Examining resuscitation skill education as a component of paramedics' practice development

by

Amy Seymour-Walsh

Thesis

Submitted to Flinders University

for the degree of

Doctor of Philosophy

College of Medicine and Public Health

November 2017

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

PRESENTATIONS ARISING FROM THIS STUDY

2015

Do two teaching methods result in different skill acquisition? Asia Pacific Medical Education Conference (APMEC) Singapore. Feb 2015.

Doing the Military Two-step or Militantly Forcing Four-step? Australia and New Zealand Association for Health Professional Educators (ANZAHPE) conference. Newcastle, NSW. Mar 2015.

Measuring the cost-effectiveness of clinical skill teaching. Rogano Conference. Glasgow, UK. Sep 2015.

How do two different skill teaching methods compare in terms of cost effectiveness? An International Medical Education Conference (AMEE) Glasgow, UK. Aug 2015.

Young investigator award. European Resuscitation Congress (ERC) Prague, Czech Republic. Nov 2015.

Does patient morbidity and mortality improve with a skill teaching strategy embedded in learning theory? European Resuscitation Congress (ERC) Prague, Czech Republic. Oct 2015.

2016

Developing a validated marking tool from expert clinician consensus. Ottawa conference. Perth, Australia. Mar 2016

3-Minute Thesis. Australian Resuscitation Outcomes Consortium (AusROC) symposium. Apr 2016.

Is our assessment of resuscitation skills robust? Examining the validation process. European Resuscitation Congress. Reykjavik, Iceland. Sep 2016.

Is the common approach to teaching ALS skills cost-effective? European Resuscitation Congress. Reykjavik, Iceland. Sep 2016.

The art of education: Why do we teach skills the way we do? European Resuscitation Congress. Reykjavik, Iceland. Sep 2016.

ACRONYMS, ABBREVIATIONS AND SYMBOLS

Acronyms

This thesis will employ a range of acronyms. Below is a table of many acronyms used, however this list is not exhaustive. Where an acronym has been used locally in a section, it may not appear in this list:

2SA	Two Stage Approach
4SA	Four Stage Approach
$4SA_m$	Four Stage Approach (modified)
AED	Automated External Defibrillator
AHA	American Heart Association
ALS	Advanced Life Support
ANOVA	Analysis of Variance
ANZCOR	Australia And New Zealand Committee on Resuscitation
ATAR	Australian Tertiary Admissions Rank
AusROC	Australian Resuscitation Outcomes Consortium
BVM	Bag Valve Mask
CASP	Critical Appraisal Skills Program
CG	Control Group
CI	Confidence Interval (95% confidence intervals used throughout this study)
CPR	Cardio Pulmonary Resuscitation
ECG	Electrocardiograph
EEG	Electroencephalograph
ERC	European Resuscitation Congress
ERC	European Resuscitation Council
ETT	Endotracheal Tube
FU	Flinders University of South Australia
GERM	Global Education Reform Movement
GRS	Global Rating Scale
GT	Grounded Theory
ICC	Intraclass Consistency Coefficient
ICP	Intensive Care Paramedic
IG	Intervention Group
IHD	International Health Development
ILCOR	International Liaison Committee on Resuscitation
IO	Intraosseous
IPPI	Integrated Procedural Performance Instrument
IPPV	Intermittent Positive Pressure Ventilation
IQR	Interquartile Range
IRCC	Inter-Rater Consistency Coefficient
IVA	Intravenous Access
IVC	Intravenous Cannula
K-S	Kolmogorov-Smirnov test
LMA	Laryngeal Mask Airway
M	Mean (statistical)

NESB	Non English Speaking Background
NGT	Nasogastric Tube
NPA	Nasopharyngeal Airway
OPA	Oropharyngeal Airway
OSP	Overall Skill Performance
PALS	Paediatric Advanced Life Support
PBL	Problem Based Learning
PPE	Personal Protective Equipment
RCT	Randomised Controlled Trial
ROSC	Return of Spontaneous Circulation
SAAS	South Australian Ambulance Service
SD	Standard Deviation
SPSS	Statistical Package for the Social Sciences
S-W	Shapiro-Wilk test
TAFE	Technical and Further Education
VF	Ventricular Fibrillation
VT	Ventricular Tachycardia

Symbols

±	Refers to a single standard deviation of data points around a mean score
α	Cronbach's alpha
κ	Kappa (a consistency coefficient)

Variable measures

Chlist_total	The performance score according to skill specific checklists developed in Chapter 4, excluding items prompted by some facilitators
Chlist_taught	As per "Chlist_total", including only the items taught in each teaching session
GRS_25	Global Rating Scale (five items amounting to a total mark of up to 25)
GRS_30	Global Rating Scale (six items amounting to a total mark of up to 30)
TTT1_4exTD	The time required to teach Stages 1 to 4 (of 4SA or all of 2SA), excluding the delay time or tangents discussed in the teaching session

Statistics

All data analysis and exploration were performed with IBM SPSS Statistics 23. In all statistical tests, a significance value of $p \leq .05$ was accepted as statistically significant.

Visual consistency

A consistent colour palette has been used throughout this theses to assist in easy identification of visual data. These colours are used as a tribute to Flinders University, from whose brand they are drawn:

2SA teaching (Chapters 3 and 5)

4SA$_m$ teaching (Chapter 3)

4SA teaching (Chapter 5)

Groups taught LMA with 4SA and IO with 2SA

Groups taught IO with 4SA and LMA with 2SA

Other undifferentiated data (for example demographic information)

Other explanatory notes

Referencing throughout is presented in the format advocated by the American Psychological Association (APA), 6th edition (American Psychological Association, 2010). As such, page numbers will be used for quotes and very specific references such as figures and tables, even when transcribed for this thesis. The key exception is the use of Australian English throughout (for example, "generalisation") except where a direct quote uses other forms (such as "generalization"), rather than American English.

Some terminology is used such as *craftsman* and *andragogy* which may connote a reference to the male gender for individual readers. For such terms, gender neutrality is implied in this thesis.

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.

Signed...

Date..

ACKNOWLEDGEMENTS

Amidst the potential isolation which a PhD can impose, I have been surrounded by a village of family, friends and colleagues, without whom this journey would have been significantly harder.

First and foremost, to my Mum and Dad: you have supported me through my victories, failures, and frustrations. You are my safe place. I am so richly blessed to be able to lean on you like I have. Another 100,000 words would be insufficient to describe how precious your support is. I thank you.

My supervisors have each helped me learn unique lessons during my candidature. When I first approached Professor Tara Brabazon for guidance, it was with professional respect, but there was no way I could anticipate the depths of personal admiration that I would soon have for her. Tara, I have treasured your friendship, encouragement, intellectual generosity, insight and expertise. You don't compromise kindness in your strength, or vice-versa, and you have shown care for my intellectual, emotional, professional and personal wellbeing. You are incredible.

Associate Professor Anna Vnuk, you have done more for me than you realise. You are a profoundly intelligent academic, and working with you has allowed me to see the beauty, discipline, value and strength of qualitative research. This has been a huge gift for me.

Professor Paul Worley, you inspired me to learn, grow and risk. Even when I misunderstood a key idea, you *never* made me feel insufficient. You *never* made me feel like the goal was unattainable, and when I struggled, you inspired me to keep searching and work it through. The implication for that was a resilience and a resolve which cannot be taught. Your encouragement was potent in my transition into (and through) this PhD.

I have also had some tremendous support from many other colleagues. Professor Lambert Schuwirth, Dr Julie Ash, Dr Svetlana King and Dr Koshila Kumar: thank-you. You have been patient, kind, and so generous. Your work in health professions education research has urged me forward more often than you know. My dear colleague Minh Nguyen has dedicated expertise and advice as we worked through statistical analysis and interpretation together. Minh, your kindness and encouragement as has been incredibly valuable.

Melanie Thorrowgood, Trevor Matthews and Kathryn Mason, I could always bring my raw human-ness to you, and you would *always* build me up. Seeing myself as you see me has spurred me onwards. And to my farmer, you are one of the most intelligent people I know. The care you have shown is a treasure to me, and you have challenged me to keep forging ahead. I am so thankful to and for you all.

This PhD was completed in three years and two months, in part due to the financial support awarded by the Australian Resuscitation Outcomes Consortium (AusROC). This generous scholarship allowed me to protect time in my weekly schedule to dedicate to research and writing.

Finally, I know that I have been upheld, in prayer to God, by my Christian brothers and sisters. You have helped me maintain perspective, and shown great care for me through the challenges and triumphs of this research. He has sustained me as His child, in part, through each of you.

My village is a great blessing to me, and includes many not mentioned here by name. I do hope you know who you are, and hope that I am able to express my thanks to you personally.

ABSTRACT

Clinical skills are a fundamental component to health professionals' practice. Training programs maintain a responsibility to health service providers and patients treated by course graduates who are tasked to care for the community. However educational institutions also experience resourcing pressure to ensure this outcome is achieved as cost-effectively as possible.

Recognising these parameters and limitations, my original contribution to knowledge is a multifaceted examination of the cost-effectiveness of an internationally advocated clinical skill instruction method. The findings in this thesis are of relevance to medical education institutions and organisations, health care providers, and clinical educators. To enact these findings, this thesis considers a commonly used four stage skill teaching method, and aims to understand its cost-effectiveness in comparison to a more traditional and arguably more natural two stage approach. Paul Worley's model for symbiotic clinical education (Prideaux, Worley, & Bligh, 2007; Worley, 2002a, 2002b; Worley, Prideaux, Strasser, Magarey, & March, 2006) is used as a theoretical framework upon which to structure definitions and order outcomes and implications for cost and effectiveness, in relation to three of the four key relationships: the clinical, institutional and community axes. A pragmatic approach is adopted throughout the course of study to ensure the research philosophy and design are informed and shaped appropriately by the research question. As such, this mixed methods program includes a post-positivist approach, a social constructionist approach, and others in between.

An initial trial was conducted to compare the two skill teaching methods, and showed no statistically significant difference in paramedic students' ability to acquire manual defibrillation skills with repeated measures analysis. Small sample size is a noted limitation for this study. This thesis then presents the development of a clinically relevant assessment tool for both intraosseous and laryngeal mask airway insertion. A modified Delphi approach has been used to understand expert pre-hospital clinicians' approach when performing these two skills, in order to construct these educational tools, as no appropriately validated specific skill performance checklists was found in the literature. These tools have then been applied to another comparative trial which compares both the acquisition and the retention (6 months later) of paramedic and nursing students in the application of these skills, with a comparison between teaching methods employed. This trial indicates that there is no difference in either acquisition or retention. The clinical skill assessment tools developed will then be critiqued, and a validation argument presented for their appropriate and accurate application in future education and assessment. Incidental data arising from the two

trials have indicated that the four stage approach may not be as easy for educators to perform as a traditional two stage approach, and an additional original contribution to knowledge presented in this thesis involves the educator's perspective and perception of such a teaching method. A qualitative study flavoured by tenets of phenomenology will allow educators to understand, for the first time, how the clinical educator's craft and practice are influenced by the four stage approach, and vice-versa.

This doctoral research builds on the work of others who have sought to compare the potential benefit of the four stage teaching method compared to traditional methods (Archer, Van Hoving, & de Villiers, 2014; Bitsika et al., 2013; Greif, Egger, Basciani, Lockey, & Vogt, 2010; Herrmann-Werner et al., 2013; Jenko, Frangež, & Manohin, 2012; Krautter et al., 2011; Lee, Boyd, & Stuart, 2007; Lund et al., 2012; Orde, Celenza, & Pinder, 2010). I have done so using a modern clinical education framework to understand the costs and benefits from a clinical education system approach in order to address the wider impact of skills education. This approach includes relevant discussion of the patient outcomes, institutional demands, clinician workload, and the human resource aspect of clinical education staff.

PROLOGUE

This PhD started when I did not consider myself a researcher, but rather a curious clinician and educator. My initial question was small. I wondered whether a teaching method which seemed likely to be more effective due to its grounding in learning theory resulted in greater skill performance acquisition and retention than a simpler teaching method. This wonder informed the study described in Chapter 3, and reflects many assumptions about knowledge, learning, and assessment which I have since challenged. That study encountered obstacles and another grew from it, another then grew from that, which gave rise to more questions, more challenges, and further critique of my own philosophical assumptions. This sense that the project adapted, diverted, re-converged and snowballed unexpectedly gives the program of research described in these pages authenticity, as it evidences the impact the main question had on the series rather than limiting the question (s) by a rigid plan which I could have otherwise set in place at the beginning. My study "grew legs" and I chose to adjust my approach, theoretical paradigm and structure accordingly.

This thesis can be likened to a journey through a park. I enter the gate, and wander down the path. But a fork in the path presents a choice: turn left and walk along the stream, or veer right and enjoy the cool shade of giant oaks? A jogger blazes past me, in perfect rhythm. He continues right, knowing just where he intended to go before he even arrived. The path, for him, was anticipated. His route was planned and he knew that his footing was stable. However, I turn left, and see ducks on the lawn, nuzzling their bills to find juicy bugs beneath the grass. It makes me wonder where the ducklings live. Why are some ducks in the water, and some on land? Does the light shimmer off the water only at a certain part of the day? I could not see that particular light from the main gate. I go on wandering, and turn to the age-old oaks in profile. "I'd have missed this view if I took the path to the right".

So it is with research. Allowing myself to recognise and research the unexpected, and to wonder *if*, or *why*, or *whether*, allowed this project to evolve with authenticity, and under guidance. I ended up on a different path to the one I anticipated, and one which was built on new assumptions which challenged and changed my approach. I can still see the oaks, but I see them from the side, not from beneath. My view (my epistemological perspective) will therefore impact what I see. I won't see what the jogger sees, but I suppose we came to this park for different reasons.

In this doctoral thesis, the study series is not compelled or resolved by a single approach or perspective. Rather, the influence of a pragmatist approach (Creswell, 2013, p. 10) will infuse the studies, by encouraging the specific research question to shape the relevant methodology for that individual chapter. In selecting an appropriate philosophical, ontological and epistemological perspective, it is crucial that a researcher not only clearly knows the question, but also examines his or her own understanding of how knowledge is created, tested and described. Additionally, researchers must address their own assumptions in order to bring to light anything which might affect the data's integrity and reliability in reference to the research question. For example, in seeking to compare two chosen skill teaching methods, my personal epistemology (including assumptions about knowledge and truth) shaped both the question and a strategy to find an answer.

My professional background as an Ambulance Paramedic influenced my initial inclination to deploy quantitative methods. As a product of health science and clinical training, I entered this research with a set of assumptions about knowledge attainment and *truth* which aligned heavily to positivist thinking. *Evidence* was considered biased unless it was developed under the scrutiny of Randomised Controlled Trial (RCT), which has been upheld as the gold standard of knowledge attainment and hypothesis testing throughout this era of empirical science. My exposure to scientific disciplines including botany and ecology encouraged me to place an emphasis on what is observable as objectively evident and reliable. This is also true for the pre-hospital paramedic setting. However, much of our practice awaits this level of evidence and the limitations of this approach soon became apparent to me. A quantitative approach was well suited to some questions, but was ill-equipped to answer others which cannot be understood as less important. In part, this thesis documents my transition from a post-positivist, experimental researcher to one who has learnt to embrace the richness afforded through the malleability and rigour of qualitative traditions.

This personal evolution is reflected in the expression used in the various chapters. Much of the chapters are written in the first person, active voice. This captures the interconnectedness between research and the researcher. By presenting my journey alongside the research which has informed, guided and challenged it, I aim to render this series of studies more authentic, reflexive and contextual. After all, awareness of the "reciprocal influence of the researcher and that which is researched" will allow the emergent data to be understood in light of the perspective which has obtained and identified it (Lamb & Huttlinger, 1989). It will be evident, however, that in the chapters which relate to a more post-positivist paradigm, the voice is more passive, reflecting an intentional separation of the researcher from the researched. The voice, therefore, is quite consistent with the assumptions of the paradigm employed for each section.

I often describe myself as an "accidental PhD candidate". This thesis outlines the development of the research I performed, in addition to my personal growth from an *accidental* to a *professional* researcher.

1 INTRODUCTION

Clinical education is complex in task, setting and resource need. It occurs at the intersection of clinical practice and professional education, both of which are impacted by many unique demands (Egan & Jaye, 2009). These may take the form of an unpredictable environment, unusual patient presentations, complex clinical cases, or subtle learning needs which may be difficult for an educator to navigate. Whether the context is an artificial setting, high fidelity simulation, or the authentic clinical context available through student placements, clinical education must often compete with institutional and organisational demands. Whether a part of a formal training program, or opportunistic teaching in the clinical setting, training tomorrow's health professionals is intimately connected to the health service, education organisations, government policy and funding initiatives, patients, and of course clinicians.

