model, the intercept parameter being estimated is the overall mean and the variance parameter is the variance in scores over both individuals and occasions about that overall mean; that is, it is the total variance in scores. The model is represented by the equation:

$$\mathrm{Self}_{\mathrm{ij}} = \beta_{00} + e_{ij}$$

where

$$e_{ij} \sim \mathrm{N}[0,\sigma^2_{\mathrm{e}}]$$

Implicit assumptions of the model are that the overall mean is an adequate representation of students' problem solving scores and that variation between students and across occasions are small. However, differences are anticipated between students and across occasions. While the model is statistically trivial in that the population mean and variance and their standard errors can be obtained from simple descriptive statistics, its purpose is to provide a baseline value of the likelihood function for the series of models. If subsequent models do not represent an improvement over this baseline one, it indicates that there is very considerable variability over individuals and assessment occasions and that this variability swamps any systematic variation between individuals and occasions.

Model 2: A variance components model

The most basic two-level model is represented by the set of equations:

$$\mathrm{Self}_{ij} = \beta_{0j} + e_{ij}$$

$$\beta_{0j} = \beta_{00} + \mu_{0j}$$

where

$$u_{0j} \sim N[0, \sigma^2_{\mu 0}]$$

$$e_{ij} \sim N[0, \sigma^2_{e}]$$

This is a variance components model with no explanatory variables. It is referred to as the 'null' model in multilevel modelling and as the 'unconditional means' model by Singer and Willett (2003, pp. 92-93). Problem solving performance is modelled at two levels, assessment occasion (Level 1, denoted by the subscript *i*) and student (Level 2, denoted the subscript *j*). However, no time related variable, such as occasion, is used in the model. Thus, the model implicitly assumes no change over time. The equations may be read as the problem solving score (Self_{ij}) of an individual student (*j*) on occasion (*i*) is the mean score for that student plus the deviation (e_{ij}) of scores on each occasion from the mean for that student. The mean score (β_{0j}) for a student (*j*) is the overall mean score for all students (β_{00}) plus the deviation of that student from the mean of all students (μ_{0j}).

The purpose of this model is to estimate the separate variance components at the two levels of the model. Significant variances at each level indicate that there is enough variance at the two levels to warrant modelling performance at both levels. An improvement in the estimate of the log-likelihood function value or deviance over the one-level model indicates that the problem is indeed a two-level problem in practice.

Variance in assessment scores occurs within students between occasions (e_{ij}) and between students (μ_j). The ratio of the between student variance to the total variance is referred to as the Variance Partition Coefficient (VPC), particularly among users of MLwiN, and as the intra-class[27] correlation (ρ) by users of HLM.

---

[27] The term intra-class is normally applied to two-level models in which individuals are clustered within groups, such as students within classes. In the case of change over time, where repeated observations are clustered within individuals, the term intra-class seems inappropriate. In this context of repeated measures of individuals, the term variance partition coefficient seems more appropriate.

If the variation between students is very small compared with the total variance, that is, if the VPC is very small, it suggests that students are responding similarly to each other and that the estimation of a multilevel model is not warranted. In the case of self-assessed scores, the VPC is 0.401, indicating that 40 per cent of the total variance in scores occurs between students and that 60 per cent occurs within students between occasions. The significant variance at both levels indicates the need for two-level modelling.

Model 3: A random intercepts, fixed slopes model

Of more interest is whether students' rates of change in problem solving performance are significantly different from zero. That is, is there a change in problem solving performance over assessment occasions? This is modelled initially using a common (fixed) change parameter for all students (β_1). This model is intermediate between Singer's and Willett's unconditional means and unconditional change models. The model is represented by the set of equations:

$$\text{Self}_{ij} = \beta_{0j} + \beta_1.\text{Occ}_1 + e_{ij}$$

$$\beta_{0j} = \beta_{00} + \mu_{0j}$$

where

$$u_{0j} \sim \text{N}[0, \sigma^2_{\mu 0}]$$

$$e_{ij} \sim \text{N}[0, \sigma^2_{e}]$$

The change parameter (β_1) is a fixed parameter – that is, common to all students – and so it is not represented at Level 2 of the model.

The time-related variable, assessment occasion (Occ1), takes the values 0, 1 and 2 for the first, second and third assessment occasions and therefore centres the time variable on the first assessment. In the models described below that include assessment occasion, the intercept term is the estimate of problem solving performance on the first assessment. Had the original occasion variable (Occas) been used, which took the values 1, 2 and 3 respectively for the three assessment occasions, the intercept would be estimated for a time (Occas = 0) that lies outside the assessment occasion range. While this does not matter in that students who perform well have a higher intercept than those who perform poorly, especially because a linear change model is being fitted, the intercept term does not have an obvious relationship to the assessment occasions that are used. It is also possible to centre on either the second or the final assessment occasions, but centring on the first leads to a conventional interpretation of the intercept term as representing initial problem solving performance.

Model 4: A random intercepts and slopes model

The final model in the sequence of structure-only models is the random intercepts and slopes model. Singer and Willett (2003, pp. 97-101) referred to this as the 'unconditional growth' model. That is, the model posits change over time, but does not explain this growth. In this model, slopes – as well as intercepts – are allowed to vary between students, as is shown by the additional subscript on the slopes parameter (β_{1j}). Singer and Willett (2003) commented that:

> Estimating these variance components allows us to distinguish level-1 variation from the two different kinds of level-2 variation and to determine whether interindividual differences in change are due to interindividual differences in true initial status or true rate of change. (p. 99)

The model is represented by the equations:

$$\mathrm{Self}_{ij} = \beta_{0j} + \beta_{1j}.\mathrm{Occ}_1 + e_{ij}$$

$$\beta_{0j} = \beta_0 + \mu_{0j}$$

$$\beta_{1j} = \beta_1 + \mu_{1j}$$

where

$$\begin{bmatrix} \mathrm{u}_{0j} \\ \mathrm{u}_{1j} \end{bmatrix} \sim N([0, \Omega_u]) : \Omega_u \begin{bmatrix} \sigma^2_{u0} & \\ \sigma_{u01} & \sigma^2_{u1} \end{bmatrix}$$

$$e_{ij} \sim \mathrm{N}[0, \sigma^2_e]$$

In this model, both intercept (β_{0j}) and slope (β_{1j}) parameters are permitted to vary between students, so these parameters do have equations at Level 2. The random components of the model are now more complex than in previous models. In addition to the occasion level of the model, there are now two variances at the student level of the model – a variance ($\sigma^2_{\mu0}$) for the intercept parameter residuals (μ_{0j}) and a variance (σ^2_{u1}) for the slope parameter residuals (μ_{1j}). In addition, there is a covariance ($\sigma_{\mu0\mu1}$) between the intercept and slope parameter residuals. In this model, that lacks explanatory variables, the magnitudes of the separate variance components for intercepts and slopes indicate whether there is significant variance available to be explained by predictors. The covariance between intercepts and slopes indicates whether there is any relationship between them, and if so, whether it is positive or negative. A positive covariance would indicate that those individuals who begin with greater problem solving ability also increase their performance at a greater rate than do other students. A negative covariance would indicate that students who begin with lower levels of performance catch up with their colleagues over time.

Significant residual variance at the occasion level indicates that there are time-varying (occasion-dependent) factors that may explain this residual variance. Significant variances in the two student level residuals (intercept and slope) indicate that there may be time-invariant, student-related factors that can explain differences in students' initial performance (intercept) and their growth in performance (slope).

Models 1 to 4 (above) represent an evolution in the complexity of the growth modelling. Other than the time-dependent variable, no explanatory variables are included, either for the intercept or the slope terms in the models. In the next three models described, explanatory variables are added.

Model 3A: Explaining intercepts in the random intercepts model

The random intercepts, fixed slopes model developed above (Model 3 in Figure 21) shows that significant variation exists in individual intercepts. The question addressed in the present model is: What variables, if any, are related to the variation in intercepts between students? Three variables are candidates, namely the critical thinking, problem solving and interpersonal understandings scores from the Tertiary Skills Assessment (TSA) instrument. These are taken as indicators of prior achievement, a construct thought likely to be related to both problem solving performance and the mark achieved by students at the end of their course. The TSA is designed to measure a set of discrete constructs. The correlations between scores on these scales for the 42 individuals for whom complete data are available are shown in Table 36, and they indicate that the sub-scale scores do not contribute to a composite ability scale. For this reason, each of the TSA sub-scales is tried, separately and in combination, as explanatory variables for individual intercepts in the growth model.

Table 36: Correlations between TSA sub-scales

	Critical thinking	Interpersonal understanding
Interpersonal understanding	0.23 (0.15)	
Problem solving	0.30 (0.67)	0.56 (<0.001)

Note: N=42; p values are shown in parentheses.

Different variables are found to be related to problem solving performance measured using the PSSAT, depending upon whether self-assessed or teacher-assessed scores are used. The model with an explanatory variable for the intercept, and which has a common parameter across all students, is represented by the equations:

$$\mathrm{Self}_{ij} = \beta_{0j} + \beta_1.\mathrm{Occ}_1 + e_{ij}$$

$\beta_{0j} = \beta_{00} + \beta_{01}.\mathrm{TSAIP}_j + u_{0j}$

where

$u_{0j} \sim N[0, \sigma^2_{\mu 0}]$

$e_{ij} \sim N[0, \sigma^2_e]$

The second of these equations tests the proposition that the variation in students' intercepts about the mean intercept is predicted by students' interpersonal understanding scale score (TSAIP_j). In the modelling of self-assessed problem solving performance, the interpersonal understanding score is found to be significant. When teacher-assessed problem solving performance is modelled (see below), both interpersonal understanding and critical thinking scores are found to be marginally significant. In neither case – self-assessed nor teacher-assessed problem solving performance – is the TSA problem solving score even marginally significant.

Model 4A: Explaining intercepts in the random intercepts and slopes model

This model is a development of Model 4 in which intercepts and slopes are permitted to vary across individuals. In the augmented model, explanatory variables for the variations in students' intercepts, but not slopes, are introduced. The equations for this model are:

$\mathrm{Self}_{ij} = \beta_{0j} + \beta_{1j}.\mathrm{Occ}_1 + e_{ij}$

$\beta_{0j} = \beta_{00} + \beta_{01}.\mathrm{TSAIP}_j + \mu_{0j}$

$\beta_{1j} = \beta_{10} + \mu_{1j}$

where

$$\begin{bmatrix} \mathrm{u}_{0j} \\ \mathrm{u}_{1j} \end{bmatrix} \sim N([0, \Omega_u]) : \Omega_u \begin{bmatrix} \sigma^2_{u0} & \\ \sigma_{u01} \sigma^2_{u1} & \end{bmatrix}$$

$e_{ij} \sim N[0, \sigma^2_e]$

In the equations shown above, the interpersonal understandings variable is shown as an explanatory variable. The regression parameter for this variable is significant for self-assessed problem solving performance. The parameters for other possible explanatory variables are non-significant for self-assessed problem solving.

Model 4B: Explaining intercepts and slopes in the random intercepts and slopes model

The final model in the sequence of growth curve models is one in which both the intercept and the growth parameters are explained by student-level variables. This model differs from the preceding one by the addition of an explanatory variable for the growth parameter. As in the previous model, candidate explanatory variables are the TSA sub-scale scores critical thinking, problem solving and interpersonal understandings. The interpersonal understandings scale scores are found to be significant for both intercepts and slopes for self-assessed problem solving performance. The finding that the interpersonal understanding parameter is significant, while the problem solving one is not, is surprising and is discussed later. For teacher-assessed problem solving performance, both interpersonal understanding and critical thinking are used as explanatory variables for intercepts and slopes. The equations for the self-assessment model are:

$$\mathrm{Self}_{ij} = \beta_{0j} + \beta_{1j}.\mathrm{Occ1} + e_{ij}$$

$$\beta_{0j} = \beta_{00} + \beta_{01}.\mathrm{TSAIP}j + \mu_{0j}$$

$$\beta_{1j} = \beta_{10} + \beta_{11}.\mathrm{TSAIP}j + \mu_{1j}$$

where

$$\begin{bmatrix} \mathrm{u}_{0j} \\ \mathrm{u}_{1j} \end{bmatrix} \sim N([0, \Omega_u]) : \Omega_u \begin{bmatrix} \sigma^2_{u0} & \\ \sigma_{u01} \sigma^2_{u1} & \end{bmatrix}$$

$$e_{ij} \sim \mathrm{N}[0, \sigma^2_{\mathrm{e}}]$$

Model Summary

Following the pattern represented in Figure 21, the models can be compared by reference to the sets of equations that describe each model. These equation sets are shown in Table 37.

Table 37: Equations for the models of problem solving performance

Model	Base models	Extended models	
		Random intercepts (model variant A)	Random intercepts and slopes (model variant B)
Model 1	$\text{Self}_{ij} = \beta_{00} + e_{ij}$		
Model 2	$\text{Self}_{ij} = \beta_{0j} + e_{ij}$ $\beta_{0j} = \beta_{00} + \mu_{0j}$		
Model 3	$\text{Self}_{ij} = \beta_{0j} + \beta_{1}.\text{Occ}_{1} + e_{ij}$ $\beta_{0j} = \beta_{00} + \mu_{0j}$	$\text{Self}_{ij} = \beta_{0j} + \beta_{1}.\text{Occ}_{1} + e_{ij}$ $\beta_{0j} = \beta_{00} + \beta_{01}.\text{TSAIP}_{j} + u_{0j}$	
Model 4	$\text{Self}_{ij} = \beta_{0j} + \beta_{1j}.\text{Occ}_{1} + e_{ij}$ $\beta_{0j} = \beta_{0} + \mu_{0j}$ $\beta_{1j} = \beta_{1} + \mu_{1j}$	$\text{Self}_{ij} = \beta_{0j} + \beta_{1j}.\text{Occ}_{1} + e_{ij}$ $\beta_{0j} = \beta_{00} + \beta_{01}.\text{TSAIP}_{j} + \mu_{0j}$ $\beta_{1j} = \beta_{10} + \mu_{1j}$	$\text{Self}_{ij} = \beta_{0j} + \beta_{1j}.\text{Occ1} + e_{ij}$ $\beta_{0j} = \beta_{00} + \beta_{01}.\text{TSAIP}j + \mu_{0j}$ $\beta_{1j} = \beta_{10} + \beta_{11}.\text{TSAIP}j + \mu_{1j}$

Comparing the Models

In order to assess each of the models described above, the regression parameter estimates and the variance component estimates of each model and the likelihood ratio tests between adjacent models (see Figure 21) are examined.

As noted above, five imputed data sets are generated and are tested with each model. In the tables summarising the model results presented below (see Table 38 for the self-assessed problem solving performance and Table 40 for teacher-assessed problem solving performance), the parameter values shown for each model are averaged from the five imputed data estimations. The complete set of results for each model and each imputation are included as tables in Appendix 11.

Results of Self-assessed Problem Solving Performance

The results of the sequence of models for self-assessed problem solving performance are shown in Table 38. The results of particular interest in the sequence of models of self-assessed problem solving ability are (a) the parameters for the explanatory variables (including assessment occasion), (b) the variances and covariances of the several residual terms in the models, and (c) the deviance, which measures the departure of the data from the modelled structure. While these elements are related it is helpful to consider them separately in an evaluation of the models.

Table 38: Results of models of self-assessed problem solving performance

Model 1: Single level, overall mean and variance							
Fixed parameters				Variance components			
Parameter	Estimate	se	p	Parameter	Estimate	se	p
β_{00}	56.16	1.31	<0.001	σ^2_{eij}	217.76	27.44	<0.001
-2LL	1035.43						
Model 2: Two-level variance components (null) model							
Fixed parameters				Variance components			
Parameter	Estimate	se	p	Parameter	Estimate	se	p
β_{00}	56.16	1.76	<0.001	σ^2_{eij}	130.38	20.12	<0.001
				$\sigma^2_{\mu0}$	87.38	29.37	<0.001
-2LL	1015.81						
Model 3: Random intercepts with fixed slope							
Fixed parameters				Variance components			
Parameter	Estimate	se	p	Parameter	Estimate	se	p
β_{00}	51.10	2.09	<0.001	σ^2_{eij}	103.86	16.03	<0.001
β_1	5.05	1.11	<0.001	$\sigma^2_{\mu0}$	96.22	29.07	<0.001
-2LL	997.30						
Model 4: Random intercepts and slopes							
Fixed parameters				Variance components			
Parameter	Estimate	se	p	Parameter	Estimate	se	p
β_{00}	51.10	2.01	<0.001	σ^2_{eij}	111.99	24.44	<0.001
β_{01}	4.85	1.06	<0.001	$\sigma^2_{\mu0}$	77.64	42.65	0.03
				$\sigma^2_{\mu1}$	-8.13	16.15	0.31
				$\sigma_{\mu0.\mu1}$	12.00	19.85	0.27
-2LL	996.34						
Model 3A: Random intercept with explanatory variable; fixed slope							
Fixed parameters				Variance components			
Parameter	Estimate	se	p	Parameter	Estimate	se	p
β_{00}	36.81	4.61	<0.001	σ^2_{eij}	103.86	16.03	<0.001
β_1	5.05	1.11	<0.001	$\sigma^2_{\mu0}$	67.70	22.98	0.002
β_2 (IP)	0.28	0.08	0.001				
-2LL	986.98						
Model 4A: Random intercepts and slopes with explanatory variables for intercepts							
Fixed parameters				Variance components			
Parameter	Estimate	se	p	Parameter	Estimate	se	p
β_{00}	34.48	4.61	<0.001	σ^2_{eij}	111.99	24.44	<0.001
β_{01}	5.05	1.06	<0.001	$\sigma^2_{\mu0}$	65.97	40.41	0.05
β_2 (IP)	0.29	0.08	0.001	$\sigma^2_{\mu1}$	-8.13	16.15	0.31
				$\sigma_{\mu0.\mu1}$	3.59	19.99	0.43
-2LL	986.00						
Model 4B: Random intercepts and slopes with explanatory variables for intercepts and slopes							
Fixed parameters				Variance components			
Parameter	Estimate	se	p	Parameter	Estimate	se	p
β_{00}	40.96	5.52	<0.001	σ^2_{eij}	111.99	24.44	<0.001
β_{01}	0.90	2.96	0.38	$\sigma^2_{\mu0}$	63.16	39.89	0.06
β_2 (IP)	0.20	0.10	0.03	$\sigma^2_{\mu1}$	-10.51	15.82	0.25
β_3(Occ.IP)	0.08	0.06	0.07	$\sigma_{\mu0.\mu1}$	6.17	19.53	0.38
-2LL	983.74						

Deviance statistics

The deviance computed for each model is shown in Table 39. This table follows the pattern of models displayed in Figure 21. The refinements from Models 1 to 3 result in successive improvements in the match between the evolving model and the observed data. The reductions in deviance from Model 1 to 2 (19.62, df=1) and from Model 2 to 3 (18.51, df=1) are highly significant ($p<0.001$). There is also a reduction in deviance of 10.32 (df=1, p=0.001) from Model 3 to 3A, in which the intercept was explained by the inclusion of interpersonal understanding score. There was, however, no significant improvement from Model 3 to 4 (0.96, 2df, p=0.62). The deviance change from Model 4 to 4A suggests an improvement in model fit. The improvement is similar to that observed from Model 3 to 3A, and for the same reason – the addition of interpersonal understanding as an explanatory variable. Further, Model 4A can be regarded as a development of Model 3A and it shows a non-significant reduction in deviance of 0.99 (df=2; p=0.61). Thus, Model 4A represents no improvement in model fit over Model 3A. It is noted that Model 4A is a development of Model 4 and it is found not to represent an improvement over Model 3.

Model 4B, in which an attempt is made to explain variations in individual slopes, shows no significant improvement in deviance (2.26, 1 df, p=0.133) over Model 4A and at the cost of model complexity. This is unsurprising, given that it attempts to explain a variation in slopes for which there was no evidence anyway, as seen in the evolution from Model 3 to 4.

Table 39: Deviance values for models of self-assessed problem solving performance

Base model	No explanatory variables	Explanatory variables for intercepts ('A' suffix)	Explanatory variables for intercepts and slopes ('B' suffix)
1 Single level	1035.432 (2)		
2 Null model	1015.813 (3)		
3 Random intercepts	997.299 (4)	986.983 (5)	
4 Random intercepts and slopes	996.339 (6)	985.995 (7)	983.735 (8)

Note: Each cell shows the deviance estimate for the model with the number of estimated parameters in parentheses below it.

Variance components

The total variance, between occasions and students, is 217.759 (Model 1). In Model 2, this variance is partitioned into between-occasions, within students (Level 1) and between-students (Level 2) components and these are 130.379 and

87.380 respectively. The variance partition coefficient (VPC) is the ratio of the between student to total variance and is 0.401. That is, 40 per cent of the total variance occurs between students and 60 per cent occurs between occasions within students. In any regression modelling, one purpose is to explain variance. In this two-level model, the two separate variance components warrant separate explanation. Both components are significantly different from zero, so there is scope to attempt to explain both variance components.

The variance between occasions is of particular interest in growth modelling, and explaining that variance by examining how much of that variance is predicted by assessment occasion is undertaken in Model 3. An assessment occasion term (Occ1) is introduced into the model and a common growth parameter is assumed for all students. The intercept is permitted to vary across students. The between occasion residual variance falls from 130.38 to 103.86, a reduction of 20.3 per cent. That is, assessment occasion accounts for 20 per cent of the between-occasion variance. This leaves a very substantial amount of random variation in the assessment scores within students over time that is not explained by growth in self-assessed problem solving performance. This unexplained variance may be attributable to factors such as the different assessment tasks in which students undertook the problem solving assessments and to variation in students' judgements about their performances.

In the final development of the basic model (Model 4), that is, without introducing explanatory variables, slopes are allowed to vary across students. This model has the same between-occasion specification as Model 3. The change in variance explained between occasions is very small, being 103.86 in Model 3 and 111.99 in Model 4. The difference between Models 3 and 4 lies in the between-student component where slopes are permitted to vary in Model 4. The residual variance between students, which in Model 3 is 96.22, is now split into three components, namely the variation in intercepts, variation in slopes, and a covariance between intercepts and slopes. These values are 77.64, -8.13 and 12.00 respectively, with the latter two values not being significantly different from zero.[28] The very small changes in variances at the between-occasion and between-student levels between Models 3 and 4 suggest that Model 4 does not provide an improved explanation of the relationship between self-assessed problem solving performance over that

[28] Negative variance can occur in estimation using maximum likelihood and iterative generalised least squares. In MLwiN, a default estimation setting makes any negative variance zero, but this may bias other parameters in the estimation and the default setting was over-ridden, allowing negative variances for individual parameters (Rasbash, Steele, Browne, & Prosser, 2004, p. 88).

represented by Model 3. This is consistent with the non-significant deviance change between these models.

Model 3A is an extension of Model 3 and in it a variable, interpersonal understanding, is introduced to explain variation in intercepts between students. Model 3A shows no change in the variance between occasions but does reveal a significant reduction in the variance between students from 102.86 to 67.70. That is, the variance in intercepts between students is predicted in part by their interpersonal understandings scores. The finding is interesting for several reasons. First, whether or not interpersonal understanding is used, neither problem solving nor critical thinking, as measured by the TSA, is significant. If the TSA and PSSAT measures of problem solving do indeed assess the same underlying construct, the TSA assessed problem solving measure can be expected to predict the application of problem solving processes, which is what the PSSAT assesses. Second, if problem solving, as measured by the PSSAT is a measure of general cognitive ability, a relationship with critical thinking can be expected. The revelation of a relationship between self-assessed problem solving and interpersonal understanding suggests that those people who rate themselves highly on the application of problem solving processes are also those who score highly on the interpersonal understandings scale. This observation warrants further discussion, and this is presented below, after consideration of the variable coefficients of the various models.

Based on the deviance estimate, Model 4 does not represent an improvement over Model 3. Nonetheless, it is useful to compare Models 4, 4A and 4B. Based on the deviance parameters, Model 4A represents an improvement over Model 4. The models have the same Level-1 (between-occasions) model but they differ in that Model 4A adds an explanatory variable (interpersonal understanding) at Level-2 for student intercepts. This results in a reduction in the residual variance (μ_{0j}) from 77.64 to 65.97. Model 4B is a further development in which the same explanatory variable is added at Level-2 for student slopes (trajectories). The deviance estimate shows no significant improvement and the change in the residual variance (μ_{0j}) is very small, from 65.97 to 63.16. The residual variance for the slopes (μ_{1j}) is negative in Models 4, 4A and 4B. In all three cases, this variance is not significantly different from zero. As noted above, individual variance estimates were allowed to fall below zero during estimation iterations in order to ensure unbiased estimates of other parameters in the models. However, the negative variance estimates suggest that these models have exhausted the information available in the data set and represent an attempt at over-interpretation of the data.

Finally, the models are compared on their interpretability. To do this, variable coefficients are considered.

Variable coefficients

The variable coefficients that are important in the model are the parameter for growth (the slope parameter β_1 in Model 3 and subsequent models) and the parameter β_{01} for the explanatory variable interpersonal understanding in Model 3A. While this parameter is also included in Models 4A and 4B, the deviance and residual variances discussed above show that Model 4 and its derivatives do not contribute to a greater understanding of the growth in self-assessed problem solving performance compared with Model 3A.

The growth parameter is 5.05 and is significant ($p<0.001$). This is the change in self-assessed performance with each additional assessment occasion, and since there are three occasions (two additional occasions) the growth in performance over the assessment period was 10.1 units. The unit (the logit) was derived from Rasch scaling of responses to the PSSAT instrument, and this was rescaled to have a mean of 50 and a standard deviation of 15 units. Growth of approximately 10 units, or two-thirds of a standard deviation, appears to be a substantial improvement. This finding is consistent with the hypothesis that repeated application of the problem solving assessment tool, designed to scaffold problem solving performance, is associated with an improvement in problem solving performance.

The parameter for interpersonal understanding on individual problem solving performance intercepts (β_{01} in Model 3A) is 0.28 ($p=0.001$). Interpersonal understanding is also a Rasch scaled score, rescaled to a mean of 50 and a standard deviation of 10 units. Thus, although the influence of interpersonal understanding on the problem solving intercept is statistically significant, it is rather small. The problem solving and critical thinking GSA variables are also used in this model (3A), but are found not to have a significant influence on the problem solving intercept.

It is worth noting that the parameter for interpersonal understanding on the change in problem-solving performance over time (i.e., the slope) is positive and marginally significant ($p=0.07$, see Model 4B, Table 38). This leads to the suggestion that greater interpersonal understanding, perhaps through the self-awareness component of this construct, might lead to greater growth in problem solving over time. This possible outcome is not confirmed in these analyses because, as pointed out above, Model 4B does not represent an improvement over other models. The possibility of greater interpersonal understanding leading to greater learning is a matter that could be investigated in further studies.

Summary and interpretation

In summary, a sequence of seven models is presented in order to develop an understanding of the change over time of students' problem solving performance. The first two models are only of interest in providing baseline values for the deviance statistic and variance components. Model 3 is the first substantively meaningful model and it shows that students' problem solving performance does improve over time. The parameters estimated for Model 3A show that, in addition to the improvement in performance over time, students' initial problem solving performance scores are related to their knowledge of interpersonal skills as measured by the GSA. The final models, (Models 4, 4A and 4B) do not add to the information found from Model 3A, although Model 4A confirms the finding of Model 3A that knowledge of interpersonal skills influences initial self-assessed problem solving performance.

The finding that problem solving performance improves over time, indicated by the positive and significant parameter (β_1 in Models 3 and 3A and β_{01} in Models 4 and 4A, Table 38) for occasion, is consistent with the hypothesis that the use of the Problem Solving Skills Assessment Tool (PSSAT) leads to an improvement in problem solving performance. This is an important finding in the context of the thesis. It supports the contention discussed in Chapter 4 that formative assessment, including teacher feedback, can lead to improved performance (Black & Wiliam, 1998a). It also supports the speculations raised in that chapter that self-assessment can play a role in improved learning (Boud, 1995b, 2002).

The interpretation of this finding is that the use of PSSAT scaffolds students' application of problem solving skills. Alternative explanations for this change can be advanced. Ramachandran (2003) argued that for a phenomenon to be regarded as 'scientific' it must meet three criteria: it must be shown to be a real effect – that is observable generally in different contexts; there must be a plausible explanation for it; and it must have broad implications; that is, it must be meaningful. To his second requirement, one might add that there should be no superior explanation. The observation of improvement has been made in several contexts using the PSSAT. A plausible explanation is that the tool scaffolds the application of generalised problem solving skills; and such an improvement does have broad implications as it suggests that problem solving performance, which is a desired generic skill, can be improved through interventions that draw attention to generalised problem solving skills. There is a possible competing explanation. It could be argued that familiarity with the instrument could lead to the observed improvement, at least in part. The design of the current experiment cannot lead to a rebuttal of this possibility. An alternative experiment, perhaps one in which a similar magnitude of change was observed using different forms of skills scaffolding on each assessment occasion would be required. The theory underlying the conception of problem solving used in this study and the support for formative assessment suggest that the primary interpretation – that repeated exposure to the explicit assessment of problem-solving processes leads to improved problem-solving performance – is the more plausible of the two.

The observation that knowledge of interpersonal understanding, but that neither problem solving nor critical thinking skills, contribute to students' initial problem solving performance is interesting and warrants explanation. The non-significance of the TSA problem solving score on initial self-assessed problem solving using the PSSAT suggests that different conceptions of problem solving are being measured. The tasks used in the two assessments is a key difference between the TSA assessment and the assessment of problem solving in this study. The TSA tasks are small-scale general ones, whereas the tasks used in this study are more substantial and are routine in the courses that students take. It could be argued that these tasks are more authentic than those used in the TSA.

The finding that interpersonal understanding is related to initial self-assessed problem solving performance may have more to do with the students' capacities for self-assessment and for applying the performance criteria embedded in the PSSAT than with their problem solving ability. However, students who are better problem solvers need be more self-aware to support monitoring and reflection, so a strong relationship between interpersonal understanding and problem solving performance could be expected. This possibility cannot be verified in the current study.

Problem solving growth: Teacher assessment

A set of models, analogous to those for self-assessment, can be developed for teacher assessment of problem solving. The descriptions of the models of teacher assessments of problem models, and their equations, are functionally the same as for the self assessment models so they are not repeated here.

Only the first two of these models are described and their results are shown in Table 40. As is the case for self assessed problem solving, Model 1 is statistically simple. It models only the mean and variance of problem solving performance over all students and all three assessment occasions. The log likelihood function has a value of 1126.43.

The variance components model (Model 2) has the same estimate for the mean as Model 1. In Model 2, the total variance is partitioned into a within-student (between assessment occasion) component and a between student component. Several aspects of this model are noted. First, the estimate of the between occasions variance (σ^2_{eij}) is 443.35 and the between-student variance ($\sigma^2_{\mu0}$) is 9.94. The variance partition coefficient (VPC) is 0.02. Effectively, all the variance in the system is between-occasion variance with very little of the variance between students. Second, the change in the log-likelihood function is very small at 0.026 with 1 degree of freedom, p=0.87. The most basic two-level model does not provide a better representation of the data than a one-level model that simply reports the overall mean and total variance. Because the variance partition model does not provide a better representation of the data than Model 1, no purpose is served by attempting to explain the variance between occasions by modelling intercept and slope parameters for individuals, which is the purpose of these models.

An explanation is required for the failure to model change in teacher assessed problem solving performance. The explanation has two components. First, in discussing the exploratory plots for teacher assessed problem solving scores (see Figure 19 and Figure 20), the point is made that the trajectories shown in those plots are non-linear. The attempt to fit a liner model to such data is unsuccessful. Second, an explanation for the non-linearity in the change trajectories is required. In part, this may be explained by the change in assessor that occurred for the third assessment occasion. That assessor appears to have been a harsher judge than the first assessor. This leads to the high between-occasion variance which swamps the between student variance and to the failure of the two-level model to represent an improvement over the simplest single-level model for the data.

Table 40: Results of models of teacher-assessed problem solving performance

Model 1: Single level, overall mean and variance							
Fixed parameters				Variance components			
Parameter	Estimate	se	p	Parameter	Estimate	se	p
β_{00}	34.034	1.824	<0.001	σ^2_{eij}	447.327	56.358	<0.001
-2LL	1126.426						
Model 2: Two-level variance components (null) model							
Fixed parameters				Variance components			
Parameter	Estimate	se	p	Parameter	Estimate	se	p
β_{00}	34.034	1.838	<0.001	σ^2_{eij}	443.352	61.073	<0.001
				$\sigma^2_{\mu0}$	9.937	42.833	0.408
-2LL	1126.220						

Summary and discussion of key findings

The Problem Solving Skills Assessment Tool

The revised version of the problem solving assessment instrument, the Problem Solving Skills Assessment Tool (PSSAT), has reasonable measurement properties. It appears to provide a sound basis for the measurement of students' use of problem solving processes as they work through discipline-based assessment tasks. The indicator that reflects the Mayer key competencies performance levels does not fit with indicators of problem solving processes. It is suspected that the key competencies indicator may depend on the affordances or demands of the tasks being undertaken rather than the problem solving procedures being used by individuals as they seek solutions to the problems they face. Confirming this speculation requires a design in which task difficulty, student ability and rater severity can be modelled separately. In such a model, an interaction between this indicator and task difficulty is expected.

Basing the PSSAT performance levels on the SOLO taxonomy has not produced the expected pattern of performance level thresholds (see Figure 15 and related discussion). The indicators of monitoring and reflection, both metacognitive, activities have higher thresholds than the cognitive representation and planning indicators. It seems that the complexity of the cognitive processes have their own levels of difficulty. This does not suggest a failure of the SOLO taxonomy. Rather, the lowest threshold for any level of monitoring and reflection might demand high levels of cognitive processing, perhaps multi-structural. A limitation of many Rasch measurement software programs, including Quest, is that if there are no observations in the lowest response category, the threshold between the next pair of response categories is assigned to the first threshold series. In the case of the PSSAT, the final Reflection indicator (Anticipates situations in which current approach may be useful) is dichotomous. This indicator, therefore, has a single threshold. However, threshold level for the performance 'Considers future broader applications of solution strategy' suggests a high cognitive demand, perhaps at the extended abstract SOLO level. (See Figure 15 for threshold locations). This possibility can be accommodated by using a rating scale format for the instrument in which all SOLO levels are represented for all indicators. However, some options would need to be marked as not available for some indicators. For low level indicators, the higher SOLO levels would need to be greyed out while for indicators like the final Reflection one, the intermediate levels would need to be marked as not available. Such an approach still requires software than can accommodate thresholds for which certain response categories are systematically missing. This requires a variant of the partial credit model and Wilson and Masters (1993) have shown that this is possible.

Differences in the use of performance levels for the cognitive and metacognitive indicators between students and teachers are apparent (see Figure 16). It appears that teachers have higher expectations than students do in judging performance on the metacognitive indicators.

Models of change in problem solving performance

In both exploratory and multilevel analyses of students' self assessments of their problem solving performance there is evidence of improvement over time. The change is approximately one half of the standard deviation in students' self assessed performance and is regarded as a moderate effect. This is the main finding from the current study and addresses a key research question.

The finding of improved problem-solving performance associated with repeated assessment and feedback cycles, although significant, is not uniformly distributed over students, as some students show no improvement while others show substantial gains (see Figure 17). What is not known is the extent to which students note and act on the feedback that is provided by lecturers and the variation in change over time may be attributed to variation in students' response to the feedback that is provided. This is a matter that could be investigated in further studies.

The intercept term for self assessed problem solving performance, which represents initial problem solving performance, is predicted by the interpersonal understanding sub-scale score from the TSA. Students who have a better understanding of interpersonal relationships appear to have better problem solving skills. However, this understanding does not lead to greater improvement in their problem solving performance over time. Neither of the TSA problem solving and critical thinking sub-scale scores predict initial self-assessed problem solving performance.

While a very substantial increase in the students' teacher assessed problem solving performance between the first and second assessment occasions is found in the study, the use of a different rater for the third assessment occasion, and who appears to be a more severe judge, leads to a strongly non-linear relationship between performance and assessment occasion. The attempt to apply a linear model to this relationship is unsuccessful and no claim can be made about change in teacher assessed problem solving performance over time.

In summary, the two studies reported in Chapters 7 and 8, provide support for several key propositions.

First, it appears that it is possible to develop a problem-solving assessment tool that is based on a theoretical conception of problem solving and to use that instrument in diverse assessment tasks to generate a measure of problem-solving performance.

Second, the instrument developed through the two studies, the PSSAT appears to have acceptable measurement properties.

Third, the instrument has been used by both students and teachers. Both groups reported finding the instrument helpful in their assessments.

Fourth, students' problem-solving performance can be enhanced through the combination of an assessment tool that draws attention to key problem-solving processes and an iterative formative assessment method through which students undertake assessment tasks, make judgments about their performance and receive feedback on those tasks.

Chapter 9: Summary, Discussion, Implications and Conclusions

A personal reflection

The investigation of generic skills reported in this thesis arose from a personal interest in these concepts on the part of the researcher who was involved in the development and implementation of 'graduate qualities' while employed in an academic professional development role in an Australian university during the late 1990s. That interest developed further through his involvement in background work undertaken by the Australian Council for Educational Research for the Australian Chamber of Commerce and Industry and the Business Council of Australia. The latter research, along with work undertaken by other contributors, culminated in the publication of *Employability skills for the future* (ACCI & BCA, 2002). Both experiences led to concerns about what teachers, trainers and lecturers and their students might understand about the ideas and relationships that were represented as the 'generic skills' that they were expected to teach and assess. In particular, the schemes appeared to lack a disciplinary basis. Further, the employability skills scheme focused on selecting skills that were endorsed by industry but it did not attend to whether the skills were assessable. This was a difficulty in relation to the eight key skills of the employability skills framework but a major impediment in relation to the 13 personal attributes that formed part of the framework.

Summary

This section presents a summary of the background to this study and summarises key findings from the study organised around the four main research questions that framed the investigation.

What is known about generic skills?

In the current study, the review of literature focusing on generic skills developments in Australia, but informed by developments overseas that appear to have influenced decision making in Australia, reveals a set of issues that have not been addressed adequately in the generic skills schemes that have been promoted.

The issues surrounding the description and implementation of generis skills are their:

- definition and selection,
- dissemination and implementation,
- assessment and reporting, and
- certification and recognition.

Of these issues, it is argued that two are central. First, the definition of generic skills, both as a class of constructs and as individual concepts, has not been as well developed as is required to ensure that shared understandings exist among stakeholders and to ensure that what is reported as achievement of these skills is consistent. Second, and deriving from clear definition, is the requirement to establish assessment methods that are valid, reliable, fair and feasible. The combination of clear definitions and appropriate assessment methods is expected to lead to reports of achievement that are informative to learners and to others with an interest in these reports, such as potential employers.

It is argued that the remaining issues, while important, depend upon definition and assessment. Many skills have been proposed as ‘generic’ and there may be insufficient time to teach and develop an extensive set of generic skills in schools, vocational education and higher education settings. The selection of those skills judged to be the most important depends upon defining the specific concepts clearly and judging which are likely to be used most in specific work and social settings and which are likely to be amenable to assessment. Similarly, their dissemination and implementation depend upon clear definitions and feasible assessment methods. Finally, the certification and reporting of their achievement depend upon credible assessment of well-defined and broadly accepted conceptions of generic skills.

In the literature reviewed in this thesis, two major reports have identified and recommended that particular sets of generic skills should be endorsed and implemented (ACCI & BCA, 2002; Mayer Committee, 1992). While the Mayer Committee made specific recommendations on assessment, the ACCI and BCA report did not. The literature review in Chapter 2 canvassed reasons for the poor record of implementation of the Mayer Committee’s recommendations. In summary, the key competencies were poorly understood in, and the recommendations for assessment were not accepted by the school and vocational education sectors (Dawe, 2002; Jasinski, 1996; Ryan, 1997).

The review of literature and, in particular, the finding that generic skills have not been adequately defined in generic skills schemes, leads to the first two research questions that guide this investigation into generic skills.

Research question 1: What are the characteristics of generic skills?

This question is answered in part through the review of the literature on generic skills. For example, the diversity of terms that have been used as labels for these skills indicates the variety of conceptions that have formed around them. They are described as skills, attributes, qualities and abilities. They are also said to be generic, key, core and transferable.

An important contribution to the debate about generic skills occurred in the OECD sponsored DeSeCo project (Rychen & Salganik, 2001). Of particular note in that project were the contributions to the definition and selection of 'key competencies' made from several disciplinary perspectives including anthropology, economics, philosophy, social psychology and sociology. The different disciplinary perspectives led to different sets of skills being proposed. This is not surprising, as each discipline has its own traditions and concerns. From an economic view, which was grounded in human capital theory and concerned with productivity, emerged one set of skills. From a sociological perspective, concerned with the individual and collective roles of 'social actors,' appeared another proposal. Some common ground was found among these proposals and three 'meta-competencies' were put forward, namely acting autonomously, using tools interactively, and joining and functioning in socially heterogeneous groups (Rychen, 2003). These meta-competencies were rather abstract they have not proved to be useful in constructing measures of generic skills performance.

Research question 2: Can a foundation for generic skills be found in cognitive psychology?

The current research occurs within the perspective of cognitive psychology. Cognitive psychology is concerned with the relationships between the thinking and behaviour of individuals and seeks to describe, explain, predict and influence behaviour. No claim is made that this discipline is superior to others, for example those that were represented in the DeSeCo project. However, it is the discipline within which the author operates.

A strength of cognitive psychology is its use of rigorous methods for the definition of constructs and for their investigation and measurement, although it has not always taken as rigorous an approach to measurement as is desirable

(Embretson, 1999; Michell, 1997). In order to generate measures of generic skills, valid and reliable assessment methods are sought.

The first two research questions are addressed through an analysis of two potential constructs underlying generic skills, namely intelligence and competence. In order to constrain the scope of the investigation to manageable proportions, problem solving is the main focus and several psychological conceptions of it are compared.

Towards a definition of generic skills and of problem solving

The research begins with a review of generic skills schemes and finds that the proposed skills are not well defined. It is argued that generic skills need to be defined as a class of constructs and the conception of each skill requires specific elucidation. Many skills have been proposed as generic in the many schemes that have been developed in many countries. Each proposed skill warrants investigation, but the scope of such an investigation is too great to be accommodated within the confines of this thesis. The focus of the thesis beyond the initial review of generic skills as a class of related constructs is limited to problem solving. Problem solving is regarded as one of three major generic skills, the other two being ‘communication’ and ‘teamwork.’ However, it is hoped that the tools and methods that are developed in this study may be used as exemplars for assessing other generic skills.

Characterising generic skills

There has been some debate and disagreement in the literature about the utility of measures of intelligence as predictors of workplace performance. Consequently, generic skills have been injected into this debate as being skills that are required for work, especially given rapid changes in the skill requirements of workplaces.