The mastery of clinical skills is a key aspect of a health professional's development (Sawyer et al., 2015), and of the profession itself. Medical and health professional education bodies therefore commit much energy, time and resources to effectively training student clinicians to safely apply clinical skills within a framework of anatomical and physiological science knowledge, rationale, and practice protocols and guidelines.

The safe and effective application of clinical skills is known to directly impact patient care (Naik & Brien, 2013). Understanding of the financial, social and physical impact of clinical errors is sparse, likely due to incomplete reporting by staff, and lack of dissemination by health services. The true impact of poorly learnt, recalled or applied psychomotor skills is difficult to determine. While this may allow health care education providers some scope to deny the potential human and financial impact of skill atrophy, a report released in 2010 which estimated the financial cost of medical error in the United States of America alone at 17-29 billion dollars annually (US Dollars) makes that denial more difficult to sustain (Oyebode, 2013).

Skill expertise requires intentional practice and maintenance (Fitts, 1964, p. 268), with ongoing exposure to sufficient clinical opportunity to perform skills also likely to impact on skill development and maintenance (Smith & Greenwood, 2012). Thus, skill mastery becomes more than just "technically correct" performance, but performance guided by wisdom and understanding of the craft of health care, including when it is better *not* to perform the skills. However, for Australian ambulance paramedics, critical resuscitation skills may be used much less frequently than is optimal for ongoing skill maintenance (Dyson et al., 2015). Safe and timely application of these skills plays a significant role in the effective out-of-hospital resuscitation attempt. However, little is understood about how these infrequently used but critically important skills are retained and applied by ambulance paramedics following initial training, or in-service training when new skills and procedures are introduced.

Added to the clinical concerns of skill maintenance are the growing financial pressures on health systems and training institutions as they strive to address multifaceted pressures within their professional contexts (Nolan, Barry, Burke, & Thomas, 2014; Squires, 2014; Thomson et al., 2014). The impact of the recent global financial crisis, changing climates and ecological threats not only have a significant impact on society and populations' health and nutrition, they have increased the burden which health systems must address, without necessarily proportionate increases to financial resources (Butler & Harley, 2010). Where a complete medical or health professional degree must be delivered in the context of reduced financial resources and competing curriculum demands, choices must be made.

Clinical skills are a hallmark of the professional practitioner, with a potentially significant impact on health service costs (Oyebode, 2013). However, teaching procedural skills is just one part of health care curricula. It must therefore be approached with reference to education cost and skill retention understanding. As I will demonstrate in this thesis, little is understood of the cost, effectiveness, and retention or time requirements for clinical skill education.

1.1 Teaching clinical skills: two common approaches

In the late 20th century, a four-stage approach (4SA) to teaching surgical skills was described by Mike Walker and Rodney Peyton (1998). This teaching strategy (outlined in Table 1) gained support both in the surgical theatre setting and in skills laboratories during a time when technological advances in medicine demanded clinical advances in practice which may not have been adequately met by skills education. An example of this is a local increase in patient surgical complications following the introduction of the less invasive laparoscopic surgical technique, resultant largely from inadequate clinician skill training (Rodney Peyton, personal communication, August 31, 2015). A need for ongoing clinical education, including skill education, was identified by this surge in complications potentially caused by poor familiarity with new techniques and technologies. Grief et al. (2010) add that a decrease in specialisation training time also prompted the development of the 4SA. The 4SA was first described in the literature in 1998 (Walker & Peyton, 1998), and has gained momentum in many areas of health professions education, including Advanced Life Support (ALS) (Australian Resuscitation Council, 2015), Paediatric Advanced Life Support (PALS) and "Teaching on the Run" courses (Lake & Hamdorf, 2004).

The four stages of the 4SA have come to be known as demonstration, deconstruction, comprehension and performance (Lake & Hamdorf, 2004), although this language was not originally used in the description, and different descriptors[1] are present elsewhere in the literature (Hamdorf & Hall, 2000; Thomas, 2012). In the first stage, the educator demonstrates the skill as it would normally be performed, in real time and without particular explanation. It is sometimes referred to as a "silent" demonstration (Benjamin, 2005) however this is not necessarily the case as incidental communication with the patient or other colleagues may occur during this step. During the second stage (deconstruction), the educator again demonstrates the skill, but does so slowly, with explanation at each step. Walker and Peyton (1998) argue that at this point the student starts to develop an understanding of the rationale and process of each step in the skill and is invited to ask questions. Stage two is common in many traditional and ad-hoc teaching strategies. Stage three (comprehension), on the other hand, is less common. Here, the student prompts the educator to perform the skill, by giving them direction at each step. Finally, in stage four (performance) the student performs the skill, verbalising what they are about to do and why as they perform the skill to allow the educator time and opportunity to intercept if required. Again, the educator will prompt or question understanding as necessary and provide feedback to the student.

A more traditional, and perhaps simpler, skill teaching strategy may be referred to as the two-stage approach (2SA), and is also outlined in Table 1. 2SA was documented as early as the early 1970s (Mackety, 1973) and even then it was already well established as a teaching method within the medical apprenticeship model. Sometimes referred to as "see one, do one"(Herrmann-Werner et al., 2013), this technique involved a demonstration of the skill, with varying levels of explanation, followed by student performance of the skill. The andragogical rationale behind 2SA is poorly documented, but it has become a way of life for many clinical students and educators.

Table 1: The Four-Stage Approach (4SA) and the Two-Stage Approach (2SA)

	4SA	2SA
Stage 1	Real-time demonstration	Slower demonstration with accompanying explanation
Stage 2	Slower demonstration with accompanying explanation	Student performs the skill
Stage 3	Student instructs the teacher, as the teacher performs the physical components of the skill	N/A
Stage 4	Student performs the skill	N/A

[1] Thomas (2012) describes the four steps as demonstration, demonstration, formulation and performance; Handorf et al. (2000) and Barelli et al. (2010) refer to demonstration, deconstruction, formulation and performance.

4SA was first described early in the educational age of "standardization and marketization" (Hargreaves & Goodson, 2006, pp. 30-31). This era of education development began in the mid-1990s, and this context may have promoted subscription to the standardised approach 4SA offers to clinical courses. The intention is reasonable: that standardised teaching leads to acceptable clinical performance and baseline competence. However, the argument that a standardised teaching input creates a standardised performance output only holds true if the students are standardised. As I will demonstrate, little evidence exists to support or refute the assumption that 4SA is a superior skill teaching method.

1.2 Theory of learning

In planning to answer these questions, my initial approach to learning has been challenged significantly by the course of my research, with an evolution from an initial focus on observed action and behaviour (Johnston, 2016) to a later emphasis on change in identity. The first three studies (Chapters 3 to 5) focus on measuring and comparing differences in skill performance, with observed action used as a proxy to assess behaviour. However, during the development of assessment tools in Chapter 4, some data indicated that clinicians' performance is not simply an action, but rather it is adaptable and flexible, based on the needs of the patient. Thus, as a student clinician learns, they learn not only action, but also the appropriate application and adaptability. This requires a deeper definition of learning as a change which drives behaviour. The final chapter of this thesis has brought me to an appreciation that clinical education is far more complex than a behavioural approach can encompass, because learning is a change in being. It encompasses a clinician's emerging professional identity which infuses practice and works out in action, but runs so much deeper than what is evident through the positivist (or post-positivist) lens. I will argue that this is true both for the educator who is learning to teach and becoming a teacher, and for the clinical student who is learning to do, and becoming a practitioner.

1.3 Theoretical Framework

The focus of this doctoral research is a series of studies performed to understand the cost-effectiveness of 4SA, in an effort to address how strongly educators should be encouraged to use the approach. Paul Worley's model of Symbiotic clinical education (Worley et al., 2006) has been adopted as a theoretical framework for this thesis to guide and order "what issues are important to examine" (Creswell, 2013, p. 64). Worley's model of Symbiosis (also sometimes referred to as the integrity model, or 4R model due to its four key relationships) will be used to give multidimensional meaning to our considerations for cost, effectiveness, and the overall picture of worth. The symbiotic model emphasises the *relationships* of the key groups of people in the educational system (for example students, clinicians and patients, universities and health services, communities and governments), positing that it is the discovery or creation of ‘win-win’ relationships between the various groups that leads to educational success. In this thesis I focused first on the different groups' clinical, institutional, and social interests, responsibilities and limitations, reflecting three of the four key relationships identified by Worley et al. (2006). As the research progressed, it became apparent that the remaining relationship in this model (the tension between a clinical student's professional expectations and personal values) is also prominent in skill development. Use of this model brought meaning to the outcome measures in the studies by highlighting the potential tensions and collaborations between the groups that arise from a student's placement in the system. This framework assists in building the case that clinical education reaches beyond the skills laboratory to the patient, health system and wider community. This framework will be described more fully in section 2.6.

1.4 Philosophical approach

As with many PhDs, this study series has challenged and moulded my understanding of the nature of knowledge, and assumptions about it. Due to a significant development in my understanding of these, describing a single overarching ontological and epistemological approach which informed the research question, methods and research strategy would be an artificial representation of the approach to this series. Instead, each study chapter will make explicit the guiding ontological, epistemological and methodological perspective used to approach the research question being asked in that study.

The extent of this research series was initially limited to a longitudinal comparative trial comparing the retention of manual defibrillation skills between students who were taught with 2SA and 4SA (adapted to a cross-sectional trial represented in Chapter 3). This post-positivist[2] approach reflected

[2] In this thesis, references to positivism and post-positivism are consistent with those described by Crotty (1998). Crotty presents positivism as reliant on objectivism, as meaning and reality exist independent of conscious experiences (pp. 5-6). Scientific knowledge is accurate, certain, and objective (p. 27), thus the world addressed by this approach is not related to personal interpretation and experience (p. 28). Post-positivism, on the other hand is described as a "less arrogant form of positivism" (p. 29). It refers to probability, and a certain level of certainty, rather than total certainty. This is consistent with the use of *p*- values to quantify the statistical certainty present in the data obtained. Lincoln, Lynham, and Guba (2011) also argue positivism as a presentation of reality as a series of laws or facts, whereas post-positivism presents reality as "imperfectly and probabilistically apprehendible" (p. 100), and they add the idea of hypothesis falsification rather than hypothesis verification, the latter of which is more prominent in positivist approaches. Thus, the statement

assumptions that reality was fixed, and an absolute truth existed that could be measured. The findings of that study, in addition to subsequent literature and conceptual exploration challenged this approach to research and knowledge that I had subscribed to, and a shift towards a more subjective, constructionist, contextual approach to reality and knowing will be increasingly evident towards the final study. This evolution is not a sacking of the former approach, but rather a widening perspective on the context and usefulness of specific approaches relating to the question at the heart of the research.

Initially the research question focussed on the comparative effectiveness of two skill teaching methods: an objective question which inclined appropriately to the post-positivist paradigm. But as the studies grew it became apparent that the actual research question behind this study was concerned with understanding how educators ought to teach, and how the needs of the health service, teaching organisations, patient and clinician are influenced by different skill teaching methods. A post-positivist approach, employing a comparative trial is only one way to address this question, hence additional strategies were later employed. This is made explicit in each study, with the acknowledgement that a single question, or problem, can be understood and approached in many different ways, depending on the researcher's background, strengths, contextual experience and perceived implications.

1.5 Thesis outline

After building a foundation upon which to base the study series in the literature, I will present a series of four discreet but related studies. Each study is connected to the rest as outlined in Figure 1.

---

and testing of a *null hypothesis* in order to seek to falsify the hypothesis is consistent with post-positivist approaches.

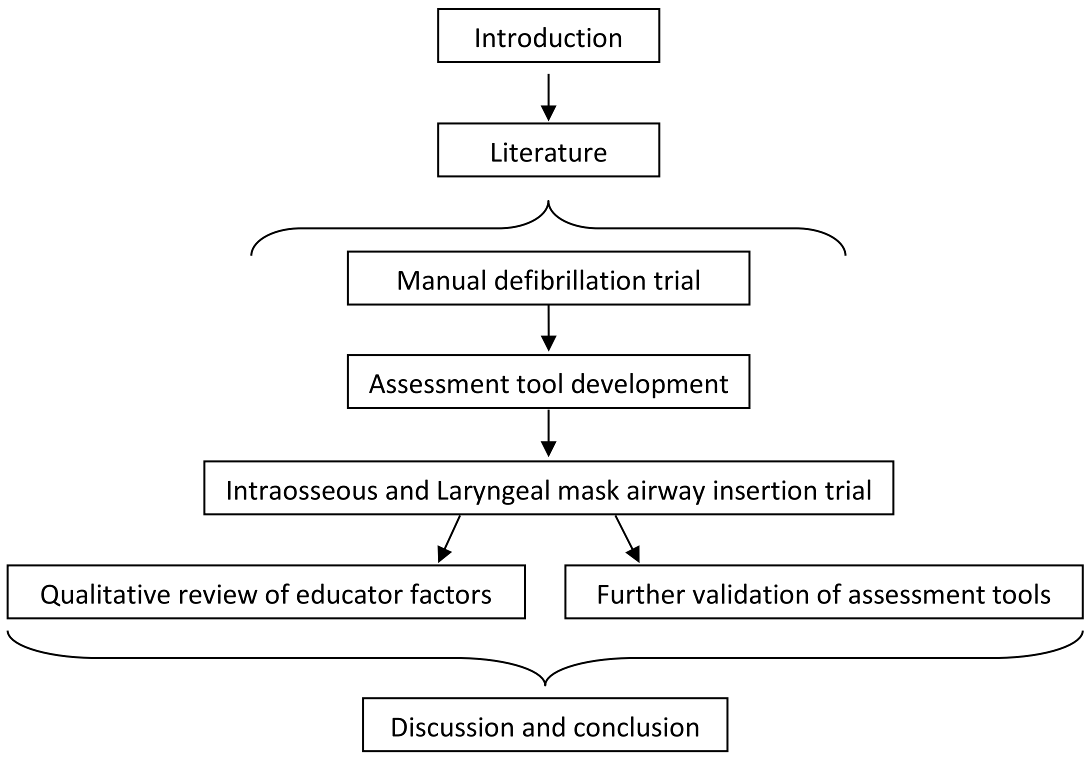


Figure 1: Relationships between the major parts of this thesis.

1.5.1 Study 1: Comparative trial

This study was designed to detect whether there is a difference in defibrillation skill acquisition related to the skill teaching method: 4SA or 2SA. The study aims to build on existing work in the area (Archer et al., 2014; Bitsika et al., 2013; Greif et al., 2010; Herrmann-Werner et al., 2013; Jenko et al., 2012; Krautter et al., 2011; Lund et al., 2012; Orde et al., 2010) with the introduction of an additional unique feature: understanding the participants' baseline performance. This aspect to the study is key to attributing performance ability accurately to the teaching methodology, rather than to pre-existing knowledge.

1.5.2 Study 2: Assessment tool development

This study aimed to address a gap in the literature to allow a scholarly and rigorous approach to Study 3. There were no skill-specific education and assessment tools for pre-hospital Intraosseous (IO) and Laryngeal Mask Airway (LMA) insertion. These two skills are used in a variety of contexts, however the unique context of pre-hospital emergency resuscitation demands particular considerations for insertion of these two medical devices. A modified Delphi approach was employed and will later be critiqued and further validated in Chapter 6. This study proposes two clinically relevant education and assessment tools, based on expert practice, specifically for use in the resuscitation education setting. Tools such as these are key to expanding education and assessment capacity beyond a dependence on clinical experts. The literature demonstrates that such non-experts can offer valid and reliable judgements on subjective assessment tools where they are carefully structured (Schuwirth & Ash, 2013). This is a significant step forward for regions and contexts which may learn, practise, teach and assess these skills with little or no supervision or feedback following achieving the authority to practise such skills.

1.5.3 Study 3: Comparative trial (retention)

The assessment tools developed in Chapter 4 will then be applied to a second trial, in order to understand not only the acquisition of IO and LMA skills with reference to baseline performance, but also retention six months following instruction. This study is key to beginning to understand the atrophy of clinical skills over time (Amaral & Troncon, 2013; G. S. Anderson, Gaetz, & Masse, 2011; Greif et al., 2015), but for the first time it provides data on anticipated patient risk as skills atrophy. Given some limitations of a single assessment of competence (Eva et al., 2016), measuring skill retention/decay over a time period and understanding the patient impact are much more realistic indicators of skill performance and actual learning (sustained change in behaviour) over time.

1.5.4 Study 4: Validation of assessment tools

This chapter is integral to understanding the validity of the checklists developed in Chapter 4, and the assessment of clinical skills generally. The Delphi process has been presented in the literature as a means to produce a validated skill performance checklist. However, the face validity afforded through the Delphi consensus process is an incomplete validity analysis (Boulkedid, Abdoul, Loustau, Sibony, & Alberti, 2011). Instead, by reviewing the outcomes, assumptions and intended uses of the assessment tools, this chapter will provide a more rigorously validated use of the checklists. This argument is informed by Kane's definition of validation procedures (2006), and presented under a series of arguments clarified by Schuwirth and van der Vleuten (2012).

1.5.5 Study 5: Qualitative review

Finally, questions arising from Chapters 3 and 5 around the educator's use of 4SA will be examined as the thesis more fully embraces the deep value of qualitative research in building new understanding. This shift in epistemological approach was informed by a new research question which focuses on why educators teach the way they do, how educators approach 4SA, and why we might have seen lower compliance to 4SA than to 2SA during the comparative studies. Clinical instructors were recruited from two pools who have training and some (albeit recent) experience in either performing or watching 4SA. Advanced Life Support (ALS) instructors who were completing their instructor training were invited to complete surveys and participate in semi-structured interviews, and clinical educators who were enrolled in postgraduate clinical education studies at

Flinders University participated in a recorded educator debrief and focus group. Thematic analysis was performed at multiple stages of data collection, and these informed further data collection. This study is the first of its kind to address the educator perspective of 4SA, with reference to the costs and benefits for a key human resource: the educators.

1.6 Summary

This educational research is rooted within the tensions and interests of the clinical practice environment. It seeks to uphold a focus on the institutional efficiency and patient impact of clinical skills education. The original contribution to knowledge proposed through this thesis is to test what has previously only been assumed: That the four-step approach (4SA) is cost effective. Through a series of distinct but related studies, factors of cost and effectiveness will be understood from a holistic and integrated perspective. Effectiveness will be understood in light of baseline performance, patient morbidity and mortality factors, skill acquisition and performance retention as such factors are of key importance to patient management and outcomes. Cost will not directly measure financial implications, but will consider time, resources and effort with. The range of questions asked in this research demands a mixed methods approach with epistemological and methodological adaptability. It is an ambitious aim to comment on patient outcomes from an educational study, and such claims will be limited to inferences from observed simulated practice.

This research is pertinent to enabling the appropriate direction of potentially scarce medical resources to the key area of skills education within the health professional education setting. In addition, it will enable understanding of the educator impact when they are asked to use specific strategies to perform their role. By better understanding the costs of 2SA and 4SA, and the measurable outcomes of the student performance from each method, educators, education bodies, and health service providers will be better equipped to give informed direction on staff teaching strategies in light of the time and resources available, with a more evidenced-based understanding of expected skill maintenance or atrophy.

2 LITERATURE

2.1 Introduction

Having described 4SA, this chapter will review the key learning principles relating to clinical skill development before identifying and critiquing the main rationale supporting the 4SA. It is tempting to complete a survey of the literature at this point, as a compelling case can be made for the strong connection of 4SA to learning principles accepted throughout the education and medical education community, however when I consider the current evidence relating to the cost-effectiveness of 4SA compared to more traditional teaching techniques, some concerns remain: The findings are inconsistent between studies, and cost factors have not been considered in the studies identified. Could it be that the solid theoretical foundation upon which 4SA is structured is not as strong as first thought?

These questions fuel the study series within this thesis, and prompt a more critical review of the learning literature to understand if it really does support 4SA in the way many educators believe it to. In this chapter, I will first consider the skill education literature, before turning attention to evidence surrounding the use of 4SA. I will critically examine the literature in light of the hypothesis that educators ought to teach with 4SA rather than 2SA on the basis of its effectiveness, and also seek information on its cost-effectiveness, in order to help course designers make informed decisions about how clinical educators should be encouraged to teach. Finally, the symbiotic clinical education model will be used to demonstrate a compelling case for the significance of the study to the wider stakeholders of the integrated clinical education system.

2.2 "Skill" development

4SA is advocated by many on the basis of a strong theoretical foundation in learning principles, some of which are described in the initial work, and accepted in later works by other authors in support of 4SA (Barelli & Scapigliati, 2010; Resuscitation Council UK, 2008). It seems appropriate therefore, to first understand the landscape of andragogical principles relating to clinical skill education, before understanding what case can be made for 4SA. What is understood by the term *skill* may vary. Do clinical skills require expertise and ability which sets the practitioner apart from the average, such as a *skilled* painter or a *skilled* sportsperson? Or is *skill* simply used to describe an action performed adequately? A professional (or clinical) skill demands a student to know, do and perform within an appropriate setting, and these elements of skill development and practice are understood to varying extents in the discussion of learning theory to follow. The term *clinical skill*, does not tend to capture the idea of elite, excelling practice. Mok and Ker (2015), for example, refer to the National Health Service definition of clinical skill, referring to "any action involved in direct patient care which impacts on clinical outcome in a measurable way" (p. 405). This definition refers to an action which effects patient care, but like most references to clinical skills, risks overlooking the idea of *expertise* in skilled practice.

2.2.1 Cognitive aspects of learning (knowing)

2.2.1.1 Cognitive limitation

Within cognitive learning perspectives, the number seven has been argued as a "measure of short-term memory capacity for processing cognition" (Saaty & Ozdemir, 2003, p. 234). Miller first noted the apparent limitations of the mind's ability to hold information in short-term memory during a musical experiment where most participants (such as those who are not "musically sophisticated") were able to hold around 2.5 units of discernible variation (bits), or around six different pitches in their head before he or she begins to get confused (G. A. Miller, 1956). On the basis of mathematical computations, nearly half a century later Saaty and Ozdemir confirmed that Miller's "seven plus or minus two is indeed a limit, a channel capacity on our ability to process information" (Saaty & Ozdemir, 2003, p. 244). The number of pieces of new information is therefore an important factor in skill education.

Nicolis and Tsuda (1985) deduce that "there is a span of absolute judgement that can distinguish seven categories and there is a span of attention that can encompass about seven objects or symbols at a glance (p. 345). Miller is quick to appreciate, however, that the appearance of the number seven in both human attributes, "the span of absolute judgement and the span of immediate memory are quite different kinds of limitations that are imposed on our ability to process information. Absolute judgment is limited by the amount of information. Immediate memory is limited by the number of items" (1956).

2.2.1.2 Cognitive load theory

Since Miller's initial work, cognitive load theory (sometimes referred to as "cognitive limit theory", "memory span" or "attention span") has become a well-established and accepted principle in educational practice. Cognitive learning theories such as that outlined by Sweller, Van Merrienboer and Paas (1998) reinforce effective learning as a function of not only the volume, but also the classification of cognitive load. The proposed classifications of cognitive load, outlined in Table 2, are intrinsic, extraneous and germane load (Sweller et al., 1998; Whelan, 2007). Intrinsic load is unavoidable, finite, and dependent on the difficulty of the content. Extraneous load depends on the way in which knowledge is presented to the learner, and germane load refers to the student's ability (and sometimes also motivation) to reorganise or construct schemas to accommodate and organise the knowledge for streamlined retrieval (Nicolis & Tsuda, 1985; Whelan, 2007). Organisation of the learning (germane load), and the schematic reorganisation used to do so will affect the compressibility of learnt information, and its storage space (Nicolis & Tsuda, 1985). The total cognitive load on the student is the sum of these three parts, and Sweller et al. (1998) suggest that the only one which the educator can address to try to free up cognitive space for the others is extraneous load.

Table 2: Cognitive Load Classifications

Type of Cognitive Load	Explanation
Intrinsic	The unavoidable demand, dependent on the complexity of the teaching content
Extraneous	Effected by the presentation of the data
Germane	Dependent on the learner's ability to organise learned data for future retrieval

Information recall is closely aligned to the effective development of schemata, not just the storing of knowledge itself. Fitts (1964) discusses the adaptive processes of the "stored-program data processing system", which seeks to efficiently perform tasks by efficiently recruiting the learnt process ("subroutines") that make up the larger routine (p. 251). In this way, the smaller components of a skill are retrieved and placed together to form the whole skill. This, Fitts goes on to argue, is consistent with advice to "chunk" knowledge into manageable packages as it is often retrieved as such and reconstructed into the wider whole. Cognitive load theory supports the teaching of new information in "chunks" which are a manageable size, of seven pieces or less.

In relation to cognitive capacity, though, Oliver Sacks asks "to what extent are we - our experiences, our reactions - shaped, predetermined, by our brains, and to what extent do we shape our own brains?" (Sacks, 2005, p. 25). This acknowledgement that neural habits reinforced over a person's lifetime make some cerebral processes automatic, when at the same time the experience creates new neuronal stimulation and potentially also new pathways, leaves us with an unanswerable dilemma: How much learning is a result of stimulus, and how much is the result of pre-existing anatomy shaped over a person's years of experiencing the world?

2.2.1.3 Neuron recruitment through observation and listening.

Observation of a procedure has been found to activate the same circuits in the cerebral cortex as would be recruited were that person actually performing the skill (Balmer & Longman, 2008); (Blandin, Lhuisset, & Proteau, 1999). The mirror neurons responsible for this have been demonstrated to discharge whether a person is doing or witnessing an action (Cattaneo & Rizzolatti, 2009). The value of demonstration is reinforced by Wong et al. (2009) in their finding that animations showing the real-time movement involved with manipulating an object results in superior performance for primary school children when compared to children who were provided with static graphics to learn the task (folding origami). These findings were specific for motor skills, indicating the benefit from viewing a performance of the action being taught.

Likewise, "listening to a verbal description of the skill can activate the same visual-motor circuit as those activated... when completing the skill". Through testing the brain regions associated with the motor action of a particular statement (such as "I kick a ball"), Tettamanti et al. (2005) identified cerebral activity in the cerebral regions responsible for the related motor action, even though that action was not occurring. Thus, observing and listening to rich bodily descriptions of psychomotor skill performance may pre-empt neuronal pathways required for motor neuron deployment during the skill action. Sacks also notes the ability of some blind people to "construct detailed visual images from verbal descriptions" (Sacks, 2005, p. 38). This is possibly an adaptation as other participants, even those who had initially lived with the ability to see, had eventually lost the ability to imagine visual images. This descriptive account confirms the notion of interconnectedness between cerebral regions, with one form of sensory input stimulating another. Sacks reinforces this idea, noting "sensory modalities can never be considered in isolation" (Sacks, 2005, p. 33).

It is noteworthy that neither of the above studies measure or "control" for people with different learning styles. The notion of kinaesthetic, auditory or visual learners has risen to great popularity despite a lack of demonstrable validity around the use of learning style assessment instruments for maximal student-instruction matching (Dembo & Howard, 2007; Riener & Willingham, 2010). Other studies suggest that cerebral activation of motor (kinaesthetic) neurones is prompted by verbal (auditory) and observational (visual) data, implying that they are interconnected and interdependent, rather than separated within the learner (Cattaneo & Rizzolatti, 2009; Tettamanti et al., 2005). Individual learning styles, therefore, will not feature as a part of this study.

2.2.1.4 Separation of cognitive and manual dexterity components of the skill

Separating the cognitive understanding of the steps of the skill from the motor components is proposed to decrease cognitive competition for the student when performing the manual dexterity components of the physical skill. By having a head start in understanding the *knowing* and *knowing how* components of the skill (later this will be discussed more explicitly with reference to Miller's pyramid), when they come to perform the skill, some educators believe this to be an advantage as a foundation is already laid. It is a similar concept to the pre-work completed in a flipped classroom model of instruction. An assumption of this argument is that the brain, like a computer processor, has a somewhat limited capacity and therefore commanding too many different tasks at once may create overload and slow the central processor.

Zhu et al. (2011) note that the Electroencephalograph (EEG) of an expert surgeon shows less co-activation of the verbal-analytic and the motor planning regions of the brain than for a novice. This is thought to be due to an increase in cerebral efficiency due to neural plasticity. This suggested increase in neural efficiency amongst expert surgeons is consistent with other distinctions in novice/expert practice, and suggests that there is more active cerebral involvement between these areas for a learner, and this is noted to “interfere with motor performance and have been implicated in skill breakdown under psychological stress... Additionally, inexperienced performers who depend on verbal-analytic processes tend to be disrupted by multitasking conditions" (Zhu et al., 2011, p. 291). This may support the use of strategies which engage the verbal-analytic area, such as asking the student to verbalise each step to the skill. Zhu et al. (2011) determined that cerebral processes differ between surgical practitioners of varying experience. This raises the question of how one defines and measures learning and performance: Is it knowing (cognitive), doing (behavioural), or performing in context (social)?

2.2.2 Behavioural aspects of learning (doing)

Instruction focussed on communicating knowledge from the educator to student paves a way for motor neuron recruitment in order to perform the manual aspects of the psychomotor skill. Fitts and Posner (1967b) argue that there are two types of skills: perceptual motor skills which involve manual dexterity and sensory perception, and language skills which relate to signs, symbols, mathematics and problem solving (p. 4). Fitts' (1964) theory of skill development outlines three phases of perceptual motor skill learning. In the first phase, the early (cognitive) phase, the learner grows to understand what is required, and completes "a few preliminary trials" of the skill (Fitts, 1964, p. 262). Fitts identifies that a particular difficulty at this stage is "response integration". Even if a learner is able to perform isolated parts of the new skill with relative ease, coordination of multiple simple actions can be highly demanding (p. 262). Fitts argues that demonstration and verbal instruction are the most useful forms of instruction at this point (Fitts & Posner, 1967a, pp. 11-12). This is comparable to the foundation of Miller's pyramid: to know. During the intermediate (associative) phase, practice of the skill is demonstrated to improve performance (Fitts, 1964, p. 266). Continual asymptotic improvement of a skill, followed by eventual plateau is characteristic of musical and sporting skill development, with expert skill rarely demonstrated within "several years of intensive, almost daily practice" (Fitts, 1964, p. 268). The learner reinforces new patterns, as integrated with existing ones, with a decreasing occurrence of errors (Fitts & Posner, 1967a). The final autonomous phase reflects cognitive autonomy, and faster skill performance with the capacity to multitask (Fitts & Posner, 1967a, p. 14). This mastery, having developed over years of deliberate practise is also known as the late phase.

I suggest that it is imperative to address some key assumptions in this widely accepted model of skill development. Firstly, practice must be guided and supervised with appropriate feedback in order to promote appropriate performance. Independent practice may provide the opportunity to reinforce erroneous choreography, which may become habitual. Thus, practice ought to reinforce and build *expertise*, rather than simply *experience*. Secondly, there is a detectable undercurrent in this outline of skill development which may assume that a skill is performed in a single way, which can therefore be improved over time. For many clinical skills, especially those in the uncontrolled pre-hospital, rescue or resuscitation environment, is that variation is almost the only reliable constant. Rationale, patient considerations, co-morbidities, setting, and even clinician posture relative to the patient dependent on incident location and patient access will all vary to impact skill application. This variability is in tension with an assumption that a skill ability, when practised consistently, will reliably improve.

Oermann et al. (2011) notes that during the autonomous phase, conscious thought is not required in order to perform each step, likening the development to that described in the fourth level of the ladder of competence (or competence model). Sometimes depicted as a ladder, and other times a wheel or matrix, Sorensen (2014, pp. 63-64) points out that the movement through various competences from novice to competence involves both *knowing* and *doing* competences (presented in Figure 2. While the origin of the concept is unclear, Adams (2016) attributes the model to "*former Gordon Training International employee, Noel Burch over 30 years ago".* It has been widely adopted across disciplines.