It may be more useful to perceive ‘generic skills’ as competences rather than as manifestations of ‘general intelligence’ (or of specific types of intelligence). As competences, they are regarded as being amenable to development through instruction and experience, whereas if they are viewed as manifestations of a traditional conception of intelligence they may be perceived as relatively fixed attributes. The comparison of intelligence and competence as potential concepts underlying generic skills does not consider intelligence and competence as polar opposites. Rather, it appears that intelligence and competence are two views of a common set of human capabilities. Sternberg’s theory of successful intelligence appears to connect theories of intelligence and competence by incorporating sets of cognitive and metacognitive processes (Sternberg et al., 2000). For the purposes of enhancing and assessing generic skills, their representation as competences has some advantages. It avoids the mistaken view of intelligence as a fixed trait and avoids the controversy about the predictive power of intelligence for workplace performance (Hunt, 1995; Jensen, 1998; McClelland, 1973; Spencer & Spencer, 1993). However, much more importantly, it presents generic skills as attributes that may be developed through instruction and experience.

Finding indicators of competence

By viewing generic skills as competences, they are recognised as latent attributes. They may not be observed directly and indicators of them need to be sought. These indicators need to be articulated so that teachers and learners are aware of and can work to produce the behaviours that are being sought in the assessment of emerging competence.

Processes may be the most direct indicators of competence in non-experts

In this study, the focus is restricted to problem solving. Several theories of problem solving are canvassed and two theories, the situative and information processing ones are compared. The most useful conception of problem solving among novices is an information-processing model. Situative theory locates problem solving performance in the social and contextual features of the situation, whereas the information processing model invokes more general processes that have the potential to be activated in novel contexts. There is no certainty that the transfer implied by the possible generalisation of these processes occurs. However, the notion of competence involves a capability to abstract relationships from situations and there is evidence that such abstraction is an essential element of adaptive competence (J. R. Anderson et al., 1996; K. F. Miller & Stigler, 1991).

Three models of problem solving (Bransford & Stein, 1984; Hayes, 1989; Polya, 1957), each of which invoked a set of processes that were similar to those recognised by Sternberg and others (2000), are reviewed in this study. The following processes are thought to capture the cognitive and metacognitive activity involved in problem solving:

- apprehending, identifying and defining the problem,
- planning an approach to the problem including selecting strategies,
- carrying out the chosen plan,
- monitoring progress towards the goal, and
- reflecting on the effectiveness of the solution attempt.

Understanding the possibilities and limitations of alternative assessment methods

Assessment is one of the two major issues that are identified as being inadequately developed in the implementation of generic skills schemes. Assessment can serve various purposes, often summarised as assessment **of** learning, assessment **for** learning and assessment **as** learning. It is also described as having formative and summative purposes. In the current study, the formative purpose is the focus.

Assessment draws attention to desired learning outcomes

The proposition that assessment is a driver for learning arises from a negation of its antithesis. Callan (2002) observed that students did not value what was not assessed. While the converse of this statement, that students do value what is assessed, does not follow logically from it, it does appear that assessing particular knowledge and skills may draw attention to them.

Formative assessment may enhance performance

Cognitive skills are believed to make a contribution to economic productivity (Hanushek & Wößmann, 2007). Developing these skills is regarded as important and formative assessment is believed to facilitate the development of skills (Black & Wiliam, 1998a). In this study, it is hypothesised that generic skills can be learned, and that they are likely to be acquired in the contexts in which students learn the substantive content of their disciplines and vocations. Formative assessment may be part of the solution to the issue of encouraging the development of generic skills.

Feedback is central to learning gains

Formative assessment necessarily involves feedback to learners about their developing skills, and for feedback to make an effective contribution to learning, students need to engage sequentially in learning tasks in which the feedback from one task feeds into their performance for the next task.

Self-assessment with feedback leads to enhanced learning

Engaging students in assessment appears to increase the learning gains that arise from assessment. In many assessment regimes, the assessment is designed and set by teachers; students generate responses to the assessment prompts; and teachers grade those responses. In such arrangements, students' roles are relatively passive. Boud (1995b; 2002) argued a case for the greater use of self-assessment. Students complete assessment tasks and grade their own work against established performance criteria. In the act of grading their own work, students are assumed to focus more closely on the performance criteria, to identify and compare salient features of their work to the standards and therefore to learn more effectively.

Feedback on students' interpretations of performance standards is expected to improve students' interpretation and application of performance standards.

A common framework is desired for performance standards

In order to establish consistent standards across a diversity of tasks, a general framework for setting performance standards is required. Several possibilities exist. The structure of Bloom's taxonomy has been evaluated (see, e.g., Hill & McGaw, 1981). The SOLO Taxonomy (Biggs & Collis, 1982) is believed to provide a fruitful framework. That framework is applied jointly with an information-processing model of problem solving to provide construct-based assessment standards.

Generic skills assessment can be integrated with substantive course assessment

Wiggins (1998) proposed that the main purpose of assessment was to improve learning in "worthwhile tasks that require enduring knowledge and skill… in authentic tasks" (p. 10). Generic skills are believed to be constructs that embody worthwhile knowledge and skill. Their development and demonstration in authentic tasks is believed to be both desirable and feasible. The main purpose identified in this study is the development of problem solving skills in contexts defined by existing learning activities.

Developing a problem solving assessment tool

The first two research questions address the first of two major issues identified in a review of the implementation of generic skills, namely deficiencies in the definition of generic skills. The second of the major issues is the assessment of generic skills. A review of approaches to assessment suggests that there are ways in which generic skills, once well-defined, can be assessed. This position leads to the third research question.

Research question 3: Can valid and reliable measures of performance in generic competences be developed?

The operational definition of problem solving, developed through a review of models of problem solving (see above), includes five processes. These processes are:

- apprehending, identifying and defining the problem,
- planning an approach to the problem including selecting strategies,
- carrying out the chosen plan,

- monitoring progress towards the goal, and
- reflecting on the effectiveness of the solution attempt.

In the Problem-Solving Assessment instrument (see Appendix 5), these processes are labelled 'representation,' 'planning,' 'execution,' 'monitoring' and 'reflection.'

For each process, a set of indicators is identified and for each indicator, performance standards are proposed (Sadler, 2005). The performance standards are based on the levels of cognitive demand described in the SOLO taxonomy (Biggs & Collis, 1982). Thus the instrument incorporates five problem solving processes; each is operationalised through several indicators, and performance on each indicator may be observed at one of the described performance standards.

An assessment protocol is developed with five key attributes. First, students undertake problem solving assessments using the tasks that are the normal tasks set for the assessment of the substantive subject content. Second, as students complete the tasks, they use the Problem Solving Assessment (PSA) tool and assess their own performance against the standards that are specified in the assessment tool. Third, students submit their problem solving self-assessment with the materials that form part of their routine assessments. Fourth, as lecturers assess students' submissions for the substantive course content, they also assess students' problem solving performance using the PSA instrument. Fifth, as a consequence of their assessment of students' work, lecturers provide feedback to students on their problem solving performance.

This component of the research focuses on establishing the measurement properties of the assessment instrument. The Rasch method is the principal technique used in this analysis. The instrument has useful measurement properties. The item separation index is 0.78, indicating that the item difficulties are adequately distributed along the measurement scale. The distribution of item difficulties is compared with the distribution of person abilities. This shows that the instrument is not as well targeted as it could be, but the participants are volunteers and there may be a bias towards high ability students. Further, the instrument might have acted as a prompt and encouraged the use of the suggested processes. These possibilities can account for the observed mis-targeting. The problem-solving process indicators show good fit and most performance levels within the indicators show the ordering that is expected, given that the performance standards are based on the ordered SOLO taxonomy.

A validation study is conducted with a separate group of students. Rasch analysis of the responses of that group showed good measurement properties, although there were some differences in item locations between the two groups.

Students' reactions to the assessment, revealed through an online survey, were generally favourable, although their responses to prompts about the clarity of the process were less favourable.

It appears that the Problem Solving Assessment instrument has adequate measurement properties and that it is feasible to assess students' problem solving performance on authentic tasks using this tool.

Simply reporting student achievement is one aspect of the investigation. A second possibility, that formative assessment accompanied by feedback on performance, may lead to improved performance is apparent in the literature on assessment (Black & Wiliam, 1998a). This possibility leads to the fourth research question.

Research question 4: Can problem solving performance be enhanced by repeated assessment and feedback cycles?

As a result of the analyses of data gathered from the administration of the Problem Solving Assessment (PSA) instrument and feedback from students, a revised version of the instrument is developed for use in the second study. This version is called the Problem Solving Skills Assessment Tool (PSSAT) and has 18 rather than 21 indicators. Support materials are also developed and the PSSAT and the teacher and student support materials are shown in Appendix 8.

In order to gather independent data on students' generic skills performance before the planned sequence of self-assessment, validation and feedback, the Tertiary Skills Assessment (TSA) instrument was administered. This 55-item multiple-choice instrument included three scales, namely interpersonal understanding, critical thinking and problem solving. The TSA has very good measurement properties, with all items showing good fit and scale reliabilities are 0.75, 0.90 and 0.96 respectively.

The revised assessment instrument, the PSSAT was recalibrated in the study. It has good measurement properties. Some differences were found in the ease with which lecturers and students endorsed indicators. Students' self-assessments were slightly more favourable than teacher assessments of metacognitive indicators while teachers' judgements on lower level cognitive indicators were slightly more favourable than those made by students. These differences suggest that students might not have been as aware of the higher levels of performance being sought through the metacognitive indicators and may need further instruction about these indicators.

The main purpose of this component of the study is to test the proposition that repeated assessment and feedback cycles leads to enhanced performance. Rasch-scaled assessment scores for teacher judgements and student self-assessments are available for three test occasions, the minimum number required to be able to estimate a growth curve (Singer & Willett, 2003, pp. 9-10).

As a consequence of a reassignment of lecturers' workloads, a replacement assessor was used for the third assignment. This introduced a new variable and it is not sensible to evaluate change in lecturer-assessed performance over time. A series of statistical models is developed using the self-assessment scores as the criterion measure. Analysis of these scores does suggest a modest gain over the three assessment occasions. The simplest model with the best explanatory power is one in which the initial problem solving performance (the intercept) varies between individuals while the change in assessment scores over time (the slope) does not vary between individuals. This is Model 3A in Chapter 8 and is represented by the following Level 1 (occasion) and Level 2 (individual) equations.

$$\mathrm{Self}_{ij} = \beta_{0j} + \beta_{1}.\mathrm{Occ}_{i} + e_{ij}$$

$$\beta_{0j} = \beta_{00} + \beta_{01}.\mathrm{TSAIP}_{j} + u_{0j}$$

These equations may be read as follows. The self-assessed score of an individual (j) on occasion (i) depends upon the intercept for that individual (β_{0j}) and a common slope (β_{1}) for all individuals by occasion (i) with a residual term (e_{ij}). At Level 2, the intercept for the individual (β_{0j}) is predicted by the overall intercept for all individuals (β_{00}) with an adjustment (β_{01}) based upon the individual's score on the TSA interpersonal understanding scale, with a Level 2 residual term (u_{0j}).

Models, in which the estimated slope value for occasion in Level 1 (β_{1j}) is regressed on predictors, require the estimation of additional parameters, but the reduction in the likelihood function is shown to be non-significant. The Level 2 residual terms are not significantly different from zero. That is, the model indicates there is no firm evidence in this data set for variations in slopes between individuals.

The change in problem solving performance between occasions is 5.05 units (with a standard error of 1.11 units) on the problem solving performance scale. (This scale is calibrated to a mean of 50 units and a standard deviation of 15 units). The growth between occasions is therefore approximately one-third of a standard deviation. Given that there are three assessment occasions with two intervals, the mean growth over the assessment period is 10.1 units or two-thirds of a standard deviation. In general, effect sizes of this magnitude are regarded as satisfactory, although no comparison studies are known in which change in problem solving scores are estimated.

The research question is answered in the affirmative. Evidence is found that self-assessed problem solving performance can be enhanced through repeated assessment and feedback cycles.

Discussion

The research design has certain limitations. The decision to conduct the research using intact classes and existing assessment tasks in students' courses for the assessment of problem solving imposes limits on generalisations that may be made on the basis of the research findings. In addition, since the inception of this research, other methods for the assessment of higher order skills, such as scientific literacy and problem solving have been implemented. It may be fruitful to contrast the methods developed in this research with those alternative approaches.

Limitations

Facets of measured problem solving performance

The development and testing of the problem solving tool (see Chapter 7) are undertaken in a naturalistic setting, although it is not a naturalistic ethnographic study (Burns, 1997, p. 301), but nor is it an experimental investigation with experimental and control groups. The decision to use a naturalistic setting, but to inject into it a new assessment domain and approach, was taken at the commencement of the study in order to assess the feasibility in realistic settings of assessing problem solving using an instrument and a set of processes designed to render the *in situ* assessment objective. The study, therefore, is conducted within intact classes being taught by the teachers and lecturers who routinely take those classes. The tasks used for the assessment of problem solving are a selection of the tasks that are used routinely for the assessment of the substantive content in those courses. Students' responses are marked by their regular teachers for their substantive content and for the application of problem solving processes. In an experimental design, assessment tasks, students and markers would be linked in a crossed design with sufficient overlap to ensure that parameters for task difficulty, student ability and marker severity could be jointly estimated. In such a design, it would be possible to evaluate the influence of task variability on students' use of problem solving processes.

Tasks may be more or less difficult with respect to their substantive content, but this aspect of difficulty may not be related linearly to problem solving difficulty. Very easy tasks may not be challenging enough to require the application of problem solving processes. On the other hand, tasks that are very demanding of content knowledge may lead to high cognitive load and inhibit the application of problem solving processes (Sweller, 1989). In the current study, teachers are asked to select tasks that provide learners with opportunities to develop and demonstrate their problem solving skills. The design of this study does not permit testing of whether the chosen tasks do uniformly provide the desired opportunity to activate problem solving processes.

A related limitation of the design is that it does not enable rater severity to be assessed. Rater effects are well known (Andrich, 1997b; Linacre, 1997) and they can be modelled by using the rater as a facet in an extension of basic Rasch models (dichotomous, rating scale or partial credit models). Wilson (2001), however, argued that the facets model (Linacre, 1989) is inadequate and that a further refinement of the model is required to account fully for rater effects.

An implication of this limitation is that precision of the reported measures of problem solving performance may be under-estimated. The item and threshold parameter estimates, as well as separating influences on performance in learner-item interactions, may also echo variations in task affordance (for problem solving) and rater severity. It is desirable that more precise item and person parameter estimates are available. It is also desirable that the influences of different tasks and variations in rater severity are available. If variations in task affordances can be traced to characteristics of those tasks, then it may be possible for instructors to select optimal tasks for learners to develop and demonstrate their problem solving abilities. It is important to know the likely variations in rater severity. If this is high, it may be necessary to improve the support materials provided to instructors. It may also enable interested stakeholders to judge the reliability of reported problem solving performance. High expected rater variability indicates that comparisons cannot be made reliably across different institutions and courses. If rater variability is low, greater confidence can be placed in reported problem solving performance.

A future study based on the one presented in Chapter 7 would use a limited set of tasks. They would be undertaken by students who would be expected to undertake those tasks routinely in their studies. Students' responses would be graded on both their substantive subject content and their problem solving processes by at least two raters from a panel who were selected because of their familiarity with the domain and who were trained in the use of the problem solving assessment tool. The assignment of students, tasks and raters would ensure that each student would undertake a selection of tasks, that each rater would mark a selection of tasks, that each student would be graded by a selection of raters and that the design would include no unique pairings of raters, tasks and students so that the three facets could be independently estimated on both the substantive subjects content and the deployment of problem solving processes. In addition to estimating student ability, task difficulty and rater severity, the two dimensions of task difficulty (substantive content and the application of problem solving processes) could be compared to see whether there is a relationship between them, and if so of what type. It is anticipated that there may be no relationship, a linear one, or a curvilinear one. Moreover, the relationship may be different for groups of differing abilities.

Teacher judgment of problem solving performance

Growth in problem solving performance (see Chapter 8) is expected to occur through repeated assessment and feedback, scaffolded for both the instructor and the learner through the use of the Problem Solving Skills Assessment Tool. Two measures are reported for each assessed activity, namely the students' self-assessed performance and the teachers' judgment of that performance. In the study reported in Chapter 8, two unanticipated situations arose. The cycle of assessment and feedback occurred on four occasions, but due to an administrative error, students' scripts, including their self-assessed and their teacher validated assessment forms were returned to them without being made available to the researcher. This left data for three assessment occasions, and while this is sufficient to estimate robust change over time, it limits the statistical model to a linear one. It is possible that the change trajectory is non-linear and that there may be a diminishing return to the assessment and feedback cycle beyond some limiting value. The imputation of such a trajectory is not possible with only three observations and it is desirable that a replication is undertaken in which four or more assessment and feedback cycles are used. It would be useful to know how many assessment and feedback cycles are needed to generate learning gains.

A second unanticipated difficulty arose in the study. Because of workload pressure, the lecturer for the course had to engage a tutor to mark the final assignment, including its substantive content and the problem solving assessment. While the researcher provided all lecturer support materials and briefed the tutor on the process, it was possible, and even likely it appeared given the distribution of marks over the three assessment occasions, that the tutor was a more severe marker than the original course lecturer. This possibility prevents the reliable estimation of growth in problem solving as judged by an instructor. For reports of problem solving performance to have credibility among stakeholders (e.g., potential employers) or for such assessments to be aggregated and used as institutional performance indicators as the Mayer Committee had recommended (Mayer Committee, 1992, p. 55), instructor rather than student judgments ought to be used. In the study reported in Chapter 8, the replacement of the marker limits the reliability of the teacher judgment data that are available for analysis.

Implications for practice and further investigation

The structure of problem solving

Problem solving is construed as being a goal directed activity that employs sets of latent cognitive and metacognitive competences. These processes are operationalised as five latent problem solving processes, namely identifying, planning, executing, monitoring and reflecting. In turn, these processes are manifested in a set of indicators and for each indicator, a set of performance levels is proposed. The performance levels are based on the SOLO taxonomy and are thought to represent levels of cognitive ability in problem solving.

Two distinct possibilities for the structure of problem solving capability emerge. Problem solving may be a continuous trait and individuals may develop over time along a continuum or it may be a trait characterised by having discrete developmental stages. If it is continuous, the SOLO-based performance levels may simply be a convenient way of identifying ordered locations along that continuum. Individuals and indicators both need to be distributed more or less continuously along that scale. If problem solving capability develops in stages, then, assuming there is sufficient variation in student ability and indicator difficulty, both individual and indicator locations need to be clustered along the scale. According to Wilson (1997), persons and items should show both clustering and ordering. It may be noted that in the study reported in Chapter 7, the data are cross-sectional rather than longitudinal and the existence of developmental levels is inferred from individuals demonstrating various levels of performance.

The structure of problem solving, that is whether it is continuous or staged, is not investigated in this thesis. Neither of the studies reported in this thesis was intended to address this question, and the data are not sufficiently precise to answer it. The present study set out to develop and test a method for the assessment of problem solving in real instructional contexts. The contexts imposed constraints on the design of the studies. It was not possible, for example to implement a balanced block design in which students, tasks and assessors were crossed. A consequence of the design is that facets that may influence performance judgments, such as task difficulty and marker severity, could not be separated. In a fully crossed design, in which all tasks are attempted by all students and their responses are graded by all assessors, the dimensions may be separated (R. L. Thorndike & Thorndike, 1997). In a stochastically crossed design, these facets could be separated using software such as Conquest (Wu et al., 1998).

Alternative assessments and measures of problem solving

Since the conception and execution of this research, two reports have appeared on the international assessment of problem solving and the generation and reporting of problem-solving performance measures from that assessment (OECD, 2004; Reeff et al., 2005). They used similar definitions of problem solving based on the same sources as used in the current study. The OECD definition used in the Programme for International Student Assessment (PISA) for 15 year-olds was:

> …an individual's capacity to use cognitive processes to confront and resolve real, cross-disciplinary situations where the solution path is not immediately obvious and where the content area or curricular areas that might be applicable are not within a single subject area of mathematics, science or reading. (p. 26)

The definition of problem solving used in the assessment of adults for the international Adult Literacy and Life Skills (ALLS) survey was:

> Problem solving is goal-directed thinking and action in situations for which no routine solution procedure is available. The problem solver has a more or less well-defined goal, but does not immediately know how to reach it. The incongruence of goals and admissible operators constitutes a problem. The understanding of the problem situation and its step-by-step transformation, based on planning and reasoning, constitute the process of problem solving. (Reeff et al., 2005, p. 197)

These definitions derive from the sources that are used in the present study. However, there are differences between the present research and the two large scale international studies. A key difference results from the constraints of a large trans-national project. In the research described in this thesis, tasks that are routinely undertaken in students' scheduled classes are used whereas, in large-scale studies tasks must be selected that are believed to be 'cross-disciplinary.'

The PISA study identified three types of problems namely, decision making, systems analysis and design and trouble shooting (OECD, 2004, pp. 28-29). For each type, a similar list of problem solving processes were identified namely, understanding the situation or given information, identifying relevant aspects of the problem, representing relationships and alternative approaches, deciding among alternatives, checking and evaluating the decision, and communicating or justifying the solution (OECD, 2004, p. 29). In the ALLS study, “analytical problem solving” was the focus and was differentiated from the “social, emotional and creative” aspects of problem solving (Reeff et al., 2005, p. 201).

In the PISA study, three levels of the problem solving scale were recognised (OECD, 2004, pp. 28-31). Individuals who performed at the highest level (Level 3) were described as “reflective, communicative problem solvers” and were said to be capable of coping with multiple relationships and during their attempts, to have iterated between their solutions and the conditions of the problem situations. Although no reference was made to the SOLO taxonomy, this description is suggestive of the multi-structural and possibly extended abstract levels of the SOLO taxonomy. Level 2 problems solvers were described as “reasoning, decision-making problem solvers” capable of dealing with well-defined problem situations. The descriptions of this performance level allude to Bloom’s taxonomy through the use of terms such as ‘analyse’ and ‘synthesise’ information. “Basic problem solvers” were described as being able to transform given information and able to locate and retrieve information relevant to aspects of the tasks, but were depicted as not being able “to deal successfully with multi-faceted problems.” This description again evokes the SOLO taxonomy and in particular its uni-structural level. A fourth performance level was also described, albeit as a residual level in that those assigned to it had failed to reach Level 1 performance standard.

It may be worth noting, in the context of generic skills assessment, that the highest problem solving level recognised in the PISA assessments involved reflecting on and communicating problem solutions. In the example problem solving tasks presented (OECD, 2003, pp. 161-170), communication did not appear in the scoring rubrics, although it was identified as one of the problem solving processes (OECD, 2003, p. 171). Communication is recognised as a distinct generic skill in many generic skills schemes, so the reference to communication as a performance level differentiator in problem solving may indicate that construct validity was not a central concern in the construction of items. Alternatively, it may reflect the recognition that generic skills are not used as discrete capabilities but are deployed as orchestrated compositions (see, Hager & Beckett, 1999).

Four levels of problem solving proficiency were postulated for the ALLS problem solving scale (Reeff et al., 2005, p. 202). These levels were based on “post-Piagetian theories of cognitive development.” Level 1 problem solvers were said to be limited to solving problems that depended on the application of domain-specific knowledge and procedures. Level 2 problem solvers could apply "rudimentary systematical [*sic*] reasoning" to one-dimensional tasks using "concrete logical operations." Level 3 problem solvers were able to use "formal operational procedures" and to deal with problems that were multi-dimensional or ill-defined. At Level 4, problem solvers were "capable of grasping a system of problem states and possible solutions as a whole." These problem solvers were able to explain how they arrived at their solutions and applied both critical thinking and metacognitive skills.

It may be noted that despite the "post-Piagetian" theoretical basis, Piagetian-like descriptions of performance were used, particularly for Levels 2 and 3.

Three aspects of the problem solving assessment evident in the PISA and ALLS studies warrant comment. First, they are based on a similar cognitive model of problem solving to the one used in the present study. Second, both use procedural definitions of problem solving activity, and this too is similar to the approach taken in the present study. In this sense they differ from many previous problem solving assessments that used unique rubrics for each assessment task and from which an underlying level of cognitive problem-solving ability was inferred (Herl et al., 1999).

Third, and of considerable interest, is the basis upon which performance levels were inferred. In the PISA study, no reference was made to the Bloom or SOLO taxonomies nor was any reference made to Piagetian perspectives even though the level descriptors are reflective of the SOLO terminology. The problem tasks were selected by expert groups who might have used such frameworks, but none was specifically identified in the text (OECD, 2003). The ALLS assessment referred to post-Piagetian theories of cognition, but descriptors appear to reflect both Piagetian levels of development and Bloom's taxonomic levels. In the PISA and ALLS assessments, the purpose is summative. Individual measures of performance are recorded and are aggregated to country levels in order to estimate country performance means. These data are then used in cross-country comparisons. In these studies, ensuring accurate grading by raters is vital, so strict rubrics are required for all selected items to ensure consistency of assessment within and between countries.

A key difference between the study reported in this thesis and the PISA and ALLS is assessment purpose. In the present study, an assessment and feedback loop is used as a driver for learning. The large-scale assessments serve a summative and a systems-evaluation purpose. It is hoped that the approach taken in the current study may complement larger scale assessment such as the PISA and ALLS studies so that there is a method of enhancing student performance through the use of feedback and then of assessing it objectively.

In the international context, a new round of assessments of higher order abilities is under development, namely the OECD sponsored Programme for the International Assessment of Adult Competencies (PIAAC) (Schleicher, 2008). It is planned that the methods used in this project will build upon those used in the ALLS project.

Testing assessment frameworks

The SOLO taxonomy (Biggs & Collis, 1982) is selected as the basis for distinguishing performance levels in the present study. The original version of Bloom's taxonomy is evaluated as an alternative, but the SOLO framework is preferred. There is a substantial literature on Bloom's taxonomy, as it was reviewed and later refined (L. W. Anderson & Krathwohl, 2001; Hill, 1987; Kreitzer & Madaus, 1994). The present study does not subject the SOLO taxonomy to the same degree of scrutiny that was applied to Bloom's taxonomy (Hill, 1987; Kropp & Stoker, 1966). It is suggested that it would be preferable to evaluate the SOLO taxonomy using modern measurement procedures, in particular, the Saltus extension of the Rasch model (Wilson, 1989b), and using cognitive tasks from domains about whose structure there is more general agreement than is evident in relation to generic skills. If the ordered structure of the SOLO taxonomy is confirmed, then it may be applied in the multi-faceted design that is suggested above to check the influences of task variability and rater severity on the measurement of problem solving. Further, it may be desirable to compare the revised version of Bloom's taxonomy and the SOLO taxonomy using a common set of tasks. Performance levels of responses to the tasks could be scored using both taxonomies.

Conclusions

Initiatives to promote generic skills have occurred in Germany since 1974 (Mertens, 1974), in Australia since 1985 (Quality of Education Review Committee, 1985) and continue to the present. Many of these initiatives are reviewed in Chapter 2. Since the conception of this research in 2000 and its execution between 2002 and 2005, many reports have been published on the assessment and reporting of generic skills achievement in the school, vocational education and training, and higher education sectors. Some reports have canvassed a variety of alternatives to assessment and reporting (e.g., Ratio Pty Ltd & Down, 2003) while others have suggested that the solution to this matter is simple and can be resolved by requiring students to develop portfolios (Allen Consulting Group & National Centre for Vocational Education Research, 2004).

The review of generic skills schemes in Chapter 2, the examination of psychological constructs that underlie generic skills and Chapter 3 and the exploration of assessment purposes and methods in Chapter 4 indicate that the problem is not a simple one and that simple solutions are not available.

The two studies reported in Chapters 7 and 8 suggest that it is possible to assess generic skills rigorously and that their assessment can be used as a means for enhancing student performance. These studies also suggest that the development of generic skills requires some additional effort by both teachers and learners.

Investigations into generic skills continue

Despite the numerous reports that have been published in Australia, most funded by government departments (see Chapter 2 for a discussion and Tables 7 and 8 for a summary of these reports), governments continue to seek solutions to the issues of generic skills development, assessment and reporting.

The Department of Education, Employment and Workplace Relations (DEEWR) commissioned a report on options for assessing and reporting the employability skills achievement of senior secondary students (Matters & Curtis, 2008). Later, it commissioned further work on the recognition of employability skills of students who are unlikely to complete Year 12 at school (Sweet, 2008). Most recently, the House of Representatives Standing Committee on Education and Training again addressed the issue of recognising students' employability skills, recommending that a passport be developed to record the employability skills achieved by young people through paid and volunteer work and recommending that a code of practice be developed for supervisors "outlining their responsibilities in assisting young people" in the development and documentation of employability skills (2009, Recommendations 2 and 3, p. 62).

Efforts to assess generic competencies continue in other countries and internationally. The Organisation for Economic Cooperation and Development (OECD), through its Education Directorate, is again investigating generic competences, having supported the DeSeCo project between 1998 and 2003. It is undertaking the Assessment of Learning Outcomes in Higher Education (AHELO) project, a feasibility study into the assessment of learning outcomes in higher education (Nusche, 2008). The outcomes in its remit are broader than generic competences and include domain-specific knowledge, contextual features of learning and the 'value-added' contribution of higher education programs to the learning of individuals. The generic competences strand of the feasibility study will be assessed using the Collegiate Learning Assessment (Benjamin et al., 2009). It is interesting to note that the CLA used a selection of common items to assess students' competences, but used generalised scoring rubrics. This is contracted with the current study in which assessment tasks are drawn from those routinely used in the students' course.

It would appear that, despite the advice contained in the many commissioned reports, no tractable solution to the issues of developing, assessing, certifying and reporting generic skills achievement has been found. It is hoped that the research reported in this thesis makes a contribution to policy debate about this issue.

The issues surrounding the learning and assessment of generic skills will not go away and there appears to be no widely accepted solution. This research shows that the central issues are the definition and assessment of generic skills. Other important issues, such as how student performance of these skills may be reported, may become much more tractable after clear definitions of these constructs are accepted and following assessment approaches that are most compatible with the particular purposes for which the skills are required.

References

21st Century Workforce Commission. (2000a). *Building America's 21st century workforce*: National Alliance of Business and US Department of Labor.

21st Century Workforce Commission. (2000b). Skills for a new century: a blueprint for lifelong learning.   Retrieved 6 June, 2000, from http://www.vpskillsummit.org/blueprint.asp

Abrahamsson, K. (1999). *Vocational education and training in Sweden*. Luxembourg: CEDEFOP.

AC Nielsen Research Services. (2000). *Employer satisfaction with graduate skills. Research report*. Canberra: Department of Education, Training and Youth Affairs.

Acock, A. C. (2005). Working with missing values. *Journal of Marriage and Family, 67*, 1012-1028.

Adams, R. J., & Khoo, S. T. (1993). *Quest: the interactive test analysis system (manual)*. Melbourne: Australian Council for Educational Research.

Adams, R. J., & Khoo, S. T. (1999). Quest: the interactive test analysis system (Version for PISA) [Rasch analysis software]. Melbourne: Australian Council for Educational Research.

Airasian, P. W. (1994a). *Classroom assessment* (2nd ed.). New York: McGraw Hill.

Airasian, P. W. (1994b). The impact of the taxonomy on testing and evaluation. In L. W. Anderson & L. A. Sosniak (Eds.), *Bloom's taxonomy. A forty year retrospective. Ninety-third yearbok of the National Society for the Study of Education. Part II* (pp. 82-102). Chicago: National Society for the Study of Education.

Alexander, P. A. (2003a). Can we get there from here? *Educational Researcher, 32*(8), 3-4.

Alexander, P. A. (2003b). The development of expertise: the journey from acclimation to proficiency. *Educational Researcher, 32*(8), 10-14.

Allen Consulting Group. (1999). *Training to compete. The training needs of Australian industry. A report to the Australian Industry Group*. North Sydney: Australian Industry Group.

Allen Consulting Group. (2004). *Development of a strategy to support the universal recognition and recording of employability skills: Issues paper*. Canberra: Allen Consulting Group.

Allen Consulting Group. (2006a). *Assessment and reporting of employability skills in Training Packages*. Canberra: Department of Education Science and Training.

Allen Consulting Group. (2006b). *World class skills for world class industries. Employers perspectives on skilling Australia*. Canberra: Australian Industry Group.

Allen Consulting Group, & National Centre for Vocational Education Research. (2004). *Development of a strategy to support the universal recognition and recording of employability skills. Directions Paper*. Melbourne: Allen Consulting Group.

Allerup, P. (1997). Rasch measurement theory. In J. P. Keeves (Ed.), *Educational research, methodology, and measurement: an international handbook* (2nd ed., pp. 863-874). Oxford: Pergamon.

Allix, N. M. (2000). The theory of multiple intelligences: a case of missing cognitive matter. *Australian Journal of Education, 44*(3), 272-293.

Alverno College Faculty (Ed.). (1994). *Student assessment-as-learning at Alverno College*. Milwaukee, Wisconsin: Alverno College Institute.

Alverno College Institute. (1996). Alverno College ability-based learning program. Milwaukee, Wisconsin: Alverno College.

American Educational Research Association, American Psychological Association, & National Council on Measurement in Education. (1999). *Standards for educational and psychological testing*. Washington, DC: American Educational Research Association.

Anderson, J. R. (1995). *Cognitive psychology and its implications* (4th ed.). New York: W H Freeman.

Anderson, J. R., Reder, L. M., & Simon, H. A. (1996). Situated learning and education. *Educational Researcher, 25*(4), 5-11.

Anderson, J. R., Reder, L. M., & Simon, H. A. (1997). Situative versus cognitive perspectives: Form versus substance. *Educational Researcher, 26*(1), 18-21.

Anderson, L. W. (1997). Attitudes, measurement of. In J. P. Keeves (Ed.), *Educational research, methodology, and measurement: an international handbook* (2nd ed., pp. 885-895). Oxford: Pergamon.

Anderson, L. W., & Krathwohl, D. R. (Eds.). (2001). *A taxonomy for learning, teaching, and assessing. A revision of Bloom's taxonomy of educational objectives* (Abridged ed.). New York: Addison Wesley Longman.

Andrich, D. (1978). A rating formulation for ordered response categories. *Psychometrika, 43*(4), 561-573.

Andrich, D. (1982). An index of person separation in latent trait theory, the traditional KR-20 index, and the Guttman scale response pattern. *Educational Research and Perspectives, 9*(1), 95-104.

Andrich, D. (1988). *Rasch models for measurement*. Newberry Park, CA: Sage Publications.

Andrich, D. (1997a). Essays: equating of marks. In J. P. Keeves (Ed.), *Educational research, methodology, and measurement: an international handbook* (2nd ed., pp. 830-836). Oxford: Pergamon.

Andrich, D. (1997b). Rating scale analysis. In J. P. Keeves (Ed.), *Educational research, methodology, and measurement: an international handbook* (2nd ed., pp. 874-880). Oxford: Pergamon.

Andrich, D. (1997c). Thurstone scales. In J. P. Keeves (Ed.), *Educational research, methodology and measurement: An international handbook* (2nd ed., pp. 819-822). Oxford: Pergamon.

Andrich, D. (1999). Rating scale analysis. In G. N. Masters & J. P. Keeves (Eds.), *Advances in measurement in educational research and testing* (pp. 110-121). Amsterdam: Pergamon.

Andrich, D. (2004). Controversy and the Rasch model: a characteristic of incompatible paradigms? In E. V. Smith & R. M. Smith (Eds.), *Introduction to Rasch measurement* (pp. 143-200). Maple Grove, Mn: JAM Press.

Arbeitstelle für vergleichende Bildungsforschung, & European Training Foundation. (1999). The European House of Education: Education and Economy - A new Partnership. Background paper prepared for a meeting of European Ministers of Education. Retrieved 29 May, 2001, from http://www.etf.eu.int/etfweb.nsf/pages/vetdown/$file/backgroundpaper-en.doc

Assessment Reform Group. (1999). *Assessment for learning: Beyond the black box*. London: Association for Achievement and Improvement through Assessment.

Assessment Reform Group. (2002). *Assessment for learning: 10 principles. Research-based principles to guide classroom practice*. Cambridge: Assessment Reform Group.

Atkinson, R. C., & Shiffrin, R. M. (1968). Human memory: A proposed system and its control processes. In K. Spence & J. Spence (Eds.), *The psychology of learning and motivation* (Vol. 2). New York: Academic Press.

Atwood, M. E., & Polson, P. G. (1976). A process model for water jug problems. *Cognitive Psychology, 8*(2), 191-216.

Aungles, P., Karmel, T., & Wu, T. (2000, 30 May 2000). Demographic and social change: implications for education funding.   Retrieved 1 June, 2000, from http://www.detya.gov.au/highered/occpaper/00B/default.htm

Australasian Curriculum Assessment and Certification Authorities. (2003). *The employability skills framework report of mapping task undertaken by ACACA agencies*. Canberra: ACACA.

Australian Bureau of Statistics. (1997). *Aspects of literacy: assessed skill levels. Australia* (Cat No. No. 4228.0). Canberra: ABS.

Australian Bureau of Statistics. (2006a). *Education and training experience (2005)* (Cat. No. 6278.0). Canberra: Australian Bureau of Statistics.

Australian Bureau of Statistics. (2006b). *Schools Australia (2005)* (Cat. No. 4221.0). Canberra: Australian Bureau of Statistics.

Australian Chamber of Commerce and Industry, & Business Council of Australia. (2002). *Employability skills for the future*. Canberra: Department of Education, Science and Training.

Australian Council for Educational Research. (2000). *Graduate Skills Assessment. Test development and progress report*. Camberwell, Vic.: ACER.

Australian Council for Educational Research. (2001a). *Graduate Skills Assessment. Summary Report.* Canberra: DETYA, Higher Education Division.

Australian Council for Educational Research. (2001b). *Graduate Skills Assessment. Summary Report. GSA Exit 2000*. Camberwell, Vic.: ACER.

Australian Council for Educational Research. (2003). *Graduate Skills Assessment. Sample questions*. Melbourne: Australian Council for Educational Research.

Australian Education Council. (1989). Common and Agreed National Goals for Schooling in Australia (The Hobart Declaration on Schooling).   Retrieved 10 October, 2000, from http://www.mceetya.edu.au/mceetya/default.asp?id=11577

Australian Education Council. (1991). *Young people's participation in post-compulsory education and training. Report of the Australian Education Council Review Committee. (Mr Brian Finn, Chair)*. Canberra: Australian Government Publishing Service.

Australian Education Council. (1992). *Putting general education to work. The key competencies report. (Mr Eric Mayer, Chair)*. Canberra: Australian Government Publishing Service.

Australian Education Council. Mayer Committee. (1992). *Key competencies. Report of the Committee to advise the Australian Education Council and Ministers of Vocational Education, Employment and Training on employment-related Key Competencies for postcompulsory education and training*. Canberra: Australian Education Council and Ministers of Vocational Education, Employment, and Training.

Australian National Training Authority. (2002). *Fresh thinking about learning and learners. A blue sky project*. Melbourne: Australian National Training Authority.

Australian National Training Authority. (2003). *Employability skills in training packages. Policy engagement forum, 28 May*. Melbourne: Australian National Training Authority.

Australian National Training Authority. (2004). *Training package development handbook*. Melbourne: Australian National Training Authority.

Australian National Training Authority, & Ratio Pty Ltd. (2004). *Certificate I in industry pathways: Stage 2 - qualification outlines*. Melbourne: ANTA.

Australian Parliament. House of Representatives Standing Committee on Education and Training. (2004). *Learning to Work. Report on the inquiry into vocational education in schools*. Canberra: Parliament of the Commonwealth of Australia.

Australian Parliament. House of Representatives Standing Committee on Education and Training. (2009). *Adolescent overload? Report of the inquiry into combining school and work: supporting successful youth transitions*. Canberra: Australian Parliament House.

Australian Treasury. (2002). *Intergenerational report 2002-03* (Budget Paper No. 5). Canberra: Australian Treasury.

Australian Vice Chancellors Committee. (1995, September, 1995). Code of practice for the public disclosure of data from the Graduate Careers Council of Australia's Graduate Destination Survey and Course Experience Questionnaire. Retrieved 12 August, 1997, from http://www.avcc.edu.au/avcc/pubs/gdsceqcp.htm

Baker, E. L., O'Neil, H. F., & Linn, R. L. (1993). Policy validity prospects for performance-based assessments. *American Psychologist, 48*(1210-1218).

Baker, F. B. (2001). *The basics of item response theory* (2nd ed.). Iowa: ERIC Clearinghouse on Assessment and Evaluation.

Baltes, P. B., Staudinger, U. M., & Lindenberger, U. (1999). Lifespan psychology. Theory and application to intellectual functioning. *Annual Review of Psychology, 50*, 471-507.

Bandura, A. (1989). Perceived self efficacy in the exercise of personal agency. *The Psychologist, 2*(10), 411-424.

Bandura, A. (1997). *Self-efficacy. The exercise of control*. New York: W H Freeman.

Banks, G. (2005). *Comparing school systems across Australia*. Paper presented at the Schooling in the 21st century: Unlocking human potential. from http://www.pc.gov.au/__data/assets/pdf_file/0007/7657/cs20050928.pdf.

Bar-On, R. (2000). Emotional and social intelligence: Insights from the Emotional Quotient Inventory. In R. Bar-On & J. D. A. Parker (Eds.), *The handbook of emotional intelligence: Theory, development, assessment, and application at home, school, and in the workplace* (pp. 363-388). San Francisco, CA: Jossey-Bass.

Beaton, A. E. (1997). Missing scores in survey research. In J. P. Keeves (Ed.), *Educational research, methodology, and measurement: an international handbook* (2nd ed., pp. 763-766). Oxford: Pergamon.

Bejar, I. I. (1983). *Achievement testing. Recent advances*. Beverly Hills: Sage Publications.

Benjamin, R., Chun, M., Hardison, C., Hong, E., Jackson, C., Kugelmass, H., et al. (2009). *Returning to learning in an age of assessment. Introducing the rationale of the Collegiate Learning Assessment* (Monograph). New York: Council for Aid to Education.

Bereiter, C. (1999). Situated cognition and how to overcome it. In D. Kirschner & J. A. Whitson (Eds.), *Situated cognition: Social, semiotic, and psychological perspectives* (pp. 281-300). Hillsdale, NJ: Erlbaum.

Bereiter, C., & Scardamalia, M. (1986). Educational relevance of the study of expertise. *Interchange, 17*(2), 10-19.

Bereiter, C., & Scardamalia, M. (1987). *The psychology of written composition*. Hillsdale, N.J.: Lawrence Erlbaum.

Bezruczko, N. (2000). Thurstone: measurement for a new science. *Popular Measurement, 3*(1), 7-11.

Biggs, J. B., & Collis, K. F. (1982). *Evaluating the quality of learning: the SOLO taxonomy (Structure of the Observed Learning Outcome)*. New York: Academic Press.

Biggs, J. B., & Moore, P. J. (1993). *The process of learning* (3rd ed.). New York: Prentice Hall.

Black, P. (2004). *The nature and value of formative assessment for learning - Draft.*Unpublished manuscript, London.

Black, P., & Wiliam, D. (1998a). Assessment and classroom learning. *Assessment in Education: Principles, Policy and Practice, 5*(1), 7-74.

Black, P., & Wiliam, D. (1998b). Inside the black box: Raising standards through classroom assessment. *Phi Delta Kappan, 80*(2), 139.