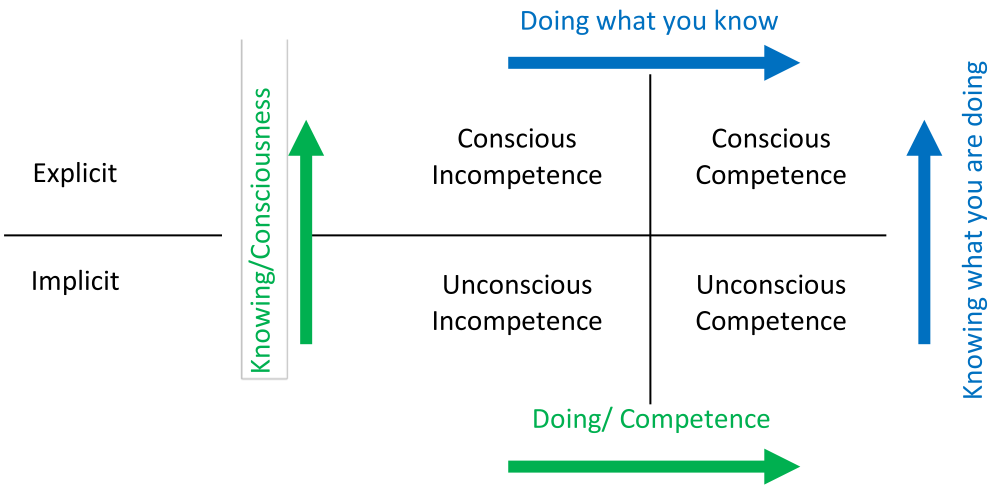


Figure 2: Conscious and unconscious competences presented as a matrix. Adapted from "Improvisation and teacher expertise: a comparative case study" (2014) by Sorensen, N. T., pp. 63-64.

The ladder of competence argues that learning begins with the realisation of incompetence. While the learner is unconsciously competent, they may not be ready to accept new information. However by increasing awareness of their own incompetence, the student prompted to learn to perform a task, and with practice less and less conscious thought is required in order to perform it. In the first stage, the learner is unconsciously incompetent. They have no experience or knowledge of what is required. At this point, learning is difficult because the student is not aware of their need to learn the skill. In the second stage, the learner has transitioned into conscious incompetence: they have knowledge of their inability to perform the task, and become more able to value the learning. In the third stage they become consciously competent, still having to think explicitly about performing the skill, and in the fourth stage through much practice, the action becomes somewhat automatic and competence is unconscious. The proposal of a fifth stage suits the competence ladder more so than the matrix, with the addition of mastery. This stage recognises the limitations of an unconsciously competent practitioner to recognise the difficulty of a procedure they find easy and automatic. The masterful practitioner, therefore, is mindful of the skill complexity without being burdened by the conscious difficulty of the task which they are expert in. This clinician is ready to engage in teaching with an understanding of the skill required.

With deliberate practice, the student moves from conscious competence (having to consider each step) to unconscious competence (automaticity). Deliberate practice, however, is not a simple case of independently repeating the task. Oermann et al. (2011) state that it involves assessment, constructive feedback, and skill improvement (however the intervention in their study was independent practice, rather than guided/facilitated practice with feedback). In their study, the group which practised the skill for just six minutes a month were able to perform some parts, but not all, of the skill (Cardio Pulmonary Resuscitation, or CPR) better than the control group which received no opportunity to practise. This study suggests that small amounts of practice can improve performance, however the type of practice suggested by Oermann et al. (2011) involves deliberate feedback aimed at improving performance. Deliberate practice, rather than practice alone, speaks to the difference of experience (through practice) and expertise (through refinement and deliberate practice). In fact, Guadagnoli and Lee (2004) claim that the "generalisability of the relationship between practise and skill is so profound that it is sometimes modelled mathematically and referred to as a law".

Dreyfus' (2004, p. 181) five-stage model of skill acquisition describes development from novice to expert in terms of the skill application context (whether the skill is performed in isolation or embedded in a clinical situation), perspective, decision reasoning and commitment. These are demonstrated in Table 3.

Table 3: Dreyfus' Five Stages of Skill Acquisition.

Skill Level	Components	Perspective	Decision	Commitment
1. Novice	Context free	None	Analytic	Detached
2. Advanced beginner	Context free and situational	None	Analytic	Detached
3. Competent	Context free and situational	Chosen	Analytic	Detached understanding and deciding; involved outcome
4. Proficient	Context free and situational	Experienced	Analytic	Involved understanding; detached deciding
5. Expert	Context free and situational	Experienced	Intuitive	Involved

Note: Components: this refers to the elements of the situation that the learner is able to perceive. These can be context free and pertaining to general aspects of the skill or situational which only relate to the specific situation that the learner is meeting. Perspective: as the learner begins to be able to recognize almost innumerable components, he or she must choose which one to focus on. He or she is then taking a perspective. Decision: The learner is making a decision on how to act in the situation he or she is in. This can be based on analytic reasoning or an intuitive decision based on experience and holistic discrimination of the particular situation. Commitment: This describes the degree to which the learner is immersed in the learning situation when it comes to understanding, deciding, and the outcome of the situation-action pairing. Adapted from "The five-stage model of adult skill acquisition" (2004) by Dreyfus, S. E., *Bulletin of science, technology & society* 24(3), p. 181.

The development of skill performance, therefore, is multifaceted. Dreyfus argues that the transition from novice to expert skill performance is underpinned by an evolution of various internal factors which Guadagnoli and Lee (2004) would argue, develop during much practice.

2.2.2.1 Retrieval for learning

Karpicke (2012) argues that learning should not be so concerned with knowledge storage, but rather efficient retrieval and knowledge restructuring. What information is recalled may not reflect the information stored, but rather how it is stored, organised, interpreted and utilised. Karpicke and Roediger (2008) found that a group of students who study material only once, but actively retrieved that information multiple times outperformed students who study the material three times, with no conscious attempt at recalling the information. This would then contribute to the hypothesis that 4SA, with the inclusion of the third step where students speak the educator through the skill performance, would result in better recall than for 2SA as it is recalled more frequently. Bearing in mind, the study to which they refer is one of knowledge rather than manual dexterity or psychomotor skill performance. While this was said to relate to long-term learning, the knowledge was re-tested a week following the end of the learning period.

One way information may be retrieved is through teaching someone else. This idea is reflected in the "see one, do one, teach one" adage popular in medical education throughout the second half of the 20th century. The notion that teaching is the best way to learn is popularised by 'inspirational' quotes such as the following:

> "We Learn …
> 10% of what we read
> 20% of what we hear
> 30% of what we see
> 50% of what we see and hear
> 70% of what we discuss
> 80% of what we experience
> 95% of what we teach others."
>
> William Glasser, as cited by Steen (2008, p. 527)

However this notion finds no solid base in educational (or medical education) literature. Bower (2009) uses this quote to support the argument that active involvement in education achieves deeper learning. His argument is in particular reference to obtaining and maintaining the attention of the learner. Despite lack of empirical evidence, this hypothesis is widely accepted, including the adapted version from Walker and Peyton, who state "it is estimated that we retain 5% of what we hear, 10% of what we see but up to 90% of what we do" (Walker & Peyton, 1998, p. 176).

Owen and Plummer (2002) argue that "Showing a video from the perspective of someone performing an intubation has been reported to increase initial intubation success rates" with the later caveat that "Learning must include acquisition, retention and retrieval, and in clinical practice, knowledge retrieval must be accompanied by the ability to apply it appropriately in different and novel situations." Owen and Plummer therefore couple visual and other teaching strategies with retrieval and clinical application. Extending the importance of retrieval to the risk of skill atrophy in the resuscitation setting, Williams (2011) argues that:

> [O]pportunities for learning and revising information about resuscitation should be the primary focus to prevent knowledge decay and enhance performance and ... courses that focus only on certification may actually inhibit learning (p. 244).

Here, Williams touches on the practical tension between competence-based assessment required to state a minimum standard for some certified training courses, and the practical ability of graduands to perform the skills expected of them following such an approach.

2.2.2.2 Acquisition and retention of skills

Smith and Greenwood (Smith & Greenwood, 2012) assert the essential nature of practise in the attainment and retention of skill proficiency:

> Practice is essential until the trainee can undertake the task automatically, allowing them to concentrate on the more complex technical and non-technical issues. Finally, once the student has mastered the skill, they will still have to practise regularly to maintain proficiency. This is especially important where the skill may be needed in an emergency which is why, for instance, resuscitation competencies need regular revalidation (p. 474).

This statement tracks the evolution from acquiring expertise to maintaining expertise, both of which must be intentional. The emergency situation may have been identified due to the stakes, or time-critical nature of the skills required; it is not clear in the text. However, the "regular revalidation" of critical skills is urged to ensure expertise retention.

Attrition of emergency skill application and knowledge has been well documented widely (Ali et al., 1996; Amaral & Troncon, 2013; Driscoll, Gwinnutt, & McNeill, 1999; Wayne, Siddall, et al., 2006; Wiles, 2015). Ali, Howard, and Williams (2002) noted no measurable skill loss at six months, but significant and increasing loss at two years, four years and then six years following training, contrasting with results reported by Wayne, Siddall, et al. (2006) who found no deterioration in resuscitation skills in their 14 month longitudinal study. Ali et al. (2002) report that skill decay was related to authentic workplace exposure, with OSCE performance of ATLS skills declining more significantly for practitioners with access to fewer than 50 trauma patients per year. Interestingly, it was the procedural skills that atrophied, rather than the methodical approach taught in the ATLS course (p.144).

Wisher, Sabol, Ellis, and Ellis (1999) argue that skill retrieval depends on three things: knowledge retrieval, cognitive processing, and precise execution of motor function. The processing required for procedural tasks which require recall of a series of steps in a task, cognitive tasks which require decision making, rationale and "troubleshooting" and perceptual motor tasks which demand precise motor control, all take place in different cerebral locations. When it comes to understanding or predicting retention, de Ruijter, Biersteker, Biert, van Goor, and Tan (2014) warn that different types of knowledge atrophy differently. In their study, procedural skills decay the most, followed by physical skills and then finally declarative knowledge. Wisher et al. (1999) argue that retention of cognitive components such as decision making and judgement tend to stay relatively stable for up to a year following training, however memory for the knowledge component (facts to be recalled) shows more prominent decay, and memory for perceptual tasks and procedural tasks shows reasonable atrophy. Wisher et al. (1999) state that:

> Many procedural tasks show this quick decline. It has been found, for example, that only 20% of civilians trained on the first aid task of giving cardio-pulmonary resuscitation (CPR) are proficient six months later (p. 8).

Likewise, Bullock (2000) argues the importance of maintaining resuscitation skills, and the role of training and health care organisations in tending to this through clinician review.

> Educationalists should be involved in curriculum design for resuscitation skills teaching. They should also be responsible for undertaking audit and research in order to establish the effectiveness of the teaching methods. Why is this important? Current data suggests (sic) that retention of both cognitive knowledge and psychomotor skills in resuscitation is significantly weaker after four to six months (p. 140).

Lammers, Byrwa, Fales, and Hale (2009) argues that continuing paramedic education is the "most effective remedy for skill atrophy". However, there are some flaws with this claim. In their study, paramedics were given initial training on paediatric emergencies, and assessed six months later using a checklist developed in order to reflect the clinical assessment modules. All participants were provided with comparable initial training, but assessments were not performed at the initial stage, so an assumption is made that participants' acquisition of the material is complete, and this may not be the case. Therefore, the study may not actually measure atrophy if it is not established that the knowledge was present at training. It is unclear how consistent the delay between training to testing was, and while continuing education is argued as the best remedy for skill attrition, this is not explicitly supported by the data obtained. While studies such as this do not provide specific suggestions for deficiency identification in order to target initial and remedial training and support, paramedic exposure to critical cases, in addition to alarming skill atrophy rates, prompt continued investigation into a foundational understanding of what learning resuscitation skills entails, with reference to ongoing practice.

Hein, Owen and Plummer (2010) studied LMA retention among first year paramedic students. Students attempted the skill 12 times immediately following the teaching, and returned six months later for reassessment. They were randomised into a control group (performed the skill during a scenario), or an intervention group (re-viewed the initial instruction video, and had the opportunity to practise the skill prior to reassessment). The initial performance was rated as success or failure by the principal author (Hein), and the retention performance scored by independent blinded assessors according to a locally devised series of Likert scales to provide a performance score out of 13. Students who had received refresher training prior to completing the LMA insertion task had superior skill performance to those who were asked to insert the LMA without refresher material. The situation of an ambulance crew receiving a dispatch to an urgent case, however, does not lend itself to staff undergoing refresher training prior to the case in order to boost the performance of potentially decaying skills.

George and Doto (2001) amend 4SA to add the contextual overview prior to the first step of 4SA to create a five-step skill teaching procedure. They assert that this is a,

> quick and easy five-step method. While going through these five steps may seem lengthy, the result is that the preceptor will spend less time observing and correcting performance problems and will ensure a better learning environment (p. 578).

George and Doto acknowledge the clinical and institutional tensions of clinical skill teaching, expressing assurance that despite criticism of longer skill teaching methods, the clinician, educator and health service will benefit in the long run as students will require less corrective instruction. Other authors, however, argue the requirement for intentional practice in aiding retention, rather than a superior initial skill teaching method (Wisher et al., 1999).

2.2.3 Social learning (being)

Having considered cognitive aspects of learning (knowing), and the impact this may have on the output action or behaviour, I will now focus on the integrated, social function of skill application and development. This aspect is central to the contextual nature of many modern learning environments and the practical expectations of clinical students.

2.2.3.1 Experiential learning

What has become known as Kolb's experiential learning cycle was originally published by Kolb as the "Lewinian experiential learning cycle" (see Figure 3), after Lewin "borrowed the concept of feedback from electrical engineering to describe a social learning and problem-solving process that generate valid information to assess deviations from desired goals" (Kolb, 1984, pp. 21-22). Yardley, Teunissen, and Dornan (2012) later attribute this to Kolb with minor changes (see Figure 4).

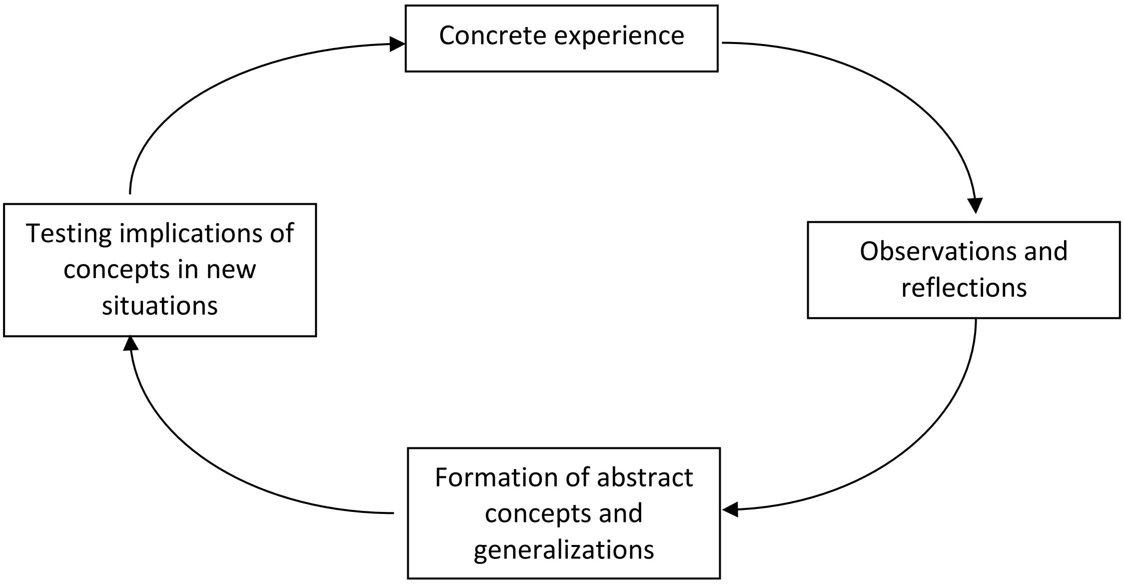


Figure 3: The Lewinian experiential learning model. Adapted from "*Experiential learning: Experience as the source of learning and development" by* Kolb, D. A., p. 21. Copyright 1984 by Prentice Hall.

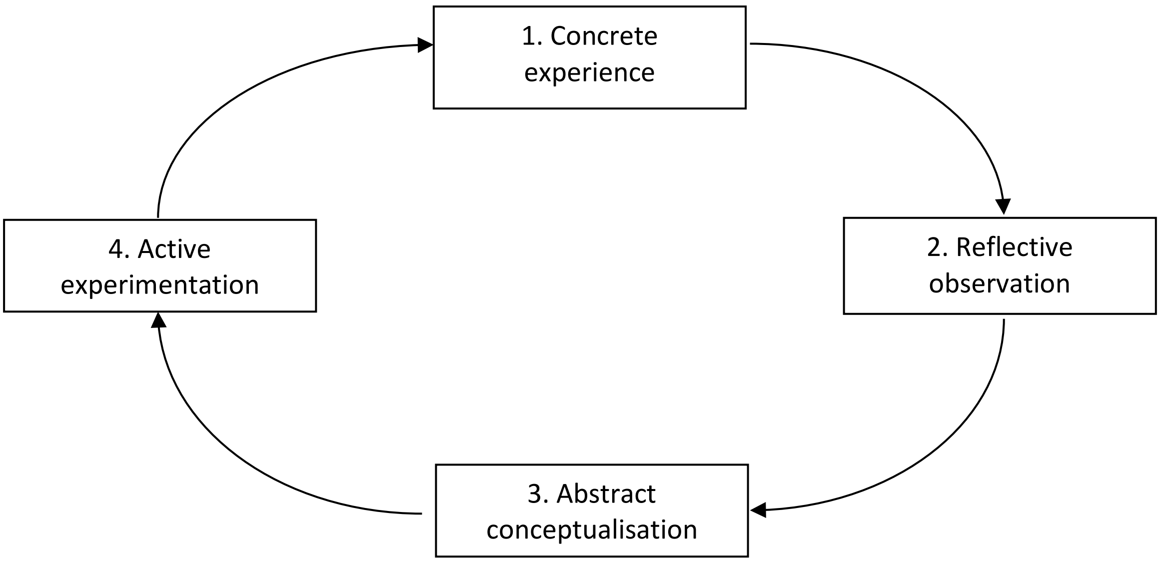


Figure 4: Kolb's learning cycle. Adapted from *"Experiential learning: AMEE guide No. 63" (2012) by Yardley, S., Teunissen, P., Dornan, T., Medical Teacher,* 34(2), p. e105.

Kolb and Lewin's four-step learning cycles may have roots aligned to those of Schon whose five-step experiential learning cycle appears in Peyton's work (Peyton, 1998, p. 33):

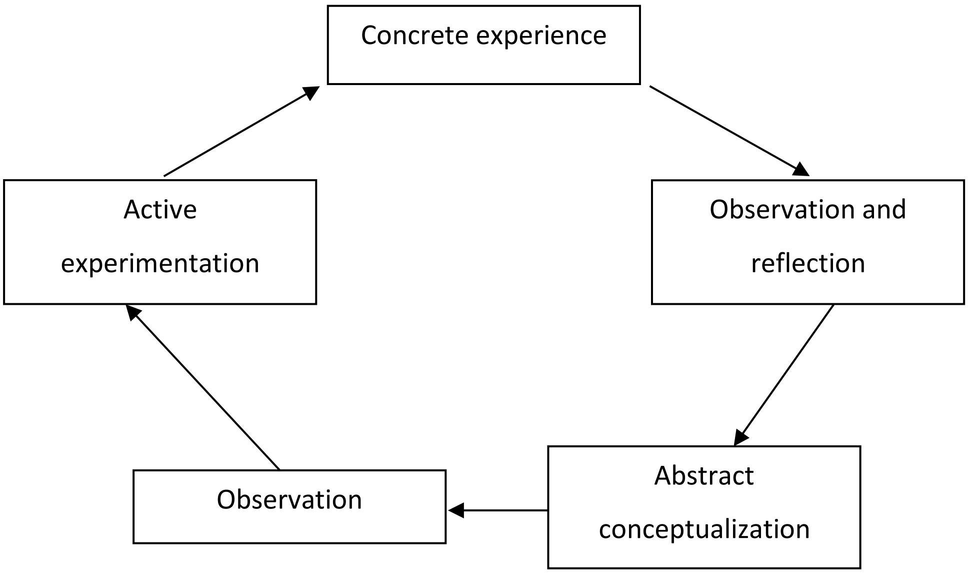


Figure 5: Schon's experiential learning cycle. Adapted from "Teaching and learning in medical practice" by Peyton, JWR., p. 33. Copyright 1998 by Manticore Europe.

Regardless of the origin, the notion of experiential learning argues that performance in a practice context is an essential part of learning, and in clinical education it occurs in either a simulated or authentic (situated) clinical environment. Bullock (2000) argues the centrality of the teaching environment for students learning psychomotor skills in his statement that:

> Learning practical skills is concerned with knowledge, skills and attitudes. However, others believe that mastery of the skill is also concerned with how the learner interacts with the teaching environment (p. 139).

The teaching environment is a key part to learning, and this true in the education environment, but also the practice environment. It may be seen as the environment in which one learns, or the environment *through* which one learns.

Situated learning and authentic learning promote application of what is learnt and have been significantly developed by Stephen Billett (1996). Billett's work bridges the cognitive and social aspects of learning by emphasising and demonstrating the importance of practice expectations and context on the learner and what is being learnt. He posits that the context in which knowledge is learnt significantly impacts whether or not it will be retrieved, and if a mismatch of learning environment (setting of knowledge construction) and application environment exists, "subsequent redeployment [of content] to other situations and settings" is unlikely (pp. 1-2). As knowledge construction is borne from problem solving, both in routine practice and within complex situations, learning is argued to be interdependent with the learning context and activity setting, rather than simply a function of the delivery style. This literature addresses learning less in cognitive terms, and more in social practice terms. Does this mean that a skill teaching approach used in the skills laboratory, taught without particular reference to the context, will access the cognitive resources (such as activation of problem solving) in order to achieve deep learning? And as a secondary but related question: How much does a unidirectional imparting of knowledge, focussed at developing memory and recall function within the student, activate problem solving processes? Do performance assessments measure how much knowledge is transferred (focus on teacher), or constructed (through problem solving by the learner)? And what does this reveal about our assumptions of learning?

Billett (2001) argues that expertise involves knowledge, how it is organised for appropriate retrieval, procedural and conceptual understanding, in addition to the social environment in which application of the knowledge occurs. He states that:

> Individuals' construction of the knowledge that comprises a situated domain of expertise is founded in interpsychological processes of how individuals act within social practice, as this interaction interdependently engages knowledge with historical, cultural and situational geneses. This interdependence between the social practice and those who act within it (Lave, 1991) may be contested (Billett, 1995b) or resisted (Hodges, 1998) (p. 3).

We must consider, then, the importance of integrating the skills taught with the practical, contextual and cultural expectations of their implementation. This necessitates reference to both the authentic work environment during clinical education as a valuable input to the social expectations of skill learning and application, and the fidelity of the teaching environment. In Brady et al.'s (2015) comparison of low and high fidelity midwifery teaching environments, low fidelity involved the use of part task trainers, in a standard birth-suite hospital bed. However an argument can be made that this setting is not entirely detached from the authentic clinical setting. Authenticity is designed to make the setting feel real to the learner, in a way which conjures the psychological and emotional responses which mirror those experienced in the professional setting (p. 525). This offers a glimpse at the importance of a student clinician's professional identity development, as a factor of clinical skill development. Key to developing a professional identity is the clinical student's acceptance of complexity and uncertainty (Bleakley & Bligh, 2008). Bleakley and Bligh (2008) argue that:

> If science informing medicine is taken as a template for medical education, then such an education must have high levels of indeterminacy and complexity, and we should frame learning as an emergent property of a dynamic, unstable system. This has clear implications for identity construction of medical students, no longer framed as a unitary agency, in control, and denying uncertainty (p. 98).

Through a development of professional identity, presumably assisted by authentic education and assessment, the clinical student learns to apply the professional self to the unpredictable problem, rather than apply a pre-determined approach. The deciding factor is the complexity and variability of the problem at hand, and the subsequent adaption required to treat any given case as unique. This level of authenticity and subsequent identity construction is based on the exchange between the student and patient, with the clinical educator acting as a "resource" for the student's learning, rather than a shaper of it. With this in mind, can educators be certain that a simulated patient can offer greater professional authenticity, psychological or emotional factors than does a manikin? Or can a clinical expert bring this context to the simulated setting through their experience of the authentic setting? And what does this say about the education of clinical skills? For invasive skills, an argument can be made for skill education to occur in a simulated setting on part task trainers, and integrating wider context and global expectations of professional development as skill proficiency increases and therefore identity may be formed. However, where a contextual approach to learning is sought, the educator will more richly contribute this if they also bring a developed professional identity as a clinician to the teaching.

Reznick and MacRae (2006) point out that learners in early stages of skill development should practise "outside the operating room", suggesting that these performances of higher error rate should take place in a simulated environment for patient safety reasons (p. 2664). They further argue that that fidelity may be of variable importance depending on the stage of the learner, with no difference in performance noted between simulation laboratory training using manikins, and either high fidelity video simulation or live animal training in two separate studies.

For initial skill learning, skills laboratory may be adequate and more cost-effective, with a transition to simulated patients where authenticity increases, and then to the authentic clinical context for further practise as proficiency grows. A strategy like this may be consistent with a spiral curriculum, first described by Bruner in 1960 in *The Process of Education*. Ronald Harden (1999) introduced the idea to Dundee medical school in the 1990's, with significant adoption of the spiral curriculum in clinical education literature since then.

Primarily used in medical school training, the spiral curriculum (as depicted in the four-stage model illustrated in Figure 6) conceptualises a student entering medical training at the bottom, and moving through various stages of a curriculum designed to integrate knowledge, skills and attitudes in a way which builds on previous learning, builds a foundation for future learning, and revisits topics of instruction with increasing difficulty (Harden, 1999, p. 142). Lake and Hamdorf (2004) add that as a student moves through the different phases of a medical degree, they learn normal physiology (phase 1), then learn explore deviations from normal (phase 2), relate this to clinical practice (phase 3), and put it into practice (phase 4). Note that this application suspends some experiential, application aspects of learning. The spiral curriculum was intended to reflect the reinforcing ability a curriculum can have on medical students during progression through their training, with reference to the role of clinical skill education. Of particular note is the introduction of professional attitudes in the very early part of the model. It is depicted with the same importance as clinical skills and theoretical understanding.

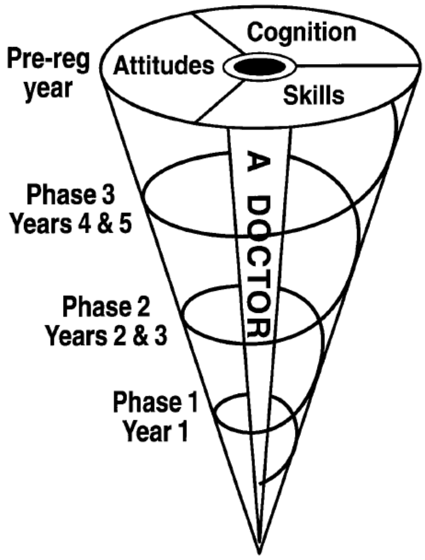


Figure 6: Harden's spiral curriculum. Retrieved from "What is a spiral curriculum?" (1999) by Harden, R.M., *Medical Teacher*, 21(2), p. 142

Miller's triangle presents a similar argument for the progression of learning. Focussed on the assessment of skills, Miller urges clinical educators and assessors to consider the depth of ability achieved by the various stages of development described in

Figure 7 (G. E. Miller, 1990, p. S63). For example, a multiple choice questionnaire may give insight into the bottom stage, but this ought not to be evidence that a student is safe to practice such knowledge. In skill education and assessment, this framework is helpful insofar as it highlights the progression of skill development from a cognitive process of knowing, to a behavioural process of doing.

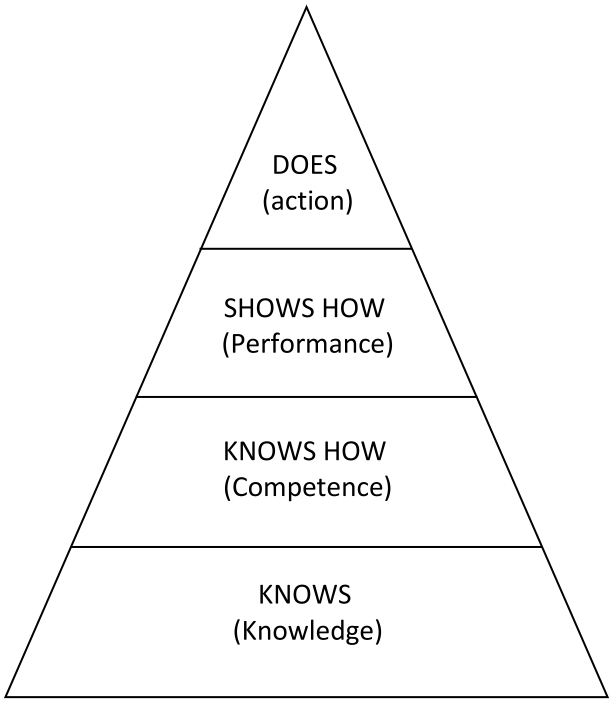


Figure 7: Miller's triangle (also referred to as *Miller's pyramid*). Adapted from "The assessment of clinical skills/competence/performance" (1990) by Miller, G. E., *Academic Medicine* 65(9) p. S63

A similar body of work which preceded Miller's triangle is Russel Ackoff's (1989) hierarchy of data, information, knowledge, understanding, and wisdom. Data are "symbols" which must be collated, processed, and interpreted, Ackoff argues, in order to be functional as information. He posits that information addresses the *what*, knowledge addresses the *how*, and understanding addresses the *why*, with wisdom addressing values and judgment. Miller takes this work further, however, by addressing the action in an authentic setting, underpinned by professional expectations and role. Similarly, the development of Ryle's separation of *knowledge how* and *knowledge that* (Ryle, 1945) is well exceeded by Miller's approach, particularly with reference to authentic practice.

Van der Vleuten and Schuwirth (2005) argue that authenticity in medical education assessment is not an exclusive property of the apex of the pyramid. Instead, "it is present at all levels of the pyramid, and in all good assessment methods". Therefore education and assessment methods which consider a clinical student's knowing, knowing how, showing how or doing should be reflective of and steeped in the clinical practice context (p. 313).

Skill education will ordinarily lead to performance, where practice (be it guided or autonomous) is aimed at moving a student from being able to *show how*, to *doing* the skill. This is analogous in some ways to the ladder of competence described earlier.

When a student is asked to retrieve information, the classification of that information is challenged and refined to be made more accessible (Karpicke, 2012). Social learning theories focussed on being the clinician, rather than doing clinical tasks would predict that when a student is asked to recall the information in the context of a professional role, learning is reinforced with reference to the contextual triggers to which it relates.

2.3 The four-stage approach (4SA)

Having considered the key literature around learning psychomotor skill development, I will now consider the hypothesis that 4SA is better than 2SA. First, 4SA will be described in greater depth, and then claims made by its developers and other scholars about its theoretical foundations will be examined, before a further critique of the evidence and literature.

The four-stage approach to teaching considered in this thesis consists of the following four stages (Walker & Peyton, 1998):

> Stage 1: teacher demonstration in real time
> Stage 2: teacher demonstration with explanation at each step
> Stage 3: student talks through the teacher through each step of the procedure while the teacher performs it, and
> Stage 4: student performs the skill, with verbalisation at each step.

2.3.1.1 The first stage

The first stage provides context for each sub-part of the skill. This is a similar rationale to the use of simulation in learning, as a means of connecting the segments of the practical clinical event. The Resuscitation Council UK (2008, p. 28) likens this stage to allowing the student to be a "fly on the wall", providing a visual representation of what the skill is like realistically. This gives the learner a "strong visual imagery which shapes learning". An advantage of this step is the opportunity given to the student to develop an overall picture of the skill from which to build their understanding in Stages 2 to 4 of the education session (Bullock, 2000, p. 141). In the same way that the theory-practice divide is shortened by placing components of clinical practice within a simulation or wider clinical event (see Figure 8), the first step of 4SA is believed to provide the overall context for each step within the skill and in doing so, provides relevance to each part of the skill (represented in Figure 9).

Step 1 of clinical event	Step 2 of clinical event	...	Step 10 of clinical event
Steps 1, 2, ...10 of clinical event sitting together within the context of the simulation as a whole.			

Figure 8: Placing the individual components of the event into the wider context of the simulated encounter. Each part of a clinical encounter gains context within the wider setting of the simulated case, thus bridging the theory-practice divide.

Step 1 of skill	Step 2 of skill	...	Step 10 of skill
Steps 1, 2, ...10 of skill sitting together within the skill/procedure as a whole.			

Figure 9: Placing the individual components of the skill into the context of the whole procedure.
Each step of a skill gains context within the wider performance of the procedure.

Bullock (2000) posits that the visual overview of the skill presented in stage one prepares a contextual foundation upon which the information provided in stage two is processed and stored. Hence, the "acquisition of new ideas and knowledge in stage two" is more pronounced (p. 141).