Bloom, B. S., Engelhart, M. D., Furst, E. J., Hill, W. H., & Krathwohl, D. R. (Eds.). (1956). *Taxonomy of educational objectives. The classification of educational goals. Handbook 1: Cognitive domain*. New York: David McKay.

Blumberger, W. (1997). Latest developments concerning post-15 (16) vocational training in Austria. New problems, old strategies and many questions. Retrieved 06 June, 2001, from http://www.ibe.co.at/Kurzberichte/Post-16_Kurz.htm

Boden, M. (1997). Promise and achievement in cognitive science. In D. M. Johnson & C. E. Erneling (Eds.), *The future of the cognitive revolution* (pp. 55-67). New York: Oxford University Press.

Boman, P., Curtis, D. D., Furlong, M. J., & Smith, D. C. (2006). Cross-validation and Rasch analyses of the Australian version of the Multidimensional School Anger Inventory - revised. *Journal of Psychoeducational Assessment, 24*(3), 225-242.

Bond, T. G., & Fox, C. M. (2001). *Applying the Rasch model. Fundamental measurement in the human sciences*. Mahwah, NJ: Lawrence Erlbaum and Associates.

Borkowski, J. G., Carr, M., & Pressley, M. (1987). 'Spontaneous' strategy use: perspectives from metacognitive theory. *Intelligence*(January), 61-75.

Borthwick, A. (1993). Key competencies - uncovering the bridge between general and vocational. In C. Collins (Ed.), *Competencies. The competencies debate in Australian education and training* (pp. 21-34). Canberra: Australian College of Education.

Boud, D. (1995a). Assessment and learning: contradictory or complementary? In P. Knight (Ed.), *Assessment for Learning in Higher Education* (pp. 35-48). London: Kogan Page.

Boud, D. (1995b). *Enhancing learning through self assessment*. London: Kogan Page.

Boud, D. (2002). Assessment: have we forgotten something? In Australian National Training Authority (Ed.), *Fresh thinking about learning and learners: a blue sky project* (pp. 42-45). Melbourne: Australian National Training Authority.

Bowden, J. A., Hart, G., King, B., Trigwell, K., & Watts, O. (2000). Generic capabilities of ATN university graduates. Retrieved 28 August, 2000, from http://www.clt.uts.edu.au/ATN.grad.cap.project.index.html

Bowden, J. A., Hart, G., Trigwell, K., & King, B. (1998). Defining, developing and assessing generic capabilities of ATN graduates (Proposal). Retrieved 28 August, 2000, from http://www.clt.uts.edu.au/DEETYA_submission.pdf

Bowden, J. A., & Masters, G. N. (1993). *Implications for higher education of a competency-based approach to education and training*. Canberra: Department of Employment, Education, and Training.

Boyer, E. L., Altbach, P. G., & Whitelaw, M. J. (1994). *The academic profession: An international perspective*. Ewing, New Jersey: Carnegie Foundation for the Advancement of Teaching.

Brainerd, C. J. (1978). *Piaget's theory of intelligence*. New Jersey: Prentice Hall.

Bransford, J. D., Sherwood, R., Vye, N., & Rieser, J. (1986). Teaching thinking and problem solving. *American Psychologist, 41*(10), 1078-1089.

Bransford, J. D., & Stein, B. S. (1984). *The ideal problem solver: a guide for improving thinking, learning, and creativity*. New York: W H Freeman.

Bransford, J. D., & Stein, B. S. (1993). *The ideal problem solver*. New York: W H Freeman.

Brookhart, S. M. (2001). Successful students' formative and summative uses of assessment information. *Assessment in Education: Principles, Policy & Practice, 8*(2), 153-169.

Brown, A. (1998). Designing effective learning programs for the development of a broad occupational competence. In W. J. Nijhof & J. N. Streumer (Eds.), *Key qualifications in work and education* (pp. 165-186). Dordrecht: Kluwer.

Brown, A. L., & Palincsar, A. S. (1989). Guided, cooperative learning and individual knowledge acquisition. In L. B. Resnick (Ed.), *Knowing, learning, and instruction: Essays in honor of Robert Glaser* (pp. 393-451). Hillsdale, NJ: Erlbaum and Associates.

Bryk, A. S., & Raudenbush, S. W. (1992). *Hierarchical linear models: applications and data analysis methods*. Newbury Park, CA: Sage.

Bryk, A. S., Raudenbush, S. W., & Congdon, R. (2005). HLM for Windows. Hierarchical linear and nonlinear modeling with HLM/2L and HLM/3L. (Version 6.03) [Multilevel analysis software]. Chicago: Scientific Software International.

Burns, R. B. (1997). *Introduction to research methods* (3rd ed.). South Melbourne, Australia: Longman.

Business Higher Education Round Table. (2003). *Developing generic skills: examples of best practice. (B-HERT News 16, Special Issue)*. Melbourne: Business/Higher Education Round Table.

Calfee, R. C. (1994). Cognitive assessment of classroom learning. *Education and Urban Society, 26*(4), 340-351.

Callan, V. J. (2002). *Generic skills. Understanding VET teachers and students views.* Adelaide: NCVER.

Camp, C. J. (1992). The Problem Solving Inventory (Test 303). In J. J. Kramer & J. C. Conoly (Eds.), *Eleventh mental measurements yearbook* (pp. 699-701). Highland Park, NJ: Gryphon Press.

Candy, P. C., Crebert, G., & O'Leary, J. (1994). *Developing lifelong learners through undergraduate education. National Board of Employment, Education and Training. Commissioned Report 28.* Canberra: Australian Government Publishing Service.

Canto-Sperber, M., & Dupuy, J.-P. (1999). *Competencies for the good life and the good society: a philosophical perspective*. Neuchatel: NFSO, OECD.

Carnevale, A. P. (1991). *America and the new economy*. San Francisco: Jossey-Bass.

Carroll, J. B. (1993). *Human cognitive abilities. A survey of factor-analytic studies*. Cambridge: Cambridge University Press.

Chapman, J. D., & Aspin, D. N. (2001). Schools and the learning community: Laying the basis for learning across the lifespan. In D. Aspin, D. Chapman, Y. Sawano & M. Hatton (Eds.), *The international handbook of lifelong learning*. The Netherlands: Kluwer.

Charles, R. I., & Lester, F. K. (1984). An evaluation of a process oriented instructional program in mathematical problem solving in grades 5 and 7. *Journal of Research in Mathematics Education, 15*(1), 15-34.

Chi, M. T. H. (2000). Self-explaining expository texts: The dual processes of generating inferences and repairing mental models. In R. Glaser (Ed.), *Advances in instructional psychology* (pp. 161-238). Mahwah, NJ: Lawrence Erlbaum Associates.

Chi, M. T. H., Feltovich, P. J., & Glaser, R. (1981). Categorization and representation of physics problems by experts and novices. *Cognitive Science, 5*, 121-152.

Chi, M. T. H., Glaser, R., & Rees, E. (1982). Expertise in problem solving. In R. J. Sternberg (Ed.), *Advances in the psychology of human intelligence* (Vol. 1). Hillsdale, N.J.: Lawrence Erlbaum.

Chiappe, D., & MacDonald, K. (2005). The evolution of domain-general mechanisms in intelligence and learning. *The Journal of General Psychology, 132*(1), 5-40.

Choppin, B. H. (1997). Objective tests. In J. P. Keeves (Ed.), *Educational research, methodology, and measurement: an international handbook* (2nd ed., pp. 771-775). Oxford: Pergamon.

Cizek, G. J. (2001). Conjectures on the rise and call of standard setting: An introduction to context and practice. In G. J. Cizek (Ed.), *Setting performance standards. Concepts, methods, and perspectives* (pp. 3-17). Mahwah, NJ: Lawrence Erlbaum.

Clanchy, J., & Ballard, B. (1995). Generic skills in the context of higher education. *Higher Education Research and Development, 14*(2), 155-166.

Clandinin, D. J., & Connelly, F. M. (2000). *Narrative inquiry: experience and story in qualitative research*. San Francisco: Josse-Bass.

Clark, D. (2002, 21 May 2000). Learning domains or Bloom's taxonomy. Retrieved 20 June, 2003, from http://www.nwlink.com/~donclark/hrd/bloom.html

Clayton, B., Blom, K., Meyers, D., & Bateman, A. (2003). *Assessing and certifying generic skills. What's happening in VET?* (Research report No. NR1008). Adelaide: NCVER.

Cleary, M., Flynn, R., & Thomasson, S. (2006). *Employability skills: from framework to practice. An introductory guide for trainers*. Canberra: DEST.

Coaldrake, P., & Stedman, L. (1998). *On the brink. Australia's universities confronting their future*. St Lucia: Queensland University Press.

Confederation of British Industry. (1998, January, 1998). Qualified to compete. Creating a world class qualifications framework. Retrieved 18 January, 2001, from http://www.cbi.org.uk/home.shtml

Confederation of British Industry. (1999, July, 1999). Making employability work. An agenda for action. Retrieved 18 January, 2001, from http://www.cbi.org.uk/home.shtml

Confederation of British Industry. (2000a, May, 2000). The CBI response to the Foundation Degrees Consultation Paper. Retrieved 18 January, 2001, from http://www.cbi.org.uk/home.shtml

Confederation of British Industry. (2000b, July, 2000). Fact not fiction. UK training performance. Retrieved 18 January, 2001, from http://www.cbi.org.uk/home.shtml

Conference Board of Canada. (1992). *Employability Skills Profile: What are employers looking for?* Ottawa: Conference Board of Canada.

Conference Board of Canada. (2000a, May, 2000). Employability skills 2000+. Retrieved 17 January, 2001, from http://www.conferenceboard.ca/nbec/pdf/esp2000.pdf

Conference Board of Canada. (2000b). Employability skills toolkit for the self-managing learner. Retrieved 04 February, 2001, from http://www.conferenceboard.ca/education/learning-tools/toolkit.htm

Craik, F. I. M., & Lockhart, R. S. (1972). Levels of processing: A framework for memory research. *Journal of Verbal Learning and Verbal Behavior, 11*, 671-684.

Creswell, J. W. (2005). *Educational research. Planning, conducting, and evaluating quantitative and qualitative research* (2nd ed.). Columbus, OH: Pearson, Prentice Hall.

Crittenden, B. (1997). The proper role of universities and some suggested reforms. In J. Sharpham & G. Harman (Eds.), *Australia's future universities* (pp. 86-103). Armidale, NSW: University of New England Press.

Cummings, R., Ho, R., & Bunic, S. (1997, July). *Assessing and reporting of student generic qualities in higher education.* Paper presented at the Higher Education: Advancing International Perspectives, Adelaide.

Curtis, D. D. (1996). Using technology. In D. D. Curtis, R. Hattam, J. P. Keeves, M. J. Lawson, R. Reynolds, A. Russell, H. Silins & J. Smyth (Eds.), *Teaching and learning the key competencies in the vocational education and training sector: research support.* Adelaide: Flinders Institute for the Study of Teaching.

Curtis, D. D. (2004a). The assessment of generic skills. In J. Gibb (Ed.), *Generic skills in vocational education and training: research readings* (pp. 136-156). Adelaide: National Centre for Vocational Education Research.

Curtis, D. D. (2004b). Person misfit in attitude surveys: Influences, impacts and implications. *International Education Journal, 5*(2), 125-143.

Curtis, D. D., & Denton, R. (2003). *The authentic performance-based assessment of problem-solving*. Adelaide: NCVER.

Curtis, D. D., & Matters, G. (2008). *Assessment and reporting of employability skills of senior secondary students*. Paper presented at the ACER Research conference Touching the future: Building skills for life and work.

Curtis, D. D., & McKenzie, P. (2002). *Employability skills for Australian industry: Literature review and framework development*. Canberra: Department of Education, Science and Training.

Dawe, S. (2002). *Focusing on generic skills in training packages*. Leabrook, SA: NCVER.

Dawkins, J. (1988). *Higher education. A policy statement* (Ministerial statement). Canberra: Australian Government Publishing Service.

Dearing, R. (1996). *Review of qualifications for 16-19 year olds*. London: SCCA.

Dellitt, J. (1993). Key competencies and schooling in South Australia. In C. Collins (Ed.), *Competencies. The competencies debate in Australian education and training* (pp. 57-67). Canberra: Australian College of Education.

Delors, J. (1996). *Learning: the treasure within. Report to UNESCO of the International Commission on Education for the Twenty-first Century*. Paris: United Nations Educational Scientific and Cultural Organisation.

Department of Education Science and Training. (2006, 2008). Technical Note No 2. Student outcome indicators for the Learning and Teaching Performance Fund. Retrieved 31 August, 2008, from http://www.dest.gov.au/sectors/higher_education/policy_issues_reviews/key_issues/assuring_quality_in_higher_education/technical_note_2.htm

Down, C. (2000, November). Key competencies in training packages. Retrieved 1 May, 2001, from http://www.avetra.org.au/papers%202000/down1.PDF

Dreyfus, H. L., & Dreyfus, S. E. (1984). Putting computers in their proper place: analysis versus intuition in the classroom. *Teachers College Record, 85*(4), 578-601.

Dunn-Rankin, P., & Zhang, S. (1997). Scaling methods. In J. P. Keeves (Ed.), *Educational research, methodology, and measurement: an international handbook* (2nd ed., pp. 790-798). Oxford: Pergamon.

Dweck, C. S. (1986). Motivational processes affecting learning. *American Psychologist*, *41*(10), 1040-1048.

Dweck, C. S., & Henderson, V. L. (1989). *Theories of intelligence: background and measures.* Paper presented at the Biennial Meeting of the Society for Research in Child Development (ED 312 057), Kansas City, MO.

Dweck, C. S., & Leggett, E. L. (1988). A social-cognitive approach to motivation and personality. *Psychological Review, 95*(2), 256-272.

Earl, L. M. (2005). *Assessment as learning: Using classroom assessment to maximise student learning*. Thousand Oaks, CA: Corwin Press Inc.

Elliott, E. S., & Dweck, C. S. (1988). Goals: An approach to motivation and achievement. *Journal of Personality and Social Psychology*, *54*(1), 5-12.

Eltis, K., & Mowbray, B. (1997). Focusing on learning in New South Wales. In J. Lokan (Ed.), *Describing learning. Implementation of curriculum profiles in Australian schools 1986- 1996* (pp. 82-106). Camberwell, Vic: ACER Press.

Embretson, S. E. (1999). Issues in the measurement of cognitive abilities. In S. E. Embretson & S. L. Hershberger (Eds.), *The new rules of measurement. What every psychologist and educator should know* (pp. 1-15). Mahwah, NJ: Lawrence Erlbaum and Associates.

Embretson, S. E., & Hershberger, S. L. (Eds.). (1999). *The new rules of measurement. What every psychologist and educator should know.* Mahwah, NJ: Lawrence Erlbaum.

Embretson, S. E., Schmidt, & McCollam, K. M. (2000). Psychometric approaches to understanding and measuring intelligence. In R. J. Sternberg (Ed.), *Handbook of intelligence* (pp. 423-444). Cambridge, UK: Cambridge University Press.

Employment and Skills Formation Council. (1992). *The Australian Vocational Certificate Training System (Carmichael Report)*. Canberra: National Board of Employment, Education and Training.

Eraut, M. (1994). *Developing professional knowledge and competence*. London: The Falmer Press.

Eraut, M. (1998). Concepts of competence. *Journal of Interprofessional Care, 12*(3), 127-139.

Erebus Consulting Partners. (2003). *The acquisition of employment related skills by secondary school students. A final report to the MCEETYA Taskforce on Transition from School.* Canberra: Department of Education, Science and Training.

Ericsson, K. A., & Simon, H. A. (1980). Verbal reports as data. *Psychological Review, 87*(3), 215-251.

Esposto, A., & Meagher, G. (2007). *The future demand for employability skills and the implications for the VET system*. Paper presented at the 10th Annual Research Conference - Evolution, Revolution or Status Quo? , from http://www.avetra.org.au/publications/12-Esposto.pdf.

Estes, W. K. (1972). Reinforcement in human behavior. *American Scientist, 60*(6), 723-729.

Faure, E., Herrera, F., Kaddoura, A.-R., Lopes, H., Petrovsky, A. V., Rahnema, M., et al. (1972). *Learning to be. The world of education today and tomorrow.* Paris: UNESCO.

Ferguson, G. A. (1954). On learning and human ability. *Canadian Journal of Psychology, 8*, 95-112.

Field, L. (2001). *Employability skills required by Australian workplaces*. Sydney: Field Learning.

Flavell, J. H. (1976). Metacognitive aspects of problem solving. In L. B. Resnick (Ed.), *The nature of intelligence* (pp. 231-236). Hillsdale, NJ: Erlbaum.

Ford, M. E. (1986). For all practical purposes: criteria for defining and evaluating practical intelligence. In R. J. Sternberg & R. K. Wagner (Eds.), *Practical intelligence. Nature and origins of competence in the everyday world* (pp. 183-200). Cambridge: Cambridge University Press.

Fratczak-Rudnicka, B., & Torney-Purta, J. (2001). *Competencies for civic and political life in democracy*. Neuchatel, Switerzerland: OECD, SFSO.

Friedman, T. L. (2005). *The world is flat. The globalized world in the twenty-first century*. London: Penguin Books.

Furst, E. J. (1994). Bloom's taxonomy: philosophical and educational issues. In L. W. Anderson & L. A. Sosniak (Eds.), *Bloom's taxonomy. A forty year retrospective. Ninety-third yearbok of the National Society for the Study of Education. Part II* (pp. 28-40). Chicago: National Society for the Study of Education.

Gardner, H. (1983). *Frames of mind: the theory of multiple intelligences*. New York: Basic Books.

Gardner, H. (1993). *Frames of mind: the theory of multiple intelligences* (2nd ed.). New York: Basic Books.

Gay, G. (2001, July, 2001). The nature of metacognition. Retrieved 17 June, 2009, from http://www.lrdc.ca/contents/

Gick, M. L., & Holyoak, K. J. (1983). Schema induction and analogical transfer. *Cognitive Psychology, 15*(1), 1-38.

Glaser, R. (1963). Instructional technology and the measurement of learning outcomes: some questions. *American Psychologist*(18), 519-521.

Goldstein, H. (1999). *Multilevel statistical models*. London: Institute of Education.

Goody, J. (1999). *Education and competences. An anthropological perspective*. Neuchatel, Switzerland: DeSeCo Report, Swiss Federal Statistics Office.

Grace, N., & Ludwig, C. (1997). Use of National Statements and Profiles in Queensland schools. In J. Lokan (Ed.), *Describing learning. Implementation of curriculum profiles in Australian schools 1986- 1996* (pp. 140-165). Camberwell, Vic: ACER Press.

Greeno, J. G. (1985). Looking across the river: views from the two banks of research and development in problem solving. In S. F. Chipman, J. W. Segal & R. Glaser (Eds.), *Research and open questions* (Vol. 2).

Greeno, J. G. (1997). On claims that answer the wrong questions. *Educational Researcher, 26*(1), 5-17.

Greeno, J. G., Collins, A. M., & Resnick, L. B. (1996). Cognition and learning. In D. C. Berliner & R. C. Calfee (Eds.), *Handbook of educational psychology* (pp. 15-46). New York: Macmillan.

Griffin, P. (2001). Performance assessment of higher order thinking. Paper presented at AERA Conference, April 10, Seattle. Retrieved 28 June, 2001, from http://www.edfac.unimelb.edu.au/ARC/recentpubs.html

Griffin, P., & Gillis, S. (1999). *Assessing higher order competencies: a multi-source approach to CBA*. from http://www.aare.edu.au.

Gulikers, J. T. M., Bastiaens, T. J., Kirschner, P. A., & Kester, L. (2006). Relations between student pereptions of assessment authenticity, study approaches and learning outcome. *Studies in Educational Evaluation, 32*, 381–400.

Gustafsson, J.-E. (1984). A unifying model for the structure of intellectual abilities. *Intelligence, 8*, 179-203.

Gustafsson, J.-E. (1997). Models of intelligence. In J. P. Keeves (Ed.), *Educational research, methodology, and measurement: an international handbook* (2nd ed., pp. 937-944). Oxford: Pergamon.

Hager, P., & Beckett, D. (1999, 7 December 2000). Making judgments as the basis for workplace learning - preliminary research findings. Retrieved 23 December, 2000, from http://www.avetra.org.au/abstracts-papers/19_hager.pdf

Hager, P., Holland, S., & Beckett, D. (2002). *Enhancing the learning and employability of graduates: the role of generic skills*. Melbourne: The Business /Higher Education Round Table.

Hämäläinen, K., & Jakku-Sihvonen, R. (2000). *More quality to the quality policy of education*. Helsinki: National Board of Education.

Hambleton, R. K. (1989). *Principles and selected applications of of item response theory.* (3rd ed.). New York: Macmillan.

Hambleton, R. K. (1994). Standard setting in criterion-referenced testing. In T. Husen & T. N. Postlethwaite (Eds.), *The international handbook of education* (2nd ed., pp. 5721-5726). Oxford: Pergamon.

Hambleton, R. K., & Swaminathan, H. (1985). *Item response theory: principles and applications*. Boston: Kluwer-Nijhoff.

Hambur, S., Rowe, K., & Luc, L. T. (2002). *Graduate Skills Assessment. Stage one validity study*. Melbourne: Australian Council for Educational Research.

Hanushek, E. A., & Wößmann, L. (2007). *The role of education quality in economic growth*. Washington, DC: The World Bank.

Harwell, M. R., & Gatti, G. G. (2001). Rescaling ordinal data to interval data in educational research. *Review of Educational Research, 71*(1), 105-131.

Haste, H. (1999). *Competencies; psychological realities. A psychological perspective* (Electronic document). Neuchatel, Switzerland: OECD, DeSeCo Project.

Hatano, G., & Oura, Y. (2003). Commentary: Reconceptualizing school learning using insight from expertise research. *Educational Researcher, 32*(8), 26-29.

Haught, P. A., Hill, L. A., Nardi, A. H., & Walls, R. T. (2000). Perceived ability and level of education as predictors of traditional and adult problem solving. *Experimental Aging Research, 26*, 89-101.

Hayes, J. R. (1989). *The complete problem solver* (2nd ed.). Hillsdale, NJ: Lawrence Erlbaum.

Heppner, P. P., & Petersen, C. H. (1982). The development and implications of a personal problem-solving inventory. *Journal of Counseling Psychology, 29*(1), 66-75.

Herl, H. E., O'Neil, H. F., Chung, G. K. W. K., Bianchi, C., Wang, S.-L., Mayer, R. E., et al. (1999). *Final report for validation of problem-solving measures* (CSE Technical Report No. 501). Los Angeles: Center for the Study of Evaluation and National Centre for Research in Evaluation, Standards, and Student Testing.

Higher Education Council, & National Board of Employment Education and Training. (1992). *Achieving quality*. Canberra: Australian Government Publishing Service.

Hill, P. W. (1987). Modeling the hierarchical structure of learning. In P. Cuttance & R. Ecob (Eds.), *Structural equation modeling by example. Applications in educational, sociological, and behavioral research* (pp. 65-85). New York: University of Cambridge Press.

Hill, P. W., & McGaw, B. (1981). Testing the simplex assumption underlying Bloom's taxonomy. *American Educational Research Journal, 18*(1), 93-101.

Hodgson, A., & Spours, K. (2000). *Qualifying for Success: Towards a framework of understanding*. London: Institute of Education.

Houghton, J., & Sheehan, P. (2000, February). A primer on the knowledge economy. Retrieved 01 June, 2000, from http://www.cfses.com/primer.htm

Howes, D. (1997). Implementing the Curriculum and Standards Framework in Victoria. In J. Lokan (Ed.), *Describing learning. Implementation of curriculum profiles in Australian schools 1986- 1996* (pp. 107-139). Camberwell, Vic: ACER Press.

Hox, J. J. (1995). *Applied multilevel analysis*. Amsterdam: TT-Publikaties.

Human Resources Development Canada. (2000, 27 September). Essential skills: Introduction. Retrieved 27 June, 2001, from http://www.hrsdc.gc.ca/en/hip/hrp/essential_skills/essential_skills_index.shtml

Hunt, E. (1995). *Will we be smart enough? A cognitive analysis of the coming workforce*. New York: Russell Sage Foundation.

Institution of Engineers Australia. (1996a). *Changing the culture: Engineering education into the future. Report summary*. Barton, ACT: Institution of Engineers Australia.

Institution of Engineers Australia. (1996b). *Changing the culture: Engineering education into the future. Review report*. Barton, ACT: Institution of Engineers Australia.

International Baccalaureate Organization. (2004). *Diploma programme assessment. Principles and practice*. Geneva, CH: International Baccalaureate Organization.

Jasinski, M. (1996). *Teaching and learning the key competencies in vocational education and training*. Adelaide: Department of Education, Employment, Training and Youth Affairs.

Jeffries, R., Polson, P. G., Razran, L., & Atwood, M. E. (1997). A process model for Missionaries-Cannibals and other river crossing problems. *Cognitive Psychology, 9*(4), 412-440.

Jensen, A. R. (1998). *The g factor. The science of mental ability*. Westport, CT: Praeger.

Kane, M., Berryman, S., Goslin, D., & Meltzer, A. (1990). *Identifying and describing the skills required by work. Report prepared for the Secretary's Commission on Achieving Necessary Skills.* Washington, DC: Pelavin and Associates.

Kearns, P. (2001). *Generic skills for the new economy. A review of research relating to generic skills*. Adelaide: National Centre for Vocational Education Research.

Keats, J. A. (1997). Classical test theory. In J. P. Keeves (Ed.), *Educational research, methodology, and measurement: an international handbook* (2nd ed., pp. 713-719). Oxford: Pergamon.

Keats, J. A. (1998). Measurement in educational research. In T. Husen & T. N. Postlethwaite (Eds.), *The international encyclopedia of education* (2nd ed., Vol. 7, pp. 3698-3707). Amsterdam: Elsevier.

Keeves, J. P. (1997a). Longitudinal research methods. In J. P. Keeves (Ed.), *Educational research, methodology, and measurement: an international handbook* (2nd ed., pp. 138-149). Oxford: Pergamon.

Keeves, J. P. (1997b). Measures of variation. In J. P. Keeves (Ed.), *Educational research, methodology, and measurement: an international handbook* (2nd ed., pp. 580-592). Oxford: Pergamon.

Keeves, J. P., & Alagumalai, S. (1999). New approaches to measurement. In G. N. Masters & J. P. Keeves (Eds.), *Advances in measurement in educational research and testing* (pp. 23-42). Amsterdam: Pergamon.

Keeves, J. P., & Kotte, D. (1996). The measurement and reporting of key competencies. In D. D. Curtis, R. Hattam, J. P. Keeves, M. J. Lawson, R. Reynolds, A. Russell, H. Silins & J. Smyth (Eds.), *Teaching and learning the key competencies in the vocational education and training sector. Research support.* (pp. 139-167). Adelaide: Flinders Institute for the Study of Teaching.

King John. (1215). Magna Carta. Retrieved 20 May, 2005, from http://www.britannia.com/history/docs/magna2.html

Klemp, G. O., & McClelland, D. C. (1986). What characterizes intelligent functioning among senior managers? In R. J. Sternberg & R. K. Wagner (Eds.), *Practical intelligence. Nature and origins of competence in the everyday world* (pp. 13-30). Cambridge: Cambridge University Press.

Kotovsky, K., Hayes, J. R., & Simon, H. A. (1985). Why are some problems hard? Evidence from Tower of Hanoi. *Cognitive Psychology, 17*(22), 248-294.

Krathwohl, D. R., Bloom, B. S., & Masia, B. B. (Eds.). (1964). *Taxonomy of educational objectives. The classification of educational goals. Handbook 2: Affective domain*. New York: David McKay.

Kreitzer, A. E., & Madaus, G. F. (1994). Empirical investigations of the hierarchical structure of the taxonomy. In L. W. Anderson & L. A. Sosniak (Eds.), *Bloom's taxonomy. A forty year retrospective. Ninety-third yearbok of the National Society for the Study of Education. Part II* (pp. 64-81). Chicago: National Society for the Study of Education.

Kropp, R. P., & Stoker, H. W. (1966). *The construction and validation of tests of the cognitive processes as described in the Taxonomy of Educational Objectives* (ERIC document No. ED 010044). Tallahassee, FL: Florida State University, Institute of Human Learning and Department of Educational Research and Testing.

Lajoie, S. P. (2003). Transitions and trajectories for studies of expertise. *Educational Researcher, 32*(8), 21-25.

Larkin, J. H. (1985). Understanding, problem representations, and skill in Physics. In S. F. Chipman, J. W. Segal & R. Glaser (Eds.), *Research and open questions* (Vol. 2).

Lave, J. (1988). *Cognition in practice: Mind, mathematics and culture in everyday life*. Cambridge, N.Y.: Cambridge University Press.

Lave, J., & Wenger, E. C. (1991). *Situated learning: legitimate peripheral participation*. Cambridge: Cambridge University Press.

Lawley, D. N. (1943). On problems connected with item selection and test construction. *Proceedings of the Royal Society of Edinburgh, 61-A*, 273-287.

Lawson, M. J. (1990). The case for instruction in the use of general problem-solving strategies in mathematics teaching: A comment on Owen and Sweller. *Journal for Research in Mathematics Education, 21*(5), 403-410.

Lawson, M. J. (1991). Testing for transfer following strategy learning. In G. Evans (Ed.), *Learning skills and cognitive training.* Melbourne: Australian Council for Educational Research.

Lawson, M. J., & Hopkins, S. (1996). Solving problems in commercial cookery. In D. D. Curtis, R. Hattam, J. P. Keeves, M. J. Lawson, R. Reynolds, A. Russell, H. Silins & J. Smyth (Eds.), *Teaching and learning the key competencies in the vocational education and training sector. Research support.* (pp. 17-43). Adelaide: Flinders Institute for the Study of Teaching.

Levy, F., & Murnane, R. J. (1999). *Are there key competencies critical to economic success? An economics perspective*. Neuchatel, Switzerland: NFSO, OECD.

Levy, M. (1987). *The Core Skills Project and work-based learning.* London: Manpower Services Commission.

Licht, B. G., & Dweck, C. S. (1984). Determinants of academic achievement: the interaction of children's achievement orientation with skill area. *Developmental Psychology, 20*(4), 628-636.

Linacre, J. M. (1989). *Many-facet Rasch measurement*. Chicago: MESA Press.

Linacre, J. M. (1997). Judgements, measurement of. In J. P. Keeves (Ed.), *Educational research, methodology, and measurement: an international handbook* (2nd ed., pp. 846-851). Oxford: Pergamon.

Linacre, J. M., & Wright, B. D. (1994). Dichotomous mean-square fit statistics. *Rasch Measurement Transactions, 8*(2), 350.

Loacker, G. (Ed.). (2000). *Self assessment at Alverno College*. Milwaukee, Wisconsin: Alverno College Institute.

Lokan, J. (Ed.). (1997). *Describing learning. Implementation of curriculum profiles in Australian schools 1986-1996*. Camberwell, Vic: ACER Press.

Lord, F. M. (1980). *Application of item response theory to practical testing problems*. Hillsdale, NJ: Lawrence Erlbaum and Associates.

Luce, R. D., & Tukey, J. W. (1964). Simultaneous conjoint measurement: A new type of fundamental measurement. *Journal of Mathematical Psychology, 1,* 1-27.

Macquarie Library (Ed.) (1981) The Macquarie Dictionary. Sydney: Macquarie Library.

Mansfield, B., & Schmidt, H. (2000). *Linking vocational education and training standards and employment requirements. An international manual*. Torino: European Training Foundation.

Marquardt, R. (1998). *Labour market policies and programmes affecting youth in Canada*. Paris: OECD.

Masters, G. N. (1982). A Rasch model for partial credit scoring. *Psychometrika, 47*(2), 149-174.

Masters, G. N. (1997). Partial credit model. In J. P. Keeves (Ed.), *Educational research, methodology, and measurement: an international handbook* (2nd ed., pp. 857-863). Oxford: Pergamon.

Masters, G. N. (1999). Partial credit model. In G. N. Masters & J. P. Keeves (Eds.), *Advances in measurement in educational research and testing* (pp. 98-109). Amsterdam: Pergamon.

Matters, G. (2005). *Designing assessment tasks for deep thinking.* Paper presented at the Curriculum Corporation Conference. Retrieved January, 2008, from http://cmslive.curriculum.edu.au/verve/_resources/Session_6.1_Matters_sml.ppt.

Matters, G., & Curtis, D. D. (2008). *Assessment and reporting of employability skills of senior secondary students*. Canberra: Department of Employment, Education, Training and Workplace Relations.

Mayer, J. D. (2001). Emotion, intelligence, and emotional intelligence. In J. P. Forgas (Ed.), *Handbook of affect and social cognition* (pp. 410-431). Mahwah, NJ: Lawrence Erlbaum Associates.

Mayer, J. D., & Salovey, P. (1993). The intelligence of emotional intelligence. *Intelligence*, *17*(4), 433-442.

Mayer, J. D., Salovey, P., & Caruso, D. (2000). Models of emotional intelligence. In R. J. Sternberg (Ed.), *Handbook of intelligence* (pp. 396-420). Cambridge, UK: Cambridge University Press.

Mayer, R. E. (1989). Human nonadversary problem solving. In K. J. Gilhooly (Ed.), *Human and machine problem solving* (pp. 39-56). New York: Plenum Press.

Mayer, R. E. (1992). *Thinking, problem solving, cognition* (2nd ed.). New York: W H Freeman.

Mayer, R. E. (1995a). Development of learning skills in problem-solving and thinking. In L. W. Anderson (Ed.), *International encyclopedia of teaching and teacher education* (2nd ed., pp. 4722-4728). Oxford: Pergamon.

Mayer, R. E. (1995b). Teaching and testing for problem-solving. In L. W. Anderson (Ed.), *International encyclopedia of teaching and teacher education* (2nd ed., pp. 4728-4731). Oxford: Pergamon.

Mayer, R. E. (2000). Intelligence and education. In R. J. Sternberg (Ed.), *Handbook of intelligence* (pp. 519-533). Cambridge, UK: Cambridge University Press.

Mayer, R. E., & Wittrock, M. C. (1996). Problem-solving transfer. In D. C. Berliner & R. C. Calfee (Eds.), *Handbook of educational psychology* (pp. 47-62). New York: Macmillan.

McAllister, S. M. (2005). *Competency based assessment of speech pathology students' performance in the workplace.* Unpublished Thesis, University of Sydney, Sydney.

McClelland, D. C. (1973). Testing for competence rather than for "intelligence". *American Psychologist, 28*, 1-14.

McCurry, D., & Bryce, J. (1997). *The school-based key competencies levels assessment project. Final report*. Canberra: DEETYA.

McCurry, D., & Bryce, J. (2000). *Victorian Board of Studies. Key Competencies levels assessment trial. Working paper 2*. Melbourne: Victorian Curriculum and Assessment Authority.

MCEETYA. (2004). Performance Measurement and Reporting Taskforce. Retrieved 15 October, 2004, from http://www.mceetya.edu.au/taskfrce/task224.htm

MCEETYA Transition from Schools Taskforce. Working Group on Key Competencies. (1996). *Key competencies for work, education and life. Report on the outcomes of the key competencies pilot phase* (No. 12/96). Canberra: DEETYA.

McGaw, B. (1997). *Shaping their future. Recommendations for the reform of the higher school certificate*. Sydney, NSW: Department of Training and Education Coordination.

McGaw, B. (2006). *Assessment fit for purpose*. Paper presented at the 32nd IAEA Annual Conference. Assessment in an Era of Rapid Change: Innovations and Best Practices. from http://www.iaea2006.seab.gov.sg/conference/download/papers/Keynote-Assessment%20fit%20for%20purpose.pdf.

McInnis, C., Griffin, P., James, R., & Coates, H. (2001). *Development of the Course Experience Questionnaire (CEQ)*. Canberra: Department of Education, Training and Youth Affairs.

McKenzie, P. (1983). Recurrent education: Economic and equity issues in Australia. *Australian Education Review*(20), 1-107.

McLarty, J. R., & Vansickle, T. R. (1997). Assessing employability skills: The Work Keys system. In H. F. O'Neil (Ed.), *Workforce readiness. Competencies and assessment* (pp. 293-325). Mahwah, NJ: Lawrence Erlbaum.

McLeish, A. (2002). *Employability skills for Australian small and medium sized enterprises*. Canberra: Department of Education Science and Training.

Mentkowski, M. (2000). *Learning that lasts. Integrating learning, development, and performance in college and beyond*. San Francisco: Jossey-Bass.

Mertens, D. (1974). Schlusselqualificationen: Thesen zur Schulung fur eine moderne Gesellschaft [Key skills: Theses on schooling for a modern society]. *Mitteilungen aus der Arbeitsmarkt und Berufsforschung, 7*(1), 36-43.

Messick, S. J. (1995). Validity of psychological assessment. Validation of inferences from persons' responses and performances as scientific inquiry into score meaning. *American Psychologist, 50*(9), 741-749.

Messick, S. J. (1998). Alternative modes of assessment, uniform standards of validity. In M. D. Hakel (Ed.), *Beyond multiple-choice: Evaluating alternatives to traditional testing for selection* (pp. 59-74). Mahwah, NJ: Lawrence Erlbaum.

Michell, J. (1997). Quantitative science and the definition of measurement in psychology. *British Journal of Psychology, 88*, 355-383.

Michell, J. (2003). Measurement: A beginner's guide. *Journal of Applied Measurement, 4*(4), 298-308.

Miller, K. F., & Stigler, J. W. (1991). Meanings of skill: effects of abacus expertise on number representation. *Cognition and Instruction, 8*(1), 29-67.

Miller, M. D., & Linn, R. L. (2000). Validation of performance-based assessments. *Applied Psychological Measurement, 24*(4), 367-378.

Milne, C. (2000). Tertiary literacies: integrating generic skills into the curriculum. In S. Fallows & C. Steven (Eds.), *Integrating key skills in higher education. Employability, transferable skills and learning for life* (pp. 87-97). London: Kogan Page.

Ministerial Council on Education Employment Training and Youth Affairs. (1999, July). The Adelaide Declaration on National Goals for Schooling in the Twenty-first Century. Retrieved 10 October, 2000, from http://www.mceetya.edu.au/mceetya/nationalgoals/natgoals.htm

Ministerial Council on Education Employment Training and Youth Affairs. (2003). Response to advice on employability skills, *MCEETYA Out-of-session resolution, December*.

Mitchell, J. (2007). *Level 1 workshop for implementing the model for assessing and reporting employability skills. Program and resources*. Retrieved 12 December 2007, from http://www.pco.com.au/esw/pre-workshop_reading.htm .

Moser, C. (1999). A fresh start – improving literacy and numeracy. Retrieved 21 May, 2001, from http://www.lifelonglearning.co.uk/mosergroup/index.htm

Moss, P. A. (1994). Can there be validity without reliability? *Educational Researcher, 23*(2), 5-12.

Mumford, M. D., Baughman, W. A., Supinski, E. P., & Anderson, L. E. (1998). A construct approach to skill assessment: Procedures for assessing complex cognitive skills. In M. D. Hakel (Ed.), *Beyond multiple-choice: Evaluating alternatives to traditional testing for selection* (pp. 75-112). Mahwah, NJ: Lawrence Erlbaum.

Murray, T. S. (2003). Reflections on international competence assessments. In D. S. Rychen & L. H. Salganik (Eds.), *Key competencies for a successful Life and a well-functioning society* (pp. 135-159). Göttingen: Hogrefe & Huber.

Muthen, L. K., & Muthen, B. O. (2006). MPlus (Version 4.0) [Statistical analysis with latent variables]. Los Angeles, CA: Muthen & Muthen.

Nash, B. E., & Korte, R. C. (1997). Validation of SCANS competencies by a national job analysis study. In H. F. O'Neil (Ed.), *Workforce readiness. Competencies and assessment* (pp. 77-102). Mahwah, NJ: Lawrence Erlbaum.

National Curriculum Council. (1990). *Core skills 16-19*. London: National Curriculum Council.

National Industry Education Forum. (2000). *The Key Competencies portfolio approach - a kit*. Canberra: Department of Education, Training and Youth Affairs.

National Quality Council. (2007). *Excellence criteria. Guide for Registered Training Organisations. Draft*. Canberra: Department of Education, Employment and Workplace Relations.

National Skills Task Force. (1998). *Towards a national skills agenda*. London: Department for Education and Employment.

National Training Board. (1991). *National competency standards. Policy and guidelines*. Canberra: National Training Board.

National Training Quality Council. (2005). *Resolution: Final advice on the incorporation of employability skills in training package qualifications and related issues*. Melbourne: National Training Quality Council.

Neisser, U., Boodoo, G., Bouchard, T. J., Boykin, A. W., Brody, N., Ceci, S. J., et al. (1996). Intelligence: knowns and unknowns. *American Psychologist, 51*(2), 77-101.

Newell, A., & Simon, H. A. (1972). *Human problem solving*. Englewood Cliffs: Prentice Hall.

Nickerson, R. S. (1994). The teaching of thinking and problem solving. In R. J. Sternberg (Ed.), *Thinking and problem solving* (pp. 409-449). San Diego: Academic Press.

Nijhof, W. J. (1998). Qualifying for the future. In W. J. Nijhof & J. N. Streumer (Eds.), *Key qualifications in work and education* (pp. 19-38). Dordrecht: Kluwer.

Norwegian Board of Education. (1997). *Core curriculum for primary, secondary and adult education in Norway*. Oslo: Norwegian Board of Education.

Nusche, D. (2008). *Assessment of learning outcomes in higher education: a comparative review of selected practices* (Education Working Paper No. 15). Paris: OECD.

O'Neil, H. F., Allred, K., & Baker, E. L. (1997). Review of workforce readiness theoretical frameworks. In H. F. O'Neil (Ed.), *Workforce readiness. Competencies and assessment* (pp. 3-25). Mahwah, NJ: Lawrence Erlbaum.

OECD. (1996). *Lifelong learning for all*. Paris: OECD.

OECD. (1998). *Country Note: Canada. Report for the Thematic Review of Transition from Initial Education to Working Life*. Paris: OECD.

OECD. (2001). *Thematic review on adult learning. Denmark background paper*. Paris: OECD.

OECD. (2003). *The PISA 2003 assessment framework - mathematics, reading, science and problem solving knowledge and skills*. Paris: OECD.

OECD. (2004). *Problem solving for tomorrow's world. First measures of cross-curricular competencies from PISA 2003*. Paris: OECD.

OECD, & Swiss Federal Statistics Office. (2000). *DeSeCo country contribution process*. Paris, Neuchatel: OECD, SFSO.

OECD and Human Resources Development Canada. (1997). *Literacy skills for the knowledge society. Further results from the International Adult Literacy Survey*. Paris: OECD.

Oliver, K. M., Russell, C., Gilli, L. M., Hughes, R. A., Schuder, T., Brown, J. L., et al. (1997). Skills for Success in Maryland: Beyond workplace readiness. In H. F. O'Neil (Ed.), *Workforce readiness. Competencies and assessment* (pp. 47-76). Mahwah, NJ: Lawrence Erlbaum.