2.3.1.2 The second stage

Information transmission from educator to student occurs at Stage 2. At this stage, Walker and Peyton (1998) comment, "the trainee comes to understand exactly what is required" (p. 175). During this stage, communication is not exclusively unilateral, as the student has the opportunity to ask questions. The explanation provided in stage two lays the groundwork for a cognitive understanding of the skill, which is expressed and checked in Stages 3 and 4.

2.3.1.3 The third stage

With two of the four teaching stages now complete, the student has still not touched either the equipment or a patient. The value of this teaching method hinges on the construction of an important cognitive foundation to maximise accuracy and depth of understanding of the performance stage, rather than prematurely progressing to the performance stage where a lack of muscle memory exists, and the student becomes easily overwhelmed and confused by attempting manual dexterity tasks without a solid cognitive base. This step is said to ensure the student's understanding (Walker & Peyton, 1998, p. 175).

During the third stage, responsibility for the skill is partially transferred to the student, and cognitive ability to perform the skill is tested, reinforced and challenged (Bullock, 2000). This is an important stage in providing a firm springboard from which physical practice of the skill may progress. Concurrently, the learner receives visual feedback on their understanding and memory of the skill, and the educator is able to pause, prompt or remind the student where steps have been missed. The educator checks the student's understanding of the skill with questions aimed at checking the underlying knowledge and rationale behind each step. This may promote reflection during learning.

2.3.1.4 The fourth stage

Bullock (2000) mark this as the completion of "transference of excellence from the expert (instructor) to the novice (candidate)" with retention of the skills dependant on practice. A key feature of the original 4SA which is often overlooked in other adaptations and later commentary is the student's articulation of their intended action prior to performing it to ensure safe performance and the opportunity for educator intervention if required (Walker & Peyton, 1998, p. 175).

With 4SA each stage seems to prepare a foundation to strengthen and stabilise each subsequent one, with the pinnacle of the theory being skill performance which is underpinned by firm understanding of the skill (see Figure 10).

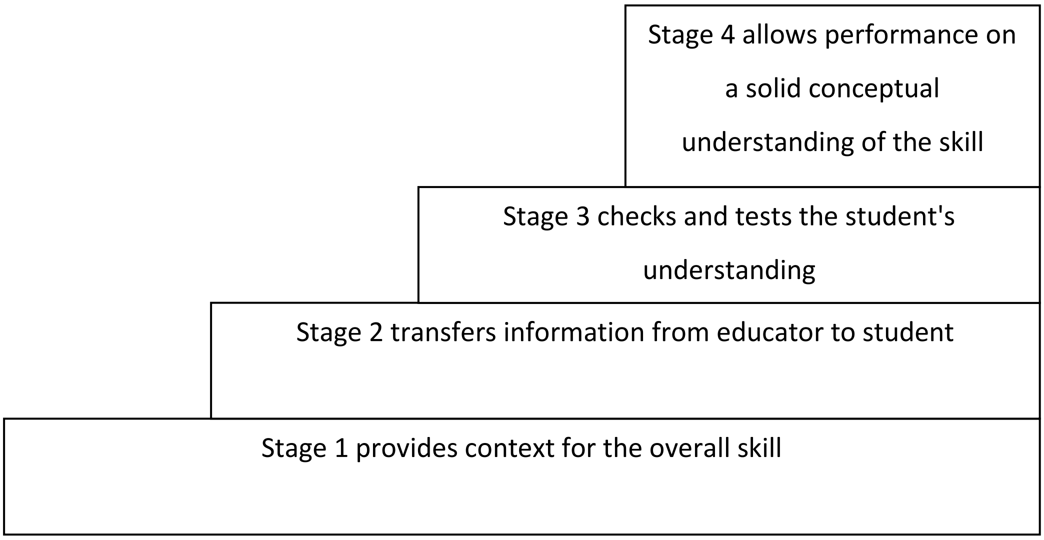


Figure 10: Progressive cognitive foundation argued by 4SA

2.3.1.5 Group adaptations

4SA has been adapted for the group teaching situation by either: selecting a single student to prompt the educator to perform the skill (Bullock, 2000, p. 140); asking the student who instructed the teacher in Stage 3 to be instructed by another student for their manual performance, thereby one student's Stage 3 is another student's Stage 4, until all students have explained and performed the skill (Nikendei et al., 2014); or asking each student in the session to contribute subsequent steps to the educator's skill performance. Lake and Hamdorf (2004) suggest Stages 1 and 2 can be performed in a large group, for example with pre-recorded audio-visual aids. Stages 3 and 4 can then be performed in smaller groups. As 4SA was originally described in a teaching context where the skill is being taught to one student, it is acknowledged that this thesis considers the *adapted* use of 4SA to group settings rather than 4SA as it was originally published.

2.3.2 Peyton and Walker's original claims about 4SA

The context of 4SA as it was first described was for one student and one clinician in an apprenticeship style collaboration within the surgical theatre (Walker & Peyton, 1998, p. 171), although it is presented as "equally valid" for other skills. Peyton states that "There is no one correct method of teaching. It is a dynamic and interactive process. The skill is in knowing when to vary the input" (Peyton, 1998, p. 14), with the later admission that there is indeed one incorrect way of teaching: "The old 'see one do one, and teach one' has no place in surgical training" (Walker & Peyton, 1998, p. 174), and later "A common mistake in teaching is to continue to oscillate between stage two to stage four, missing out on stage three which is one of the most important parts of the process, particularly when it comes to more complex procedures which will be discussed later" (Peyton, 2008).

Peyton is not alone in his argument that the traditional skill teaching approach of *see one, do one, teach one* is insufficient and inherently unsafe. Historically, this approach often involved a simple observation of the skill (without explanation), performance of the skill (possibly without supervision or feedback), and then subsequent teaching of that skill to a more novice peer or student, however in modern medical education this would most likely be adapted to include explanation, supervision and feedback. Limited knowledge of the skill background, rationale, or anatomical foundation is transferred during 4SA, so this may be complemented with prior training or skills laboratory time, outside of the 4SA approach in-theatre (Walker & Peyton, 1998, p. 173). Stage 1 gives the trainee an overview of all the steps, and the final fluency with which to expect a proficient clinician to perform the skill (Walker & Peyton, 1998), as depicted visually in 2.3.1.1. This is said to allow a conceptualisation of the skill, with reference to Schon's experiential learning cycle earlier in Peyton's text (Peyton, 1998, p. 33).

With much learning theory infused into the earlier, supportive section of his book, Peyton (1998) is not explicit about the underpinning theoretical arguments in support of 4SA, however, many reasonable practical arguments are suggested. 4SA is presented within a book which infuses a contextual, philosophical approach to teaching and learning, although this is not considered in other authors' approaches to 4SA. The text in which 4SA first appears presents a student- and patient-centred approach to clinical education. Much of the advice presented seems to build a philosophy of education, rather than a series of strategies to use. 4SA has become a template connected to some aspects of learning theory, however it tends to be advocated in isolation from the philosophy presented more widely in the original text. This may be due to the intentional segregation of learning theory in the initial chapters of Peyton's book, followed later by chapters of a more practical focus (where 4SA can be found). It is 4SA, having been isolated from the underpinning philosophy as it is now referred to in the literature, which will be the focus of this research.

2.3.3 Others' claims about 4SA

In their editorial, Barelli and Scapigliati (2010) refer to the graduated nature of 4SA as a tool to move a student clinician from being consciously incompetent to being consciously competent, and opens the doorway for progress to unconscious competence following further practice. Despite "no evidence of better skill acquisition", Barelli and Scapigliati encourage its continued use until further research contributes conclusive data on its use.

Bullock (2000) presents 4SA as a teaching strategy supported by cognitive learning theory. Through helping students receive and process the information delivered, 4SA may be argued as a technique which allows improved cognitive learning of a skill. Bullock states that:

> ...the current teaching methodology referred to as the 'four stage teaching approach', utilises broader contemporary thinking, and is influenced by models of teaching that relate to how information is processed. Although all attempts to educate are about information processing, some techniques are specifically orientated to developing the learner's aptitude to acquire and operate on information received (p. 140).

Cognition is but one aspect of skill development, however. In considering movement up the ladder of competence, Barelli and Scapigliati (2010) argue that models such as 4SA may aid stepwise progression towards competence. Specifically how 4SA helps students process information is not explicit, although Bullock (2000) refers to knowledge transference from expert (instructor) to novice (student), resulting in changed behaviour. The experiential and situated learning theories may not be seen as relevant for this more cognitive and behavioural perspective on learning, despite their undeniable relevance for a learning approach which focuses on the student *becoming* a clinician.

Lake and Hamdorf (2004) present 4SA as a favourable option to 2SA. They argue that 2SA may fail in teaching or ensuring competence of the skill, but 4SA makes the skill more manageable by breaking it down, allows the learner to vocalise the steps, and provides repetition and more opportunity for correction (p. 327). Lake and Hamdorf (2004) also advocate the adaptation of 4SA. For skills like suturing, a single wound requiring four sutures is necessary to complete the 4SA model, however skills like an Intramuscular Injection (IMI) are more likely to require four separate patients. In the latter case, Lake and Hamdorf suggest that the educator may use a real patient for Stages 1 and 4, but refer to simulation equipment for Stages 2 and 3. Benjamin (2005) advocated for 4SA as a useful teaching technique in the skills laboratory, also citing the risk of surgeons who may assume their trainees have greater ability than they actually have at novice level, and the cycle of learning can be overlooked. This reflects the unconsciously competent practitioner, and may explain poor subscription to 4SA in some clinical education settings. Hamdorf and Hall (2000) present 4SA as a method for teaching surgical skills, but do not discuss its efficacy. In an editorial, Toouli (2006) asserts that watching an expert demonstration is but one part of the learning process. He goes on to argue that 4SA offers a more thorough approach to learning clinical skills than a demonstration alone which, without follow-up teaching, is "entertainment value" at best (p. 164).

2.4 Evidence comparing 2SA and 4SA

2.4.1 Controlled comparative trials - a systematic review

From the perspective of effectiveness, the literature discussed so far could support the prediction that 4SA will be effective in skill learning. 4SA is, after all, more repetitive, and there may be arguable merit behind allowing the student the opportunity to perform cognitive aspects of learning prior to performing the procedure. In order to better understand this, I conducted a systematic review of the literature to understand what is known about the cost-effectiveness of 4SA compared to 2SA. In the review, I was interested in data relating not only to the financial cost (for example teaching wages, consumable resources), but also non-financial costs (for example time required to teach).

I searched eleven databases relevant to clinical education for trials where the acquisition and/or retention of clinical skills was compared for 2SA instruction (specifically Stages 2 and 4 of 4SA), and 4SA. The search strategy used is outlined in Appendix 10.1, including the number of articles found in each database. Outcomes of interest in this review were skill acquisition, skill retention and teaching cost (in terms of time and resources). The search strategy and search terms are described in Table 45 of Appendix 10.1.1. Google Scholar (Google, 2013) was used to search for other studies which reference the studies already identified.

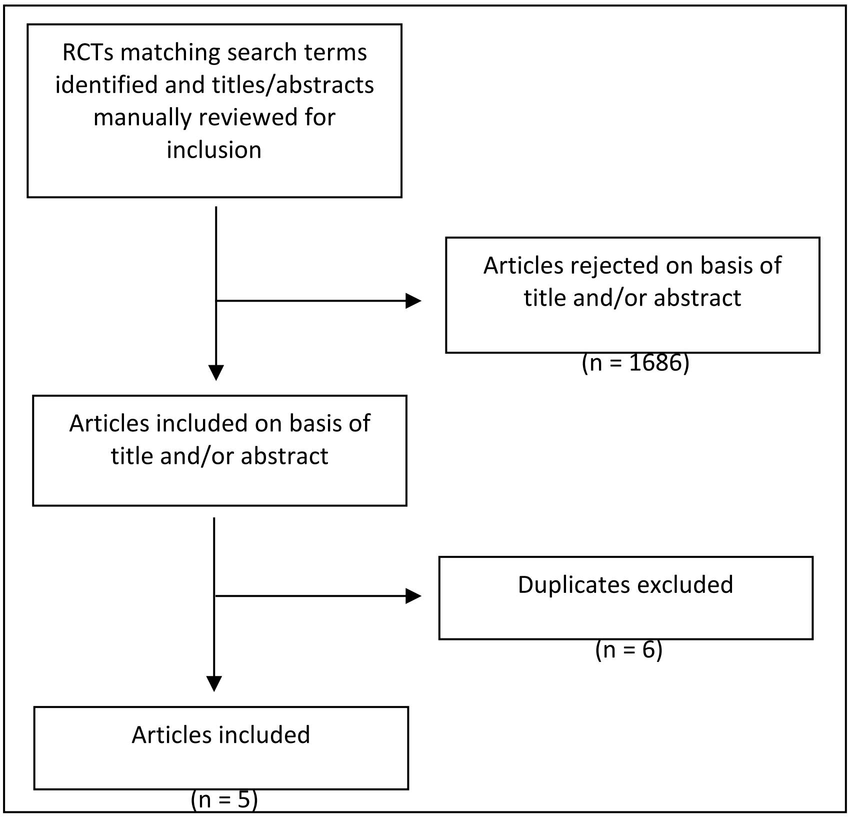


Figure 11: Systematic literature review search strategy results for randomised controlled trials (RCTs) comparing the two-stage and four-stage teaching approaches

Five studies met the stated inclusion criteria (Archer et al., 2014; Bitsika et al., 2013; Jenko et al., 2012; Krautter et al., 2011; Orde et al., 2010), and four other studies were identified which offered significant insight to the question at hand (Greif et al., 2010; Herrmann-Werner et al., 2013; Lee et al., 2007; Murphy, Neequaye, Kreckler, & Hands, 2008).

2.4.1.1 Skill acquisition

Krautter et al. (2011) and Jenko et al. (2012) presented the strongest studies when reviewed using the Critical Appraisal Skills Programme (CASP) checklist for RCTs (Critical Appraisal Skills Programme, 2013). Jenko et al. (2012) report 4SA as statistically superior to 2SA in only one of 14 performance measures (number of chest compressions/minute) with the 4SA students performing closer to the optimal range than those taught CPR with 2SA. Krautter et al. (2011) conducted a study of 34 medical students learning gastric tube insertion. 17 were taught using 2SA, and the remaining 17 were taught with 4SA. The study identified comparable performance (marked as a percentage) in the two groups when scored with a binary checklist (88.1±7.5% for 4SA, and 85.2±11.3% for 2SA, $p<$ .781), but when an Integrated Procedural Performance Instrument (IPPI) and global communication scale rating were used, those taught with 4SA performed significantly better. The most impressive component was noted in communication aspects of the skill, with an effect size of 60 to 100% between the two groups. Time to teach the skills was greater for 4SA (though not statistically significant with 4SA taking 605±65 seconds and 2SA requiring 572±79 seconds, $p<.122$), but students taught with 4SA were able to perform the skill faster (168±30 seconds, compared to 242±53 seconds for 2SA, $p<0.01$). Students taught with 4SA obtained better scores on global rating scoresheets (IPPI scales were used) regarding overall score, technical aspects and communication features, but not for skill-specific binary-style mark sheets. The study also examines the students' acceptance of the teaching model, and researchers report 4SA to be well accepted. Researchers are reluctant to apply these results more broadly to other skills, particularly in other settings or with more complex skills.

Lee et al. (2007) compared performance of students taught Intraosseous (IO) insertion by watching a 10 minute video and 10 minutes of practice in the simulation laboratory (this arm was considered 2SA), with students who were taught the skill with a 20 minute face-to-face 4SA session and facilitated practice. In this interprofessional group, the video supported teaching session was shown to be superior to the face-to-face 4SA (mean score 7.56±1.65 compared to 6.00±1.84, $p<0.01$). Forbes et al. (2016) reviewed the use of video in teaching clinical skills, and found overall agreement that video-supported skill teaching sessions are as effective or more effective than standard face-to-face teaching, and such studies could support an argument to incorporate video media into 4SA as a skill teaching strategy.

Greif et al. (2010) also compared 4SA to a video-supported two-stage approach, in addition to two other teaching approaches based on various permutations of 4SA's four steps. Anaesthesiologists were invited to teach the undergraduate medical students cricothyroidotomy in one of the four teaching methods. All groups were found to perform the skill with statistically comparable times. By the fourth attempt, half of all participants in each group could perform the skill in under 60 seconds, and by the 8th attempt, this increased to 80%. Overall, there was no statistical difference found between the four groups. The authors, however, were careful not to completely reject 4SA for clinical teaching:

> The 4-stage approach was only tested on one skill in one specific setting. Further investigations in more complex skills under different settings might show the superiority of the 4-stage approach in clinical skill training. The 4-stage approach is still a well-structured, interactive teaching method based on sound educational theory. If there is enough time for teaching with it, the 4-stage approach might be applied for skill training, but we still do not have evidence that it is superior to other skill training methods (p. 1696).