Osterlind, S. J. (1998). *Constructing test items: Multiple choice, constructed response, performance, and other formats* (2nd ed.). Boston: Kluwer Academic Publishers.

Owen, E., & Sweller, J. (1989). Should problem solving be used as a learning device in mathematics? *Journal for Research in Mathematics Education, 20*(3), 322-328.

Oxford University Press. (1997). New Shorter Oxford English Dictionary (Version 1.03). Oxford: Oxford University Press.

Palumbo, D. B., & Vargas, E. A. (1988). Problem solving: a behavioral interpretation. *Computers in the Schools, 4*(3&4), 17-27.

Paris, S. G., Wixson, K. K., & Palincsar, A. S. (1986). Instructional approaches to reading comprehension. In E. Rothkopf (Ed.), *Review of research in education*. Washington, D.C.: American Educational Research Association.

Payne, J. (1999). *All things to all people: Changing perceptions of 'skill' among Britain's policy makers since the 1950s and their implications* (SKOPE Research Paper No. 1). Warwick, UK: SKOPE, University of Warwick.

Pellegrino, J., Chudowsky, N., & Glaser, R. (Eds.). (2001). *Knowing what students know. The science and design of educational assessment. A report of the National Research Council*. Washington, DC: National Academy Press.

Perrenoud, P. (1999). *The key to social fields: Essay on the competences of an autonomous actor. Or how to avoid being abused, alienated, dominated or exploited when you are neither rich nor powerful: A sociological perspective*. Neuchatel: NFSO, OECD.

Piskaty, G., Elsik, M., Blumberger, W., & Thonabauer, C. (2000). *Vocational education and training in Austria*. Luxembourg: CEDEFOP.

Pitman, J., Matters, G., & Nuyen, A. (1998). *A case for testing generic skills*. Paper presented at the 24th Annual Conference of the International Association for Educational Assessment.

Polya, G. (1957). *How to solve it. A new aspect of mathematical method* (2nd ed.). Harmondsworth: Penguin.

Popper, K. (1978). Three worlds [Electronic Version]. *The Tanner Lectures on Human Values*. Retrieved 7 August 2009 from http://www.tannerlectures.utah.edu/lectures/documents/popper80.pdf.

Postlethwaite, T. N. (1994). Validity vs. utility: personal experiences with the taxonomy. In L. W. Anderson & L. A. Sosniak (Eds.), *Bloom's taxonomy. A forty year retrospective. Ninety-third yearbok of the National Society for the Study of Education. Part II* (pp. 174-180). Chicago: National Society for the Study of Education.

Powell, R., Smith, R., & Reakes, A. (2003). *Basic skills and key skills: A review of international literature. Final report*. London: NFER.

Pressley, M., & McCormick, C. B. (1995). *Advanced educational psychology for educators, researchers, and policymakers*. New York: Harper Collins.

Pumphrey, J., & Slater, J. (n.d.). *An assessment of generic skills needs*. London: The Council for Administration.

Quality of Education Review Committee. (1985). *Quality of education in Australia. Report of the Review Committee. (Prof Peter Karmel, Chair)*. Canberra: Australian Government Publishing Service.

Queensland Department of Education. (1997). *Assessing and reporting the key competencies of students of post-compulsory age through 'work experience'*. Canberra: Department of Education, Employment, Training and Youth Affairs.

Queensland Department of Education. (2003). *New Basics Project. New Basics reseach program*. Brisbane.

Queensland Department of Education and Training. (2004). *Work futures: Employability skills and attributes*. Brisbane: Queensland Department of Education and Training.

Ramachandran, V. S. (2003). Purple numbers and sharp cheese. Lecture 4. *The Reith Lectures.* Retrieved June, 2005, from http://www.bbc.co.uk/radio4/reith2003/lecture4.shtml

Ramprasad, A. (1983). On the definition of feedback. *Behavioural Science*(28), 4-13.

Ramsden, P. (1991). Report on the Course Experience Questionnaire trial. In R. Linke (Ed.), *Performance indicators in higher education* (Vol. 2). Canberra: Commonwealth Department of Employment, Education and Training.

Ramsden, P. (1998, 29 - 30 November 1998). *The CEQ: Looking back - and forward.* Paper presented at the Course Experience Questionnaire Symposium 1998, University of New South Wales, Sydney.

Rasbash, J., Browne, W., Goldstein, H., Yang, M., Plewis, I., Healy, M., et al. (2002). *A user's guide to MLwiN*. London: Centre for Multilevel Modelling, Institute of Education, University of London.

Rasbash, J., Browne, W., Healy, M., Cameron, B., & Charlton, C. (2005). MLwiN (Version 2.02) [Multilevel analysis software]. London: Multilevel Models Project.

Rasbash, J., Steele, F., Browne, W., & Prosser, B. (2004). *A user's guide to MLwiN Version 2.0*. London: Centre for Multilevel Modelling, Institute of Education, University of London.

Rasch, G. (1960, 1980). *Probabalistic models for some intelligence and attainment tests*. Copenhagen, Chicago: Danmarks Pegagogiske Institut, University of Chicago Press.

Ratio Pty Ltd, & Down, C. (2003). *Employability skills in training packages. Final report*. Melbourne: Australian National Training Authority.

Raudenbush, S. W., & Bryk, A. S. (1997). Hierarchical linear models. In J. P. Keeves (Ed.), *Educational research, methodology, and measurement: an international handbook* (2nd ed., pp. 2590-2596). Oxford: Pergamon.

Raudenbush, S. W., Bryk, A. S., & Congdon, R. (2005). HLM for Windows (Version 6.03). Chicago: Scientific Software International.

Ray, J. (2001). *Apprenticeship in Australia. An historical perspective.* Adelaide: National Centre for Vocational Education Research.

Reber, A. S. (1985). *The Penguin dictionary of psychology*. Harmondsworth: Penguin.

Reeff, J.-P., Zabal, A., & Klieme, E. (2005). ALL problem solving framework. In T. S. Murray, Y. Clermont & M. Binkley (Eds.), *Measuring adult literacy and life skills: New frameworks for assessment* (pp. 192-227). Ottawa: Statistics Canada.

Reich, R. B. (1991). *The work of nations. Preparing ourselves for 21st century capitalism*. New York: Alfred A Knopf.

Rescorla, R. A. (1988). Pavlovian conditioning. It's not what you think. *American Psychologist, 43*(3), 151-160.

Resnick, L. B. (1987). Learning in school and out. *Educational Researcher, 16*(9), 13-20.

Reynolds, R., & van Eyk, B. (1996). Training and key competencies - other providers. In D. D. Curtis, R. Hattam, J. P. Keeves, M. J. Lawson, R. Reynolds, A. Russell, H. Silins & J. Smyth (Eds.), *Teaching and learning the key competencies in the vocational education and training sector. Research support.* (pp. 107-118). Adelaide: Flinders Institute for the Study of Teaching.

Roe, J. (2003). Employability skills: social engineering or skilling youth for employment? Retrieved 9 October, 2003, from http://www.utlc.org.au/Training/JRoePaper.pdf

Rogers, B., McDonald, J., & Sizer, T. (1993). Standards and school reform: Asking the essential questions. Retrieved 5 July, 2007, from http://www.essentialschools.org/cs/resources/view/ces_res/106

Rohwer, W. D., & Sloane, K. (1994). Psychological perspectives. In L. W. Anderson & L. A. Sosniak (Eds.), *Bloom's taxonomy. A forty year retrospective. Ninety-third yearbok of the National Society for the Study of Education. Part II* (pp. 41-63). Chicago: National Society for the Study of Education.

Ruiz, D., & Newell, A. (1989). *Tower-noticing triggers strategy-change in the Tower of Hanoi: a Soar model*. Paper presented at the 11th Annual Conference of the Cognitive Science Society.

Ryan, C. (1997). *NSW key competencies pilot project report*. Sydney: NSW Department of Education and Training Coordination.

Rychen, D. S. (2003). A frame of reference for defining and selecting key competencies in an international context. In D. S. Rychen, L. H. Salganik & M. E. McLaughlin (Eds.), *Contributions to the second DeSeCo symposium, Geneva, 11-13 February, 2002* (pp. 107-116). Neuchatel, Switerzerland: SFSO.

Rychen, D. S., & Salganik, L. H. (2000). *Definition and selection of key competencies. A report to the INES General Assembly*. Neuchatel, Switzerland: OECD.

Rychen, D. S., & Salganik, L. H. (Eds.). (2001). *Defining and selecting key competencies*. Seattle: Hogrefe & Huber Publishers.

SACE Board. (2009). *SACE moderation procedures*. Adelaide: Government of South Australia, SACE Board.

Sadler, D. R. (1986). *Subjectivity, objectivity, and teachers' qualitative judgments* (Discussion Paper No. 5). Brisbane: Queensland Board of Secondary School Studies.

Sadler, D. R. (1989). Formative assessment and the design of instructional systems. *Instructional Science, 18*(2), 119-144.

Sadler, D. R. (2003, 6-7 November). *How criteria-based grading misses the point*. Paper presented at the Effective Teaching and Learning Conference, Southbank Campus, Griffith University, Brisbane.

Sadler, D. R. (2005). Interpretations of criteria-based assessment and grading in higher education. *Assessment and Evaluation in Higher Education, 30*(2), 175-194.

Salganik, L. H. (2002). Highlights from current assessments. In D. S. Rychen, L. H. Salganik & M. E. McLaughlin (Eds.), *Contributions to the Second DeSeCo Symposium, 11-13 February, 2002* (pp. 195-196). Neuchatel, Switzerland: SFSO.

Salganik, L. H., Rychen, D. S., Moser, U., & Konstant, J. W. (1999). *Definition and Selection of Competencies. Projects on Competencies in the OECD Context. Analysis of Theoretical and Conceptual Foundations*. Neuchatel, Switzerland: OECD.

Salovey, P., Bedell, B. T., Detweiler, J. B., & Mayer, J. D. (2000). Current directions in emotional intelligence research. In M. Lewis & J. M. Haviland-Jones (Eds.), *Handbook of emotions* (2nd ed., Vol. 1, pp. 720). New York: The Guilford Press.

Saxe, G. B. (1988). The mathematics of child street vendors. *Child Development, 59*, 1415-1425.

Saxe, G. B. (1996). Studying cognitive development in a sociocultural context: The development of a practice-based approach. In A. Colby & D. Jessor (Eds.), *Ethnographic approaches to the study of human development* (pp. 291-323). Chicago: University of Chicago Press.

Schafer, J. L. (2000). NORM: Multiple imputation of incomplete multivariate data under a normal model (Version 2.03) [Missing data imputation]: Pennsylvania State University.

Schleicher, A. (2008). PIAAC: A new strategy for assessing adult competencies. *International Review of Education, 54*(5), 627-650.

Schofield, K., & McDonald, R. (2004). *Moving on... Report of the high level review of training packages*. Melbourne: Australian National Training Authority.

Scholten, I., Keeves, J. P., & Lawson, M. J. (2002). Validation of a free response test of deep learning about the normal swallowing process. *Higher Education, 44*(2), 233-255.

Scribner, S. (1984). Everyday cognition: its development in social contexts. In B. Rogoff & J. Lave (Eds.), *Studying working intelligence* (pp. 9-40). Cambridge, MA: Harvard University Press.

Shavelson, R. J., Gao, X., & Baxter, G. P. (1993). *Sampling variability of performance assessments* (CSE Technical Report No. 361). Los Angeles: University of California, CRESST.

Sheridan, B., Andrich, D., & Luo, G. (1997). RUMM (Version 2.7Q) [Statistical analysis software]. Perth: RUMM Laboratory.

Singapore Workforce Development Agency. (2003). The seven CREST skills. Retrieved 8 October, 2004, from http://www.wda.gov.sg/crest/about_skills.html

Singer, J. D., & Willett, J. B. (2003). *Applied longitudinal data analysis: Modeling change and event occurrence*. New York: Oxford University Press.

Smith, E., & Comyn, P. (2003). *The development of employability skills in novice workers through employment*. Adelaide: NCVER.

Smith, E. V., & Smith, R. M. (Eds.). (2004). *Introduction to Rasch measurement*. Maple Grove, Mn: JAM Press.

Smith, R. M. (1994a). A comparison of the power of Rasch total- and between-item fit statistics to detect measurement disturbances. *Educational and Psychological Measurement, 54*(1), 42-55.

Smith, R. M. (1994b). Detecting item bias in the Rasch rating scale model. *Educational and Psychological Measurement, 54*(4), 886-896.

Smith, R. M. (2000). Fit analysis in latent trait measurement models. *Journal of Applied Measurement, 1*(2), 199-218.

Soanes, C., & Stevenson, A. (Eds.). (2005). *Oxford dictionary of English* (2nd ed., rev. ed.). Oxford: Oxford University Press.

South African Qualifications Authority. (1997). Critical cross-field outcomes. *SAQA Bulletin, 1*(1).

South Australia. Department of Education Training and Employment. (2001). *South Australian Curriculum Standards and Accountability Framework* (Vol. 2001): Department of Education Training and Employment.

Spencer, L. M., & Spencer, S. M. (1993). *Competence at work. Models for superior performance*. New York: John Wiley & Sons.

SPSS Inc. (1995). SPSS for Windows (Version 6.1.3) [Statistical analysis program]. Chicago: SPSS Inc.

Stanley, G. (1993). The psychology of competency-based education. In C. Collins (Ed.), *Competencies. The competencies debate in Australian education and training* (pp. 145-153). Canberra: Australian College of Education.

Stanwick, J. (2005). *Australian Qualifications Framework lower-level qualifications: Pathways to where for young people?* Adelaide: National Centre for Vocational Education Research.

Stanwick, J. (2006). *Outcomes from higher-level vocational education and training qualifications*. Adelaide: NCVER.

Starick, R. (2006). *Imputation in Longitudinal Surveys: The Case of HILDA* (Cat No. 1352.0.55.075). Canberra: ABS.

Stein, B. S., & Bransford, J. D. (1979). Constraints on effective elaboration: Effects of precision and subject generation. *Journal of Verbal Learning and Verbal Behavior, 18*, 769-777.

Sternberg, R. J. (1983). *How can we teach intelligence?* (ERIC Document No. ED 242 700).

Sternberg, R. J. (1985a). *Beyond IQ. A triarchic theory of human intelligence*. London: Cambridge University Press.

Sternberg, R. J. (1985b). Human intelligence: the model is the message. *Science, 230*(4730), 1111-1118.

Sternberg, R. J. (1995). *In search of the human mind*. Fort Worth: Harcourt Brace.

Sternberg, R. J. (1998). Abilities are forms of developing expertise. *Educational Researcher, 27*(3), 11-20.

Sternberg, R. J. (2003). What is an "expert student?" *Educational Researcher, 32*(8), 5-9.

Sternberg, R. J. (Ed.). (2000). *Handbook of intelligence*. Cambridge, UK: Cambridge University Press.

Sternberg, R. J., Forsythe, G. B., Hedlund, J., Horvath, J. A., Wagner, R. K., Williams, W. M., et al. (2000). *Practical intelligence in everyday life*. Cambridge, UK: Cambridge University Press.

Sternberg, R. J., & Wagner, R. K. (Eds.). (1986). *Practical intelligence. Nature and origins of competence in the everyday world*. Cambridge: Cambridge University Press.

Stevens, S. S. (1951). Mathematics, measurement, and psychophysics. In S. S. Stevens (Ed.), *Handbook of experimental psychology* (pp. 1-49). New York: John Wiley.

Stobart, G. (2004). *The formative use of summative assessment: Possibilities and limits*. Paper presented at the 30th Annual conference of the International Association for Educational Assessment.

Stocking, M. L. (1997). Item response theory. In J. P. Keeves (Ed.), *Educational research, methodology, and measurement: an international handbook* (2nd ed., pp. 836-840). Oxford: Pergamon.

Streumer, J. N., & Bjorkquist, D. C. (1998). Moving beyond traditional vocational education and training: Emerging issues. In W. J. Nijhof & J. N. Streumer (Eds.), *Key qualifications in work and education* (pp. 249-264). Dordrecht: Kluwer.

Sweet, R. (2008). *The Job Ready Certificate. A discussion paper*. Canberra: DEEWR.

Sweller, J. (1989). Cognitive technology: some procedures for facilitating learning and problem solving in mathematics and science. *Journal of Educational Psychology, 81*(4), 457-466.

Sweller, J. (1990). On the limited evidence for the effectiveness of teaching general problem-solving strategies. *Journal for Research in Mathematics Education, 21*(5), 411-415.

Sweller, J., Mawer, R. F., & Howe, W. (1982). Consequences of history-cued and means-end strategies in problem solving. *American Journal of Psychology, 95*(3), 455-483.

Tabachnick, B. G., & Fidell, L. S. (2007). *Using multivariate statistics* (5th ed.). Boston, MA: Pearson Education.

Tang, C. (1994). Assessment and student learning: Effects of modes of assessment on students' preparation strategies. In G. Gibbs (Ed.), *Improving student learning: Theory and practice* (pp. 151-170). Oxford, UK: Oxford Brookes University, The Oxford Centre for Staff Development.

Taras, M. (2003). To feedback or not to feedback in student self-assessment. *Assessment & Evaluation in Higher Education, 28*(5), 549-565.

The Secretary's Commission on Achieving Necessary Skills. (1991). *What work requires of schools. A SCANS report for America 2000*. Washington, DC: U.S. Department of Labor.

The Secretary's Commission on Achieving Necessary Skills. (1992a). *Blueprint for action: Building community coalitions*. Washington, DC: U.S. Department of Labor.

The Secretary's Commission on Achieving Necessary Skills. (1992b). *Learning a living: A blueprint for high performance. A SCANS report for America 2000*. Washington, DC: U.S. Department of Labor.

The Secretary's Commission on Achieving Necessary Skills. (1993). *Teaching the SCANS competencies*. Washington, DC: U.S. Department of Labor.

Thomson, S., Cresswell, J., & de Bortoli, L. (2004). *Facing the future: A focus on mathematical literacy among Australian 15-year-old students in PISA 2003*. Melbourne: ACER and OECD.

Thomson, S., Wernert, N., Underwood, C., & Nicholas, M. (2008). *TIMSS 07: Taking a closer look at mathematics and science in Australia*. Melbourne: IEA, ACER.

Thorndike, R. L., & Thorndike, R. M. (1997). Reliability. In J. P. Keeves (Ed.), *Educational research, methodology, and measurement: an international handbook* (2nd ed., pp. 775-790). Oxford: Pergamon.

Thorndike, R. M. (1999). IRT and intelligence testing: past, present, and future. In S. E. Embretson & S. L. Hershberger (Eds.), *The new rules of measurement. What every psychologist and educator should know* (pp. 17-35). Mahwah, NJ: Lawrence Erlbaum and Associates.

Trier, U. (2001). *12 countries contributing to DeSeCo. A summary report*. Neuchatel, Switzerland: Swiss Federal Statistical Office.

Trinchero, R. (2006). Competenze trasversali? Unpublished Conference paper. Unito.

Troper, J., & Smith, C. (1997). Workplace readiness portfolios. In H. F. O'Neil (Ed.), *Workforce readiness. Competencies and assessment* (pp. 357-382). Mahwah, NJ: Lawrence Erlbaum.

Tuijnman, A. C. (1986). *Recurrent education and socioeconomic success. A theoretical and longitudinal analysis*. Stockholm, Sweden: Institute of International Education, University of Stockholm.

Tuijnman, A. C. (1996). Concepts, theories, and methods. Introduction. In A. C. Tuijnman (Ed.), *International encyclopedia of adult education and training* (2nd ed.). Oxford: Elsevier Science.

Turner, D. (2002). *Employability skills development in the United Kingdom*. Adelaide: National Centre for Vocational Education Research.

UNESCO Institute for Education. (1997). *Adult education. The Hamburg Declaration and The Agenda for the Future*. Hamburg: UNESCO Institute for Education.

University of South Australia. (2000, December, 2000). Graduate qualities. Retrieved 21 March, 2001, from http://www.unisanet.unisa.edu.au/extracourseinformation/gradquals.asp

US Department of Labor. (1986). *Dictionary of occupational titles* (4th ed.). Washington: US Department of Labor.

Viertel, E., & Grootings, P. (2001). *Review and lessons learned of Phare Vocational Education and Training Reform programmes 1993 - 1998*. Torino: European Training Foundation.

von Eye, A., & Spiel, C. (1997). Human development: research methodology. In J. P. Keeves (Ed.), *Educational research, methodology, and measurement: an international handbook* (2nd ed., pp. 134-138). Oxford: Pergamon.

Wagner, R. K., & Sternberg, R. J. (1986). Tacit knowledge and intelligence in the everyday world. In R. J. Sternberg & R. K. Wagner (Eds.), *Practical intelligence. Nature and origins of competence in the everyday world* (pp. 51-83). Cambridge: Cambridge University Press.

Washer, P. (2007). Revisiting Key Skills: A practical framework for higher education. *Quality in Higher Education, 13*(1), 57-67.

Weinert, F. E. (1999). *Concepts of competence*. Neuchatel, Switzerland: OECD, DeSeCo Project.

Weiss, D. J., & Yoes, M. E. (1991). Item response theory. In R. K. Hambleton & J. N. Zaal (Eds.), *Advances in educational and psychological testing: theory and applications* (pp. 69-95). Boston: Kluwer Academic Publishers.

Werner, M. C. (1995). *Australian key competencies in an international perspective*. Adelaide: NCVER.

West, R. C. (1998). *Learning for Life. Final Report. Review of Higher Education Financing and Policy*. Canberra: Department of Employment, Education, Training and Youth Affairs.

Wiggins, G. P. (1989). A true test: Toward more authentic and equitable assessment. *Phi Delta Kappan, 70*(9), 703-713.

Wiggins, G. P. (1998). *Educative assessment. Designing assessment to inform and improve student performance*. San Francisco: Jossey-Bass.

Wiliam, D. (1998). *Construct-referenced assessment of authentic tasks: alternatives to norms and criteria*. Paper presented at the Testing and Evaluation: Confronting the Challenges of Rapid Social Change conference.

Willett, J. B. (1997). Measurement of change. In J. P. Keeves (Ed.), *Educational research, methodology, and measurement: an international handbook* (2nd ed., pp. 671-678). Oxford: Pergamon.

Wilson, M. R. (1989a). Empirical examination of a learning hierarchy using an Item Response Theory model. *Journal of Experimental Education, 57*(4), 357-371.

Wilson, M. R. (1989b). Saltus: a psychometric model of discontinuity in cognitive development. *Psychological Bulletin, 105*, 276-289.

Wilson, M. R. (1997). Measurement of developmental levels. In J. P. Keeves (Ed.), *Educational research, methodology, and measurement: an international handbook* (2nd ed., pp. 908-915). Oxford: Pergamon.

Wilson, M. R. (1999). Measurement of developmental levels. In G. N. Masters & J. P. Keeves (Eds.), *Advances in measurement in educational research and testing* (pp. 151-163). Amsterdam: Pergamon.

Wilson, M. R. (2001). The rater bundle model. *Journal of Educational and Behavioral Statistics, 26*(3), 283-306.

Wilson, M. R. (2005). *Constructing measures: An item response modelling approach*. Mahwah, NJ: Lawrence Erlbaum.

Wilson, M. R., & Masters, G. N. (1993). The partial credit model and null categories. *Psychometrika, 58*(1), 87-99.

Wilson, M. R., & Sloane, K. (2000). From principles to practice: an embedded assessment system. *Applied Measurement in Education, 13*(2), 181-208.

Wise, L., Chia, W. J., & Rudner, L. M. (1990). *Identifying necessary job skills: a review of previous approaches. Report prepared for the Secretary's Commission on Achieving Necessary Skills.* Washington: Pelavin Associates.

Wright, B. D. (1988). Rasch measurement models. In J. P. Keeves (Ed.), *Educational research, methodology and measurement: An international handbook* (pp. 85-97). Oxford: Pergamon.

Wright, B. D. (1997). A history of social science measurement. *MESA Memo, 62*, 1-28.

Wright, B. D. (1998). Georg Rasch. The man behind the model. *Popular Measurement, 1*(1), 15-22.

Wright, B. D. (1999a). Fundamental measurement for psychology. In S. E. Embretson & S. L. Hershberger (Eds.), *The new rules of measurement. What every psychologist and educator should know* (pp. 65-104). Mahwah, NJ: Lawrence Erlbaum and Associates.

Wright, B. D. (1999b). Rasch measurement models. In G. N. Masters & J. P. Keeves (Eds.), *Advances in measurement in educational research and testing* (pp. 85-97). Amsterdam: Pergamon.

Wright, B. D., & Masters, G. N. (1982). *Rating scale analysis*. Chicago: MESA Press.

Wright, B. D., & Mok, M. M. (2004). An overview of the family of Rasch models. In E. V. Smith & R. M. Smith (Eds.), *Introduction to Rasch measurement* (pp. 1-24). Maple Grove, Mn: JAM Press.

Wright, B. D., & Stone, M. H. (1999). *Measurement essentials* (2nd ed.). Willmington: Wide Range Inc.

Wu, M. L., Adams, R. J., & Wilson, M. R. (1998). ConQuest generalised item response modelling software (Version 1.0) [Statistical analysis software]. Melbourne: Australian Council for Educational Research.

Yekovich, F. R., Thompson, M. A., & Walker, C. H. (1991). Generation and verification of inferences by experts and trained nonexperts. *American Educational Research Journal, 28*(1), 189-200.

Zeller, R. A. (1997). Validity. In J. P. Keeves (Ed.), *Educational research, methodology, and measurement: an international handbook* (2nd ed., pp. 822-829). Oxford: Pergamon.

Zhu, W. (1996). Should total scores from a rating scale be used directly? *Research Quarterly for Exercise and Sport, 67*(3), 363-372.

Appendices

Appendix 1: Ethics Approvals

Approval from Flinders University

Below is a copy of the letter from the Social and Behavioural Research Ethics Committee indicating that the research has been approved.

FLINDERS UNIVERSITY
ADELAIDE • AUSTRALIA

Social and Behavioural Research Ethics Committee
Faculty of Social Sciences

GPO Box 2100
Adelaide 5001 Australia

Telephone: (+61 8) 8201 3513
Fax: (+61 8) 8201 3756
Email: Lesley.Wyndram@flinders.edu.au

SBRE 2413

7 December 2001

Mr David Curtis
11 Scott Ave
Flinders Park SA 5025

Dear Mr Curtis

Project 2413 Generic Competences in Education

Further to my letter dated 14 November 2001, I am pleased to inform you that approval of the above project has been confirmed following receipt of the additional information you submitted on 6 December.

May I draw to your attention that, in order to comply with monitoring requirements of the *National Statement on Ethical Conduct in Research Involving Humans* an annual and/or final report must be submitted. A copy of the report pro-forma is available from the SBREC website http://www/ssn/flinders.edu.au/ethics/ethics/

Yours sincerely

Lesley Wyndram
Secretary
SOCIAL AND BEHAVIOURAL RESEARCH ETHICS COMMITTEE

c.c. Simon Macdonald, Office of Research
Prof J Keeves

(esr\letter\2413finapp)

Location: Sturt Road, Bedford Park, South Australia.

Approval from the Department of Education Training and Employment

Below is a copy of the letter from the South Australian Department of Education, Training and Employment (DETE) Research Unit to the researcher indicating that the research has been approved.

DEPARTMENT OF EDUCATION
TRAINING AND EMPLOYMENT

10th floor, Education Centre
31 Flinders Street
Adelaide 5000
South Australia
GPO Box 1152
Adelaide 5001

Strategic Planning and Information
Research Unit

Tel: 8226 2472
Fax: 8226 3448
Sharrock.Kylie@saugov.sa.gov.au

DETE 0454/01.k

12 February 2002

Mr David Curtis
11 Scott Avenue
Flinders Park 5025

Dear Mr Curtis

Thank you for your letter requesting approval for your project *'Assessment of Generic Competences in Education'*. Following consideration by a senior DETE consultant, I am pleased to approve your application.

Your project has been reviewed with respect to protection from harm, informed consent, confidentiality and suitability of arrangements and has been found to fit within DETE guidelines.

Please note that it will still be necessary to obtain the agreement of the principals of schools involved. Enclosed is a letter that should be shown to them.

Please supply the department with a copy of the final report, which will be circulated to interested staff and then made available to DETE educators for future reference.

I wish you well with your project.

Sandra Lloyd

Sandra Lloyd
EXECUTIVE DIRECTOR STRATEGIC PLANNING AND INFORMATION

Below is a copy of the letter from the South Australian Department of Education, Training and Employment (DETE) Research Unit to DETE site managers indicating that the research has been approved.

DEPARTMENT OF EDUCATION TRAINING AND EMPLOYMENT

10th floor, Education Centre
31 Flinders Street
Adelaide 5000
South Australia
GPO Box 1152
Adelaide 5001

Strategic Planning and Information
Research Unit

Tel: 8226 2472
Fax: 8226 3448
Sharrock.Kylie@saugov.sa.gov.au

DETE 0454/01.k

12 February 2002

Dear Principal/Director/Site Manager

The research project entitled '*Assessment of Generic Competences in Education*' being conducted by Mr David Curtis from Flinders University has been reviewed centrally and approval granted for access to DETE sites. However the researcher will still need to obtain your agreement to proceed with this research at your particular site. Consequently, your approval for access to your site is sought.

Once approval has been given at the local level, it is important to ensure that the researcher fulfils their responsibilities in obtaining informed consent as agreed, that individuals' confidentiality is preserved, and that safety precautions are in place.

Researchers are encouraged to provide feedback to sites they use in their research, and you may want to make this one of the conditions for accessing your site. To ensure maximum benefits to DETE, researchers are also asked to supply the department with a copy of their final report, which will be circulated to interested staff and then made available to DETE educators for future reference.

If you have any queries regarding research issues, please feel free to contact the Knowledge/Research Officer on telephone (08) 8226 2472 for further information.

Sandra Lloyd

Sandra Lloyd
EXECUTIVE DIRECTOR, STRATEGIC PLANNING AND INFORMATION

Appendix 2: Overseas Generic Skills Schemes

The information presented in this Appendix is based upon material reported in the body and one of the appendices of the report *Employability skills for Australian industry: Literature review and framework development* (Curtis & McKenzie, 2002).

The United States

Initial steps – the early 1990s

In the US, the Report of The Secretary's Commission on Achieving Necessary Skills (1991), (subsequently referred to as the SCANS Report) was released in June 1991. The Commission’s remit was to identify the skills required for employment, to propose levels of proficiency in them, to suggest effective ways to assess them, and to disseminate its findings. On the basis of analyses of the skills required in a range of jobs and in-depth interviews with workers from five major industry groups (Kane et al., 1990; Wise et al., 1990), the report did define what it called “workplace know-how” which comprised a set of five workplace competencies and three foundation elements.

The workplace competencies were an ability to productively use:

- resources;
- interpersonal skills;
- information;
- systems; and
- technology.

The foundation comprised three elements:

- basis skills;
- thinking skills; and
- personal qualities.

The Commission proposed five proficiency levels for workplace know-how: preparatory, work-ready, intermediate, advanced, and specialist. Since its major focus was on adolescents leaving school and either moving into the workforce or into further education, the Commission was most concerned with the second of these levels – entry level for the workforce. The higher levels are intended to differentiate the performance of experienced workers. Members noted that approximately half of all school leavers would not achieve the work-ready level and that this would continue to generate a problem for enterprises seeking to become or remain world competitive, high performance workplaces. The

Commission also acknowledged that many current members of the workforce were deficient in both the competencies and the foundation attributes and indicated that, in order to rectify this situation, lifelong learning would have to become a reality in the United States (The Secretary's Commission on Achieving Necessary Skills, 1991, p. 20).

Although the report canvassed the issue of assessment of workplace know-how, it was not able to make firm recommendations on this matter. A later report did include examples of assessment practices, but, again, no recommendations for particular approaches were made (The Secretary's Commission on Achieving Necessary Skills, 1993). It can be noted that the situation in the United States is rather different from that in Australia. The Report’s authors were conscious of the heavy testing and assessment burden borne by schools in the United States and that workplace know-how was considered not to be amenable to the type of testing that was common in the that country’s education systems. The report recommended that assessment of workplace know-how should be undertaken at Years 8 and 12 so that the initial testing could yield information on areas requiring particular attention, and again at the end of schooling as an appropriate time to certify students’ competences as they entered the workforce or sought further education.

Dissemination of its work was part of the Commission’s brief. This was accomplished though the release of a series of reports (The Secretary's Commission on Achieving Necessary Skills, 1992a, 1992b, 1993). However, these activities did not achieve the degree of implementation that was implicit in the Commission’s charter. Instead, matters of implementation appear to have become caught up in the strong schools reform agenda in the United States.

The Commission identified a challenge. It saw the need to establish a dialogue – based upon a common language – between employers and schools, to set proficiency standards for the components of workplace know-how, and to certify students’ achievements (The Secretary's Commission on Achieving Necessary Skills, 1991, p. 5). Their definition of workplace know-how and the dissemination of the SCANS reports substantially achieved the first aim. Although tentative proficiency standards have been suggested, the lack of a viable assessment strategy has meant that they have not been widely implemented, and therefore a necessary condition for certification is wanting. It might also be the case that the federal SCANS initiative was proposed at a time when similar initiatives were being developed in individual states (e.g, McLarty & Vansickle, 1997; Oliver et al., 1997).

Following the release of the initial SCANS report, a series of subsequent documents appeared. In *A Blueprint for Action* (The Secretary's Commission on Achieving Necessary Skills, 1992a), an implementation strategy was proposed. This was to involve community based action, using the original SCANS report as a starting point and establishing local networks of parents, teachers, students, employers, and unions, and was to be supported through a SCANS Toolkit which comprised a 1-800 number to provide access to SCANS information and resources, a list of Department of Labor regional contacts, contact details of interested organisations, and a bibliography of relevant publications.

In *Teaching the SCANS competencies* (The Secretary's Commission on Achieving Necessary Skills, 1993) case studies and examples of ways in which workplace know-how has been taught and assessed in schools were presented. This document contributed to the implementation of the SCANS competencies. A section on assessment of workplace know-how by John Wirt identified three problems in the assessment of work-related skills. Students' prior knowledge differed and so even in a common context, their performances would vary, leading to concerns about the validity of assessment. Wirt expressed some doubt about the validity of assessing personal qualities. He also questioned the feasibility of assessing students' understandings of complex conceptions such as systems design, relationships, and performance using standard assessment techniques. These concerns suggested that the matter of assessment required further analysis and that some of the competencies might require elaboration with more precise specification of levels of performance.

O'Neil, Allred & Baker (1997) reviewed major schemes in the United States to identify workforce readiness skills, including SCANS and the Michigan Employability Skills Task Force. Most of these generic skills schemes were based on the views of industry leaders and some educators. Importantly, the review by O'Neil et al. showed that they been validated in a series of studies that had examined the tasks routinely undertaken by workers in a wide range of jobs. Common features in the United States skills frameworks included: a core of academic skills, higher order thinking skills, adapting to change, problem-solving, creativity, decision-making, learning how to learn, interpersonal and team skills (communication, cooperation, negotiation and conflict resolution), leadership, dealing with diversity; and personal characteristics and attitudes.

The United States context is rather different from Australia's and from that of the United Kingdom (see below). In the United States, there is much greater local autonomy and a reduced role for central government in policy implementation.

One consequence of this is the lack of a national qualifications framework and therefore some policy implementation options that exist in Australia and the United Kingdom are not available in the US. This may explain the greater emphasis on dissemination rather than implementation strategies for the SCANS workplace know-how skills.

The rationale for SCANS work and the structure of workplace competencies that it proposed were influential not just in the United States but in a range of other countries including Australia. The rationale was essentially that in a highly decentralised school system oriented to general education such as existed in the United States, there needed to be a structure for curriculum development and assessment that provided students with broad skills needed for the workplace. As in many other countries, the early 1990s was a period of high youth unemployment in the United States, and there were considerable concerns both about this and the general competitiveness of American industry.

Subsequent activity – the late 1990s

The work of the SCANS Commission continued through a series of projects run through the *SCANS 2000 Center* at Johns Hopkins University (**http://www.scans.jhu.edu/**). Three themes were evident in their projects: school to work transitions and school reform, welfare to work transitions, and skills development for incumbent workers. Each of these reflects a concern with different aspects of the implementation of workplace know-how.

The first of these themes included a project that reflected the major focus of the original SCANS report, being mainly focused on the work-related skills of school leavers. A range of CD-ROM based simulations were developed for use in high school and college courses to embed the development of the workplace know-how competencies in students' course work. Each had a set of tasks for students to complete for assessment. Further work was planned on assessment and certification of workplace know-how.

The welfare to work theme, a departure from the initial SCANS focus, reflected a concern, also seen in other OECD member countries, that low initial skill levels led to low paid work and to poor employment security (See, for example, OECD and Human Resources Development Canada, 1997). These people were less likely than more skilled individuals to undertake further study, and tended to remain among the lowest skilled members of their communities, and consequently to remain the most at risk of marginalisation.

The skills development for incumbent workers program was foreshadowed in the original SCANS report which referred to the need to develop a lifelong learning approach (The Secretary's Commission on Achieving Necessary Skills, 1991, p. 20). The report recognised that globalisation would push advanced economies towards high skill, knowledge intensive industries; that this move would create a skills shortage as industry sought highly skilled individuals; and it would limit opportunities for workers who lacked the advanced sets of skills being sought. The Career Transcript and Career Management Project sought to combine a record of achievement of workplace know-how skills with other education and training certification and to link these with Career Management Accounts. Career Management Accounts (Individual Training Accounts) were vehicles for dislocated workers to seek career guidance and to fund individually developed education and training programs to enhance the skill base of workers.

21st Century Workforce Commission

In another initiative that sought to address the skills required by industry, the Vice President established the *21st Century Workforce Commission* (21st Century Workforce Commission, 2000b). This Commission was established in response to concerns about America's future competitiveness as a result of technological change and globalisation. It sought to realise the objective of lifelong learning as an element of national competitive advantage. The Commission included representatives from business, labour unions, education, and government. Its task was to synthesise information on workplace learning in order to enhance the skill base of American workers.

The Commission identified four goals. They were:

- Delivered education, training, and learning that were tied to high standards, led to useful credentials, and met labour market needs.
- Improved access to financial resources for lifetime learning for all Americans, including those in low-wage jobs.
- Promoted learning at a time and place, and in manner that would meet workers' needs and interests.
- Increased awareness and motivation to participate in education, training, and learning.

Their report sought to build upon the generic skills thrust of SCANS. These were subsumed under its first goal, which also recommended a system of vocationally relevant, industry endorsed, nationally recognised qualifications. The situation in Australia with the Australian National Training Authority, Industry Training

Advisory Bodies, training packages, and Registered Training Organisations appears to be a close approximation to what was being recommended.

Strategies to increase the level of financial support for lifelong learning included taxation incentives for both employers and employees to provide and undertake education and training programs and the provision of subsidised loans for tuition.

In order to promote learning arrangements to suit the needs of learners, strategies included the use of information and communications technologies, the establishment of partnerships between educational providers and employers, and making available flexible work/study arrangements.

In order to make employees more aware of the value of education and training, information on job skill requirements was to be promoted more widely, to both enterprises and individuals.

The Commission also released a report *Building America's 21st Century Workforce* (21st Century Workforce Commission, 2000a). This report focused on changes in the United States economy that were likely to result from globalisation and the impact of information technology. It anticipated a skills gap as the United States became more dependent on information technology, and it proposed nine so-called 'Keys to Success' in the emerging economic environment:

- Building 21st Century literacy
- Exercising leadership through partnerships
- Forming learning linkages for youth
- Identifying pathways into IT jobs
- Increasing acquisition of IT skills
- Expanding continuous learning
- Shaping a flexible delivery policy
- Raising student achievement
- Making technology access and internet connectivity universal

This report represented a departure from the SCANS emphasis on employment related generic skills towards skills that were specific to the needs of information technology. 21st century literacy subsumed some elements of workplace know-how including thinking skills, teamwork, and proficiency in using technology. However, the core thrust of the document was toward engaging workers in information technology, and other elements of the proposal included strategies for achieving this objective. In common with the original SCANS approach, the implementation strategies included elements of community partnerships involving schools and business and school reform. This document revealed a greater commitment to lifelong learning.

Views of US industry

The US Chamber of Commerce: The US Chamber of Commerce has undertaken to promote a skills-based portable document process in which individuals can record and continuously update their record of skills achievement. (This is similar to the Department of Labor's Career Transcript System).

The National Alliance of Business: The lack of a national US qualifications framework led the NAB to develop a qualifications framework for certification in several industries, including information technology. Numerous certificates were developed by many providers and this led to confusion in the industry about what constituted an appropriate qualification. The NAB sought to establish a set of criteria to recognise both the appropriateness and quality of the certification available. The NAB took a lead in the development of the Workforce Investment Act (WIA). This legislation has been enacted and local business-led Workforce Investment Boards have been established. These, in effect, acted as brokers, identifying education and training needs, promoting workforce development in local business sectors, and engaging and monitoring the performance of education and training providers in meeting these needs.

The National Association of Manufacturers: The NAM encouraged member enterprises to invest three per cent of payroll in education and training for their workforce. The NAM also established, along with its training partner General Physics, a virtual university (NAMVU) to provide effective and convenient courses for employees. NAMVU courses result in certificates, although the lack of a national qualifications framework left some doubt about the value of such certificates in the employment market. The NAM used the virtual university as a means of providing basic adult education and the General Eduction Diploma certification in order to enhance the skill levels of existing workers.

National Retail Federation: The NRF established standards in retail and personal service industries and developed both school-to-work and unemployment-to-work transition programs.

US industry has been represented by a variety of organisations, and many of these are involved in projects building upon the work of SCANS. The industry initiatives described above reflect the absence of strong national approaches to education and training and possibly a reduced commitment to the original SCANS workplace know-how skills, since many of the initiatives would appear to be directed at addressing current skills shortages without the benefit of a skills recognition framework. There would appear to be a return to a focus on job-specific skills independent of a broader skills environment. This stands in contrast with the situation that has emerged in the United Kingdom.

Developments in the United Kingdom

In the United Kingdom, the situation in Scotland was slightly different from that in England, Wales, and Northern Ireland. However, there was an intention to maintain a high level of commonality, and a single United Kingdom approach to generic employment related skills is presented in this section, referring to differences where they occur.

The first point of difference lies in the terms used to describe employment related generic skills. In Scotland, the term core skills was used, while in England, Wales, and Northern Ireland, the terms key skills and basic skills are used and have replaced the originally used term core skills.

Core skills were introduced in England in the early 1980s through the Manpower Services Commission and were tied to the Youth Training Scheme in which unemployed young people were offered work placements and some limited off-the-job training (M. Levy, 1987). The original list of core skills was quite broad, including 103 generic skills in four broad categories. Later, the National Curriculum Council (1990) refined the list of core skills to include:

- communication
- problem solving
- personal skills
- numeracy
- information technology
- competence in a modern (foreign) language

In Scotland, competence in a modern foreign language was excluded.

These skills were to be integrated into instruction for students in the 16-19 year age group. Thus, the core skills were framed primarily as entry level skills for the workforce and did not form part of a lifelong learning agenda.

A significant shift in focus occurred in 1996 with the publication of the Dearing Report (Dearing, 1996; Moser, 1999). In it, core skills were revised relabelled key skills. They were similar to the original list of core skills, with the removal of ‘competence in a modern foreign language’ and the inclusion of ‘improving own learning and performance.’ (It can be noted that the removal of a requirement to learn a foreign language is counter to trends in the European Union). A second important change was a focus on the education and training needs of adults as well as young people entering the workforce. The inclusion of adults as learners and reference to ‘improving own learning’ reflected a move towards lifelong learning. The literacy and numeracy components of the key skills were recognised as basic skills at Entry Level and at Levels 1 and 2 of the National Qualifications Framework.

In Scotland, the core skills list was similar, but omitted ‘improving own learning and performance.’

In England, Wales, and Northern Ireland a Key Skills Certificate was available. This was based upon the first three of the key skills which are assessed at the five levels of the National Qualifications Framework. Assessment for this qualification used both a portfolio of learning or work tasks and an externally administered test in each of the key skill areas.

In Scotland, the Scottish Qualifications Authority established the Scottish Qualifications Certificate. This Certificate was a comprehensive record of each learner’s achievements. It included school and college level awards and was intended to include a Core Skills Profile, although it did not include university qualifications.

Throughout the United Kingdom, core, key, and basic skills are very closely specified at each level and extensive documentation is available on these skills and their levels of performance. Multiple pathways were provided from school to work including opportunities for students aged 14 or over to undertake substantial work-based learning as part of their school education. They were able to substitute studies leading to occupational qualifications for some of the otherwise compulsory areas of learning, and students included key skills within their vocational study. The National Qualifications Framework provided a mechanism for equating the levels of performance across the alternative pathways and qualification types that might be undertaken by learners.

The Learning and Skills Council in England and Wales introduced a new model for funding learning for those aged 16 and above that included a concept of an ‘entitlement’ to an education and training place. This entitlement was established for 16-19 year-olds, and was largely concerned with ensuring an increased supply of appropriate programs, and providing funded places in them. The British Chamber of Commerce criticised limiting the entitlement to those aged 19 years or less, arguing that all 16-24 year-olds should be entitled to free tuition in pursuing qualifications up to Level 3, as recommended by the Skills Task Force. The funding reforms attempted to embody the notion of funding following the learner, and included greater incentives for providers of education and training programs to achieve agreed learning outcomes.

Views of British industry

The Confederation of British Industry (CBI) was a prime mover in seeking to have the initial set of core skills recognised in the United Kingdom. Since that time, many changes occurred in the education and training sectors in the United Kingdom. These included major revisions to assessment and the qualifications framework for secondary students, as well as changes to the national qualifications framework that covers vocational and higher education awards. There were also changes to the administrative arrangements and bodies responsible for the oversight of this framework, the qualifications covered by it, and associated curricula. These changes occurred in the context of a commitment to realise the rhetoric of lifelong learning.

The CBI argued for a single nationally coherent qualifications framework that encompassed academic, broad vocational, and employment-specific education and training (Confederation of British Industry, 1998). This framework recognised five levels of achievement for each of the three categories of study and acted to facilitate learning pathways through and between the categories of learning. The CBI argued that the framework needed to be national in coverage, coherent in that it encompassed all awards, appropriately assessed, and broadly based using all six Key Skills (rather than a subset of them as was the case for the Key Skills Certificate) as the basis for this breadth of coverage.

In several documents, the CBI endorsed the key skills as forming a sound basis for the ongoing employability of workers and for recent school, college, and university graduates (Confederation of British Industry, 1999, 2000a, 2000b). However, the CBI also wanted the notion of employability skills to be expanded to include a broader set of individual assets, including values and attitudes related to employment and a broader set of generic skills including modern foreign languages. Further, they wanted individuals to be able to capitalise on their skills by being able to represent themselves and their skills effectively.

However, the CBI recognised that basic skills are also required by many, especially some mature workers whose initial education did not equip them with these skills and whose continued employment in a changing economy was likely to be under threat. In summary, the CBI continued to endorse the key skills, and believed that they needed to be implemented at all levels of education within a coherent framework.

Developments in Canada

Canada is a highly decentralised country that lacks a strong tradition of close engagement between employers and educational institutions. Nevertheless, during the 1990s Canadian industry began to take major initiatives to bridge this gap. The Employability Skills Profile (ESP) played a key role in this regard. The ESP was seen by many as the conceptualising tool that had been needed to encourage school systems to redefine their goals and relationships with the world of work and its methods (OECD, 1998).

The general policy thrust in Canada had been towards fewer demand-side labour-market measures that used to be central to government strategies – for example, direct job creation, wage subsidies, and tax incentives to employers – and much greater emphasis on measures designed to improve the skills and overall employability of workers (Marquardt, 1998). This was true for all age groups, but it was especially so in the case of youth. Governments in Canada had increasingly favoured measures that encouraged further formal education or that facilitated school to work transitions.

In this environment, there was evidence that younger workers who left full-time education with high level qualifications were engaging in a high level of self-initiated, career-oriented formal education and training without necessarily drawing on support from employers (Marquardt, 1998). Some observers therefore suggested that it was as important to develop policies that supported the efforts of individual young workers to develop their skills through self-initiated training as it was to promote employer-sponsored training. The Canadian emphasis on employability-related skills was increasingly aimed at developing the skills and attitudes required to be an effective lifelong learner.

In the early 1990s the Conference Board of Canada sponsored a series of projects that attempted to respond to the question of educators: “What are employers looking for?” (Conference Board of Canada, 1992). The Conference Board was a forum for leaders from business, education, government and the community, that sought to address concerns about education in Canada. The projects were organised through the National Business and Education Center, an auxiliary of the Board.

Through research and consultation with employers of all sizes, the Board developed an Employability Skills Profile that identified the generic academic, personal management, and teamwork skills that were required, to varying degrees, in every job (Conference Board of Canada, 1992). Three broad domains of employability skills were identified:

- Academic skills: those skills which provide the basic foundations to get, keep and progress in a job and to achieve the best results.
- Personal management: The combination of skills, attitudes and behaviours required to get, keep and progress in a job and to achieve the best results.
- Teamwork skills: those skills needed to work with others in a job and to achieve the best results.

A noteworthy development in Canada was the introduction of the 'Employability Skills Toolkit for the Self-Managing Learner.' The Toolkit was developed by the Conference Board in consultation with schools, provincial ministries of education, trainers and HR professionals. The Toolkit provided information on what employability skills were, and the ways that they could be developed and demonstrated at home, in education, work and the community. It was a resource that could be used by learners in developing a portfolio of their generic employability skills.

The DeSeCo Project

The DeSeCo project (The Definition and Selection of Competencies) was an OECD project developed under the umbrella of the Indicators of National Education Systems (INES) program (Salganik et al., 1999). The INES program sought to develop indicators of system outputs – the competencies developed by students in preparation for life beyond formal education. In establishing the DeSeCo project, there was a concern to ensure that the effectiveness of education systems was measured using a broader range of indicators than was available from subject-specific assessments that had been a feature of earlier attempts to compare the outcomes of educational programs. Salganik et al. (1999) claimed that earlier projects had been developed without the benefit of a thorough and sound theoretical and conceptual basis. The DeSeCo project sought to build upon the work done in the International Adult Literacy Survey (IALS), the Cross Curricular Competencies Project (CCC), and the Human Capital Indicators Project (HCI) (Salganik et al., 1999).

Specifically, the DeSeCo project set out to establish sound and broadly based theoretical conceptions of competencies. It recognised that these competencies had to apply to school and work settings but equally to life situations beyond those areas. Rychen and Salganik (2000) noted that the various national attempts to develop definitions of generic skills can be characterised as:

- boosting productivity and market competitiveness;
- developing an adaptive and qualified labour force; and
- creating an environment for innovation in a world dominated by global competition. (Rychen & Salganik, 2000, p. 3)

These were characteristics of generic employability skills. In a broader conception, generic skills are seen to be important because they also contribute to:

- increasing individual understanding of public policy issues and participation in democratic processes and institutions;
- social cohesion and justice; and
- strengthening human rights and autonomy as counterweights to increasing global inequality of opportunities and individual marginalization. (Rychen & Salganik, 2000, p. 3)

In order to achieve a broad theoretical consensus, the project commissioned a series of expert papers from individuals and groups from the disciplines psychology, sociology, economics, anthropology, politics and philosophy. While there was considerable agreement that there are generally applicable competences that are relevant at least to developed western economies and societies, Goody, an anthropologist, argued against the dominant view within the DeSeCo project. He argued that schools, a focus of the project, were not the only social institutions through which individuals developed competences and that the roles of family and friends, among others, should be taken into account in defining competences. He also argued that the cultural context of individuals defines what is valued, and that because of cultural differences, it is not feasible to define universal key competences (Goody, 1999).

Haste (1999), in arguing from a psychological perspective, identified five ‘key competencies’. They were:

- technological competence;
- dealing with ambiguity and diversity
- finding and sustaining community links;
- management of motivation, emotion, and desire; and
- agency and responsibility.

What was particularly interesting in Haste’s list of key competencies was that they represented a higher level of abstraction than those described in any of the schemes described in this report. Each competence was defined very broadly. For example, technological competence involved the meta-competence of tool use, a preparedness to acquire new skills and to relinquish those that are no longer needed, and an ability to deal reflexively with new developments. Second, they gave greater emphasis to inter- and intra-personal attributes than did most of the previously discussed schemes. Third, they introduced a values perspective in talking about individuals’ agency and responsibility.

The DeSeCo project focused upon the definition of competences from multi-disciplinary perspectives but did not develop methods for assessing and measuring them. However, this was clearly the primary intention of the project. The director of the DeSeCo project, Heinz Gilomen, in the foreword to Salganik et al. (1999), referred to the changing social, economic, and political circumstances brought about by globalisation and new technologies. He observed that the future well-being of individuals, enterprises, and societies depended increasingly on high levels of knowledge, skill, and competence, and that there was a consequent imperative for policy-makers to ensure that the social institutions that are responsible for these outcomes are able to deliver them. There was thus a need for soundly based instruments to monitor their performance.

Developments in Europe

The European situation was far from homogeneous. Even among EU Member countries, there were considerable differences in many facets of economic and social activity. There was also variation in the organisation of educational programs among member countries. For example, in Germany and Austria there was a substantial separation between general academic and vocational education streams from an early age, while in Sweden this division became less apparent as the previously segregated vocational and general secondary schools were integrated in the early 1970s (Abrahamsson, 1999). In Austria, the view expressed by industrialists (Piskaty et al., 2000), that this separation, which had been productive in the past when Austria was a major manufacturing nation, led to narrow skills specialisation and was no longer able to produce individuals with the broader and more flexible skills that were being required by emerging knowledge industries. There was also evidence that the traditionally very strong apprenticeship system was breaking down in Austria, partly because the companies that have provided it were finding apprenticeships less attractive and partly because of a narrow and possibly ‘old’ range of trades for which apprenticeships were available (Piskaty et al., 2000). Some of these matters, for example the availability of apprenticeship places, remained concerns in vocational education and training in Australia.

One of the concerns of European countries was the preservation of their unique languages and cultures within an encompassing European economic union. The issue of language preservation and maintenance would appear to have driven the requirement for proficiency in a second European language as a core skill.

Many of the industrialised countries of Europe suffered economic reversals as some manufacturing activity has been relocated to developing economies. Young people bore the brunt of labour market downturns in Europe as in Australia, and in both economies, alternative pathways from education to work continue to be negotiated.

An important influence in much of Europe was the pressure to establish lifelong learning. This arose from several sources. There was recognition in many countries, for example Sweden (Abrahamsson, 1999), that a substantial proportion of the established workforce was poorly qualified and had a narrow and industrial skills base. There was also recognition that, with static but ageing populations in much of Europe, most of the future workers in the emerging knowledge economy were those already in the workforce. Thus, there was an obvious need continuously to upgrade the skills of existing workers, and this led to calls for continuing vocational education.

The recognition of the demographic problem did not appear to be as strong in Australia until 2003 with the release of the *Intergenerational Report* (Australian Treasury, 2002), although the implications were made apparent in an earlier report that would appear not to have been widely recognised (Aungles, Karmel, & Wu, 2000).

Sweden

Abrahamsson (1999, pp.49-91) outlined a history of the development of vocational education and training in Sweden that was remarkably similar to the British system. His report showed a transition from a guild system, through an informal and privately supported apprenticeship system, to one in which vocational education was supported and provided by the state and with little evidence of the partnerships between education and industry that characterised the German and Austrian apprenticeship model. Major challenges to Sweden's industrial position were followed by changes to the education system in order to make it responsive to the needs of industry. The report documented changes to the Swedish education system in the early 1970s to merge the previously separate general academic and vocational streams to make the system both more relevant to the needs of industry for more flexible workers, and also to make vocational options more attractive to students.

One of the differences between Sweden and Australia was a concern that, as a result of successfully meeting targets for youth participation in education, an equity gap opened between younger and older Swedes, with older workers having lower qualifications and therefore reduced workforce opportunities. Legislation was enacted to provide older and less qualified workers with access to study leave and forms of study support to enable them to upgrade their skills (Abrahamsson, 1999, p. 52). Bridging the equity gap was one of the drivers for policies to implement continuing vocational education and more generally, lifelong learning arrangements.

Workplace-based education and training became a feature of continuing vocational training in Sweden. Participation in this form of training varied from 23 to 42 per cent during the 1990s, which placed Sweden well above other European countries and above the United States (Abrahamsson, 1999, p.84). However, as in many other countries, those workers who were already well qualified participated in this form of education and training at a much greater rate than those who were poorly qualified. Thus this form of education did not address fully the equity concerns in Sweden.

There was a change in the focus of education and training for employment to one which placed greater emphasis on flexibility and the skills that underpin it.

> Policy makers are increasingly underlining the importance of general education and generic competencies. In practice, this leads to more policy attention on broad programmes instead of early specialisation e.g. an apprenticeship model adapted to a certain vocation. (Abrahamsson, 1999, p. 121)

The high speed of labour market transformation and job turnover had an impact on the need for skills and competence of the work force. A significant increase in the provision of competence development at work was of crucial importance for the security and wealth of employees, but also for Sweden’s economic survival in a context of increasing international competition. Competence was no longer just a question of occupational skills. It also comprised the capacity to solve problems, to learn and adapt to changes and to communicate. Social skills were becoming more important in order to work in teams and projects and to meet customers or subcontractors. The success story of an enterprise, to a large extent, depended on its capacity to adapt its production systems. It was a challenge that called for a flexible work organisation and highly skilled employees. (Abrahamsson, 1999, p.119)

Abrahamsson also recognised the need for enterprise skills: “There is also a debate today on the need for creating the spirit of enterprise in upper secondary education programmes” (p.120). Thus there is evidence of a shift to a more broadly conceived skills base as an outcome of all educational programs.

Austria

The situation in Austria is of interest because of its dependence on a well established and successful apprenticeship model of vocational education. However, support for this model declined because it entry into apprenticeships depended upon students making early choices between an academic general secondary education and a vocationally oriented one (Piskaty et al., 2000). Employers were reluctant to continue their involvement with these apprenticeship arrangements because of:

- the tendency for more training time to be spent at school to the detriment of time spent within the company;
- the large amount of administrative red tape to be handled, which is particularly onerous for employers taking on apprentices for the first time;
- the high cost of providing apprenticeship training;
- too stringent and outdated regulations on what activities apprentices are not allowed to carry out during their training. (Piskaty et al., 2000, p.104)

This indicated a conflict between the immediate interests of employers, who might place greater value on job-related skills, and policy-makers who perceived the medium term requirements for more generally applicable competences. However, there were other tensions. Apprenticeships became less popular than they had been in the past because they were oriented to established manufacturing industries rather than to the newer industries. Blumberger (1997) showed that although there had been a steady decline in the number of places being sought, there was an even greater decline in the number of apprenticeship places available. Arrangements were established to overcome both supply and demand aspects of apprenticeships (Piskaty et al., 2000, p.104).

Curricula were also modified to place greater emphasis on generic skills: “Most curricula now reflect the importance attached to strengthening personal development and social skills.” (p. 106)

The European Training Foundation

The European Training Foundation (ETF) had primary responsibility for oversight of the Phare program under which considerable aid was provided to Central and Eastern European countries and also to some other non-European states. The objectives of this program were to enhance the relevance, efficiency, and capacity of vocational education and training systems in target countries. Under the program, many small projects with quite specific targets were funded in recipient countries. Major themes that informed the program included:

- the role of the state and the social partners in supporting the links between education and training and the economy;
- the contribution of the world of work to education and training;
- education and training to underpin economic growth processes;
- supporting people at the interface between education/training and work; and
- the role of teacher training in linking education and training and the economy. (Arbeitstelle für vergleichende Bildungsforschung & European Training Foundation, 1999)

The trends that informed the projects that were supported under the program include:

- the demand for new higher level and core qualifications;
- the quest for effectiveness and quality of education provision;
- new approaches to the governance and financing of education;
- diversification of education provision and its tailoring to individual needs;
- enhanced responsibility of institutions and individuals for the outcomes of the education process; and
- a reappraisal of the interaction between education and economic change and development. (Arbeitstelle für vergleichende Bildungsforschung & European Training Foundation, 1999)

These trends were evident in other more developed countries. However, in Central and Eastern Europe, there were important differences. For the candidate countries wishing to join the EU, the pace of change was great and the organisation and infrastructure on which this change had to be built was poorly established. In many European Union member countries, for example Germany, Austria, and France, there were traditional partnerships between state-funded education systems and private companies. This 'social partnership' was unknown in many candidate countries, as there was little private enterprise.

In their review of the Phare program, Viertel and Grootings (2001) noted that there had been a need to reform curriculum content and delivery and to make education more responsive to the needs of the emerging market economy (p. 31). As part of the process of curriculum renewal, a training manual on the preparation of training curricula was developed (Mansfield & Schmidt, 2000).

The Phare program review (Viertel & Grootings, 2001) also noted several approaches that might have relevance to change management in Australian education. One that was reported upon was the use of pilot schools as focal points for specific reforms followed by a deliberate dissemination process to mainstream the changes. In general, the review noted that this model was unsuccessful either because resources were not made available for the dissemination or because there was a lack of "political will" to mandate the dissemination process. The authors cited an exemplary model of this approach in Lithuania in which pilot schools had been partnered with non-pilot schools and in which a cascade or 'each-one-teach-one' model of dissemination had been implemented (p. 27). The risks that successful pilot ventures might fail to transfer, but the benefits when they are disseminated, suggest that for change to be successful, specific and deliberate approaches to dissemination and change management would be required.

Within an overall cautious review of the program, Viertel and Grootings (2001) made many observations about the characteristics of successful projects. These included:

- the integration of work and learning;
- the establishment of structures to facilitate the transition from school to work;
- the postponement of career choices to a later age;
- the de-specialisation of education and training programmes;
- increased possibilities to switch horizontally between educational paths and to progress vertically along the educational path;
- an increased autonomy and innovative capacity at school level;
- a shift from input to output control mechanisms;
- the development of continuing vocational education by giving various incentives to encourage the investment in training by both employers and individuals; and
- even more radically, the development of lifelong learning systems allowing to go back and forth between or combine education, training and work during the whole life period of an individual. (Viertel & Grootings, 2001, p.36)

Some of the reforms that they advocated, such as a national transparent qualifications structure, are well established in Australia. Support for continuing vocational education and lifelong learning are seen as desirable outcomes for developed European Union countries and for beneficiary candidate countries.

Implications for Generic Skills in Australian Education

Attempts to introduce generic skills into the education and training systems of other countries have some implications for Australia. There are, however, some important contextual differences that must be considered before seeking to adapt initiatives or project outcomes from other countries to the Australian situation.

Among the reasons for the introduction of generic skills schemes into other countries have included a concern with rising youth unemployment. This was especially apparent in the 1980s and 1990s. A common view was that compulsory education was inadequate to prepare young people for the demands of modern work arrangements and that broader skills would be required by young people. A second, and related concern, was that developed economies were experiencing increasing competition and that employment in many industries, especially manufacturing, was in decline. The decline was precipitated by the relocation of manufacturing to emergent low-wage economies and by the introduction of technology that required fewer, but high-skill workers. These changes led to less predictability in the types of jobs that were expected to be required and therefore to a need for education and training systems to prepare young people for a range of occupations rather than for particular ones. This preparation had to be broad and include generic rather than occupation-specific skills.

Some important contextual differences are apparent. The SCANS initiative in the United States is interesting because of the way in which the elements of workplace know-how were defined – through analyses of the tasks undertaken by individuals in a wide range of jobs. Although there were examples of how they could be taught and assessed, the skills that comprised workplace know-how appear not to have been implemented widely in the United States. This would appear to have arisen because the education systems of the United States are highly decentralised compared with the Australian situation. SCANS was a federal initiative, but state and school district boards were responsible for curriculum, so there would appear to have been limited authority for federal initiatives. A second and related contextual factor would appear to have been the lack of a national qualifications structure in the United States. The observation that a variety of industry organisations attempted to implement their versions of qualifications and skills within them would suggest that there had been limited success in attempts to generate a national framework. Without a national qualifications framework, especially for vocational education, there was limited scope to implement a common approach to skills development, whether those skills were vocationally specific or generic.

The situation in Canada was similar in several respects to that of the United States. Provinces had responsibility for education and training. Generic skills initiatives were pursued by national government agencies and with the support of the national peak employer organisation (Conference Board of Canada, 2000a; Human Resources Development Canada, 2000). It would appear that the lack of central authority led to the implementation of the employability skills initiative through the provision of resources rather than through a qualifications system.

The context of the United Kingdom was similar to that in Australia. While there was a unified national education system in the United Kingdom, Scotland pursued similar but slightly different initiatives compared with the remainder of the country. The United Kingdom had a national qualifications system and through this, the core skills and key skills initiatives were able to be developed and implemented. Implementation failed in some sectors of the United Kingdom education system (Payne, 1999, pp. 12-16). The failure of the Key Skills Certificate could be attributed to what was and was not assessed and to the assessment model (Turner, 2002). The assessment scheme involved a combination of a standardised test and a moderated work- or school-based component. The model was unwieldy, expensive and provided little information that was not available from other indications of student achievement.

The work completed under the DeSeCo project provided very useful source material in reconsidering the definition of generic employability skills in the Australian context.

The contexts of several European countries revealed that some influences were similar to those that have been apparent in Australia. The decline in employment in some traditional industries in Austria and Sweden also occurred in Australia. However, their training systems were different from Australia's. The need for generic skills as a component of their strong vocationally-specific skills training has been recognised. These countries, especially Sweden, also experienced an equity challenge. While the focus in Australia and the United Kingdom was on young people and the skills they would require for the emerging work context, in Sweden, older workers were found to have missed the opportunities for education and training that younger cohorts had enjoyed. This may also be a problem in Australia, but one that has not received the same level of attention as has the need to skill younger workers.

Appendix 3: Examples of Generic Skills Assessments

In this appendix, examples are presented to illustrate the models of generic skills assessment that are described in Chapter 4.

Standardised assessment

Graduate Skills Assessment

The Graduate Skills Assessment (GSA) was developed through extensive consultations with university staff and employers about what characteristics of graduates those staff believed to be the most important. A list of 17 skills and attributes was identified through these consultations (ACER 2001a, p. 27). In order to limit the scope of the trial of the instrument, four of these attributes were selected for development, namely communication (argument and report writing), problem solving, interpersonal understanding and critical thinking. These constructs were ranked as the most important by university staff and the first three were the top three among employer rankings. The employer's fourth ranked skill was analytic thinking, and it was regarded as being sufficiently close to critical thinking to justify the inclusion of that construct. It is worth noting that the 'big three' generic skills of communication, teamwork (interpersonal understanding) and problem solving were identified as the highest priorities among university staff and employers.

It was intended to develop the scope of the instrument further by adding some of the other identified attributes including basic skills, management skills, information technology skills, research skills and additional items to discriminate among high achievers. There was also an intention to develop the test for online delivery. This would improve the cost-effectiveness of the testing. Thus far, these developments have not occurred. The test is secure, but sample items are made available to candidates so they are familiar with the format of the test and the types of questions that will be asked (ACER 2003). In this way, no candidate can be advantaged by special coaching as all students have access to the same information.

In several phases, the test instrument was administered to several thousand students from 19 Australian universities and to a sample of 400 senior secondary students. The three scales based on multiple-choice formats (problem solving, critical thinking and interpersonal understandings) had internal scale reliabilities in excess of 0.8, indicating that the scales were coherent representations of the constructs. Some differences in scores on the various scales were identified by sex, with females performing better than males on the written communication component but less well on the problem solving items. Differences were also detected by field of study. These differences are likely to reflect differences in general cognitive abilities, indicated by the range of tertiary entrance scores in various fields of study. Some of the differences in the relative strengths are likely also to reflect personal attributes associated with career preferences. For example, the relatively good performance of nursing students on the interpersonal understandings scale is consistent with the choice of this career by people with well developed interpersonal skills.

In summary, the GSA has good psychometric properties. The model has the potential to be developed to include other constructs and the technology exists to make the test available online. Scoring the written elements of the test (the report and argument writing tasks) is one of the more expensive aspects of the testing. In the future, the GSA model could be developed to assess the employability skills of senior secondary students. Additional scales will need to be developed to assess all eight employability skills.

PISA Problem Solving Assessment

The Programme for International Student Assessment (PISA) assessment of problem solving was conducted in 2003, after the present study was undertaken. The first publication of the problem solving data appeared in 2004 (OECD, 2004). It is included in this review because it sets a benchmark for large scale standardised assessment.

PISA is a major cross-national program in which students near the end of their compulsory schooling (15-year-olds) are tested on a range of so-called ‘literacies’. The notion of literacies is an important orientation of the PISA testing. A key point is that the testing is less focused on specific school curriculum content and more on students’ abilities to understand, apply, interpret and draw inferences from given information. There are alternative testing programs, for example Trends in International Mathematics and Science Study (TIMMS) conducted by the International Association for the Evaluation of Educational Achievement (IEA) that concentrate on the achievement of curriculum content. There is, no doubt, a strong correlation between mathematical literacy as measured in PISA and achievement on curriculum-based tests, since a learner who is capable of reasoning with numeric data is likely to use that ability in their school mathematics tests. The PISA tests are repeated in a three-year cycle, with tests having been run in 2000, 2003 and 2006. In 2003, students were tested on reading comprehension, mathematical literacy, scientific literacy and problem solving.

The PISA tests of problem solving are the particular focus of this review because the problems are designed to test students' abilities to apprehend and understand problems, to identify relevant given information, to apply that information, to reason deductively and inductively with it, to reflect on tentative solutions and to communicate their findings. The development of problem solving as a construct and of problem tasks that exercise the range of problem solving processes were described in *Problem solving for tomorrow's world* (OECD, 2004, pp. 25-39). There, sample problems were presented and their relationships to the component processes and the cognitive demands of each were explained.

In advance of the testing, three levels of student performance were anticipated. These levels were described as standards that characterised students' approaches to problem solving, given their observed behaviour on the assessment tasks. Level 3 problem solvers were considered able to identify relevant information in the task description, to reason about it, to identify possible solutions, to reflect on those solutions and make judgments about them, and to communicate their results. These learners were described as 'reflective, communicative problem solvers.' Learners at Level 2 were considered able to identify information and reason using it, but were less likely to monitor their solution attempts, reflect on them and communicate their solutions clearly. They were described as 'reasoning, decision making problem solvers.' At Level 1, 'basic' problem solvers were considered able to identify given information and to use it, but were unlikely to draw inferences from it. In the cross-national study, a proportion of young people were identified as operating below Level 1.

For each problem task that was developed, an analysis of the task revealed the processes that would be involved in its successful completion. Further, the shortcomings in unsuccessful solution attempts were diagnosed. Thus, the skills that differentiated successful and unsuccessful problem solvers were identified. Scoring keys were developed for each problem, so that various incomplete or inaccurate solutions could be given partial credit for the skills that were evinced.

This approach to task design and scoring, using a principle-based analysis of the skills that were implicit in problem solving, enabled well-targeted assessment and informative feedback. Although the primary purpose of PISA is national comparisons, individual learners can be informed about their level of problem solving skills. Further, the assessment scoring process is transparent, so that learners, given this feedback, can be guided towards the next level in their quest for improvement.

Because the primary purposes of the PISA testing are international (and jurisdictional) comparisons and comparisons among population sub-groups, the results of large numbers of students are aggregated to provide estimates of national and sub-group means. It is not necessary that the scores assigned to individual students are very precise, since the results of students within defined groups are averaged. In reporting individual achievement, especially for high stakes purposes such as selection and sorting by potential employers, it is necessary to ensure that the precision of individual measures is consistent with these purposes. Thus, if items such as those developed for the PISA testing are used for employability skills assessments, each candidate needs to respond to a greater number of items than they do in the PISA assessment.

The standards-based approach to defining performance levels for generic skills was developed for the schools sector. A particular advantage of the set of performance standards defined for problem solving in PISA was that teachers could provide feedback to students who had been graded at Level 1 or Level 2 and advise them about what the students would need to do to improve their performance and achieve the next level. Such a practice would lead to improved aggregate achievement of employability skills within cohorts of young people completing senior secondary schooling.

Employability Skills Profiler[29]

The Employability Skills Profiler (ESP), funded by DEWR, was developed and trialled by Chandler MacLeod during 2006 and was implemented during 2007. Its purpose was to assess the employability skills of unemployed persons who were clients of Jobs Network and Disability Employment Network service providers, to develop a profile of clients' employability skills, and to match their profile to the skills profiles of jobs.

> The Employability Skills Profiler (ESP) objectively assesses a job seeker's generic or transferable skills and shows how well the job seeker's skills fit with the skills required by over 1000 job types. (DEWR, 2007)

The tool was described in some detail by Curtis and Grant as 'a tool for the objective assessment of employability skills'. It has been used to measure both the skill levels of individuals and the skill requirements of jobs, and therefore, to match individuals to jobs.

[29] Other than those sources specifically acknowledged, the information presented on the ESP was derived from two sources. One was a paper presented by Rob Curtis and Lieschen Grant (both DEWR staff members) at the 2005 NCVER Research Conference at Wodonga. The paper was not published, but extensive notes were taken by the author of this thesis at the conference. The other source was a telephone conversation on 2 August 2007 between the author and Mr Kevin Chandler, Executive Director of Chandler MacLeod.

Chandler MacLeod analysed the employability skills and concluded that the eight skills could not be assessed coherently but that they could be represented as 36 constructs. The 36 constructs were mapped onto a variety of ability and personality measures for which existing instruments were available. It can be noted that the ability and personality instruments treated these constructs as individual traits that were relatively stable over time. This is inconsistent with a view of employability skills as elements of human capital that can be and need to be augmented.

Subject matter experts identified the 36 employability skill elements required for more than 1100 jobs. The skills identified were classified as either required or desirable and a system of weighting was used to assess the level of performance level required for each skill. Five performance levels were identified for each component.

The profile of skills established through the administration of the ESP is being used to match people to the requirements of positions. It is used to enable individuals to identify skills strengths and skills gaps and to enable employers to match people to jobs.

Because of the match between individuals and job requirements, it enables the fit with the preferred jobs to be assessed, and therefore can be used as a career guidance tool taking into account local skills shortages. The ESP can be customised by employers to meet the specific skills needs of particular job roles. It can be used to develop an individual training program to improve job and person fit.

In addition to the job-match report, the ESP also resulted in a paper report for individuals focusing on individuals' strengths and was intended to be 'a feel-good report.'

The proposed matching of person and jobs highlights one of the definitional issues raised in the brief discussion above about what is meant by a skill and by the degree of abstraction with which generic employability skills are described. The skills requirements of particular jobs are, by definition, job-specific. In the set of skills identified for a particular position, it is possible to distinguish those skills that are highly job-specific and those that are common to many jobs. Those that are common can be classified into groups corresponding to the eight employability skills. However, when the job-specific skills are removed and a profile of the employability skills is developed, any given profile is likely to apply to many jobs. Thus a claim for the utility of the ESP as a tool both for matching unemployed persons to jobs and for reporting on the achievement of generic skills must be questioned.

The online methodology and the ability to generate reports quickly from the system are advantageous. The output of such testing can be valuable as a diagnostic tool. A disadvantage of the system is that it uses standardised ability and personality instruments. The eight employability skills were mapped onto elements of existing psychological tests that assess what were presumed to be fixed traits. This represents a re-definition of the employability skills. The assumption that they are fixed traits does not augur well for their development.

Key Skills Certificate

A set of key skills was defined in the United Kingdom. The skills were:

- Communication
- Problem solving
- Working with others
- Application of number
- Information technology
- Improving one's learning and performance.

A Key Skills Certificate for 16–19-year-old students was introduced into England and Wales in 2000. This qualification targeted communication, application of number and information technology skills – the basic key skills. Those skills not covered by this qualification were re-labelled 'wider key skills'. Initially, all 16–19-year-old students in any form of education and training were expected to gain this qualification. The qualification was voluntary, although there were perceptions that it was compulsory and this led to an initial uptake that subsequently waned (Hodgson & Spours, 2000). By 2002, few of the students enrolled in academic courses, mainly at selective schools, participated – presumably because young people enrolled in academic tracks were regarded as competent in these skills, which were being developed adequately through the academic programs. Uptake of the program was greater in colleges offering vocational qualifications (Powell et al., 2003), and this differentiation would appear to have indicated that the qualification might have signalled 'low ability' rather than affirming the achievement of 'employment-related skills.'

It can be noted that the policy context of the UK was rather different from that of Australia. Two reasons for introducing the Key Skills Certificate were to broaden the upper level secondary curriculum, in which students often took only three A Level subjects, and to bridge an apparent divide between academic and vocational study (Hodgson & Spours, 2000). If the qualification has had an effect, it might have been to exacerbate that divide.

Students who completed the qualification at Level 3 were awarded 'a very generous' (Powell et al., 2003) 60 points towards their university entrance score (UCAS). This encouraged students in academic programs to undertake the Key Skills Certificate, but not all universities used the tertiary entrance score – the more selective universities were less likely than others to recognise it. Thus, students and their parents had received mixed messages about the value of the qualification.

Hodgson and Spours (2000) summarised the problems of the Key Skills Certificate as having 'too many complications, too little currency and [being] too difficult to achieve'. Turner (2002) identified two main concerns with the Key Skills Certificate. He noted "the unsatisfactory experience of both educators and employers regarding the assessment of the three basic key skills as they were developed into a national qualification. Testing procedures became too complicated and there is real concern that [the qualification had become] assessment and not learning dominated." The Certificate was based on two components – a standardised test that was centrally administered and local performance assessments conducted by teachers. The teacher assessed component was subject to moderation. The central problem would appear to have been the complexity of the moderation processes of tests of the basic key skills and portfolio evidence of these skills leading to the award of the qualification. Employers regarded the basic key skills as of lower importance than the wider key skills that were not assessed and were hesitant "to place great store on a certificate" or even a portfolio of evidence as proof of "having these skills" (Turner, 2002, p. 17). To address this lack of breadth, Hodgson and Spours (2000, p. 30) suggested that the wider key skills should become the focus of the qualification.

In summary, the Key Skills Certificate was perceived to be too narrow and a test of basic skills only. This added little to what employers and higher education providers would know about candidates based on other school achievement information. The qualification was complex, in that it required an externally set test and a portfolio of activities that was assessed by the provider. The two sources of information then had to be moderated. Finally, neither employers nor further education providers placed much weight on the qualification, so there was little point in students doing it or schools offering it.

The lessons for Australia from the Key Skills Certificate is that the report of achievement must have credibility with employers (and others), it must provide information that is not already conveyed by other achievement results, it needs to provide information about complex skills and the assessment regime must be relatively simple. This suggests that moderation may add a level of complexity that may not be acceptable in Australian jurisdictions.

Summary of standardised testing

A common feature of the four examples of standardised assessment is their focus on cognitive skills. In the Graduate Skills Assessment (GSA), the assessment of teamwork used interpersonal understandings, a knowledge component of teamwork, as a proxy for the complete construct. This suggests that it is more difficult to measure constructs that have a more affective than cognitive character.

The difficulties experienced in the United Kingdom over the Key Skills Certificate suggest that the employability skills should not be restricted to a narrow focus on basic skills and that the skills that are more difficult to assess, such as teamwork, should be included in any assessment of employability skills. The complexity of the Key Skills Certificate assessment, attributable to the moderation of internal and external assessments, needs to be avoided. Both internal and external assessments may occur, but they should be recognised as serving different purposes and reported separately. Engendering acceptance for reports of employability skills achievement by employers and other education providers is important in ensuring that students, parents, teachers and schools are prepared to invest in developing, teaching, assessing and reporting employability skills.

Standardised testing is being used to assess skills having strong cognitive components, but with the exception of the Employability Skills Profiler, not those of a more affective character. The Employability Skills Profiler does not assess the more affective skills such as teamwork; rather it uses elements of personality instruments to assess dispositions, which it must be assumed, are proxies for the target skills. It is possible to assess attitudes, dispositions, motivations and values, but it is more difficult to assess these than it is to assess cognitive skills.

The PISA and GSA assessments meet reliability criteria. If the PISA assessment of problem solving is to be used as a model for the assessment of this and other employability skills, a pilot study to determine the precision of estimates of student achievement for different numbers of test items is required. The three performance levels (with a fourth implied) of the PISA problem solving assessment provide useful information to and about learners. Similarly informative descriptors were developed for the PISA Reading Literacy and Mathematics Literacy assessments, with five and six levels identified, respectively (Thomson et al., 2004, pp. 94 & 43). The identification of knowledge components of some of the so-called ‘hard-to-measure’ skills such as the interpersonal understandings component of teamwork may provide a partial solution to the testing of the less tractable employability skills.

A central problem highlighted in the Employability Skills Profiler is the specificity of the skills assessed. General education is designed to prepare students for personal development, community and civic engagement and workforce participation. With the possible exception of vocational education in senior secondary schooling, the workforce participation objective is necessarily broad, as students may move into very varied work roles. The notion that only specific elements of employability skills and not the employability skills themselves can be assessed as coherent constructs suggests that the current operationalisation of the skills through their facets may be flawed. The construction of these skills as fixed traits is antithetical to the developmental role of general education.

The Key Skills Certificate in the United Kingdom provides a warning. It is apparent that the qualification did not add to what was known about students who were performing well academically and the test ceased to be used for these young people. The criticism that it was too narrowly focused and that the wider key skills were of greater interest to industry suggested that efforts needed to be directed at assessing those constructs that clearly were difficult to assess.

Common assessment tasks

Validation of problem solving measures for the International Life Skills Survey

Herl et al. (1999) investigated the feasibility of measuring problem-solving performance using a set of common tasks. The tasks were presented diagrammatically. One was a bicycle pump and another, the human circulatory system.

Problem-solving ability was defined as “cognitive processing directed at achieving a goal when no solution method is obvious to the problem-solver” (after R. E. Mayer & Wittrock, 1996). In practice, problem solving was taken to comprise content knowledge, domain-dependent problem-solving strategies and self regulation. Self regulation was further subdivided into metacognition (planning and self monitoring) and motivation (effort and self efficacy). Content knowledge and problem-solving strategies were argued to be domain-specific, while metacognition was thought to be general.

Content understanding, comprising conceptual and procedural knowledge, was assessed using a concept mapping task in which students were asked to label diagrams of a bicycle pump and the human respiratory system and to show how the various components related to each other to enable the system to function. Raters were trained to assess the concept maps generated by students. The concept maps were constrained by providing a diagram of the system (bicycle pump or circulatory system) and providing stickers that students used to label the diagrams.

Problem solving was assessed through the administration of three tasks – two trouble shooting tasks and one design task. The trouble-shooting (diagnostic) tasks presented a problem such as “You push down on the handle of the pump, but no air comes out of the hose.” Responses to these prompts were evaluated using rubrics that included plausible reasons developed by experts. The design problem asked students to generate ways of increasing the efficiency of the pump. The number of plausible reasons (for the problem) or plausible design suggestions (for improving the pump) generated was the problem solving score.

Self regulation was assessed using a 32 item 4 subscale instrument developed by O'Neil (Herl et al., 1999, pp. 40-41).

Content knowledge and problem solving were moderately correlated, but both were weakly negatively correlated with self regulation.

Herl et al. (1999) concluded that the use of concept mapping to assess content knowledge was useful, although the generalisability coefficients were low. Herl et al. claimed that the use of two problems was enough to measure problem solving reliably. However, the study was a trial for the proposed International Life Skills Survey and the intention was to measure problem solving at a population level. Individual person estimates may be subject to substantial measurement error, but if it is assumed that measurement error is unbiased, then individual measures can be averaged over a sample of participants to generate a precise population estimate. At the individual level, the relatively high errors of measurement limit the value of the assessment. Scores at the individual level would be too imprecise to warrant reporting. Shavelson, Gao and Baxter (1993) had shown that task sampling variability made a substantial contribution to variance in scores, and that to generate reliable estimates at the individual level would require a much larger number of tasks. Further, problem solving scores varied with gender, suggesting the possibility that the assessment tasks were gender biased. They claimed that providing scoring information to participants had little effect on scores. However, it seemed that the information provided was not particularly useful, that is, it did not explain the criteria used to make judgments.

Performance assessment

Alverno College

Alverno College is a recognised leader in the implementation of generic skills (abilities) in post-school education. Their eight abilities (generic skills) are embedded in the teaching and learning of discipline-based courses and feature in the assessment of those courses. Six performance levels for each of the abilities are recognised and students are required to achieve Level 4 on abilities assessments in all studies and Level 6 in their major studies (Alverno College Institute, 1996). It is worth noting that the eight abilities are the focus of instruction and assessment within discipline-based courses and that the abilities are assessed through the assignments that students undertake as part of their courses. That is, Alverno College has made generic skills the core of their programs and they build curriculum specific knowledge and skill around the generic skills. This is the inverse of what has been happening in many other education systems, where generic skills have been added onto subject based curricula.

Adding generic skills to existing curriculum or asserting the presence of these skills within existing subject based curriculum appears to be the standard approach to the incorporation of generic skills within a curriculum. This approach was taken in South Australia under the South Australian Curriculum Standards and Accountability (SACSA) Framework (South Australia. Department of Education Training and Employment, 2001). A set of generic skills (essential learnings) were specified. In turn, they were mapped to the Key Competencies, and in the margins of subject curriculum documents, icons representing the Key Competencies were placed to show where, in the curriculum, those generic skills were exemplified. A similar approach was revealed in a national mapping project of Australian state and territory curriculum documents (Australasian Curriculum Assessment and Certification Authorities, 2003).

At Alverno College, in addition to the assessment within courses, students undertake a number of interdisciplinary ‘integrative assignments’. These are assessed by panels of College faculty and community members who have undertaken training in the Alverno assessment framework. Alverno appears to enjoy considerable community support for such activities. The integrative assignments provide an opportunity to focus specifically upon the abilities (Mentkowski, 2000).