Lund et al. (2012) conducted an RCT comparing how professionally (technically and with reference to communication) students performed IVA when they undergo skills laboratory training using 4SA with those who experience bedside training (2SA in the authentic clinician-patient environment). The study reported that 2SA took on average 8.4 minutes less to teach than 4SA, but that this was not statistically significant (*p*= .065). Students taught with 4SA felt their teaching session was less authentic than their peers who participated in the bedside teaching session, however the 4SA trained students reported more confidence that they would remember the steps to the skill. 4SA students also stated greater benefit from the feedback during their session. Students who learned IVA with 4SA took as many attempts as those taught with 2SA to insert it correctly, but they demonstrated more successful attempts, less time required to perform, significantly better IPPI ratings, better technical skills, better communication skills and generally higher competence in overall skill scoring. In a binary checklist, 4SA students performed more steps correctly (Lund et al., 2012, pp. 6-7).

Orde et al. (2010) conducted a study comparing LMA insertion into manikins with students taught with 2SA and 4SA. The outcome measured in this study was the number of students in each group who were able to successfully insert an LMA into a manikin within 30 seconds. This method of measurement becomes important with time-critical skills such as establishing a patent airway. Time to reassessment of LMA performance retention varied from 24 to 147 days, according to participant availability. Upon re-performance, time to insert the device increased and the number of students able to perform the skill under 30 seconds decreased (in both groups). The group taught with 4SA had a significantly lower number of actions not performed (*M*=0.31 for 4SA compared to 0.64 for 2SA; *p*=.02), but no other differences between the groups were statistically significant (including proportion of students completing the skill within 30 seconds, time taken for insertion, proportion of students completing the skill without error, number of actions performed incorrectly, and proportion of students completing the skill both within 30 seconds and without error).

2.4.1.2 Skill retention

Orde et al. (2010) found that 4SA produced superior retention in only one of six criteria with no difference observed in the remaining criteria. Two further studies reported no difference in retention between the two teaching methods (Archer et al., 2014; Bitsika et al., 2013). Bitsika et al. (2013) compared student acquisition and retention (at 45 days) of intravenous cannulation skills of nursing students taught with 2SA and 4SA. While some differences in initial performance were noted (students who performed IVA without any errors did so faster in the 2SA cohort than the 4SA cohort, overall performed fewer steps incorrectly, however omitted more steps than the 4SA cohort), no differences were noted just 45 days later.

Archer et al. (2014) compare manual defibrillation performance on manikins by medical students taught with the 2SA and 4SA. A third cohort was taught with a 5 step approach which included a preliminary step of outlining the context in which the skill ought to take place. Immediately following the teaching, students taught with 2SA out-performed those taught with 4SA (76.6% compared to 73%, *p*=.02), however no statistical differences were found during the re-test two months later (*p*=.46).

Herrmann-Werner et al. (2013) compared 2SA with 4SA in the simulation setting, with nasogastric tube (NGT) and IV Cannula (IVC) insertion. For both groups of students, a thorough explanation and demonstration was provided in the skills lab, however for the 2SA cohort, practise was not permitted prior to assessment. The 4SA cohort experienced the skill teaching in a scenario and role play, performed the skill and received feedback, so there were a number of additional variables beyond the teaching strategy. Immediate and retained performance (at 3 and 6 months) appeared superior for 4SA for both skills.

2.4.1.3 Teaching cost

No studies explicitly measured cost, but three considered the time required to teach. Orde et al. (2010) reported 4SA to take approximately 50% more time than 2SA, but did not undertake analysis to detect whether this difference was statistically significant. Krautter et al. (2011) reported no significant difference in time to teach between the two methods. Lund identified no additional teaching time was taken to teach with 4SA (Lund et al., 2012).

2.4.1.4 Overall critical analysis of the evidence

In the five studies identified in the review, there was limited, and conflicting, evidence of effectiveness and no data on which to make judgements around the cost-effectiveness. All studies addressed a clearly stated research question with randomised allocation of subjects, however there were some methodological issues with the identified studies. Common limitations I identified include the absence of baseline skill performance measurement, and the resulting assumption that initial skill ability is negligible. The assumption that the initial skill base is negligible results in the attribution of all acquisition and retention to the teaching intervention. Other limitations included variable time delay between initial reacting and retention reassessment, potential bias due to poss of participants to follow up, unequal opportunity to practice the skill, and variation between different educators were identified.

4SA was developed with apparently sound andragogical basis. However, evidence from these studies does not consistently support the use of this approach which would appear to take more time and therefore add to the cost of skill teaching. The learning theory mentioned will therefore be more critically examined in Section 2.4.3.

2.4.2 Non-RCT evidence:

Having considered the current empirical evidence comparing 4SA and 2SA, I will now consider the wider reference to advice surrounding 4SA in the literature.

2.4.2.1 Documented benefits of 4SA

2.4.2.1.1 The students like it

Positive verbal feedback from the "efficacy of the skill stations and scenario practice", coupled with "significant knowledge and skills gain" were reported following 4SA implementation when teaching BLS skills to dentists and dental care practitioners (Balmer & Longman, 2008).

2.4.2.1.2 4SA is a structured approach

It is not the teaching technique which breaks down a skill to be taught in a more manageable way. This idea sometimes referred to as *chunking* relates to delivering limited packets of information and is not an explicit concept within 4SA, nor is it exclusive to 4SA(Owen & Plummer, 2002). Indeed, 4SA may offer to an educator the insight that a skill is too complicated to be taught in a single session before the performance stage.

2.4.2.1.3 Skill complexity considerations

While Orde et al. (2010) measure no statistical benefit from teaching with 4SA, Barelli and Scapigliati (2010) ponder possible benefit using this technique for more complicated skills rather than the relatively simple skill of LMA insertion (with the exception of skills which are "too complex"). It is unclear what is meant by "too complex". Indeed, a skill's complexity relates not only to the skill being performed, but the experience of the provider, baseline knowledge of the provider, the environment in which the skill is performed, the provider's physical ability to perform the skill, performer's age, etc. LMA insertion is quite complex for a novice who has never before managed an airway in an unconscious patient, but a relatively simple skill for an anaesthetic registrar who has performed the skill a thousand times (Barelli & Scapigliati, 2010).

2.4.2.1.4 Student self-ratings and attitude

When 4SA was used to teach emergency medical care to dentists, improvements were evident in participants' attitude towards these skills (Sopka et al., 2012) in addition to an improvement in perceived ability to perform the skills safely (p. 182).

2.4.2.1.5 4SA easing the modern medical educator's workload

In an editorial published in the Journal of Surgical Education, Stevens and Davies (2012) offer 4SA as a solution to the difficulties modern workload demands have placed on the apprenticeship model of 2SA in medical education (p. 135).

4SA is presented as a simple skill teaching model (Wall, 1999), which may be used for remedial students (Thomas, 2012), specifically medical students learning suturing skills. Wang et al. (2004) also measured efficiency of 4SA using suturing as the skill. Overall time to perform the skill was measured, and students' performance was compared against their initial performance (baseline). This study identified a 52%, 22%, 7% and 6% improvement in scores for fourth year medical students, and first, second, and third year dermatology residents respectively. All improvements were statistically significant (p=.01 to .04) with an overall improvement of 24% (p= .001). Interestingly, the post instruction score for fourth year medical students (57.6% ±1.0) was comparable to the post instruction score of the third year residents (58.9%±0.4). Wang et al. (2004) indicate that the actual learning may be different for the two groups (assuming different baseline). This study argues the need for a homogenous group on entry, with assessments performed to identify baseline skill level, and identify those who claim to have learned the skill before. It goes on to note significant improvement of the taught skill when taught with 4SA, and this is seen more powerfully for novice practitioners than relative experts. This study is unable to offer comment on the effectiveness of 4SA as a teaching method compared to other strategies, as there was not a control arm. The study noted no significant change to the time taken to perform the skill before and after instruction (Wang et al., 2004).

2.4.2.2 Documented costs of 4SA

Orde et al. (2010) notes criticism 2SA has received, specifically in terms of “inadequate skills acquisition and retention”. They also report 4SA to take longer to teach than 2SA, citing Bullock (2000) (though the cited paper does not measure time to teach). Barelli and Scapigliati (2010) also comment that 4SA is relatively time consuming (p. 1607), but it is unclear where this comment is validated. It is later commented that Orde et al. measure an increased teaching time (3 min for 2SA compared to 4.4 min for 4SA, p. 1607) but the initial study reports no statistical significance for this difference. One of the common assumptions this identifies is that 4SA takes longer to teach, and while this remains to be established (Greif et al., 2010), it is an acceptable assumption. The impact of this cost on teaching organisations is still unclear in the literature.

Wearne (2011) identifies current criticism of 4SA, including increased time requirements, repetitive methodology, lack of account for learner's previous knowledge, and the possibility it can give the illusion of competence.

From the literature, 4SA is predominantly advocated in the setting of resuscitation skill teaching, therefore it seems appropriate to compare 2SA and 4SA in a professional group such as paramedics who are most likely to use resuscitation skills regularly which has not yet occurred. Additionally, there are currently no studies which have addressed this question with reference to baseline skill performance to measure existing knowledge. Furthermore, much disagreement in the studies identified in the section above. Possible explanations for the lack of clear coherence between the expected outcome (based on the learning theory around 4SA) and the experimental findings identified may lie in a misunderstanding of the learning literature used as a basis for 4SA.

2.4.3 Further interrogation of the learning literature specifically relating to 4SA

In terms of the cognitive demands of learning, 4SA is not directly aimed at decreasing the cognitive load. Miller's "magic number seven" (G. A. Miller, 1956) inherently supports the importance of measuring a participant's existing knowledge or topic familiarity prior to the teaching session. Aside from allowing the student to connect the new learning with previously stored knowledge, this also means that if they begin with, say five pieces of knowledge (which a student may need to hold in their memory during learning at a potential cost to new information), the new learning can be built around that. Walker and Peyton (1998) address this to an extent, with the suggestion that after Stage 1 the student is asked if this is "the precise technique they have employed in the past and are comfortable with". If not, and the trainee or educator consider the differences to be too significant, the remaining three steps are advised to be carried out during the teaching session (p. 177). This ensures that the "one safe method" (p. 172) is taught, although there is a brief mention to approaching a skill session such as a surgical procedure in smaller stages to allow for comprehension prior to proceeding, this is not built into the 4SA or any description of it.

Decreasing the germane and extraneous cognitive load for the learner potentially creates more cognitive availability for the intrinsic load (Sweller et al., 1998; Whelan, 2007), and using 4SA to teach could be a means to achieve this. Providing a real-time overview of what the skill will look like once proficiency is achieved (Stage 1) is said to provide an overall context for the individual components of the skill, and may therefore allow the student to build some cognitive schema to organise the individual components of the new knowledge, or identify which areas of pre-existing knowledge the new aspects can be connected to, thus decreasing the germane load (Nicolis & Tsuda, 1985). The extrinsic load may be lessened through different means of teaching; the student sees the skill, then has it explained to them, then checks their cognitive understanding, then performs the skill. This stepwise approach takes the student smaller steps each time, rather than from being unconsciously incompetent to expecting performance after only a demonstration. Though an argument could be made based on different "learning styles" such as kinaesthetic, auditory and visual learning, despite the popularity of such theories, it will not be entertained here due to the lack of evidence in educational literature.

Observation of the skill is important for the recruitment of mirror neurones even without performance, and 4SA as a teaching strategy presents more opportunity to achieve this than 2SA. During 4SA, the student has three opportunities to see the skill occur: once at real speed, once slowly, and once in response to their own direction. The student being taught by 2SA in this session would have only one opportunity to watch the skill. Therefore, the literature suggests that the student who is being taught with 4SA may have more profound activation of the neurons required to pre-empt the motor function relating to the skill than the student being taught with 2SA. Likewise, the student taught with 4SA will have heard a description of the procedure twice, compared to once for the 2SA student (once in Stage 2 from the educator, and once in Stage 3 from themselves). This, Tettamanti et al. (2005) predict, will increase cerebral activity in the motor cortex relating to the actions stated, even if that action is not occurring.

Separating the cognitive and motor aspects of the skill has some credibility for the purposes of analysis. However the finding that motor neurons are pre-empted through visual and auditory inputs (Cattaneo & Rizzolatti, 2009; Tettamanti et al., 2005) suggests that the learner's neuroanatomy creates pathways for motor function before their hands even touch a medical instrument or patient. This could be seen as the body's natural priming of motor pathways which the brain is preparing to involve in learning for assumed maximal efficiency, and therefore restricting, or not encouraging this integration early in the learning process may not recruit the maximal cerebral response. Zhu et al. (2011) determined that different teaching strategies can activate different neural pathways, and even when these are not measurable in terms of performance of a manual skill, these "nonessential" coactivations may impact other aspects of related skill application.

In 1964, Fitts (1964) heavily criticised the separation of manual dexterity and cognitive understanding as artificial and without purpose, citing the "processes which underlie skilled perceptual motor performance" as "very similar to those which underlie behaviour as well as those which underlie language behaviour as well as those are involved in problem solving and concept formation ... [therefore] no advantage would result from treating motor and verbal learning as separate topics" (p. 243). That said, Fitts separates the two to an extent, with discussion of the *intermediate phase* of skill learning including the co-recruitment of both motor and verbal processes (p. 263). This relationship (between recognition of a symbol and stating the letter with which it begins) may be criticised in its application to justify Stage 3 of the 4SA as the verbalisation *of a letter* rather than verbalisation *of understanding* and therefore consolidation of procedural understanding. Fitts (1964) identifies the term "skill" as one which refers not merely to completion of a task, but to an action where "receptor-effector feedback processes are highly organised, both spatially and temporally." This definition speaks to the involvement of muscle coordination, automaticity, perception, which are argued to be different processing mechanisms (p. 244) to those involved when a skill is being performed correctly but in a clinicians early stages of skill expertise development.

There is profound agreement in the literature that expertise requires practice to obtain and maintain. This highlights an inconsistency in the language around learning clinical skills. Where skill implies expertise, and expertise is a highly regarded fine-tuning of ability, dependent on fine motor ability and control, knowledge application, and professional standards, we soon see that in a single session, no one is likely to leave the session *skilled* in practice, or with skills. Dreyfus' model of skill acquisition (Dreyfus, 2004) helps identify the way we tend to use the term "skill" differently. Dreyfus' model identifies different competence levels, such that in a skill session an educator would typically aim for a novice level of ability, rather than a skilled practitioner. So when we teach medical skills, we are not aiming for that student to exit the session as a *skilled* person, but rather one who can adequately perform an action (or series of actions) under supervision, which will germinate into skill through intentional practice and reflection.

Retrieval is used in 4SA more than for 2SA, so it would be expected that recall is superior for students taught with 4SA to some extent as a result (Karpicke, 2012; Karpicke & Roediger, 2008). Where this retrieval occurs may impact on recall, however, with situational learning theories such as Billetts crediting rich social learning to the authentic environment. While the clinical environment of the operating theatre was the context initially proposed for 4SA to be used, it tends to be advocated now in skills laboratory settings which may seek to simulate the authentic clinical setting, but will encounter limitations in this. Hence, the social aspects of learning will be limited and impacted by the setting, rather than the strategy. With reference to Raph Morgan's work, Newton, Billett, Jolly, and Ockerby (2009) argue that:

> simulation offers the prospect of creating an environment that could require translation. The extent to which this can happen is premised upon the congruence between the lab experiences and what is practised in the clinical setting as well as the sequencing of the experience relative to the university curriculum (p. 3).

Learning a skill in isolation from its clinical context is argued to be less effective, on the basis of social learning theory, so the importation of the value of the practice context will depend on the educator, as a clinician. The increase of simulation in clinical education testifies to educators' agreement to this, however barriers exist to achieving this "higher fidelity" education space. Such barriers include competition for clinical placements, time, space, training equipment and inadequately trained simulation facilitators (Brady et al., 2013). Incorporating some authenticity into clinical skills education, however, can be achieved on a much less global scale than a full case simulation. Experiential learning involves observation of the skill, but hinges on an experience of attempting it, reflecting, conceptualising the experience, forming strategies for adaptation, and actively experimenting based on the above, to complete they cycle again. Theories such as that published by Kolb are widely accepted in medical education, however in 4SA there is little to no opportunity for the student to self-reflect (though they may reflect on feedback given by the educator), and adaptation based on reflection and abstract conceptualisation is not present in the method. By Peyton's own admission (Peyton, 2008), 4SA is about "certainty" in practice, not skill adaptation or exploration. As such, it provides the learner with some experience in the skill, but this cannot be considered true experiential learning when a single skill application strategy is the aim.