Alverno emphasises assessment as learning and students are encouraged to build what are referred to as ‘diagnostic digital portfolios’ of their assessments and to include reflections on their work and on the feedback they have received on their assignments. The portfolios are built from the outcomes of the performance assessments and are not themselves objects of assessment, although they can be assessed (Loacker, 2000).

The assessment of creative problem solving

Mumford, Baughman, Supinski and Anderson (1998) argued that the search for simplicity and standardisation in traditional testing made it difficult to help people develop and apply particular skills in complex real-world situations and they advocated performance assessment as a method for enabling that application. They asserted "the performance assessment model is attractive… because it is able to capture complex, integrated performance skills” (p.79).

However, they acknowledged that performance assessment suffers from several methodological and substantive problems and they pointed out that subjective evaluations can be both costly and vulnerable to rater variability. Raters might not always follow prescribed rules for assessing subjects. Their evaluations might be influenced by situational factors and the characteristics of candidates. Further, even though judges may agree, their agreement might not reflect the skills and other targets of the assessment.

The complexity of the tasks assessed often made performance assessment expensive and scoring difficult. When performance assessments were task specific the results could not easily be generalised to new tasks and other domains or situations.

They drew attention to the need for careful definition of the target skills. They noted that most performance assessment used a content validation model for developing assessment tasks; that is, the assessment was specific to the tasks that students undertook, and this limited the generality of any inferences than could be drawn. An alternative was a construct based approach.

> In a construct based framework, measure development would begin with the definition of the skills contributing to performance and an analysis of how these skills are applied in performance. A variety of techniques, including expert novice comparisons, think-out-loud protocols, critical incident analysis, or a review of the relevant literature might be used to identify specific skills and characteristics of skilled performance. What is essential is that such procedural analysis allows an *a-priori*, substantive specification of

> performance relevant skills and the manner in which they are applied in real-world contexts. (p. 82)

They suggested the development of low fidelity exercises that elicited the performance-relevant skills of high-level cognitive constructs (p. 83). These low fidelity exercises appeared to be an attempt to find a middle ground between the items used in traditional testing and the complex tasks often used in performance assessments. They claimed four distinct advantages to the use of these simple low fidelity exercises. First, simple exercises that consumed less time reduce the total cost of assessment. Second, simple limited exercises reduced the need for training raters. Third, these simplified structured exercises made objective scoring of performance possible. Fourth, by organising assessments around underlying constructs, more general principles were formulated. The use of these low fidelity tasks brought to performance assessment some of the constrained characteristics of standardised testing, such as greater objectivity, and led to improved reliability and feasibility compared with previous approaches to performance assessment.

Grading performance assessment has been based on holistic judgments of the performance. But Mumford et al. suggested "response requirements might be structured to call for application of crucial components of skilled performance" (p. 83) and it is "…essential, however, that people's responses to the task reflect these component constructs or attributes of skilled performance" where the components were defined in terms of the construct that were the target of the assessment.

> The construct based approach retains an element of the performance assessment model in the sense that people's responses are obtained in ill-defined performance settings, settings in which they must structure their own activities. The activities called for, however, as with traditional tests, are designed to elicit or reflect use of the underlying attributes held to represent crucial components of skilled performance. (p. 83).

Mumford et al. (1998) recommended that scoring not be based on overall performance judgments. Instead, they proposed that it be based on indices that reflected the application of the attributes that were believed to underlie skilled performance. The use of an *a priori* scoring approach required careful analysis of performance requirements in terms of the target construct, but it allowed scoring to be based on criteria reflecting the underlying construct rather than normative comparisons. Further, because the scoring strategy was based on relevant criteria, it allowed specific feedback to be provided to learners and it enhanced the value of these measures as developmental tools. This was a very significant point. It alluded to the desirability of using performance assessment as a device for developing or enhancing skills.

> When considering these observations, we are clearly recommending a construct-, rather than a content-based, approach to performance assessment.

> Unfortunately, the construct based approach to skill assessment has not traditionally been considered in the design of performance assessment systems. As a result, evidence is not available about the validity of this approach. (p. 84)

The authors applied their performance assessment model to the assessment of creative problem solving.

They defined creative problem solving as comprising four component skills, namely problem construction, information encoding, category selection, and category combination and reorganisation. They reported some evidence to support these four sub skills as being discrete components of creative problem solving. They also showed that overall task performance was predicted partly by measures of general ability but that this prediction was enhanced by the use of the four measures of creative problem solving. Thus, they were able to show that their construct had discriminant validity from general ability.

They concluded: "... the construct-based approach to skill assessment can yield valid measures" (p.106).

> The development of objective construct-based scoring systems for skill assessment is very important for two reasons. First, the availability of objective scoring procedures reduces the cost of assessment. Second, such objective scoring procedures are not subject to the rater error problems that plague so many assessment efforts. (p. 106).
>
> Until the field starts to apply construct based approaches in skill assessment, it is difficult to see how performance assessment will move past stamp collecting and emerge as a true science. (Mumford et al., 1998, p. 107)

Assessment centres

Assessment centres are used for the selection of personnel, usually for graduate level positions. An assessment centre is a process in which candidates are assessed using a variety of methods, including standardised tests and personality inventories. However, performance assessments are very commonly used. Typically, the assessment centre is of one or two days duration. Candidates are informed a few days before the assessment about the types of activities they are likely to undertake and the competencies on which they are to be assessed. Candidates usually participate in four to eight activities, and in each they are assessed on up to three competencies. Each competency is rated using a behavioural rating scale over three activities.

Typically, these performance assessment activities involve groups of candidates who are assigned a series of tasks. The group is given a scenario such as planning a product launch. There are constraints, such as a limited budget, and the group needs to select a venue, plan the sequence of events, provide refreshments, send invitations and arrange for media coverage. Several trained raters observe the group and they have target constructs, such as teamwork, communication and leadership. For each construct, there are a series of indicators and the raters assess individual candidates on each indicator for each target skill. Raters meet and compare their scores for candidates on the target skills.

Assessment centres are held to be valid activities. There is legal precedence for their use. In a company that used these activities to select those existing employees who were to be retained following down-sizing, a court upheld the appropriateness of the assessment centre methodology as being fair.

Coalition of Essential Schools

The Coalition of Essential Schools (CES) is a network of schools who share a commitment to educational reform built around ten common principles. These principles included a commitment to an assessment approach involving ‘multiple assessments based on performance of authentic tasks’– that is, performance assessment of ‘real world’ tasks and projects. A high school diploma was awarded following the demonstration of ‘mastery’ through an ‘exhibition’ of performances.

The [high school] diploma was awarded upon a successful final demonstration of mastery for graduation – an ‘Exhibition.’ As the diploma was awarded when earned, the school's program proceeded with no strict age grading and with no system of credits earned by ‘time spent’ in class. The emphasis was on the students' demonstration that they were able to do important things (Rogers, McDonald, & Sizer, 1993).

An interesting claim was made about the location of deliberation and decisions of how social and economic trends influenced educational policies and practices. Members of the CES believed that individual teachers were more responsive to ‘economic signals and shifts in cultural values’ than central bureaucracies or testing organisations and were better able to interact with students in setting and interpreting standards for assessment (Rogers et al., 1993).

The teacher-generated assessment model was used within the CES for routine assessment of courses. It can be used for employability skills if teachers judged them to be worthy assessment constructs.

Teacher judgment

Victorian Board of Studies Key Competencies levels assessment trial

Teachers were asked to make ‘on-balance’ judgments of students’ demonstration of generic skills based on teachers’ observations of students in both classroom and co-curricular activities.

Collective teacher judgment has been shown to work well in the school sector where teachers knew students’ attributes well through frequent and close observation (McCurry & Bryce, 1997, 2000). McCurry and Bryce established small panels of teachers and provided them with sufficient training in the key competencies to enable them to make consistent judgments of students’ attainment of key competencies. This training and the observation of students, both in classroom-based and co-curricular activities, enabled teachers to make sufficiently consistent judgments to discriminate eight performance levels.

Holistic judgments by individual teachers were based on a summary of that teacher’s observations of students in the classroom and co-curricular activities that the teacher had supervised. Clearly, the judgments made by individual teachers were limited compared with those made by groups of teachers. Individual teacher judgments were likely to reflect a smaller range of observed activities that were encompassed by a group of teachers and the group judgment was likely to even out less favourable and more favourable judgments of individual teachers.

Consistency of judgments within panels of teachers has been demonstrated. What has not been shown, at least in the context of generic skills, was that this consistency extended across school boundaries. Anticipated levels of performance in all assessments, even multiple-choice tests, were informed by normative views of student performance. The range of difficulty of items or the standards that were prescribed in scoring rubrics were based, initially at least, on normative expectations of students. It was very likely that the norms that frame teachers’ expectations within a school are informed by teachers’ experiences in that school. The experiences of teachers in other schools would likely lead to different performance expectations. This meant that students from different schools might be judged against different standards, and so individual achievement, assessed using this method, did not provide a basis for broad comparison.

Embedded development and assessment

One of the enduring problems, especially in the VET sector, has been misunderstanding of the relationship between alternative approaches to the development and assessment of employability skills. The view has been abroad that employability skills were either embedded (and therefore not overtly assessed), or delivered separately (and therefore explicitly assessed). In practice, there are other possibilities, and in particular, it was possible to have embedded delivery and explicit assessment. An attempt to clarify this dilemma is presented in Table 41.

Table 41: Possible arrangements for the delivery and assessment of employability skills

	Delivery	
Assessment	Embedded	Separate
Inferred	Generic skills are taught within existing subjects/courses. Performance is inferred from the achievement of unit/course objectives.	Generic skills are taught in separate subjects/modules. Since there is no other content, achievement is assumed. This is a non-assessment model.
Explicit	Generic skills are taught within existing subjects/courses. Performance is assessed explicitly using criteria or standards derived from descriptions of generic skills constructs.	Generic skills are taught in separate subjects/modules. Assessment of generic skills is undertaken within these modules.

It was possible to separate delivery and assessment strategies. In a study undertaken by the Queensland Department of Employment and Training, both integrated and separate delivery and assessment approaches were trialled (Queensland DET, 2004). Although the authors indicated a preference for separate delivery, they did report that a high proportion of the students who were enrolled in the discrete units ceased to attend classes. It seemed that students did not value this approach. Other research, in the higher education sector, found that if students perceived the assessment to be authentic (and embedded in existing and valued learning) they were more inclined to use deep learning strategies (Gulikers, Bastiaens, Kirschner, & Kester, 2006). These findings suggested that the embedded or integrated delivery of employability skills was likely to enhance the learning of these skills.

If the teaching and learning of employability skills were embedded within existing subjects, it was possible to have either inferred assessment of them, or explicit assessment, and this could take several forms, including standardised testing and performance assessment. In review of generic skills assessment options in the VET sector, inferred assessment was regarded as a “minimal solution, likely to lead to minimal results” (Ratio Pty Ltd & Down, 2003). If the employability skills were delivered in separate subjects, there might be no assessment of them, that is, their delivery could be assumed to have achieved the purpose of informing learners about them, or they might be assessed explicitly in some of the ways outlined in this Appendix.

A mapping of elements of the employability skills against the curriculum of existing subjects was undertaken to show that the employability skills were contained within the curriculum. The assumption was made that achieving those curriculum outcomes, demonstrated through existing assessment, could only occur if the mapped employability skill had been exercised. If this assumption was valid, it should be the case that students who mastered vocationally relevant content would demonstrate the mapped employability skills in practice. Surveys of employers, however, had not supported this hypothesised observation (AC Nielsen Research Services, 2000). It seems that explicit attention to employability skills in both teaching and assessment is required in order to achieve the outcomes desired for employability skills.

Portfolio construction

A portfolio is an assemblage of evidence of activities through which certain abilities are attested. The content of a portfolio can be quite varied. It can include: objective documentary evidence of achievement such as a certificate awarded following successful completion of a course; letters and testimonials of performance; photographs or videos of an individual undertaking a relevant activity; an individual journal of activities, perhaps with reflective comments.

The portfolio may be assembled and presented simply as a compendium of evidence or it can be subject to assessment as a body of evidence. In most cases reviewed, the emphasis was on the construction of the portfolio and it was not assessed. In such cases, if it was regarded as an act of assessment, it needed to be judged as an instance of self-assessment.

Portfolio assessment was recommended as the preferred method of assessing and reporting achievement of the employability skills across the schools, VET and higher education sectors (Allen Consulting Group & National Centre for Vocational Education Research, 2004). Although many other assessment and reporting options had been raised during consultations, the final report only dealt with portfolios. The reasons for this were clear as the criteria proposed to evaluate options were: effective, efficient, accountable, and transparent (Allen Consulting Group, 2004). The efficiency criterion seemed to have weighed heavily in their deliberations and portfolios were recommended as having the minimum impact on the need for professional development of teaching staff and the minimum assessment load. Neither specific portfolio templates, nor any particular methods for assessing or verifying portfolio content were suggested. Indeed a degree of latitude was also suggested in interpreting the employability skills themselves.

The review by the Allen Consulting Group (2004) failed to acknowledge the literature on portfolio assessment. First, the construction of a portfolio was not an act of assessment. It produced an artefact that was available for assessment and the act of creating it might have been a learning experience for the individual. The assessment of portfolios was a time-consuming task and that assessment was beset by low validity and reliability (Troper & Smith, 1997). Portfolios might have had low validity because the quality of a portfolio was a result of more than the target construct (demonstrated performance on an employability skill). The form of the portfolio also influenced raters' judgments.

Portfolios do serve a useful purpose. They are repositories for detailed evidence of experience and performance. Alverno College encourages students to develop a so-called 'diagnostic digital portfolio' in which students record the results of their own reflections and other assessments, including externally judged integrative assessments (Alverno College Faculty, 1994).

Many examples of portfolio templates and tools are available (Allen Consulting Group & National Centre for Vocational Education Research, 2004, pp. 22-35).

Appendix 4: The Problem Solving Inventory

The Problem Solving Inventory used in the current study was adapted from the version of the instrument by Heppner and Peterson (1982). The adapted version is shown here.

Name: ______________________________

Problem-Solving Inventory

Please read each of the statements below then check the box on the right that you think most closely applies to your problem-solving in Electronics and Information Technology.
There are no right or wrong answers.

	Almost Always	Often	Seldom	Almost Never
When a solution to a problem was unsuccessful, I examine why it didn't work.	☐	☐	☐	☐
When I am confronted with a complex problem, I develop a strategy to collect information so I can define exactly what the problem is.	☐	☐	☐	☐
When my first efforts to solve a problem fail, I become uneasy about my ability to handle the situation.	☐	☐	☐	☐
After I have solved a problem, I do <u>not</u> analyse what went right or what went wrong.	☐	☐	☐	☐
I am usually able to think up creative and effective alternatives to solve a problem.	☐	☐	☐	☐
After I have tried to solve a problem with a certain course of action, I take time and compare the actual outcome to what I thought should have happened.	☐	☐	☐	☐
When I have a problem, I think up as many possible ways to handle it as I can until I can't come up with any more ideas.	☐	☐	☐	☐
When confronted with a problem, I consistently examine my feelings to find out what is going on in the problem situation.	☐	☐	☐	☐
I have the ability to solve most problems even though initially no solution is immediately apparent.	☐	☐	☐	☐
Many problems I face are too complex for me to solve.	☐	☐	☐	☐
I make decisions and am happy with them later.	☐	☐	☐	☐
When confronted with a problem, I tend to do the first thing that I can think of to solve it.	☐	☐	☐	☐
Sometimes I do <u>not</u> stop and take time to deal with problems, but just kind of muddle ahead.	☐	☐	☐	☐
When deciding on an idea or a possible solution to a problem, I take time to consider the chances of each alternative being successful.	☐	☐	☐	☐
When confronted with a problem, I stop and think about it before deciding on the next step.	☐	☐	☐	☐
I generally go with the first good idea that comes into my head.	☐	☐	☐	☐

	Almost Always	Often	Seldom	Almost Never
When making a decision, I weigh the consequences of each alternative and compare them with each other.	☐	☐	☐	☐
When I make plans to solve a problem, I am almost certain that I can make them work.	☐	☐	☐	☐
I try to predict the overall result of carrying out a particular course of action.	☐	☐	☐	☐
When I try to think up possible solutions to a problem, I do <u>not</u> come up with very many alternatives.	☐	☐	☐	☐
Given enough time and effort, I believe I can solve most problems that confront me.	☐	☐	☐	☐
When faced with a novel situation, I have confidence that I can handle problems that may arise.	☐	☐	☐	☐
Even though I work on a problem, sometimes I feel like I am groping or wandering, and not getting down to the real issue.	☐	☐	☐	☐
I make snap judgements and later regret them.	☐	☐	☐	☐
I trust my ability to solve new and difficult problems.	☐	☐	☐	☐
I have a systematic method for comparing alternatives and making decisions.	☐	☐	☐	☐
When confronted with a problem, I usually examine what sort of external things my environment may be contributing to the problem.	☐	☐	☐	☐
When I am confused by a problem, one of the first things I do is survey the situation and consider all relevant pieces of information.	☐	☐	☐	☐
Sometimes I get so charged up emotionally that I am <u>un</u>able to consider many ways of dealing with problems.	☐	☐	☐	☐
After making a decision, the outcome I expected usually matches the actual outcome.	☐	☐	☐	☐
When confronted with a problem, I am <u>un</u>sure of whether I can handle the situation.	☐	☐	☐	☐
When I become aware of a problem, one of the first things I do is to try to find out exactly what the problem is.	☐	☐	☐	☐

Thank you for your participation

Appendix 5: The Problem Solving Assessment

The Problem Solving Assessment (PSA) was developed from an initial draft to the version used in the present study and revised following analysis of the results of the study. All versions of the instrument are included in this appendix. The development of the instrument is described in Chapter 7 and is summarised briefly here.

The first version bears the title 'Problem Solving Assessment: Draft.' This version was discussed by a focus group of teaching staff in the School of Electronic and Information Technology at the Torrens Valley Institute of Technical and Further Education. It was modified based on their feedback.

The second version bears the title 'Demonstration of Problem Solving Performance: V2.' This version was used in the study reported in Chapter 7. The format was altered from the draft so that each indicator was accompanied by a list of performance level descriptors. This provided more information to students and teachers and avoided the appearance of an incomplete rating scale. One indicator was added to the Execution section – a reflection of the Key Competencies performance levels. These levels were being recommended in some training packages and compatibility with that practice was desired. Other changes in wording were made. For example, the word 'efficacy' was changed to 'efficiency and effectiveness' as teaching staff advised that many students would not understand that term.

Following analyses of data collected in the study and further consultation with teaching staff, a third version of the instrument was developed. This is titles 'Demonstration of Problem Solving Performance: V3.'This version was used by staff after completion of the study. The number of indicators was reduced from 21 to 18 by combining several indicators. The format was changed so that all indicators and performance level descriptors fitted on one side of an A4 sheet with performance level guidelines on the reverse side.

Problem Solving Assessment: Draft

Student Name ______________________ **ID** __________

Module ______________________ **Task** __________

Assessor ______________________ **Date** __________

Problem-Solving Process	Indicator	Score				
Representation	Forms a correct understanding of the problem	0	1	2		X
	Recognises relevant given information	0	1	2		X
	Identifies the need for additional information	0	1			X
	Recalls relevant information	0	1	2		X
	Sets a realistic goal	0	1			X
Planning	Plans an approach to the problem	0	1	2	3	X
	Recalls previous relevant or similar problem tasks	0	1			X
	Identifies appropriate sub-goals	0	1			X
	Checks that required equipment is available	0	1			X
	Sets an appropriate time frame	0	1			X
Execution	Begins to follow the set plan	0	1			X
	Activates relevant knowledge	0	1	2		X
	Uses relevant skills	0	1	2		X
Monitoring	Checks that set plan leads toward problem goal	0	1			X
	Diagnoses causes of deviations from expected progress	0	1			X
	Makes adjustments to execution processes	0	1			X
	Revises original plan	0	1			X
	Checks problem representation	0	1			X
Reflection	Reviews efficacy of problem approach	0	1	2		X
	Compares current problem with previously encountered ones	0	1	2		X
	Anticipates situations in which current approach might be useful	0	1			X

Demonstration of Problem Solving Performance: V2

Name ______________________ **ID** ______________________

Module ______________________ **Task** ______________________

Facilitator ______________________ **Date** ______________________

(SA = Self Assessment, V = Facilitator Validation)

Defining the Problem		SA	V
Forms a correct understanding of the problem	Misunderstands the problem	☐	☐
	Forms a partial understanding using some given information	☐	☐
	Forms a complete understanding using all relevant factors	☐	☐
Recognises significance of given information	Fails to recognise the significance of information that is given	☐	☐
	Recognises the significance of information that is given	☐	☐
Identifies the need for additional information	Does not recognise a need for further information	☐	☐
	Identifies specific additional information required	☐	☐
Recalls relevant information	No recall, recalls irrelevant information, inaccurate recall of	☐	☐
	Accurately recalls relevant information as isolated elements	☐	☐
	Accurately recalls relevant information as integrated elements	☐	☐
Sets a realistic goal	Does not identify a clear goal	☐	☐
	Establishes a clear goal	☐	☐

Planning an Approach		SA	V
Plans an approach to the problem	Does not engage in planning activity	☐	☐
	Undertakes enough planning to begin the solution	☐	☐
	Plans all stages required to achieve the goal	☐	☐
	Considers several alternative approaches and selects the best plan to achieve the goal	☐	☐
Recalls previous relevant or similar problem tasks	Does not refer to previous problems	☐	☐
	Recalls previous problems that may be relevant	☐	☐
Identifies appropriate sub-goals	Does not identify sub-goals	☐	☐
	Breaks the task into smaller sub-goals	☐	☐
Checks equipment availability	Does not check for required equipment in advance	☐	☐
	Does check that needed resources are available	☐	☐
Sets an appropriate time frame	Does not consider time frame	☐	☐
	Estimates how long the problem solution should take	☐	☐

For each indicator, eg 'Sets a realistic goal', select the criterion description that most closely matches task performance and check its box.
If you believe that an indicator has been performed, but there is no direct evidence of it, leave the boxes for that indicator blank.

Carrying out a Plan		SA	V
Begins to follow the set plan	Begins to work on the problem without a clear system	☐	☐
	Works systematically and follows set plan	☐	☐
Activates relevant knowledge	Does not activate relevant knowledge or activates incorrect information	☐	☐
	Activates some, but not all, relevant knowledge	☐	☐
	Activates and uses accurate relevant knowledge	☐	☐
Uses relevant skills	Does not apply relevant skills	☐	☐
	Activates some relevant skills	☐	☐
	Applies skills accurately	☐	☐
Application of strategies	1. Applies an established procedure	☐	☐
Key Competency Performance Level	2. Selects and implements a strategy from those available	☐	☐
	3. Adapts an existing, or creates a new, procedure	☐	☐

Monitoring Progress		SA	V
Checks progress towards goal	Does not check progress against sub-goals or final goal	☐	☐
	Periodically checks solution progress against sub-goals and final goal	☐	☐
Responds to unexpected problems along the way	Does not recognise unexpected problems	☐	☐
	Identifies unexpected problems	☐	☐
	Diagnoses causes of unexpected problems	☐	☐
	Makes adjustments to rectify suspected causes of unexpected problems	☐	☐
Reviews original plan	Does not review original plan	☐	☐
	Reviews original plan	☐	☐
	Reviews and revises original plan or verifies that original plan is appropriate	☐	☐
Checks original understanding and definition of problem	Does not check original understanding and definition of problem	☐	☐
	Reviews understanding and definition of problem	☐	☐

Reflecting on the Result		SA	V
Reviews efficiency and effectiveness of problem approach	Does not reflect on solution efficiency and effectiveness	☐	☐
	Reviews efficiency and effectiveness of solution strategy	☐	☐
	Identifies improvements to solution strategy	☐	☐
Compares current problem with previously encountered ones	No comparisons with previous related tasks	☐	☐
	Compares current task with previous ones	☐	☐
	Notes ways in which current experience might have helped past problems	☐	☐
Anticipates situations in which current problem approach might be useful	No anticipation of future applications of solution strategy	☐	☐
	Considers future broader applications of solution strategy	☐	☐

Scoring Instructions

For each indicator, select the level of performance that matches the student's performance. The first option for each indicator is the 'did not perform' level. Other options indicate levels above this.

If you believe that the student has implicitly performed on this indicator, but has not presented obvious evidence, leave the boxes for that indicator blank.

Defining the Problem

Forms a correct understanding of the problem. How complete is the student's understanding of the problem? Has s/he formed a simple understanding based on one aspect of it (say resistance using Ohm's Law) or a more complex one based on several aspects (say resistance and heat dissipation and …)?

Recognises significance of given information. Has the student noted the significance of information that is given? Does the student recognise that the information that has been given and what they already know are not enough to enable them to proceed?

Identifies the need for additional information. Does the student recognise that they may need to find additional information to help them solve the problem? They may need to check technical manuals or other reference sources to identify specific additional information.

Recalls relevant information. Does the student recognise that they already know things that are important in solving the problem? Does s/he recall that information accurately? Is the information recalled as separate pieces of information or does the student show an understanding of how it all relates together in relation to the problem task?

Sets a realistic goal. Does the student identify a given goal or establish their own clear and achievable goal? This is how they will know that they have reached an appropriate solution.

Planning an Approach

Plans an approach to the problem. Does the student reveal any planning? Sometimes this will be short term planning, ie just enough to get started but not enough to get to a final solution. This is fine for novices, because it will be easier for them to set a short term goal and to achieve that before going onto the next phase of the problem. A more experienced student will set a series of goals that will guide them from the beginning of their solution attempt to the goal.

Recalls previous relevant or similar problem tasks. Good problem-solvers think back over their past experiences to find a problem, similar to the current one, that they have encountered previously. This will help them to develop a solution strategy. Can the student tell you about similar past problems? Can they tell you in what ways past problems were both similar to and different from the current one? Can they use this to help develop a solution strategy?

Identifies appropriate sub-goals. Many problems involve sub-problems or phases in their solution attempt. Can they set a final goal and then establish milestones or sub-goals along the way?

Checks equipment availability. Does the student check that they have all the resources they will need as they work through the problem?

Sets an appropriate time frame. Is the student clear about how long the problem should take to solve? Is their estimate reasonable? Do they integrate this with other commitments they may have?

Carrying out a Plan

Begins to follow the set plan. Does the student begin to work systematically? If the student has a clear plan, does s/he show that they are following it? In cases where experienced students did not seem to develop a plan, is there evidence that they are working systematically?

Activates relevant knowledge. If the student needed to recall information, has s/he been able to apply that knowledge correctly in the current situation?

Uses relevant skills. Does the student apply appropriate skills and procedures and has s/he been able to apply them appropriately in the current situation?

Application of strategies. Did the student apply an established procedure (eg experiment procedure or flowchart)? Did they select and implement an appropriate strategy from those available? Did they adapt or manipulate an existing, or create a new and appropriate approach to the problem?

Monitoring Progress

Checks progress towards goal. Does the student pause, or do they give any indication that during the solution attempt, they paused and checked their progress against sub-goals or the final goal to see that they were getting closer to a solution?

Response to unexpected problems along the way. If an unexpected problem arose, and they were not able to proceed as expected, do they show any evidence that they have looked for the cause? For example, do they show the need to retest existing components or recheck a circuit design? Do they seek additional information to verify their understanding of the problem? If the solution is going according to plan, especially if they are following a set procedure, this indicator may not be apparent.

Reviews original plan. If the original plan does not appear to be working, does the student review it and adopt a new approach or set new goals? If the solution is going according to plan, does the student review the plan to verify that it cannot be improved.

Checks original understanding and definition of problem. If the student has found that the solution attempt is not working, do they go back to the definition of the problem and pause to think again about the nature of the problem? Here, a student who had formed an incorrect understanding or definition of the problem might think about other possible dimensions of the problem and redefine the problem.

For each indicator, eg 'Sets a realistic goal', select the criterion description that most closely matches task performance and check its box.

If you believe that an indicator has been performed, but there is no direct evidence of it, leave the boxes for that indicator blank.

Reflecting on the Result

Reviews efficiency and effectiveness of problem approach. Having solved the problem, has the student considered how they might have been more efficient? Would the student make any major changes to the way they went about this activity if they had to do it again?

Compares current problem with previously encountered ones. Does the student compare the just-completed problem with others that had been attempted earlier? Can they identify techniques in this solution that would have helped with earlier problems? Were there techniques that were used previously that would have helped in the current problem?

Anticipates situations in which current approach might be useful. Does the student show evidence that what they have just done could be used in other real world situations? Can they identify a specific task where the current approach is likely to be applicable?

Demonstration of Problem Solving Performance: V3

Name ____________________ **ID** __________

Module ____________________ **Task** __________

Facilitator ____________________ **Date** __________

(SA = Self Assessment, V = Facilitator Validation)

Defining the Problem		SA	V
Forms a correct understanding of the problem	Misunderstands the problem	☐	☐
	Forms a partial understanding using some given information	☐	☐
	Forms a complete understanding using all relevant factors	☐	☐
Recognises significance of given information or the need for new information	Fails to recognise the significance of information that is given	☐	☐
	Recognises the significance of most information that is given	☐	☐
	Recognises the significance of all information that is given	☐	☐
	Identifies specific additional information required	☐	☐
Recalls relevant information	No recall, recalls irrelevant information, inaccurate recall of information	☐	☐
	Accurately recalls relevant information as isolated elements	☐	☐
	Accurately recalls relevant information as integrated elements	☐	☐
Sets a realistic goal	Does not identify a clear goal	☐	☐
	Identifies relevance of given goals	☐	☐
	Establishes a clear goal	☐	☐

Planning an Approach		SA	V
Plans an approach to the problem	Does not engage in planning activity	☐	☐
	Undertakes enough planning to begin the solution	☐	☐
	Plans all stages required to achieve the goal	☐	☐
	Considers several alternative approaches and selects the best plan to achieve the goal	☐	☐
Recalls previous relevant or similar problem tasks	Does not refer to previous problems	☐	☐
	Recalls previous problems that may be relevant	☐	☐
Identifies appropriate sub-goals	Does not identify sub-goals	☐	☐
	Breaks the task into smaller sub-goals	☐	☐
Sets an appropriate time frame	Does not consider time frame	☐	☐
	Estimates how long the problem solution should take	☐	☐
	Integrates other commitments into time frame established for solving this problem	☐	☐

Carrying out a Plan		SA	V
Begins to follow the set plan	Begins to work on the problem without a clear system	☐	☐
	Works systematically, but without reference to plan	☐	☐
	Works systematically and follows set plan	☐	☐
Activates relevant knowledge	Does not activate relevant knowledge or activates incorrect information	☐	☐
	Activates some, but not all, relevant knowledge	☐	☐
	Activates and uses accurate relevant knowledge	☐	☐
Application of strategies Key Competency Performance Level	1. Applies an established procedure	☐	☐
	2. Selects and implements a strategy from those available	☐	☐
	3. Adapts an existing, or creates a new, procedure	☐	☐

Monitoring Progress		SA	V
Checks progress towards goal	Does not check progress against sub-goals or final goal	☐	☐
	Periodically checks solution progress against sub-goals and final goal	☐	☐
Responds to unexpected problems along the way	Does not recognise unexpected problems	☐	☐
	Identifies unexpected problems	☐	☐
	Diagnoses causes of unexpected problems	☐	☐
	Makes adjustments to rectify suspected causes of unexpected problems	☐	☐
Reviews original plan	Does not review original plan	☐	☐
	Reviews original plan	☐	☐
	Reviews and revises original plan or verifies that original plan is appropriate	☐	☐
Checks original understanding and definition of problem	Does not check original understanding and definition of problem	☐	☐
	Reviews understanding and definition of problem Reviews and revises understanding and definition of problem or verifies that original definition is correct	☐	☐

Reflecting on the Result		SA	V
Reviews efficiency and effectiveness of problem approach	Does not reflect on solution efficiency and effectiveness	☐	☐
	Reviews efficiency and effectiveness of solution strategy	☐	☐
	Identifies improvements to solution strategy	☐	☐
Compares current problem with previously encountered ones	No comparisons with previous related tasks	☐	☐
	Compares current task with previous ones	☐	☐
	Notes ways in which current experience might have helped past problems	☐	☐
Anticipates situations in which current problem approach might be useful	No anticipation of future applications of solution strategy	☐	☐
	Considers future broader applications of solution strategy	☐	☐

Scoring Instructions

For each indicator, select the level of performance that matches the student's performance. The first option for each indicator is the 'did not perform' level. Other options indicate levels above this.

If you believe that the student has implicitly performed on this indicator, but has not presented obvious evidence, leave the boxes for that indicator blank.

Defining the Problem

Forms a correct understanding of the problem. How complete is the student's understanding of the problem? Has s/he formed a simple understanding based on one aspect of it (say resistance using Ohm's Law) or a more complex one based on several aspects (say resistance and heat dissipation and …)?

Recognises significance of given information. Has the student noted the significance of information that is given? Does the student recognise that the information that has been given and what they already know are not enough to enable them to proceed. They may need to check technical manuals or other reference sources to identify specific additional information.

Recalls relevant information. Does the student recognise that they already know things that are important in solving the problem? Does s/he recall that information accurately? Is the information recalled as separate pieces of information or does the student show an understanding of how it all relates together in relation to the problem task?

Sets a realistic goal. Does the student identify a given goal or establish their own clear and achievable goal? This is how they will know that they have reached an appropriate solution.

Planning an Approach

Plans an approach to the problem. Does the student reveal any planning? Sometimes this will be short term planning, ie just enough to get started but not enough to get to a final solution. This is fine for novices, because it will be easier for them to set a short term goal and to achieve that before going onto the next phase of the problem. A more experienced student will set a series of goals that will guide them from the beginning of their solution attempt to the goal.

Recalls previous relevant or similar problem tasks. Good problem-solvers think back over their past experiences to find a problem, similar to the current one, that they have encountered previously. This will help them to develop a solution strategy. Can the student tell you about similar past problems? Can they tell you in what ways past problems were both similar to and different from the current one? Can they use this to help develop a solution strategy?

Identifies appropriate sub-goals. Many problems involve sub-problems or phases in their solution attempt. Can they set a final goal and then establish milestones or sub-goals along the way?

Sets an appropriate time frame. Is the student clear about how long the problem should take to solve? Is their estimate reasonable? Do they integrate this with other commitments they may have?

Carrying out a Plan

Begins to follow the set plan. Does the student begin to work systematically? If the student has a clear plan, does s/he show that they are following it? In cases where experienced students did not seem to develop a plan, is there evidence that they are working systematically?

Activates relevant knowledge. If the student needed to recall information, has s/he been able to apply that knowledge correctly in the current situation?

Application of strategies. Did the student apply an established procedure (eg experiment procedure or flowchart)? Did they select and implement an appropriate strategy from those available? Did they adapt or manipulate an existing, or create a new, problem solving process?

Monitoring Progress

Checks progress towards goal. Does the student pause, or do they give any indication that during the solution attempt, they paused and checked their progress against sub-goals or the final goal to see that they were getting closer to a solution?

Response to unexpected problems along the way. If an unexpected problem arose, and they were not able to proceed as expected, do they show any evidence that they have looked for the cause? For example, do they show the need to retest existing components or recheck a circuit design? Do they seek additional information to verify their understanding of the problem? If the solution is going according to plan, especially if they are following a set procedure, this indicator may not be apparent.

Reviews original plan. If the original plan does not appear to be working, does the student review it and adopt a new approach or set new goals? If the solution is going according to plan, does the student review the plan to verify that it cannot be improved.

Checks original understanding and definition of problem. If the student has found that the solution attempt is not working, do they go back to the definition of the problem and pause to think again about the nature of the problem? Here, a student who had formed an incorrect understanding or definition of the problem might think about other possible dimensions of the problem and redefine the problem.

Reflecting on the Result

Reviews efficiency and effectiveness of problem approach. Having solved the problem, has the student considered how they might have been more efficient? Would the student make any major changes to the way they went about this activity if they had to do it again?

Compares current problem with previously encountered ones. Does the student compare the just-completed problem with others that had been attempted earlier? Can they identify techniques in this solution that would have helped with earlier problems? Were there techniques that were used previously that would have helped in the current problem?

Anticipates situations in which current approach might be useful. Does the student show evidence that what they have just done could be used in other real world situations? Can they identify a specific task where the current approach is likely to be applicable?

Appendix 6: Student Evaluation of the Problem Solving Assessment

The following comments about the Problem Solving Assessment were volunteered anonymously by Electronics and Information Technology students using an on-line form. The comments are presented as they were entered by students, except that where lecturers were named, the name has been replaced by [Teacher]. Comments have been classified according to the issues raised.

No.	Comment	Classification
1	K C not only make future job employers aware of your skills but it makes you aware of your skills and gives you the extra confidence in a job interview situation	recognition confidence
2	I did find my first attempt the most difficult and time consuming but now there is greater information available and I have completed a few more I find them an easy way to improve my 'soft' skills and analyse my methodologies.	processes improvement
3	When they were first introduced I did a few but then completely forgot about them due to the complexity of understanding the performance levels. Now due to the fact that [Teacher] is insisting we gain some while doing OHSW for Supervisors, I have gained some more. It is only with [Teacher]'s help and prompting that I have gained a little understanding of the terminology that decides the level to go for. When it gets too difficult to understand what the different levels are about, I just decide not to go for them. Doing these assessments have partly helped me understand and improve my skills because it put a name to them but I did not improve the skills, in the ones I have done anyway. I find it difficult to differentiate between some of the levels.	performance levels
4	I got kicked in the guts with KC a few times. I think that its too hard and pointless to do. I think Experience will win over KC. I don't think I will be doing anymore, because it takes too much time.	strong negative sentiment
5	I think it is great that this Tafe campus offers these great recognisation of skills.	recognition
6	Understand the importance of Key Competencies. At the moment I'm a new student trying to work my way through the first prerequisite modules.	
7	The only Key Competency Assessment I did did leave me feeling more confused than informed. However, since all the guest speakers etc I have a better idea of what it's all about.	processes
8	Yes, after the 1st assessment with [Teacher] (to induct students in the process).	
9	First Problem Solving Assessment should be made compulsory for the Diploma. The second assessment can be kept for the Cert(?)	
10	Does require a bit of work and discussion to follow through with the principle of the Key Competencies. eg Having it explained by a facilitator.	
11	As soon as I have time I will get around to it.	

Continued…

No.	Comment	Classification
12	Maybe more info can be given on how long it takes to get KC's. I wasn't ever told that you have to apply before starting a module. I think that they should be done at the easy stages of the courses.	processes
13	I am self employed and I can not see myself work for someone else but any key competencies will help in studies and my self confidence.	
14	Having the same process as the new Problem Solving Assessment available for the other Key Competencies would be beneficial.	
15	Some of the wording in the key competency checklists are difficult to understand and it is hard to know what exactly they require.	wording
16	The only issue I could comment on is that at the time I completed the assessment it was unclear exactly what parts of the curriculum were applicable to key competency assessment. It was difficult to know when a key competency could be attempted.	processes
17	Some of the wording was too technical, I think that the use of plain words would make the process of choosing which box easier to decide.	wording
No.	Comment	
18	Very helpful in breaking down the problem solving method into identifiable sections. Helped me better understand the process.	scaffolding
19	The original way of assessment for key competencies is not that clear on what is exactly required for each level of competency. The new 'problem solving' trial has helped much more in that it breaks down each section of the process and allows for detailed discussion each time.	scaffolding
20	I think the key competencies is a good way of being recognised for things that you do but are not necessarily recognised in any other way.	recognition
21	Own summary of PS skill to: • compare to next assessment might be helpful • be aware of areas for improvement How can we improve areas?	lack of guidance (feedback)
22	I reckon key competencies assessment is a good idea because it allows students to get recognised for the work they do.	recognition

Appendix 7: Recommended E&IT Assessment Tasks

Table 42: Electronics and Information Technology recommended assessment tasks

Module	Assessment Activity	Description
Applied Electricity 1 (AM)	Practical Activity 2.3 – Electrical Circuits	Introductory practical activity involving connection of simple serial and parallel circuits and basic current measurements. After which the student is asked to answer a practical circuit-breaker question based on their understanding of this prac. Procedure and circuit diagrams provided. Practical wiring of circuit Concepts of series and parallel circuits Appreciation of circuit current flow
Intro to Electricity & Electronics (PH)	Practical Activity 2.5 – Basic Circuit Measurements	Connection of basic circuit and measurement and calculation using ohms law and power formulae. Procedure and equip list provided. Practical wiring of circuit Application of formulae Concepts involved in circuit operation Relationships between basic circuit parameters
Electrical Fundamentals (TP)	Prac Test 1	Involves determining the resistance of an unknown resistor by measuring and calculating voltage and current values of a simple resistive circuit. Relies on understanding and application of Ohm's law and ability to use an analogue meter for voltage and current readings.
Electrical Fundamentals (TP)	Distinction Activity	The exercise you will be set, requires using theoretical knowledge and analytical skills to fault-find a simple series/parallel combination circuit and present findings in a formal report.
AC Principles (GS)	Practical Test	Performed under test conditions in test room. Determine value of unknown resistor in a series RC circuit.
Soldering – Basic (DB)	Project kit	To apply soldering techniques developed through the module to construct a kit using parts, parts list, circuits etc supplied. This includes some known ‘unanticipated’ (from students’ perspective) problems. Use parts list Identify all components (parts list supplied includes some outdated parts references!) Practical construction
Power Supply Principles (GS)	Topic 11 – Faultfinding Techniques	The practical unit for this exercise is a grey plastic jiffy box with a pcb mounted on it containing a dual power supply circuit. Various faults are explored to demonstrate an understanding of how power supplies work, some of the most likely faults which can occur and the symptoms they display.

Digital Electronics 1 (BG)	Practical Activity 5.1 – Digital Displays	Practically wire-up and produce truth table for a 7 segment display incorporating a BCD decoder/driver chip. Procedure provided but relies on understanding of basic electrical and digital concepts and putting them into practice. Practical wiring of circuit Calculation of limit resistors Operation of circuit to develop truth table Demonstrate LT & RBI functions
Digital Subsystems 1 (BG)	Practical Activity 2.2 – Switch Debouncing	Design and construct circuit. No procedure given. Info available from text references given or from student's own knowledge. Involves advanced use of Digital Oscilloscope. Clear understanding of the task Theoretical circuit design Practical construction of circuit Understand what results to look for to verify achievement of debouncing Use of digital oscilloscope
Microprocessor Fundamentals (RD)	Practical Activity 5.3 – Using a RAM Chip	No procedure given. Required to develop own procedure to design and wire up a RAM chip circuit and verify that data can be written to and read from various address locations. Clearly understand operation of RAM chip and all its pins Design suitable test circuit Design procedure to write to and read from RAM
PC Systems Support (KS)	Practical Activity 6.1 – Application/Hardware Installation	Procedure given. Install and configure software and hardware. Determine hardware system requirements for various software before installation. Determine system requirements Hardware and software installation and configuration IRQ conflicts Appropriate cables and orientation
Single User Operating Systems (JR)	Topic 4.1 – Configuring Operating System	Built and configure a PC system to customer requirements. Procedure given. Understanding of system specifications required for customer Determine system requirements Hardware and software installation and configuration
Single User Operating Systems (JR)	Credit Activity (Based on Topic 4.1 – Configuring Operating System but with more complex requirements)	Built and configure a PC system to customer requirements. Procedure given. Understanding of system specifications required for customer Determine system requirements Hardware and software installation and configuration Greater system requirements than Prac 4.1
Hardcopy (GS)	Practical Activity 5.1.1 – Practical Application	EITHER Install and service a printer for a client OR faultfind a laser printer as directed

Intro to Programming (GM)	Final Programming Project – Game Program (5-in-a-row)	Pull together individual concepts, programming structures and modules of code developed through module into a final working project. Procedure not provided but information, suggestions and hints supplied. Complex programming task to integrate a range of program modules written by student with a given program module to achieve functional game. Effective use of programming development software High level conceptual thinking
Embedded Systems (RD)	Practical Activity 2.3 – Interfacing a Microcontroller Development Board to a PC	Practically connect and operate a development board with a PC. Three approaches are outlined based on each of the 3 performance levels for the Key Competency 'Solving Problems' Selection of range of different resources Appropriate connection and setup of all resources Software control of development system

Appendix 8: The Problem Solving Skills Assessment Tool (PSSAT)

The PSSAT is a revised version of the Problem Solving Assessment (PSA) instrument shown in Appendix 5.

Problem Solving Skills Assessment Tool

Name		**ID**	
Unit		**Task**	
Lecturer		**Date**	

Self Assessment **SA** | Lecturer Validation **LV**

Defining the Problem		**SA**	**Evidence of Definition**	**LV**
Forms a correct understanding of the problem	Misunderstands the problem	☐		☐
	Forms a partial understanding using some given information	☐		☐
	Forms a complete understanding using all relevant factors	☐		☐
Recognises significance of given information or the need for new information	Fails to recognise the significance of information that is given	☐		☐
	Recognises the significance of most information that is given	☐		☐
	Recognises the significance of all information that is given	☐		☐
	Identifies specific additional information required	☐		☐
Recalls relevant information	No recall, recalls irrelevant information, inaccurate recall of information	☐		☐
	Accurately recalls relevant information as isolated elements	☐		☐
	Accurately recalls relevant information as integrated elements	☐		☐
Sets a realistic goal	Does not identify a clear goal	☐		☐
	Identifies relevance of given goals	☐		☐
	Establishes a clear goal	☐		☐

Planning an Approach		**SA**	**Evidence of Planning**	**LV**
Plans an approach to the problem	Does not engage in planning activity	☐		☐
	Undertakes enough planning to begin the solution	☐		☐
	Plans all stages required to achieve the goal	☐		☐
	Considers several alternative approaches and selects the best plan to achieve the goal	☐		☐
Recalls previous relevant or similar problem tasks	Does not refer to previous problems	☐		☐
	Recalls previous problems that may be relevant	☐		☐
Identifies appropriate sub-goals	Does not identify sub-goals	☐		☐
	Breaks the task into smaller sub-goals	☐		☐
Sets an appropriate time frame	Does not consider time frame	☐		☐
	Estimates how long the problem solution should take	☐		☐
	Integrates other commitments into time frame established for solving this problem	☐		☐

Please turn over...

Carrying out the Plan		**SA**	**Evidence of Planned Execution**	**LV**
Begins to follow the set plan	Begins to work on the problem without a clear system	☐		☐
	Works systematically, but without reference to plan	☐		☐
	Works systematically and follows set plan	☐		☐
Activates relevant knowledge	Does not activate relevant knowledge or activates incorrect information	☐		☐
	Activates some, but not all, relevant knowledge	☐		☐
	Activates and uses accurate relevant knowledge	☐		☐
Application of strategies	Applies a given or an established procedure	☐		☐
	Selects and implements a procedure from suggested ones	☐		☐
	Adapts an existing, or creates a new, procedure	☐		☐

Monitoring Progress		**SA**	**Evidence of Monitoring**	**LV**
Checks progress towards goal	Does not check progress against sub-goals or final goal	☐		☐
	Periodically checks solution progress against sub-goals and final goal	☐		☐
Responds to unexpected problems along the way	Does not recognise problems that arise	☐		☐
	Identifies unexpected problems	☐		☐
	Diagnoses causes of unexpected problems	☐		☐
	Makes adjustments to rectify suspected causes of unexpected problems	☐		☐
Reviews original plan	Does not review original plan	☐		☐
	Reviews original plan	☐		☐
	Reviews and revises original plan or verifies that original plan is appropriate	☐		☐
Checks original understanding and definition of problem	Does not check original understanding and definition of problem	☐		☐
	Reviews understanding and definition of problem	☐		☐
	Reviews and revises understanding of problem or verifies original definition	☐		☐

Reflecting on the Result		**SA**	**Evidence of Reflection**	**LV**
Reviews efficiency and effectiveness of problem approach	Does not reflect on solution efficiency and effectiveness	☐		☐
	Reviews efficiency and effectiveness of solution strategy	☐		☐
	Identifies improvements to solution strategy	☐		☐
Compares current problem with previously encountered ones	No comparisons with previous related tasks	☐		☐
	Compares current task with previous ones	☐		☐
	Notes ways in which current experience might have helped past problems	☐		☐
Anticipates situations in which current approach might be useful	No anticipation of future applications of solution strategy	☐		☐
	Considers future broader applications of solution strategy	☐		☐

PSSAT R3 10/03

Appendix 9: Teacher Information Package

Employability Skills Assessment Project

Using the Problem-Solving Skills Assessment Tool

David D Curtis

December 2003

Contents

Executive Summary

This booklet was prepared for lecturers who will use the Problem-Solving Skills Assessment Tool (PSSAT) in their classes as part of the Employability Skills Assessment Project. This project is part of a research program being undertaken by David Curtis towards his PhD in the School of Education at Flinders University.

The processes that accompany the use of the assessment tool are described. Teaching staff are responsible for selecting assessment tasks, which are already part of the assessment program of the course, that most effectively will enable students to develop and demonstrate their problem-solving ability. The opportunity to have their problem-solving ability assessed through these tasks is promoted among students. Students undertake self-assessment in which they examine their performance, decide on the performance levels that describe their achievements, and identify evidence for their performance. Teachers examine the evidence that has been presented and make a judgment about the performance level that is compatible with the evidence. This is the level of performance that is recorded for each indicator. Feedback is provided to students to enable them to make more effective judgments of their own performance and to improve it.

The process of performing, assessing, validating, commenting and thereby enhancing problem-solving performance is shown in an annotated diagram. (See page 4).

A copy of the Problem-Solving Assessment Tool is included. (See pages 5 and 6).

The document outlines the development of the PSSAT and explains the theories of problem-solving and assessment upon which it is based.

Introduction

Problem-solving is one of the most widely recognised generic skills. In Australia, these skills have been called key competencies and, recently, employability skills. In this booklet, we will use the term employability skills.

Purpose

The Problem-Solving Skills Assessment Tool (PSSAT) has been developed to promote problem-solving and to assess it. This tool is planned as a template for the development of additional tools to enhance and assess other employability skills.

It has been developed in cooperation with lecturers in the TAFE sector. It has been through several drafts and has been modified based on feedback from lecturers and students. It has been used in two programs in the TAFE sector. We now want to use it a wider range of courses.

Key Elements of the Problem-Solving Skills Assessment Tool

Assessment using the PSSAT is:

- Authentic — assessment occurs in real tasks that learners do anyway in context
- Performance-based — assessment based on what learners show they can do
- Evidence-based — learners present evidence of their ability
- Criterion-referenced — learners are assessed against explicit performance criteria
- Multi-level — indicators have several explicit performance levels

This document describes how the tool might be used and outlines its development.

More Than Just Assessment

The Problem-Solving Skills Assessment Tool (PSSAT) was designed and developed to achieve several purposes. It is an assessment tool and it was developed to enable assessments of problem-solving to be conducted. However, it is not only an assessment tool. It was also developed to encourage the growth of this important set of skills. The development of problem-solving ability is planned in two ways. Learners must know what problem-solving is, and the tool draws attention to the major steps in problem-solving and to particular skills that together make up problem-solving. In this way they will learn what problem-solving is.

A very good way for learners to understand problem-solving is for them to use general problem-solving processes in tasks that they must undertake as part of their routine learning experiences and to analyse and reflect on their use of those processes. Thus, the instrument was designed so that students could use it in a self-assessment activity. However, self-assessment alone, without corresponding judgments by teachers, will not lead to records of performance that are acceptable to employers. Further, self-assessment, without support from mentors, teachers or lecturers, is unlikely to lead to the high levels of learning that are expected. Thus an important part of learning is for learners to receive feedback from their teachers on their performance and on their own assessment of their learning. The PSSAT was therefore also designed as a tool to support the judgments made by teachers and as a framework to enable them to provide informative feedback to learners.

At first glance, the PSSAT may appear to be a typical multiple-choice assessment instrument. However, it is not a tick-and-flick checklist. People, teachers and learners, who use it only as a check list will miss out on its developmental benefits.

Let's now look at the processes of problem-solving. Then we will examine the Problem-Solving Assessment Tool, and see how it can be used. Finally, we will review its development.

From Problem-Solving Processes to Performance Assessment

Problem-solving can be thought of has having five major stages.

- Recognising, understanding and defining the problem
- Developing a plan to solve the problem

- Executing the plan
- Monitoring performance
- Reflecting on the performance and consolidating learning.

These major stages, along with descriptions of them and relationships among them, are shown in the figure on page 4 (The Five Steps: Component Processes of Problem-Solving).

Performance Indicators

In order to use these steps to evaluate problem-solving performance, each step has been expanded to form a set of indicators. For example, if someone is going to be effective in solving a problem, they will have to recognise that a problem exists, they will have to understand what the problem really is, use any relevant information that is given, and set a goal. Look now at the PSSAT on pages 5 and 6. Notice that the first main stage of problem-solving, defining the problem, has a set of indicators that reflect effective practice in this stage of problem-solving. Each of the five main problem-solving processes is broken down into a similar set of indicators.

Performance Levels

For each indicator, a set of performance levels is defined. Indicators do not all have the same number of performance levels. This has been done because for some indicators, it is easy to make relatively fine judgments and up to four levels can be discriminated. For other indicators, it is more difficult to make these fine distinctions, and for some only two performance levels are described. We have tried to strike a balance between making fine judgments in order to give good feedback to learners and being able to complete an assessment quickly, and therefore economically for both learners undertaking a self-assessment and for lecturers completing an assessment of performance.

Performance Assessment

In order for learners to assess their own performance, and for assessors to make consistent judgments of the evidence that learners present, a set of prompts has been prepared. These are designed to facilitate judgments of the performance level that best fits the evidence that is presented. See the prompts on page 7.

Preparing for Assessment

We suggest that lecturers examine the set of assignments for each unit and select a few that provide good opportunities for student to develop and test their problem-solving ability. Students should then be told which assignments are the best ones for a problem-solving assessment, and the availability of this employability skills assessment should be promoted among students. Restricting the number of assignments available for employability skills assessments will help to minimise the assessment load on staff.

The Assessment Procedure

The assessment process has been developed to stimulate learning, to produce reliable assessments of performance, and to do so as efficiently as possible. For these reasons, the assessment process comprises several key steps.

First, learners complete a self-assessment activity, using the same tool and prompts as their lecturers will in making final performance judgments. An important outcome of this exercise is the identification of evidence in support of the performance levels that students believe they have attained.

Students then present their completed PSSAT form to their lecturer. Their form indicates the evidence that they have identified. The lecturer makes a judgment about the performance level that is consistent with the evidence presented. Lecturers do not need to seek further evidence.

Finally, lecturers should provide feedback to learners. The most critical aspect of this occurs if the learner has misunderstood the performance criteria or the evidence requirements. It is important that learners develop the ability to make accurate self assessments. If they can do this, they have a basis for continuing to develop their ability well beyond their course.

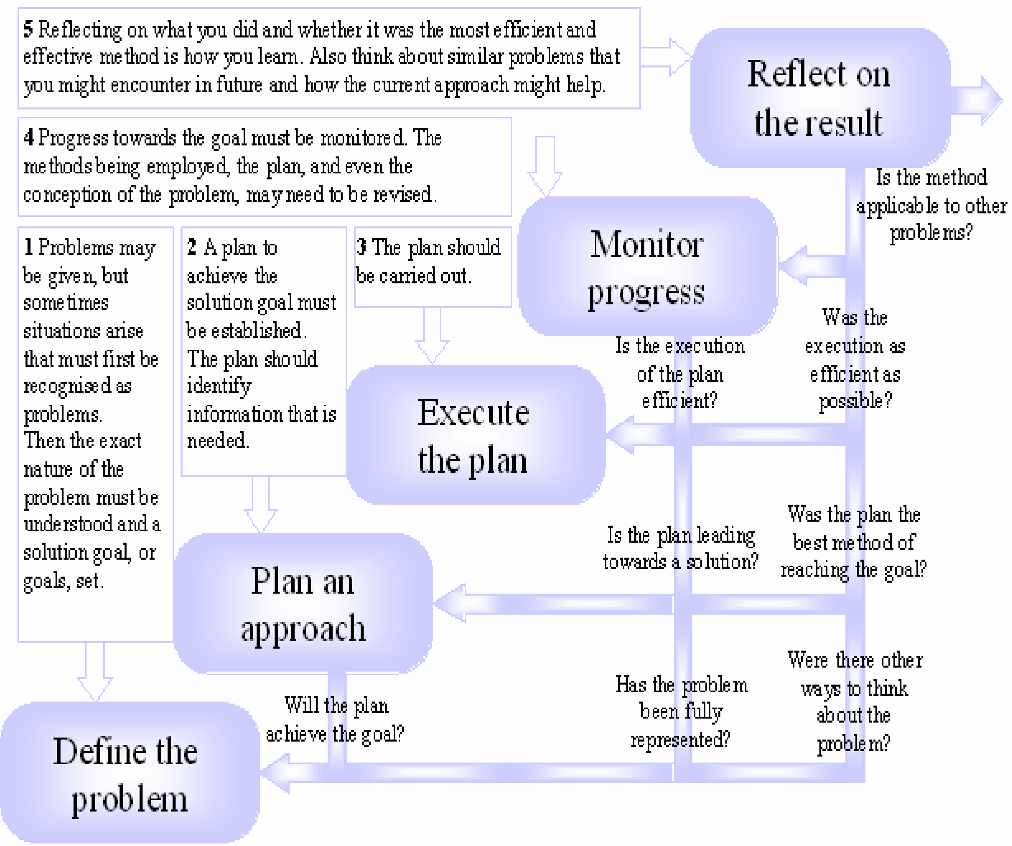


Figure 1: The Five Steps: Component Processes of Problem-Solving

Problem Solving Skills Assessment Tool

Name ____________________ **ID** __________

Unit ____________________ **Task** __________

Lecturer ____________________ **Date** __________

Self Assessment **SA** Lecturer Validation **LV**

Defining the Problem		**SA**	**Evidence of Definition**	**LV**
Forms a correct understanding of the problem	Misunderstands the problem	☐		☐
	Forms a partial understanding using some given information	☐		☐
	Forms a complete understanding using all relevant factors	☐		☐
Recognises significance of given information or the need for new information	Fails to recognise the significance of information that is given	☐		☐
	Recognises the significance of most information that is given	☐		☐
	Recognises the significance of all information that is given	☐		☐
	Identifies specific additional information required	☐		☐
Recalls relevant information	No recall, recalls irrelevant information, inaccurate recall of information	☐		☐
	Accurately recalls relevant information as isolated elements	☐		☐
	Accurately recalls relevant information as integrated elements	☐		☐
Sets a realistic goal	Does not identify a clear goal	☐		☐
	Identifies relevance of given goals	☐		☐
	Establishes a clear goal	☐		☐

Planning an Approach		**SA**	**Evidence of Planning**	**LV**
Plans an approach to the problem	Does not engage in planning activity	☐		☐
	Undertakes enough planning to begin the solution	☐		☐
	Plans all stages required to achieve the goal	☐		☐
	Considers several alternative approaches and selects the best plan to achieve the goal	☐		☐
Recalls previous relevant or similar problem tasks	Does not refer to previous problems	☐		☐
	Recalls previous problems that may be relevant	☐		☐
Identifies appropriate sub-goals	Does not identify sub-goals	☐		☐
	Breaks the task into smaller sub-goals	☐		☐
Sets an appropriate time frame	Does not consider time frame	☐		☐
	Estimates how long the problem solution should take	☐		☐
	Integrates other commitments into time frame established for solving this problem	☐		☐

Using the Problem-Solving Skills Assessment Tool

Carrying out the Plan		SA	Evidence of Planned Execution	LV
Begins to follow the set plan	Begins to work on the problem without a clear system	☐		☐
	Works systematically, but without reference to plan	☐		☐
	Works systematically and follows set plan	☐		☐
Activates relevant knowledge	Does not activate relevant knowledge or activates incorrect information	☐		☐
	Activates some, but not all, relevant knowledge	☐		☐
	Activates and uses accurate relevant knowledge	☐		☐
Application of strategies	Applies a given or an established procedure	☐		☐
	Selects and implements a procedure from suggested ones	☐		☐
	Adapts an existing, or creates a new, procedure	☐		☐

Monitoring Progress		SA	Evidence of Monitoring	LV
Checks progress towards goal	Does not check progress against sub-goals or final goal	☐		☐
	Periodically checks solution progress against sub-goals and final goal	☐		☐
Responds to unexpected problems along the way	Does not recognise problems that arise	☐		☐
	Identifies unexpected problems	☐		☐
	Diagnoses causes of unexpected problems	☐		☐
	Makes adjustments to rectify suspected causes of unexpected problems	☐		☐
Reviews original plan	Does not review original plan	☐		☐
	Reviews original plan	☐		☐
	Reviews and revises original plan or verifies that original plan is appropriate	☐		☐
Checks original understanding and definition of problem	Does not check original understanding and definition of problem	☐		☐
	Reviews understanding and definition of problem	☐		☐
	Reviews and revises understanding of problem or verifies original definition	☐		☐

Reflecting on the Result		SA	Evidence of Reflection	LV
Reviews efficiency and effectiveness of problem approach	Does not reflect on solution efficiency and effectiveness	☐		☐
	Reviews efficiency and effectiveness of solution strategy	☐		☐
	Identifies improvements to solution strategy	☐		☐
Compares current problem with previously encountered ones	No comparisons with previous related tasks	☐		☐
	Compares current task with previous ones	☐		☐
	Notes ways in which current experience might have helped past problems	☐		☐
Anticipates situations in which current approach might be useful	No anticipation of future applications of solution strategy	☐		☐
	Considers future broader applications of solution strategy	☐		☐

Page 6

Scoring Instructions

For each indicator, examine the evidence presented by the candidate and select the performance description that most closely matches the candidate's evidence. The first option for each indicator is the 'did not perform' level. Other options indicate performance above this level.

If you believe that the candidate has implicitly performed on this indicator, but has not presented obvious evidence, leave the boxes for that indicator blank.

Prompts for Indicators

Defining the Problem

Forms a correct understanding of the problem. How complete is the candidate's understanding of the problem? Has s/he formed a simple understanding based on one aspect of it (say resistance using Ohm's Law) or a more complex one based on several aspects (say resistance and heat dissipation and …)?

Recognises significance of given information. Has the candidate noted the significance of information that is given? Does the candidate recognise that the information that has been given and what they already know are not enough to enable them to proceed? They may need to check procedure manuals or other reference sources to identify specific additional information.

Recalls relevant information. Does the candidate recognise that they already know things that are important in solving the problem? Does s/he recall that information accurately? Is the information recalled as separate pieces of information or does the candidate show an understanding of how it all relates together in relation to the problem task?

Sets a realistic goal. Does the candidate identify a given goal or establish their own clear and achievable goal? Is the goal stated clearly enough so that they will know that they have reached an appropriate solution?

Planning an Approach

Plans an approach to the problem. Does the candidate reveal any planning? Sometimes this will be short term planning, ie just enough to get started but not enough to get to a final solution. This is fine for novices, because it will be easier for them to set a short term goal and to achieve that before going onto the next phase of the problem. A more experienced candidate will set a series of goals that will guide them from the beginning of their solution attempt to the goal.

Recalls previous relevant or similar problem tasks. Good problem-solvers think back over their past experiences to find a problem, similar to the current one, that they have encountered previously. This will help them to develop a solution strategy. Can the candidate tell you about similar past problems? Can they tell you in what ways past problems were both similar to and different from the current one? Can they use this to help develop a solution strategy?

Identifies appropriate sub-goals. Many problems involve sub-problems or phases in their solution attempt. Can they set a final goal and then establish milestones or sub-goals along the way?

Sets an appropriate time frame. Is the candidate clear about how long the problem should take to solve? Is their estimate reasonable? Do they integrate this with other commitments they may have?

Carrying out a Plan

Begins to follow the set plan. Does the candidate begin to work systematically? If the candidate has a clear plan, do they show that they are following it? In cases where experienced candidates did not seem to develop a plan, is there evidence that they are working systematically?

Activates relevant knowledge. If the candidate needed to recall information, has s/he been able to apply that knowledge correctly in the current situation?

Application of strategies. Did the candidate apply an established procedure (eg experiment procedure or flow-chart)? Did they select and implement an appropriate strategy from several available? Did they adapt or manipulate an existing, or create a new, problem-solving process?

Monitoring Progress

Checks progress towards goal. Does the candidate pause, or do they give any indication that during the solution attempt, they paused and checked their progress against sub-goals or the final goal to see that they were getting closer to a solution?

Response to unexpected problems along the way. If an unexpected problem arose, and they were not able to proceed as expected, do they show any evidence that they have looked for the cause? For example, do they show the need to recheck information that was given or assumed? Do they seek additional information to verify their understanding of the problem? If the solution is going according to plan, especially if they are following a set procedure, this indicator may not be apparent.

Reviews original plan. If the original plan does not appear to be working, does the candidate review it and adopt a new approach or set new goals? If the solution is going according to plan, does the candidate review the plan to seek improvements?

Checks original understanding and definition of problem. If the candidate has found that the solution attempt is not working, do they go back to the definition of the problem and pause to think again about the nature of the problem? Here, a candidate who had formed an incorrect understanding or definition of the problem might think about other possible dimensions of the problem and redefine the problem.

Reflecting on the Result

Reviews efficiency and effectiveness of problem approach. Having solved the problem, has the candidate considered how they might have been more efficient? Would the candidate make any major changes to the way they went about this activity if they had to do it again?

Compares current problem with previously encountered ones. Does the candidate compare the just-completed problem with others that had been attempted earlier? Can they identify techniques in this solution that would have helped with earlier problems? Were there techniques that were used previously that would have helped in the current problem?

Anticipates situations in which current approach might be useful. Does the candidate show evidence that what they have just done could be used in other real world situations? Can they identify a specific task where the current approach is likely to be applicable?

The Development of the Problem-Solving Skills Assessment Tool

The Problem-Solving Skills Assessment Tool and the assessment process in which it is embedded are expected to provide a sound basis for enhancing problem-solving performance and for reporting on its achievement.

The Context for Assessing Problem-Solving

In an evolving workforce, the demands on individuals change and it has been recognised that people require both vocationally specific skills that are in demand immediately and a more general set of skills that will enable individuals to adapt to the changing demands of work. The more general set of skills have been called key competencies, generic skills, employability skills, and many other names. In Australia, the terms key competencies and employability skills are used commonly. The skills that are recognised under these labels vary, but problem-solving is almost universally included. Effective problem-solving is closely related to an ability to learn from experience and therefore to contribute to a high performance work environment. Assessing problem-solving, and therefore being able to report on it, is expected to enhance the contribution that education and training will have on Australia's workforce performance.

The Approach

The approach used in developing the PSSAT was tested in the project *The Authentic Performance-Based Assessment of Problem-Solving* (Curtis & Denton, 2003). This method had three key stages:

- Definition of major problem-solving processes
- Selection of indicators for each process
- Identification of performance levels and criteria for each indicator.

The approach taken in developing this assessment tool involved locating generally accepted descriptions of problem-solving. Descriptions were sought that portrayed the **processes** that would be used by people who are competent in this area. In many situations where individuals are being assessed, it is common to examine and score the **products** of the assessed task. For example, if a problem is set as an assessment task, the score given depends on the learner getting the correct answer, and it is the final answer that is judged as right or wrong. The reason for focusing on problem-solving processes in the PSSAT is that we are seeking an indication of the likely future performance of individuals in new situations rather than simply reporting their performance on past tasks. It is argued that the processes that learners demonstrate are more likely than past products to be indicative of future performance.

Definition of Major Problem-Solving Processes

For over a century problem-solving has been a preoccupation of educators and many authors have written about it and have provided a diverse range of descriptions of it. There are some major differences among the various descriptions. Among several approaches described by Greeno, Collins & Resnick (1996), two models of problem-solving appear to be particularly useful in describing real world problem-solving. One is based on the idea that there is a set of very general processes that can be applied in a range of contexts. This is often called the information processing model (Newell & Simon, 1972). It has some strengths in that a set of general processes can be applied to solving a diversity of problems. While this model has had considerable success in describing problem-solving, there is also some counter evidence to it. Even when people have demonstrated that they have used some general processes to solve a problem, they do not automatically apply the same processes to similar problems for which they would work equally well. The second approach to problem-solving considered here is the situated model. People who advocate this approach argue that with experience in a domain, individuals develop specialised procedures for that domain and that they use them in preference to more generalised procedures that they know (Lave, 1988; Scribner, 1986). While this theory describes how experienced individuals work, it does not indicate how novices develop the expertise that allows them to use their specialised techniques. Because the focus of education and training programs is on enabling inexperienced individuals to develop sufficient competence to permit them to begin productive work in a field, the information processing model has been adopted as the basis of the development and assessment program in this project.

Among information processing theorists, there is some diversity in the number of processes implicated in problem-solving and in the labels given to them. For example, Polya (1957) listed four major processes, Bransford & Stein (1984) described five, while others have suggested more. However, there are similarities among the various schemes. Where Polya suggested that the first step is to understand the problem, Bransford and Stein advocated two components to this phase – identify the problem and define it precisely. In arriving at a modal view of problem-solving, the following processes are proposed:

- Defining the problem
- Planning an approach
- Carrying out the plan
- Monitoring progress
- Reflecting on the result

These major processes form the basis of the PSSAT. An annotated diagram that shows these processes is presented on page 4 as The Five Steps: Component Processes of Problem-Solving.

Selection of Performance Indicators

Having identified major processes, the next stage involved establishing a set of indicators for each process. The indicators were selected from actions that an effective problem-solver might perform in enacting the major problem-solving processes. These are not the only actions that such a person might undertake but are a sample of the possible indicators of each of the major processes. No doubt other indicators are available and could be used in addition to or instead of those that have been selected.

Definition of Performance Levels and Criteria

For each indicator, a set of performance levels were defined. These levels are based on the SOLO (Structure of the Observed Learning Outcome) taxonomy (Biggs & Collis, 1982) which included up to five performance levels based on the complexity of the thinking that underlies each action. At the lowest level, the person being assessed does not show evidence of a particular indicator or presents a behaviour that shows a substantially flawed application of the indicator. The highest level available is the behaviour that would be expected of a person who is able to deduce new relationships by extending the information given. A summary of the performance levels of the SOLO taxonomy is shown in Table 1.

It is important to note that the criteria for each successive level within an indicator reflect a higher level of performance. Initially, items are scored using the numbers 1, 2, 3..., but, it is possible that the third level of one indicator might require a higher level of ability than the third level of another. The methods of analysis used for the PSSAT enable these differences to be factored into a final score.

Not all indicators use all five possible performance levels of the SOLO taxonomy. It may be possible to refine the performance criteria for each indicator to allow five levels to be differentiated. However, the instrument has been designed for efficient administration. If very fine judgments are required to discriminate small differences in performance between adjacent levels on an indicator, it will take more time to use the tool than if differences between levels are readily apparent. For this reason, the smallest number of levels consistent with ease of judgment have been identified for each indicator. Past experience has shown that having only two levels can reduce the usefulness of some indicators. The number of levels chosen for each indicator reflects a balance between the competing factors of maximising information and minimising assessor load, and different numbers of levels are recognised for different indicators.

From the teacher's and the student's perspectives, it is important to make judgments using the specified criteria without being concerned about recording the highest possible performance level. The PSSAT is not meant to be a competitive test in which individuals are compared with each other. Rather, individuals are encouraged to use the instrument on several occasions (at least four) and to compare their performances on successive occasions. Over time, as learners recognise their actual performance levels and identify their strengths and weaknesses, and with the feedback provided by their teachers, they will improve in those areas where their performance is incomplete. In this way, assessment is used as a means of improving performance.

Table 1: Performance Levels of Indicators Using the SOLO Taxonomy

SOLO Level	Description	Score
Pre-structural	No knowledge, inaccurate recall, or does not use relevant knowledge	0
Uni-structural	Uses relevant knowledge/skill elements in isolation	1
Multi-structural	Uses relevant knowledge/skill elements in combination	2
Relational	Can generalise using knowledge within the problem situation	3
Extended abstract	Can extend what has been found through the problem to other situations	4

Using the Problem-Solving Skills Assessment Tool

How the PSSAT is used is a decision for teachers and learners. It can be used as a summative assessment instrument, but it was developed to be used in a formative assessment process. A formative assessment approach is outlined.

Task Selection

The PSSAT was designed to be used in authentic assessment tasks. These are tasks that learners normally undertake in their learning rather than tasks specifically constructed to test their problem-solving ability. Many tasks that learners routinely undertake in their courses include opportunities to solve problems. But learning tasks are also opportunities for learners to demonstrate a wide range of generic abilities such as problem-solving, team work, communication and others. Some tasks will be better suited to the demonstration and development of certain generic abilities. It seems prudent to match generic abilities with tasks that provide the greatest opportunity to develop and demonstrate that set of skills.

Training packages, and other course specification documents, list a range of assessment activities. It is suggested that a range of tasks that offer potential for problem-solving assessment be selected and promoted to learners for this purpose. Because the PSSAT is intended to be used on multiple occasions – at least four during a course – it would be sensible to suggest tasks in the early, middle and late stages of a course or sequence of units. Learners could select tasks for problem-solving assessment from the suggested range. In this way, they will have the opportunity to develop and to demonstrate their skills in tasks in which they have a personal investment through choice.

A key task for lecturers is to select tasks that they believe provide opportunities for students to develop and demonstrate their problem-solving ability. The list of selected tasks should be actively promoted to students through in-class announcements, posters, and email messages to make students aware of the opportunities available to them.

Self-Assessment

Self-assessment alone will not have credibility with employers, but self-assessment is an important element of learners' abilities to develop their skills. In traditional forms of assessment, students undertake tasks and submit their work for grading. Their teacher assigns a mark and makes comments about the work, but the student is not directly involved in making judgments about the quality of the work they produce. Boud (2002) has argued that ignoring self-assessment denies learners a powerful opportunity to be involved in monitoring their own learning and consequently they fail to develop important 'learning to learn' skills.

In the self-assessment approach that we are advocating, students are required to take an active role in assessment. The purpose is not simply to get students to assign the same grade as the teacher, but to use the same criteria and develop the ability to reflect on and to make informed judgments of their own work. They do this by using the PSSAT as a guide to presenting evidence in support of the judgments that they have made about their performance level on each indicator. The student's task is to examine the criteria, to make a judgment about their performance level and to identify supportive evidence.

The evidence may be apparent in the processes that students have used or it may exist in the product that they have developed. The forms of evidence that are acceptable to their teachers need to be made clear to students before they undertake the assignment. In order to make the next phase of assessment as efficient as possible, it is suggested that learners be required to summarise evidence for their performance in writing, perhaps by referring to evidence that is documented in their assignment or that is apparent in a product.

Self-assessment has another important advantage. It helps to make the assessment process more efficient than traditional assessment practices because it locates responsibility for gathering evidence of performance with the learner. This leaves the assessor with the task of making judgments about the value of the evidence against specified performance criteria. In conventional assessment approaches, the assessor has to examine the learner's work and locate evidence before judgment is exercised.

A series of prompts, expressed as questions, and that relate to each of the performance indicators are provided (see page 7). These are written both for learners, who should use them to guide self-assessment, and for teachers who should use them in validating performance.

Validation and Feedback

The assessment process that we are advocating has two components, namely self-assessment and its complement, teacher validation. When the learner presents evidence in support of the judgment of their own performance, the role of the teacher is to examine the evidence to see whether it does support the level of performance that the student claims. Where it does, the learner's judgment is affirmed. Where it does not, whether it is higher or lower than the teacher's opinion, it is assumed that the learner has not yet understood the problem-solving process or the evidence requirements adequately and that teacher feedback is required. This takes two forms. The learner must be informed about how that indicator contributes to problem-solving and about the evidence required to support a claim for the level of performance in question.

It is through a combination of self-assessment followed by teacher validation and appropriate feedback that learners will develop a more complete understanding of problem-solving.

A further benefit of this dual assessment approach is that learners develop their ability to become reflective learners and continue the process of learning to learn, which is now identified as a component of many generic skills sets, including the ACCI/BCA proposal (Australian Chamber of Commerce and Industry & Business Council of Australia, 2002).

Recording and Certifying Results

When a learner participates in a problem-solving assessment using the PSSAT, a record should be made of the assessment. This record should identify the learner, the teacher, the assessment task that was used, the date, and the score as assessed by the teacher on the assessment. It would also be sensible for both the teacher and the learner to retain a completed copy of the PSSAT form. The learner can use copies of completed PSSAT as part of a portfolio that shows evidence of their developing expertise.

When the learner has completed several of these assessments and is seeking a formal record of her/his performance, it is desirable to provide a Statement of Attainment. This statement should identify the final assessment on each of the employability skills attempted. In this sense, the PSSAT provides a final, summative assessment result. For the learner's benefit, it is also desirable that the Statement, perhaps on the reverse side, should include the sequence of assessment results. This will show the learner's progress. It is most likely that this sequence will reveal a set of increasing scores. The final score and the rate of increase together indicate more about the learner's potential than a final score alone.

Limitations and Restrictions

This version of the PSSAT is still regarded as a draft. It has been the subject of a pilot and a validation study, but further work is required to demonstrate that it is an effective instrument across a wide range of disciplines. Users should therefore be aware of its draft status and should only use the instrument in situations where they believe that it will make a useful contribution to student learning and where its use can lead to valid interpretations of students' emerging abilities.

It is being distributed on a limited basis for evaluation and comment by those who might be interested in using an instrument like this for the assessment of problem-solving. You may chose to use it to assess students' work, but if you do, you should be aware of its trial status and its limitations.

The PSSAT is part of research being conducted into the assessment, measurement and reporting of employability skills in a PhD program at Flinders University. Because the research has not been published, the PSSAT may not be used in research and this work may not be cited without the author's permission.

Its developer, David Curtis, is keen to receive feedback on the PSSAT. Comments about the PSSAT should be directed to:

David D Curtis
Centre for Lifelong Learning and Development and
School of Education, Sturt Campus, Flinders University
Sturt Road, Bedford Park SA 5042
Ph. (08) 8201 3941 Fax (08) 8340 9119
Email david.curtis@flinders.edu.au

References

Australian Chamber of Commerce and Industry, & Business Council of Australia. (2002). *Employability Skills for the Future*. Canberra: Department of Education, Science and Training.

Biggs, J. B., & Collis, K. F. (1982). *Evaluating the quality of learning: the SOLO taxonomy (Structure of the Observed Learning Outcome)*. New York: Academic Press.

Boud, D. (2002). Assessment: have we forgotten something? In Australian National Training Authority (Ed.), *Fresh thinking about learning and learners: a blue sky project* (pp. 42-45). Melbourne: Australian National Training Authority.

Bransford, J. D., & Stein, B. S. (1984). *The ideal problem solver: a guide for improving thinking, learning, and creativity*. New York: W H Freeman.

Curtis, D. D., & Denton, R. (2003). *The authentic performance-based assessment of problem-solving*. Adelaide: NCVER.

Greeno, J. G., Collins, A. M., & Resnick, L. B. (1996). Cognition and learning. In D. C. Berliner & R. C. Calfee (Eds.), *Handbook of educational psychology* (pp. 15-46). New York: Macmillan.

Lave, J. (1988). *Cognition in practice: Mind, mathematics and culture in everyday life*. Cambridge, N.Y.: Cambridge University Press.

Newell, A., & Simon, H. A. (1972). *Human problem solving*. Englewood Cliffs: Prentice Hall.

Polya, G. (1957). *How to solve it. A new aspect of mathematical method* (2nd ed.). Harmondsworth: Penguin.

Scribner, S. (1986). Thinking in action: some characteristics of practical thought. In R. J. Sternberg & R. K. Wagner (Eds.), *Practical intelligence. Nature and origins of competence in the everyday world* (pp. 13-30). Cambridge: Cambridge University Press.

Appendix 10: Student Information Package

Contents

Leaflet	Contents
What's in this package?	This is a description of the package contents, including a brief summary of the leaflets in it and advice on using the package
What's this project about?	This describes the project, emphasising likely employment advantages of developing and demonstrating generic skills.
What's in it for me?	This leaflet emphasises the development of generic skills and their employment implications. It also includes information on the TSA and on the student incentives.
Participation consent form	This is the consent form as approved by the Research Ethics Committee.
Personal details	This form seeks basic demographic information from participants.
Other employability skills	This leaflet provides information on other employability skills. In particular, it emphasises communication and teamwork as they and problem solving are 'the big three.'
How do I get an assessment?	This leaflet describes the problem solving process. In particular, it shows how the PSSAT is used, it describes the indicators and performance criteria, it describes self assessment and lecturer validation, and it presents the problem solving model diagrammatically.
The Problem Solving Skills Assessment Tool	This is the assessment instrument that students use in assessing their problem solving performance and having it validated by their lecturer.

Employability Skills Assessment Project

What's In This package?

This package was developed for you. It contains information about the Employability Skills Assessment Project. We hope that you will take part in it. It also contains other information that we think you will find useful.

What's This Project About?

This leaflet explains what is involved in the project and why it is being done. It also explains what employability skills are and what employers are looking for when they recruit new staff and when they promote existing staff.

What's In It For Me?

What do you have to do and what will you get out of participating in this project? This leaflet answers these questions. We want you to improve your skills, to get recognition for doing this, to help out in a research program, and we will reward those who do take part.

Participation Consent Form

This is the form that you will need to sign if you plan to take part in the project.

Personal Details

We want to collect a few details from you. The information that is collected will remain confidential.

Other Employability Skills

The project will focus on problem-solving, but employers are also looking for people with other skills. Two other important sets of skills are communication and team work. This leaflet will help you understand what employers are looking for and how to show that you have these abilities.

This is not part of the project, just something extra that you might find useful.

How Do I Get an Assessment?

This leaflet explains the process that you will follow in getting a problem-solving assessment. It tells you how to do a self-assessment, how to gather and present evidence of your ability, and then how your lecturer will give you feedback on your performance. It includes a diagram that will help you understand what problem-solving is and how you can develop your problem-solving ability.

The Problem-Solving Skills Assessment Tool

The device that we have developed for this project is called the Problem-Solving Skills Assessment Tool (PSSAT). It was developed to help you improve your performance, assess your own problem-solving ability, and to enable lecturers to provide consistent assessments of this important ability. Additional copies of this are available from your lecturers.

Using the Package

We have provided this information in a folder because we want you to use the folder as a place to keep information about your developing abilities. If you already have a resumé or a collection of documents, for example references and job reports, keep this folder with that material. We think that you will find this information useful in applying for jobs.

Employability Skills Assessment Project

What's This Project About?

This leaflet explains the purpose of the project, outlines what is involved in being part of it, and then shows what employability skills are and why employers value them.

Employability Skills

In addition to your job specific skills, employers are now expecting people to have a range of other more general abilities. These include communication, team work and problem-solving skills.

For more information employability skills, see *Employability skills: What are they?* (Page 3)

This project has been developed to show that employability skills can be improved and they can be assessed and reported in ways that employers will find helpful when they recruit new staff.

Problem-Solving

At this stage of the project, emphasis is being placed on problem-solving. In a pilot stage of the project, a new method for improving people's problem-solving skills and for assessing them was developed.

The Development and Assessment Procedure

This method does not involve any new assessment tasks. Instead, it uses the existing assignments that you normally undertake as your work through the units in your course. When you do an assignment in your course, you have it assessed by your lecturer as you have always done. Then, you review it and think about the problem-solving processes that you followed while you were doing it. To help you with the review, we have developed an assessment tool, the Problem-Solving Skills Assessment Tool (PSSAT), to guide you through the review. This is a type of self-assessment. You assess your own problem-solving ability by using the tool and as you do this, you identify evidence of your performance.

To find out how to assess your own performance using the PSSAT, see *How do I get an assessment.*

Then you present your self-assessment, along with supportive evidence, to your lecturer. Using the same assessment tool as you used, your lecturer will look at your evidence and they will make their professional judgment of your performance. They will tell you what their assessment is. If there is any difference between your judgment and theirs, they will explain how they reached their conclusion.

By using the assessment tool as a guide, you should learn about the main steps in problem-solving.

To view the main steps in problem-solving, see the diagram in the leaflet *How do I get an assessment?*

By assessing problem-solving ability, in an assigned task that is part of your vocational course, you are developing your general skills in a work-related context. This is what employers want to see.

By gathering evidence yourself, you are able to point to examples of your emerging ability and therefore to talk about your ability. This is part of the process of learning and developing these important skills. Being able to describe job-related problem-solving performance will also help you in interviews for jobs or promotion.

By getting feedback from your lecturer, you will also increase your knowledge about problem-solving, in a work-related context. This feedback should also help you to improve your performance, so next time you do an employability skills assessment, you should be even better at it. It is a process that you need to do a few times to get the best overall result that you can report to employers.

Page 1

How Long will this Take?

The first time you review an assessment task for your problem-solving ability, it could take up to an hour. This is because you will need to read the leaflet *How do I get an assessment?*

> **We think that this is worth doing, and the benefits of doing it are explained in *What's in it for me?***

After this it will be much quicker. In the past, students have been able to review their assignments and identify evidence in about 10 minutes.

We want you to do four problem-solving assessments over the whole year.

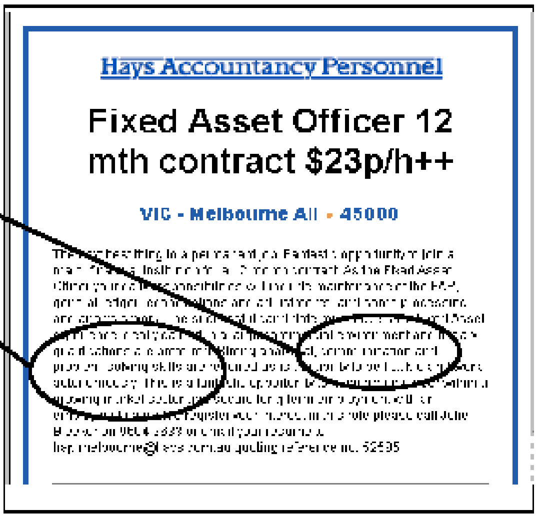
We also want you to do the Tertiary Skills Assessment. This is a written test of your problem-solving, critical thinking, and team work skills.

What Employers Want

The main employer organisations in Australia are the Business Council of Australia, the Australian Chamber of Commerce and Industry, and the Australian Industry Group. They have been putting pressure on governments and education and training providers to improve the skills of the Australian workforce.

But it is not just these big organisations that want these skills to be improved. Ordinary employers want to hire people who have them.

In the advertisement shown below, you will see the value that individual employers place on employability skills when they want to hire staff.



In the advertisement shown on the right, which appeared on the CareerOne web site, the employer is looking for someone with "strong communication and problem solving skills".

Would you be able to provide evidence of your communication and problem solving abilities?

This is just one of many advertisements that appear in newspapers every day. These skills are in high demand.

Put It To The Test

Look in local papers at job advertisements in your chosen area of work. What are employers looking for when they recruit new staff? What do they look for if they are promoting staff to more senior positions?

Go to the CareerOne web site (http://www.careerone.com.au) and try some searches yourself. Search by job title in your area and read the job descriptions that you find. Search for communication, team work, problem solving and other employability skills. Think of other words used to describe these skills, like negotiation and interpersonal skills. How many jobs can you find? What are employers looking for?

Employability Skills: What Are They?

When you look at job advertisements, you will see statements like:

Your excellent interpersonal skills, problem solving ability, sense of urgency and the ability to drive and deliver results at the highest level are expected.	You will have a high level of computer literacy and be able to demonstrate strong customer service and problem solving skills.	Time management and people management skills. Excellent written and verbal communications skills. Good problem solving / resolution skills. Advanced computer skills.

In addition to vocationally specific skills and knowledge, employers understand the importance of recruiting people who have a range of more general skills. These skills are valued in most industries and job roles. Nurses, motor mechanics, accountants, web designers, and most other positions require people who can communicate effectively with co-workers and clients, are team players, can solve problems, and are adaptable as work situations change.

These additional skills are sometimes called generic skills, key competencies, or **employability skills**.

Recently, several major Australian employer organisations, including the Business Council of Australia and the Australian Chamber of Commerce and Industry, released a report in which they listed a set of employability skills. Their list was:

communication skills that contribute to productive and harmonious relations between employees and customers

team work skills that contribute to productive working relationships and outcomes

problem-solving skills that contribute to productive outcomes

initiative and enterprise skills that contribute to innovative outcomes

planning and organising skills that contribute to long-term and short-term strategic planning

self-management skills that contribute to employee satisfaction and growth

learning skills that contribute to ongoing improvement and expansion in employee and company operations and outcomes

technology skills that contribute to effective execution of tasks

Why Are Employability Skills Valued?

People with good job specific knowledge and skills and who also have these employability skills contribute to productivity in what are called high-performance work places. People with these skills are valued because of the contribution that they can make to greater productivity.

People with these skills are also sought because they are the team leaders, supervisors and managers of the future. Companies with an eye to their own future are looking for people who have these special abilities.

How Do I Get These Skills?

You probably already have many of them – to some extent at least. You have developed them at home, at school, through sport and other social activities, and through your past work experience. Employers will value these skills however you got them, but experience has shown that skills are most effective if they are developed in situations similar to the ones where they will be needed. So if you need good team work skills, employers will value them more if you can show that you can apply these skills in a work-like situation.

This is why we are getting people to practice and develop employability skills in routine course assessment tasks, because these tasks are designed to relate directly to work contexts. If you pay attention to these employability skills when you undertake routine course assessment tasks, you will develop them and then be able to describe your abilities in ways that employers will accept.

The Problem-Solving Skills Assessment Tool, and the procedure that we are suggesting you use to develop and assess your skills, is designed to make it easy for you to improve your skills in a work-like situation.

What do I Need to Know?

All the abilities that you develop, whether they are job-specific or more general ones like team work and problem-solving, involve a combination of knowledge, skill, and attitude.

To be effective in any role, whether it is work or sport, you need to have a positive attitude to the role and about your capacity to perform it.

Most tasks that you undertake involve skilful performance. This is most apparent in sport where people are able to perform effectively and with style. In manual work, it is easy to admire people who are skilful at craft work, but this sort of skill also applies in jobs that do not require manual dexterity. Some sales people are quite skilful at finding out what customers are looking for and in helping them get it.

Underlying all good performances is knowledge. You need job specific knowledge, such as the properties of metals if you are going to weld them, and you need to know what employability skills are. If you are going to be an effective team player, you need to know what team work is. If you want to be a good problem-solver, you need to know what problem-solving is.

The Problem-Solving Skills Assessment Tool is designed to let you know the skills that contribute to problem-solving ability, to help you put them into practice, and then to be able to describe your problem-solving ability.

Where Can I Get More Information about Employability Skills?

This project focuses on problem-solving. However, eight employability skills have been listed.

More information about team work and communication is presented in the leaflet *Other Employability Skills*. Together with problem-solving, these make up the three most commonly sought employability skills. This leaflet will help you to understand what is required when employers ask for these skills. It will also help you provide evidence of your communication and team work abilities. However, these are not part of the project. They are provided to help you in applying for jobs and promotion.

An Australian Chamber of Commerce and Industry brochure on employability skills is available from the ACCI web site at:
http://www.acci.asn.au/text_files/education_employment_training/Employability_Skills.pdf

Employability Skills Assessment Project

What's In It For Me?

Getting assessments of your employability skills is worth doing for at least two reasons. You can develop these skills by deliberate practice and employers are seeking people with these skills. There is a third reason. If you participate in this project, you will be entered into a draw for a range of rewards. Read on.

What Do I Have To Do?

If you are prepared to take part in the project, you will need to:

- Complete the Participation Consent form (2 minutes)
- Complete the Personal Details form (3 minutes)
- Take the Tertiary Skills Assessment near the beginning of term 1 (2 hours)
- Undertake a self assessment of your problem solving ability on four assignments that you will be doing for your course (1 hour, first time, 15 minutes each after that)
- Take the Tertiary Skills Assessment near the end of the year (2 hours)

This is a bit of extra work for you, but we think that you will benefit personally and that other students will benefit in future because of your participation.

What Do I Gain?

Enhancing Your Employment Prospects

Employers are beginning to advertise for people with these important skills. Job descriptions are seeking people with good communication skills, good team work skills, and good problem-solving skills. There are other employability skills, but these are the big three. Many people have these abilities to some extent, but many are not able to talk about their abilities, and few can present evidence that they have deliberately developed these capabilities and had them assessed. Having these skills assessed will enable you to present evidence to prospective employers that you have them.

Employability skills assessment not only make future employers aware of your skills but it makes you aware of your skills and gives you the extra confidence in a job interview.

Being Able to Talk about Your Skills

In past trials of our approach to employability skills development, students told us that, even though they had these skills, working on them through their regular assignments helped them to talk about their skills in the context of the work that they expected to do. On the right, read what one of the students said:

> **"I thought that I had good problem solving skills, but doing these assessments has helped me understand and improve my skills. I can also talk about my skills much better than I could before."**

Developing Employability Skills

Doing employability skills assessments will enable you to develop these important abilities. They are assessed in ways that encourage their development. You are asked to assess yourself in order to assemble evidence of your ability, and then to present this evidence to a lecturer. This is quite different from many other forms of assessment. Exams at school are often designed to find out both what you know and what you don't know, you don't know what is going to be in an exam, and you only get one shot at them. There is no secret about what is meant by an employability skill. The assessment describes exactly what your assessor is looking for, so you know

Page 1

exactly what evidence to provide. The first time you elect to have an employability skill assessed, you will find it interesting to have to assess yourself in order to be able to provide or describe the evidence of your ability to an assessor. This will help you to focus on the components of each employability skill, and in turn, this experience will help you to develop and describe the components. Next time you seek an employability skill assessment, you will know much more about it and be able to describe it better and therefore, you are likely to gain a higher level assessment. You should aim to get four assessments of each employability skill during your course.

When you apply for a job that requires high levels of these skills, not only will you be able to provide evidence that you have these skills because of the results of the assessment, you will be able to demonstrate these skills by referring to the assessments themselves as evidence, and you will be able to talk about the components of employability skills. You may even know more about the detail of employability skills than the employer!

The Tertiary Skills Assessment

In this project, you are being asked to do the Tertiary Skills Assessment (TSA). This is similar to the Graduate Skills Assessment (GSA) that many university students do before they graduate. It is a multiple choice test of the employability skills problem-solving, critical thinking and interpersonal understandings.

Some major employers are requiring university graduates to have a score on the GSA. At this stage, it is early days for the TSA, but having a result on this test may well be an advantage in seeking employment. It would be sensible to keep your TSA score along with the other documents that make up your portfolio.

The Feel-Good Factor

Like you, I am a student. I need people to participate in this research so that I can complete my studies. I hope that you will help me to achieve my goal and that you will feel good about doing me a favour.

An Added Bonus

To encourage you to participate in this project, and to reward you for your help with it, if you agree to take part, you will be entered into a draw for a range of prizes.

The main prize is a $250 voucher and two prizes of $100 vouchers for clothing from a major fashion store chain. There are also many other prizes, varying from $10 to $30 vouchers from fast food chains, music stores, cinemas, and clothing stores. We will announce final details when all participating stores and chains have agreed. Winners will be contacted through their TAFE campus. Our aim is to ensure that all participants will be rewarded for their efforts.

Employability Skills Assessment Project

Participation Consent Form

The Employability Skills Assessment Project is an attempt to find a new way to assess and report on individuals' problem-solving skills in real-world situations. Problem-solving skills are to be assessed as part of routine course assessments.

I hereby give my consent to David D Curtis, a researcher in the Centre For Lifelong Learning and Development and the School of Education at Flinders University, and whose signature appears below, to access my problem-solving assessment summaries.

I understand that my participation in the research will include:

- Completion of a form outlining some individual details (name, age, gender, prior education, and work background);
- Completion of the Tertiary Skills Assessment twice; and
- The assessment of problem-solving in routine course assessment tasks on four occasions.

I also understand that, for the purposes of this research, the grades obtained in my studies to date can be accessed for use in this research project.

I give permission for the use of these data, and of other information which I have agreed may be obtained or requested, in the writing up of the study, subject to the following condition:

- That my identity and the results that I achieve in the assessments remain confidential.

I understand that I will be given a report on my Tertiary Skills Assessment results and on the Problem-Solving Assessments for each of the tasks submitted.

My participation in this study is voluntary, and I understand that I may withdraw from the study at any time.

Signatures

Researcher

Name David D Curtis

Signature David D Curtis Date 03 January 2004

Participant

Name

Signature Date

If you require further information about the project, please feel free to contact:

The researcher
David D Curtis
Phone (08) 8201 3941
email david.curtis@flinders.edu.au

His research supervisor
Professor John P Keeves
Phone (08) 8201 2392
email john.keeves@flinders.edu.au

Employability Skills Assessment Project

Personal Details

Name:	
Age (years)	
Gender	❑ Male ❑ Female
Current Course	
Educational Experience	**School Education** (Please indicate the highest level of school education completed) Less than Year 10 ❑ Year 10 ❑ Year 11 ❑ Year 12 ❑ **Post-School Education** (Please indicate the highest level of post-school education completed) No post-school education other than current course ❑ Partial completion of another TAFE course ❑ Completion of another TAFE course ❑ Completion of a university course ❑
Work Experience	Please indicate the number of full time years of work experience in any field, including unpaid work No previous work experience ❑ Up to 1 year ❑ 1 to 5 years ❑ More than 5 years ❑

Thank you for providing the above information. Remember that your details are confidential and they will not be revealed with your name to anyone. When this information is used in the research project, your name will be replaced with a code that will be known only to the researcher.

Employability Skills Assessment Project

Other Employability Skills

Employability Skills

Australian employers recognise a core set of eight employability skills. They are:

communication	planning and organising
team work	self-management
problem-solving	learning
initiative and enterprise	technology

These skills are all important, but three are listed very commonly in job advertisements. The big three are communication, team work and problem-solving. The main focus of this research project is problem-solving and through your involvement in the project, you will develop your problem-solving ability to a high level. The information in this leaflet is provided to guide you in developing and being able to provide evidence of your communication and team work skills.

Communication Skills

Communication is vitally important in all workplaces. You need to communicate with managers and supervisors, co-workers, and clients. Communication takes many forms and many modes. It is impossible to provide a simple check list of things to do in an area as complex as communication in its great variety of forms. The way you communicate is also the way you enable others to develop their perceptions of you. So if you want to create favourable impressions, you must be able to communicate clearly and you must do this in ways that are appropriate in all situations.

The guidelines about communication skills presented here are designed to help you gather evidence of your ability in this area.

Communicative competence is the ability to ensure that communication is both appropriate and effective in a given context. This definition needs to be developed and this will be done by referring to a framework of organising ideas about communication.

Purpose

All communication has a purpose. It may be to inform, to persuade, to instruct, or to seek information. There are many other purposes. You may have a main purpose, for example to inform others that you know enough to be able to do a job for them, but you need to do this in a way that creates a favourable impression of you as a person.

In many roles, for example if you work in retail, you will find that you have several related purposes. To make a sale and keep your clients happy, you need to listen to them to work out what they want, then you need to provide some alternatives, that is to inform them of their options, and then perhaps to persuade that the option you are recommending is the best available.

You need to be clear about your purposes in all acts of communication. If you are asked about your communication ability, you need to show that you are aware of the need to establish a purpose and then choose communication strategies that help to achieve it.

Audience

Being sensitive to your audience is important in communicating effectively. You need to know what you audience wants, needs, or expects. If your purpose is to inform, do you know what information they need or want? Do you know the form or forms of information that will have most appeal? For example, will they want the information in spoken, written, or graphic forms? Whatever the form, do you know the amount of information they want and

Page 1

do you know the level to pitch it.

Sensitivity to cultural aspects of your audience is important. You need to make sure that you take into account the gender and ethnicity of your audience. In one-to-one conversations, how far apart you stand and whether and how much you make eye contact influences how comfortable others are, and this depends on people's customs.

Medium and Mode

The medium through which you communicate, for example text on paper, telephone or computer mediated communication, often dictates the mode of communication, that is whether it is spoken, written, or graphic. However, you have some choices. If you are preparing a brochure to advertise a product range, you will use both written text and graphic images to fully communicate your messages to your audience.

How will you communicate in a given situation? Sometimes you have little choice. If you are replying to a letter, you will almost certainly have to write. If you receive a telephone call, you will talk to the caller, but you will not be able to 'read' their body language to get feedback on the progress of the communication.

Different modes of communication lead to different expectations. Written expression tends to be more formal and there is an expectation that you will be concise and very correct in your use of words and sentence construction. Electronic communication, say email, although written, is much less formal and often even more concise. Spoken communication is less formal than written, and often less precise. We tend to use many more words than we do in writing and listeners are more accepting of unfinished sentences ambiguities than they would be if the same ideas were expressed in a letter.

Different communication media and different modes of communication serve a range of purposes and create quite different expectations about the formality of the communication. Your tasks is to be flexible in that you are able to use a wide range of media and modes. You must also be adaptable and able to change the style of your communication to suit the range of purposes, your audience and their needs.

Non-verbal Communication

The term 'non-verbal communication' is usually taken to mean body language. This covers the way you stand or sit when talking to others, the way you use hand gestures to convey meaning, and the way you use facial expressions and nods or shakes of the head to signal understanding or agreement.

Gestures and other forms of non-verbal communication behaviours are not shared across all cultures. A nod signifies agreement in most European cultures, but in North Africa, it may mean the opposite. In western cultures, making eye contact during conversation is seen as necessary. If you avoid it, you may create the impression that you are not trustworthy. In some other cultures, maintaining eye contact may lead to the perception that you are too pushy.

Clearly, non-verbal communication is important in face-to-face conversations between individuals and in groups. The absence of this channel of communication over the telephone or in email is quite important. A word, said with a wink or a smile in conversation, could foster understanding. The same word, in an email message without the smile, could cause offence.

Formalities

Every form of communication, written, spoken or graphic, has its own sets of rules. It is important to be aware of the rules and customs of the medium, the mode, and the purpose of communication. There are times when these conventions can be relaxed, but other times when they must be followed. Being an effective communicator is about knowing when to follow and when, and by how much, you can relax these communication conventions.

Team Work and Interpersonal Skills

Team work is recognized as one of the key factors that contribute to success in enterprises and it is listed as a requirement in many job advertisements.

Team work is sometimes contrasted with individualism. They are not opposites. You are not either an individualist or a team player: you need to be both. In different situations, you may need to emphasise one or other of these skill sets. The most capable employees are those people who are able to adapt easily and apply the skill set most relevant to situations as they arise.

Your Team Work Skills

Team work is made up of many component skills. Team work skills are listed in the table over the page. Have a look at the table now. You may find this list of group skills helpful in assembling evidence to support a claim that you have well developed team work skills.

In the table, team work skills are listed under five categories: planning, contributing, seeking input, monitoring and reflection, and social interaction. For each category, several specific skills are listed and an example of each skill is taken from a team who were working online from different parts of the country.

Planning

Effective individuals plan what they are going to do, and effective teams do the same thing. To contribute to making a group effective, you need to show that you can help with planning group activities.

Contributing

Groups are only effective when everyone in the tem contributes. You can contribute by doing tasks that are assigned to you, by helping others, by giving helpful feedback to others, by sharing what you know or have found out with other team members, and by challenging others to explain their ideas.

Seeking input

Groups also work well when people know that they are needed and that their contributions are valued. This aspect of group work occurs when you need help or feedback and you seek input from your team-mates. It also occurs when you encourage others by motivating them to make an extra effort or to keep going when the task becomes difficult.

Monitoring and reflecting

Another category of group skills includes monitoring group performance and thinking about ways of making it even more efficient. Sometime this involves looking back to see what worked, what did not, and what could be improved. It also involves looking forwards. If a similar situation arose in future, how could you use what you have just done to make sure that future team efforts are effective.

Social interaction

A final set of skills is called social interaction. When you are working in a team, you need to remain focused on the task, but sometimes tense situations arise and it may be necessary to calm the situation. Conflicts do arise in work situations and it can be easy to let a complex work problem interfere with the need to get on well together for the long term cohesion of the group. A good way to keep the group happy is to share a joke or to spend a little time talking about non-work issues. Too much time spent on a social focus will detract from group productivity, but a lack of effort to ensure that the group gets on well will also jeopardise group output. Getting this balance right can be difficult.

Employability Skills Assessment Project

Team Work and Group Skills

Behaviour categories	Description	Example
Planning	Group skills: a generic code applied to expressions that encourage group activity and cohesiveness.	I know that Tan and Beth have given you good advice, but I think it's also worth knowing that you need patience.
	Organising work: Planning group work; setting shared tasks and deadlines.	Can we do this by Friday? I just want to set a time-line for myself. Is everyone OK with that?
	Initiating activities: Setting up activities such as chat sessions to discuss the progress and organisation of group work.	I would like to have a chat about that. What about this Friday at 7.30pm SA time?
Contributing	Help giving: Responding to questions and requests from others.	To access the chat room, click on virtual chat in the blackboard; chat screen will come up; click on enter...
	Feedback giving: Providing feedback on proposals from others.	I like your idea of a generic booklet and everyone contributing aspects of interesting internet services...
	Exchanging resources and information to assist other group members.	"With the implementation of an internet service … there has been a major shift in the communication function in business."
	Sharing knowledge: Sharing existing knowledge and information with others.	I think we also need to give thought to the following. 1. The issues of quality/efficiency in teaching and learning...
	Challenging others: Challenging the contributions of other members and seeking to engage in debate.	John, I liked your idea for the first part of the presentation, but I think that we need to be more direct after that.
	Explaining or elaborating: Supporting one's own position (possibly following a challenge).	OK, that's a fair point. But I think that we are n danger of being seen as 'too pushy'. The idea is strong enough to speak for itself.
Seeking Input	Help seeking: Seeking assistance from others.	Does anyone know how to edit/add/append data on the student pages?
	Feedback seeking: Seeking feedback to a position advanced.	What do you think about answering the questions that Beth has put forward?
	Advocating effort: Urging others to contribute to the group effort.	John, haven't heard from you for a while. Are you still with us?
Monitoring and Reflection	Monitoring group effort: Comments about the group's processes and achievements.	I believe the overall contribution and collaboration of working as a group requires an increase within itself as part of our learning.
	Reflecting on medium: Comments about the effectiveness of the medium in supporting group activities.	The email for the discussion group seems to work OK for me. You know it has gone through because you actually receive your email back almost straight away if it has worked.
Social Interaction	Social interaction: Conversation about social matters that are unrelated to the group task. This activity helps to 'break the ice'. (It is important in ensuring that people 'get on' but it must not become so common that it wastes time.)	Regarding the planned meeting - my weekend is pretty hectic - I have my family flying in from Greece … so the Greek festivities will be in full swing.

Think of a group activity in which you have been involved. Compare what you did with the skills listed in the table. Can you point to examples of things that you did that show your team work skills?

Page 4

How do I Get an Assessment?

Being assessed for problem-solving does not involve doing any new assignments. All you need to do is review what you did in an assignment that you have just completed for your course. We will explain the process for doing this now.

Choose an Assignment

Your lecturer has already reviewed the assignments that are required in the units that you are doing and they have recommended some that give you your best opportunity of developing and demonstrating your problem-solving ability.

A list of the assignments that are good ones for showing your problem-solving ability is available from your lecturer. They have probably already told you which are the best ones.

Decide which one you will use before you begin the assignment. This will make it easier for you to collect the evidence that you need to show your skill.

Do You Know What Problem-Solving Is?

The first time you do a problem-solving assessment, follow the diagram on the final page of this leaflet. It shows the five main steps in problem-solving. Read the descriptions of each step.

Now look at the Problem-Solving Skills Assessment Tool (PSSAT). Notice that for each major step, such as defining the problem, there are several indicators. These are the processes that good problem-solvers carry out. You have probably done some of them.

Look again at 'The five steps' diagram. Notice that there are paths marked by arrows connecting the five main steps. This is because problem-solving is not just about doing five set steps in order. It is about looking back and checking what you have done. See for example, the pathway with the question 'Is the plan leading towards a solution?' While you are solving the problem, you need to think about whether you are getting closer to a solution. If you are, that's great. If not, you may need to review the plan of action.

Assess Yourself

That's right. We want you to assess your own ability. In fact, it's quite important that you do. Assessing yourself is an effective way of improving your performance.

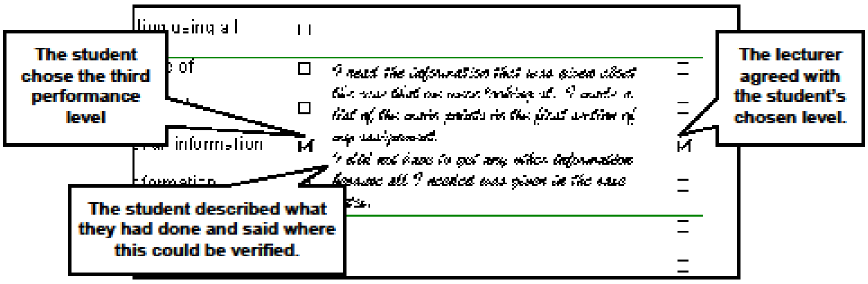
To assess your own problem-solving ability, have a copy of your recently completed assignment with you, a copy of the Problem-Solving Skills Assessment Tool (PSSAT), and a copy of the sheet *Prompts for indicators*.

These prompts are a series of questions that should direct your attention to what you might have done for each indicator on the PSSAT. These prompts correspond to the performance levels for each indicator on the PSSAT.

If you can answer 'yes' to these questions, then you can tick the corresponding level on the PSSAT in the SA (self-assessment) column. When you do answer 'yes' to a question, then you should make a note, in the space provided on the PSSAT, of the evidence that you can use to show that you did achieve that level on that indicator.

Below is an example of a completed section of the PSSAT. The student who completed this chose the third performance level, said what they had done, and told their lecturer where to look to verify this evidence. In this case, their lecturer agreed with the performance level chosen.

Employability Skills Assessment Project



This evidence is important, because that is what your lecturer will look at when they review your self-assessment. For more information, see *Gathering evidence* (below).

It is more useful for you to make fair and accurate judgments of your performance than that you try to justify a high level indicator. Some problem situations, those that are easier to solve, just do not give you the opportunity to perform at the higher levels. This is not a reflection on you, it is just a feature of the particular problem that you solved.

Give your completed PSSAT form to your lecturer with your completed assignment.

Lecturer Validation

The next stage in the process is the validation of your evidence by your lecturer. This step is an important one because the credibility of the assessment depends on an assessment by a lecturer.

Your lecturer will look at the evidence that you have referred to on your completed PSSAT. They will look at your assignment to verify that the evidence is supported.

If they agree with your self-assessment, they will tick the same box, in the LV (lecturer validation) column, as you have done. However, they may disagree with your assessment. They may think that you have performed at a lower or higher level, and they will tick the box that they think is consistent with the evidence that you have presented.

Feedback

This assessment activity is not just about getting a score: it is about improving your performance. If your lecturer has disagreed with your self-assessment, they will tell you why. It could be that you have misunderstood the criteria in the PSSAT or that you have not presented adequate evidence.

Do not be concerned if the lecturer's assessment is different from your own. Instead, learn from the feedback so that next time you do an assessment, you will understand the criteria better or you will be able to present more compelling evidence.

Gathering evidence

For each of the five steps in problem-solving, there are several 'indicators'. Read each indicator and think about the processes that you have followed in solving a problem and identify evidence of those processes. Each indicator has from two to four levels associated with it. Your task is to find a level that most closely matches what you did and then to refer to evidence that shows what you did as you solved the problem. The questions presented on page 3 will help you to find evidence and select the right level for it.

Page 2

Employability Skills Assessment Project

Prompts for Indicators

Defining the Problem

Forms a correct understanding of the problem. Did you develop a complete understanding of the problem? Did you form a simple understanding based on one aspect of it or a more complex one based on several aspects?

Recognises significance of given information. Did you use all the relevant information that was given? Did you determine whether the information that was given and what you already knew were not enough to enable you to solve the problem? Did you check procedure manuals or other reference sources to identify specific additional information that was needed?

Recalls relevant information. Did you realise that you already knew things that are important in solving the problem? Did you recall that information accurately? Did you recall the information as separate pieces of knowledge or did you show how it all related together in relation to the problem task?

Sets a realistic goal. Did you identify a goal that was given or did you establish your own clear and achievable goal? Was the goal stated clearly enough so that you would know that you had reached an appropriate solution?

Planning an Approach

Plans an approach to the problem. Did you show any planning? This may have been short term planning, ie just enough to get started but not enough to get to a final solution. Or did you plan many steps, enough to take you from the beginning right to the final problem solution?

Recalls previous relevant or similar problem tasks. Good problem-solvers think back over their past experiences to find a problem, similar to the current one, that they have encountered previously. Did you recall similar past problems to guide your solution to the problem you have just solved? In what ways were past problems both similar to and different from the current one? Did you use this experience to develop a solution strategy?

Identifies appropriate sub-goals. Many problems involve sub-problems or phases in their solution attempt. Did you set a final goal and then establish milestones or sub-goals along the way?

Sets an appropriate time frame. Did you estimate how long the problem should take to solve? Was your estimate reasonable? Did you in include this along with other commitments you had?

Carrying out a Plan

Begins to follow the set plan. Did you begin to work systematically? If you had a clear plan, Did you show that you followed it?

Activates relevant knowledge. Did you need to recall information? Did you apply that knowledge correctly in the current situation?

Application of strategies. Did you apply a given procedure? Did you select and implement an appropriate strategy from several available? Did you adapt an existing, or create a new, problem-solving process?

Monitoring Progress

Checks progress towards goal. Did you pause to check your progress against sub-goals or the final goal to see that you were on target to reach a solution?

Response to unexpected problems along the way. If an unexpected difficulty arose, and you were not able to proceed as planned, did you search for the cause? For example, did you recheck information that was given or assumed? Did you seek additional information to verify your understanding of the problem? If the solution went according to plan, especially if you followed a set procedure, you may not have any evidence for this indicator.

Reviews original plan. If the original plan did not appear to work, did you review it and adopt a new approach or set new goals? If the solution did not go according to your plan, did you revise the plan to seek improvements?

Checks original understanding and definition of problem. If you found that the solution attempt was not working, did you go back to the definition of the problem and pause to think again about what the problem is really about?

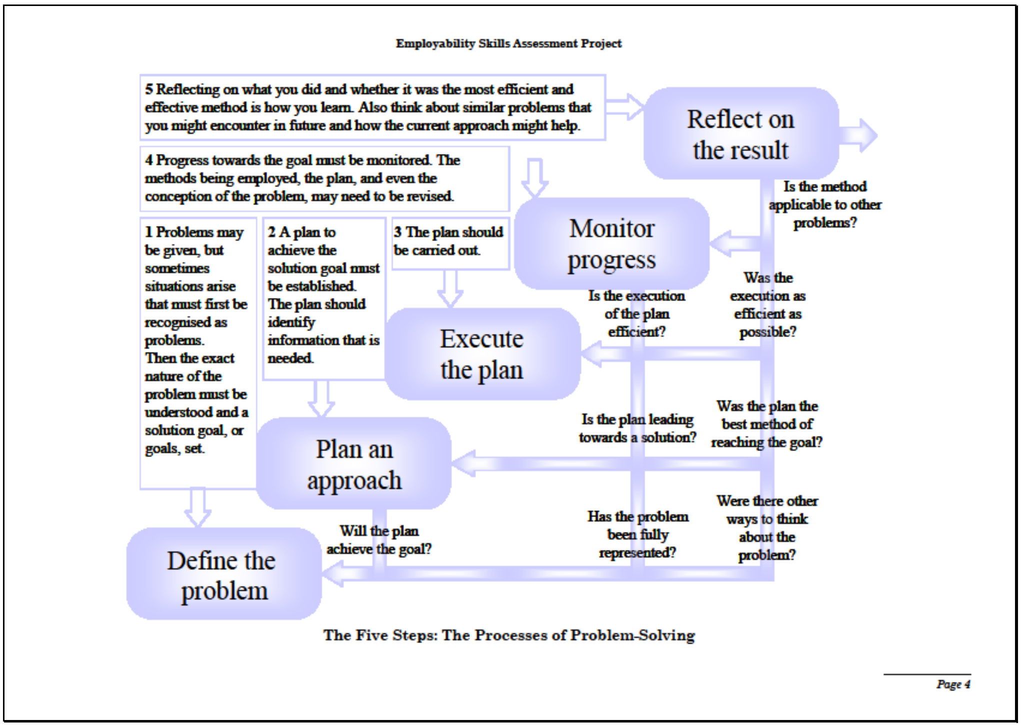
Reflecting on the Result

Reviews efficiency and effectiveness of problem approach. Having solved the problem, have you thought about how you might have been more efficient? Would you make any major changes to the way they went about this activity if you had to do it again?

Compares current problem with previously encountered ones. Have you compared the just-completed problem with others that you had attempted earlier? Can you identify techniques in this solution that would have helped you with earlier problems? Are there techniques that you have used previously that would have helped you with the current problem?

Anticipates situations in which current approach might be useful. What have you learned from the present problem? Have you thought about ways in which you could use what you have just learned in future problems? Can you identify a specific future task where the current approach is likely to be useful?

Page 3



The Five Steps: The Processes of Problem-Solving

Problem Solving Skills Assessment Tool

Name ____________________ **ID** __________

Unit ____________________ **Task** __________

Lecturer ____________________ **Date** __________

Self Assessment **SA** Lecturer Validation **LV**

Defining the Problem		**SA**	**Evidence of Definition**	**LV**
Forms a correct understanding of the problem	Misunderstands the problem	☐		☐
	Forms a partial understanding using some given information	☐		☐
	Forms a complete understanding using all relevant factors	☐		☐
Recognises significance of given information or the need for new information	Fails to recognise the significance of information that is given	☐		☐
	Recognises the significance of most information that is given	☐		☐
	Recognises the significance of all information that is given	☐		☐
	Identifies specific additional information required	☐		☐
Recalls relevant information	No recall, recalls irrelevant information, inaccurate recall of information	☐		☐
	Accurately recalls relevant information as isolated elements	☐		☐
	Accurately recalls relevant information as integrated elements	☐		☐
Sets a realistic goal	Does not identify a clear goal	☐		☐
	Identifies relevance of given goals	☐		☐
	Establishes a clear goal	☐		☐

Planning an Approach		**SA**	**Evidence of Planning**	**LV**
Plans an approach to the problem	Does not engage in planning activity	☐		☐
	Undertakes enough planning to begin the solution	☐		☐
	Plans all stages required to achieve the goal	☐		☐
	Considers several alternative approaches and selects the best plan to achieve the goal	☐		☐
Recalls previous relevant or similar problem tasks	Does not refer to previous problems	☐		☐
	Recalls previous problems that may be relevant	☐		☐
Identifies appropriate sub-goals	Does not identify sub-goals	☐		☐
	Breaks the task into smaller sub-goals	☐		☐
Sets an appropriate time frame	Does not consider time frame	☐		☐
	Estimates how long the problem solution should take	☐		☐
	Integrates other commitments into time frame established for solving this problem	☐		☐

Please turn over…

Carrying out the Plan		SA	Evidence of Planned Execution	LV
Begins to follow the set plan	Begins to work on the problem without a clear system	☐		☐
	Works systematically, but without reference to plan	☐		☐
	Works systematically and follows set plan	☐		☐
Activates relevant knowledge	Does not activate relevant knowledge or activates incorrect information	☐		☐
	Activates some, but not all, relevant knowledge	☐		☐
	Activates and uses accurate relevant knowledge	☐		☐
Application of strategies	Applies a given or an established procedure	☐		☐
	Selects and implements a procedure from suggested ones	☐		☐
	Adapts an existing, or creates a new, procedure	☐		☐

Monitoring Progress		SA	Evidence of Monitoring	LV
Checks progress towards goal	Does not check progress against sub-goals or final goal	☐		☐
	Periodically checks solution progress against sub-goals and final goal	☐		☐
Responds to unexpected problems along the way	Does not recognise problems that arise	☐		☐
	Identifies unexpected problems	☐		☐
	Diagnoses causes of unexpected problems	☐		☐
	Makes adjustments to rectify suspected causes of unexpected problems	☐		☐
Reviews original plan	Does not review original plan	☐		☐
	Reviews original plan	☐		☐
	Reviews and revises original plan or verifies that original plan is appropriate	☐		☐
Checks original understanding and definition of problem	Does not check original understanding and definition of problem	☐		☐
	Reviews understanding and definition of problem	☐		☐
	Reviews and revises understanding of problem or verifies original definition	☐		☐

Reflecting on the Result		SA	Evidence of Reflection	LV
Reviews efficiency and effectiveness of problem approach	Does not reflect on solution efficiency and effectiveness	☐		☐
	Reviews efficiency and effectiveness of solution strategy	☐		☐
	Identifies improvements to solution strategy	☐		☐
Compares current problem with previously encountered ones	No comparisons with previous related tasks	☐		☐
	Compares current task with previous ones	☐		☐
	Notes ways in which current experience might have helped past problems	☐		☐
Anticipates situations in which current approach might be useful	No anticipation of future applications of solution strategy	☐		☐
	Considers future broader applications of solution strategy	☐		☐

PSSAT R3 10/03

Appendix 11: Results of Analyses of Multiple Imputation Data Files for Self-Assessed Problem Solving

Model 1: One-level mean and variance

		Fixed parameters	
Imputation	Deviance	Intercept	se
1	1037.03	55.92	1.32
2	1054.64	54.96	1.42
3	1024.56	56.59	1.26
4	1028.15	57.01	1.28
5	1032.78	56.29	1.30
Mean	1035.43	56.15	1.31

		Variance components	
Imputation		e_{ij}	se
1		219.76	27.69
2		252.72	31.84
3		199.05	25.08
4		204.80	25.80
5		212.47	26.77
Mean		217.76	27.44

Model 2: Two-level variance components model

		Fixed parameters	
Imputation	Deviance	Intercept	se
1	1013.40	55.92	1.82
2	1046.34	54.96	1.76
3	1008.83	56.59	1.66
4	1001.74	57.01	1.78
5	1008.76	56.29	1.80
Mean	1015.81	56.15	1.76

		Variance components			
Imputation		e_{ij}	se	μ_{0j}	se
1		120.12	18.54	99.63	31.10
2		184.47	28.47	68.25	29.86
3		124.97	19.28	74.08	26.06
4		106.94	16.50	97.86	29.65
5		115.38	17.80	97.08	30.17
Mean		130.38	20.12	87.38	29.37

Model 3: Two-level random intercepts, common slope

		Fixed parameters			
Imputation	Deviance	Intercept	se	Slope	se
1	992.08	50.73	2.11	5.19	1.05
2	1020.97	48.03	2.17	6.93	1.27
3	993.90	52.08	2.00	4.51	1.12
4	984.45	52.55	2.05	4.46	1.02
5	995.09	52.13	2.10	4.16	1.08
Mean	997.30	51.10	2.09	5.05	1.11

		Variance components			
Imputation		e_{ij}	se	μ_{0j}	se
1		93.20	14.38	108.61	30.85
2		136.39	21.05	84.27	29.17
3		104.63	16.15	80.86	25.82
4		87.05	13.43	104.49	29.48
5		98.05	15.13	102.86	30.01
Mean		103.86	16.03	96.22	29.06

Model 3A: Two-level random intercepts with explanatory variable and common slope

		Fixed parameters					
Imputation	Deviance	Intercept	se	Slope	se	IP	se
1	979.61	35.40	4.45	5.19	1.05	0.29	0.08
2	1010.81	33.82	4.65	6.93	1.27	0.29	0.09
3	983.90	38.67	4.40	4.51	1.12	0.27	0.08
4	974.56	37.61	4.85	4.46	1.02	0.30	0.09
5	986.04	38.55	4.70	4.16	1.08	0.27	0.08
Mean	986.98	36.81	4.61	5.05	1.11	0.28	0.08

	Variance components			
Imputation	e_{ij}	se	μ_{0j}	se
1	93.20	14.38	72.72	23.15
2	136.39	21.05	56.39	23.31
3	104.63	16.15	56.33	20.62
4	87.05	13.43	76.47	23.45
5	98.05	15.13	76.59	24.37
Mean	103.86	16.03	67.70	22.98

Model 4: Two-level random intercepts and slopes

		Fixed parameters			
Imputation	Deviance	Intercept	se	Slope	se
1	991.93	50.73	2.13	4.19	1.08
2	1020.93	48.03	2.19	6.93	1.26
3	992.57	52.08	1.86	4.51	1.04
4	983.75	52.55	1.94	4.46	0.98
5	992.51	52.13	1.95	4.16	0.95
Mean	996.34	51.10	2.01	4.85	1.06

	Variance components							
Imputation	e_{ij}	se	μ_{0j} j	se	μ_{1j}	se	Cov	se
1	87.74	19.15	116.91	44.43	5.46	14.41	-5.97	19.34
2	138.90	30.31	85.60	50.68	-2.50	21.05	0.17	25.79
3	119.15	26.00	45.57	38.32	-14.52	16.30	22.49	18.90
4	93.96	20.50	80.38	38.61	-6.91	13.47	14.36	16.79
5	120.21	26.23	59.73	41.18	-22.16	15.51	28.95	18.43
Mean	111.99	24.44	77.64	42.65	-8.13	16.15	12.00	19.85

Model 4A: Two-level random intercepts with explanatory variables and random slopes

		Fixed parameters					
Imputation	Deviance	Intercept	se	Slope	se	IP	se
1	978.71	34.66	4.47	5.19	1.08	0.31	0.08
2	1010.10	33.16	4.67	6.93	1.26	0.30	0.08
3	983.08	38.73	4.37	4.51	1.04	0.27	0.08
4	974.26	37.40	4.85	4.46	0.98	0.30	0.09
5	983.83	38.43	4.67	4.16	0.95	0.27	0.08
Mean	986.00	36.48	4.61	5.05	1.06	0.29	0.08

	Variance components							
Imputation	e_{ij}	se	μ_{0j}	se	μ_{1j}	se	Cov	se
1	87.74	19.15	99.60	40.92	5.46	14.41	-15.22	19.43
2	138.90	30.31	74.72	48.63	-2.50	21.05	-8.30	26.02
3	119.15	26.00	35.55	36.55	-14.52	16.30	15.23	18.99
4	93.96	20.50	70.80	36.75	-6.91	13.47	5.14	17.00
5	120.21	26.23	49.17	39.22	-22.16	15.51	21.10	18.52
Mean	111.99	24.44	65.97	40.41	-8.13	16.15	3.59	19.99

Model 4B: Two-level random intercepts and random slopes with explanatory variables

		Fixed parameters							
Imputation	Deviance	Intercept	se	Slope	se	IP-int	se	IP-slope	se
1	976.82	39.17	5.52	1.42	2.91	0.22	0.10	0.07	0.05
2	1008.60	37.94	6.06	2.81	3.56	0.20	0.12	0.08	0.07
3	981.01	42.65	5.13	0.53	2.92	0.19	0.10	0.08	0.05
4	971.05	42.46	5.58	-0.38	2.81	0.20	0.10	0.10	0.05
5	981.19	42.57	5.30	0.13	2.61	0.19	0.10	0.08	0.05
Mean	983.73	40.96	5.52	0.90	2.96	0.20	0.10	0.08	0.05

		Variance components							
Imputation		e_{ij}	se	μ_{0j} j	se	μ_{1j}	se	Cov	se
1		87.74	19.15	96.50	40.31	3.29	14.06	-12.62	18.92
2		138.90	30.31	71.55	48.05	-4.85	20.70	-5.57	25.53
3		119.15	26.00	33.45	36.17	-16.69	16.02	17.37	18.63
4		93.96	20.50	67.59	36.13	-9.86	13.07	8.22	16.43
5		120.21	26.23	46.73	38.80	-24.47	15.25	23.47	18.13
Mean		111.99	24.44	63.16	39.89	-10.51	15.82	6.17	19.53

Appendix 12: Results of Analyses of Multiple Imputation Data Files for Teacher Assessed Problem Solving

Model 1: Single level

		Fixed parameters	
Imputation	Deviance	Intercept	se
1	1129.33	33.46	1.91
2	1118.25	34.77	1.82
3	1121.25	34.94	1.85
4	1126.84	33.90	1.89
5	1136.46	33.11	1.96
Mean	1126.43	34.03	1.88

	Variance components	
Imputation	e_{ij}	se
1	457.17	57.60
2	418.70	52.75
3	428.77	54.02
4	448.21	56.47
5	483.79	60.95
Mean	447.33	56.36

Model 2: Two-level null model

		Fixed parameters	
Imputation	Deviance	Intercept	se
1	1129.33	33.46	1.91
2	1118.25	34.77	1.82
3	1121.25	34.94	1.85
4	1126.85	33.90	1.89
5	1136.29	33.11	2.03
Mean	1126.39	34.03	1.90

	Variance components			
Imputation	eij	se	μ_{0j}	se
1	455.37	70.27	1.80	40.89
2	418.70	52.75	0.00	0.00
3	428.77	54.02	0.00	0.00
4	448.21	56.47	0.00	0.00
5	465.72	71.86	18.07	44.77
Mean	443.35	61.07	3.97	17.13
Mean excluding negative variances	460.54	71.06	9.94	42.83