Some authors attribute each step of 4SA to a level on Miller's pyramid (G. E. Miller, 1990), and at first thought this appears to be the case, with a progression from knowing to doing throughout the four steps. When we reconsider Miller's triangle, however, it is important to recall two aspects: the triangle (or pyramid) is of primary reference to guiding assessment in medical education, and has since been adapted to guide education; and the pinnacle of the triangle, the doing, is subtitled action, but Miller intended that it refer to the contextual practice within a setting of professional expectation "when functioning independently within a clinical practice" (G. E. Miller, 1990). The *doing* is an integration of performance into practice, which is embedded within a professional context, and in that regard, no once-off skill teaching method is likely to be appropriate alone. Some claim that 4SA moves the learner from unconsciously incompetent to consciously competent, but this too is a significant assumption, dependent on the meaning of competence. Will a student become competent in a skill following a single training session?

2.5 Theory of learning

In this review of the literature, I have addressed aspects of cognitive learning theory (how we come to know), behavioural learning theory (how we come to do), and social learning theory (how we come to do in context). However that is not representative of my initial approach to learning. The context in which I have engaged in clinical education and practised pre-hospital care is one which had a much more prevalent focus on knowing and doing. This infused the early stages of my research question, outcome measures of interest, and study design. Within this approach to learning, I understood learning as an observable change in behaviour (applied action or ability) as impacted by the educational input or stimulus.

During the course of the first two comparative trials (described in Chapters 3 and 5), this approach to learning is reflected in the study design which aimed to compare the outcome of two different teaching stimuli (2SA and 4SA), although in Chapters 6 and 7 this is challenged by further data and analysis. The validation of the assessment tools developed in Chapter 4 is discussed in Chapter 6 based on the data gained in Chapter 5. When examining the assumptions inherent in any assessment tool, I recognised that what constitutes learning became much less tangible and measurable than I had previously accepted. This is supported by Zhu et al.'s study which found that cerebral processes which were different in two learning groups were statistically different, but the accuracy of the output was not measured to be different (Zhu et al., 2011). Zhu et al. (2011) also argues that expert and novice performance can be identified by the extent to which coherence between motor planning and verbal analytic regions of the cerebral cortex is observed on an EEG. The final study in this thesis sees a transition from learning being understood as a change in behaviour or applied ability, towards learning as embedded within a social (professional) setting, involving a changed ability due to adaptation of the student's professional identity formation. Professional learning involves an adaptation of being and identity, which will manifest in an approach to which skills and ability are attached.

By the end of this thesis, I will argue that inferring an extent of learning from a finite action is flawed. Educators grapple with their dual roles as they seek to inspire, help students perceive, and facilitate problem solving. These responsibilities culminate in Chapter 7, through understanding both the teacher's perceived identity (and related roles), and the student's identity. This impacts what the educator's teaching aims. An evolution from learning being represented by an objective action, to learning involving a socio-professional challenge to identity will be evident. Learning impacts who we are, which then impacts what we do, and why and how we do it. Likewise, expert teaching comes from who we are as educators, not simply what we do.

2.6 Theoretical framework

Having explored current learning theory relevant to clinical skills education, and how 4SA may align or be challenged by this, grounding this study's development, processes and findings in current educational theory became a priority. Worley's model of symbiotic clinician education (Worley, 2003) was therefore used as a theoretical framework. The model will now be described and its application to the study's outcomes explained.

2.6.1 The symbiosis model description

The notion of *Symbiotic clinical education* (Worley et al., 2006) aides an understanding of the student's context within a wider collection of key stakeholders in the clinical education setting. Among other things, the concept of symbiosis provides direction for a holistic understanding of the costs and benefits of clinical education within the context of a complex and dynamic system of multifaceted demands (Prideaux et al., 2007, p. 114).

The framework is constructed from four interconnected axes (see Figure 12). The clinical axis focuses on the student at the centre of the clinician-patient relationship. The institutional axis considers the demands of both the health care provider as well as those of the research/education institution. The social axis considers the needs of the community in conjunction with the responsibilities of the government (or funder) in providing appropriate health care to meet such needs, and the personal - professional axis aims to address the student's personal considerations (such as moral and ethical values, and self-care) with the responsibilities and identity of the professional role. The student lies at the centre of all of these relationships. Clinical learning activities can be considered in light of these to enhance the mutual benefit to each party in a given axis.

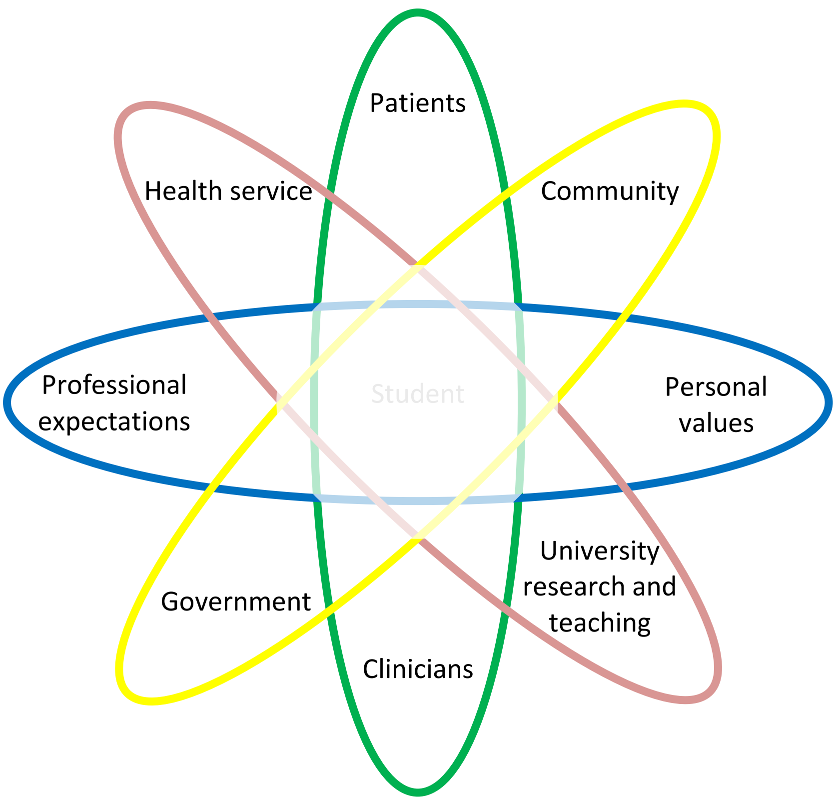


Figure 12: Worley's model of symbiotic clinical education. Adapted from "The immediate academic impact on medical students of basing an entire clinical year in rural general practice" (2003) by Worley, P. S., p. 230.

2.6.2 Symbiosis

The Framework of Symbiotic Clinical Education provides a context for assessing the costs and benefits of skills education in a holistic way, positing that all clinical education sits within this set of four relationships. The framework recognises that there are inter-relationships between the different axes in the system and explicitly identifies the nexus of education and health service provision, making it highly relevant to this research program. The symbiosis model is a series of relationships which recognise the student's place in the clinical education system. The model suggests that there can be inherent tensions between the needs and perspectives of each party in each relationship axis, and, when the presence of the student results in turning these tensions into symbiotic benefits, successful learning is more likely. Likewise, if the presence of the student results in an imbalance of benefit, then successful learning is less likely to be sustained.

2.6.3 Clinical axis

The clinical axis places the student into the patient-clinician relationship. This allows the clinician to model to the student, and may start out with a larger amount of input from the clinician. As the student develops, though, the clinician is able to give more and more autonomy to this developing professional, and the clinician's workload is reduced whilst still meeting the patients' needs. If an initial investment is willing to be made in a helpful way, the presence of a student can be a real benefit to the clinical workforce. Walters, Prideaux, Worley, and Greenhill (2011) describe an evolution of preceptor models to account for the growth in students' independent and autonomous ability in the general practice setting. Where the student begins as an observer of the doctor-patient relationship, they eventually become more involved in patient care and management with the supervising doctor first modelling, then orchestrating, and finally advising this relationship as the student gains ability and trust. Eventually, the student leads the consultation, with support from the supervisor when required (p. 460).

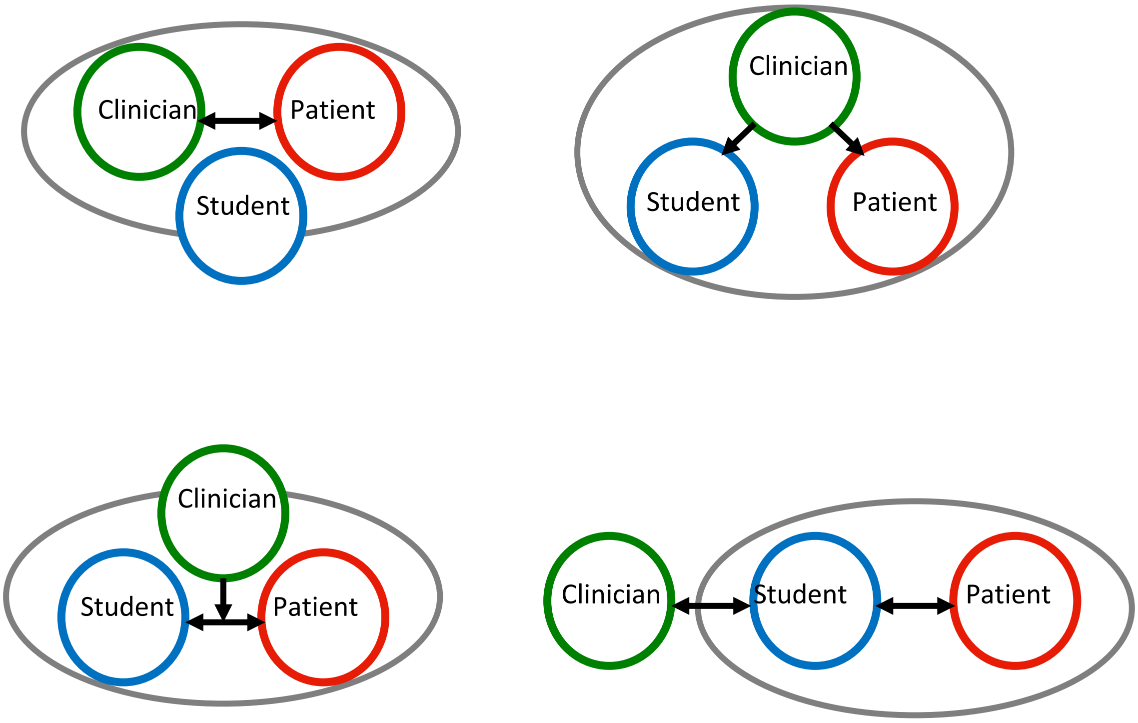


Figure 13: Walters' models of preceptor-student relationships: Top left: The student-observer model places the focus on the patient-clinician relationship, with the student observing. Top right: The teacher-healer model places the clinician as expert, with interaction between patient and student highly reliant upon this expertise; Bottom left: The doctor-orchestrator model allows the student and patient to develop a more authentic relationship, with guidance from the clinician where necessary; Bottom right: The doctor-advisor model allows the student to manage *their* patient relatively autonomously, with access to the clinician if necessary. Adapted from "Demonstrating the value of longitudinal integrated placements to general practice preceptors" by Walters, L., Prideaux, D., Worley, P., Greenhill, J., *Medical Education*, 45 (5), p. 461. Copyright 2011 by Blackwell Publishing.

In a symbiotic model, students are involved in the care of relevant patients, facilitated by enthusiastic teachers who support the student-patient relationship. Exposure to patients in an authentic setting allows students to develop illness scripts of various medical symptoms and disease (Worley, 2003).

2.6.3.1 Institutional Axis

This axis reflects the relationship between the demands of the health service and the teaching institution (Worley, 2003). The health service provides a context for authentic learning, and in return, the developing student becomes more and more a member of the healthcare team, alleviating the workload of the team as a whole. Thus, the student feels that they are making a worthwhile contribution to the team, while developing their practice under guidance.

The teaching institution may be a training organisation, private facility, university, Technical and Further Education (TAFE) or another which provides teaching support to the student, even if they are based in a regional clinical context. This support may be achieved remotely. Rural placements are a key exemplar of this relationship, as a health workforce skilled and experienced in the unique dynamic of rural life and work is required for adequate health provision in isolated areas.

The placement of the student in the centre of this relationship allows the student to experience a variety of cases within the health service, and to apply the knowledge and science taught by the education organisation, and the placement of tertiary students in a health practice may give local community approval to that service (Worley, 2003).

The student is a member of both the teaching and research institution, which strives to teach current, evidence based practice, as well as contribute to the body of knowledge through research. In the health setting, the student is able to see the implementation of this care, as well as bring memories of the patients' faces to research activity as a context and motivator for new evidence. This axis addresses the intersection of science and practice, for which problem-based learning (PBL), authentic learning and experiential learning have gained a foothold in modern medical education over the last 10 or so years. A strong relationship in this axis allows the student to bring the science to the art of practice.

2.6.3.2 Social Axis

The social axis places the student amongst the needs of the community and the government. Where a student is placed in the community to integrate learning from the teaching institution into health care application, the needs of the community emerge (for example through demographic and epidemiological trends or socioeconomic patient factors). The government plays a role in addressing the demographic, geographic, cultural and epidemiological needs of the community through policy formation, funding allocation, workforce planning, implementation of initiatives and prevention strategies. The health service and teaching institution have a responsibility, when interacting in the community, to engage consistently with such government policy and assist future strategies. In this way, students may become advocates of a community's needs, and their presence will demonstrate the impact of government policy for the community.

This axis moves the student from a focus on the individual patient (in identifying a cure or treatment plan), and towards a focus on the social and community determinants of health (with an intent for health education, prevention of illness, and access for potentially marginalised minority groups) (Worley, 2003).

2.6.3.3 Personal Axis

Finally, the personal-professional axis speaks to the relationship between the student's personal morals, values and expectations, and the professional demands and identity associated with their future role. As ethical dilemmas arise in the clinical placement setting, the student's personal values must be addressed in line with assumptions and expectations about clinical practice and its impact on the clinician as a person. Where the professional demands conflict with a set of personal values, the relationship between clinician and student takes on a mentoring role, beyond simply that of the clinical mentor (Worley, 2003).

The issues at play within this axis may have a marked effect on the student's career aspirations within health professions, as it encompasses deeply personal factors such as managing clinical error, work-life integration, and the emotional growth of a person who is becoming identified by their chosen profession.

2.6.4 Applying Symbiotic Clinical Education to this research

The relationships between the different parts of the four axes is "as important as the quality of each individual factor" (Worley, 2003, p. 229). When we consider a skill-teaching strategy within this model, we very quickly see that as the axes are described, an isolated strategy is just that: isolated, and therefore not integrated within this series of relationships. The underpinning philosophies supporting the anticipated successes of a truly symbiotic clinical education model may be less obvious in the micro level teaching session. Therefore, I have adapted Worley's model in order to understand who/what the key stakeholders are in a clinical skills teaching program, rather than understanding how this aligns to the rich relationships in the whole of clinical education as intended by the original model. Three axes feature most prominently when addressing the question of cost effective clinical skill education: The clinical, social and institutional. These axes are considered below with a consideration to the key outcomes of the study, namely teaching effectiveness (in terms of skill retention), teaching cost (in terms of teacher comfort and time required), and student confidence relating to skill performance. In this study, I use the theoretical framework of Symbiosis to assist in organising components of a cost-benefit review of 2SA and 4SA. It allows me to consider the research question in a more holistic approach than reporting only on a monetary cost of education.

2.6.5 Clinical axis (Clinician - student - patient)

2.6.5.1 Patient considerations

Student performance and confidence are both integral to patient care. With improved skill acquisition and retention, the clinical care provided to the patient is improved. This may aid recovery, prevent complications, and reduce mortality. The patient benefits from improved skill retention as improved performance should lead to decreased mortality and morbidity. When a student is taught a rarely practised skill and assessed as competent, assumptions may be made about ongoing competence, however competence in a skill session, observed structured clinical examination (OSCE), or during isolated practice does not guarantee appropriate skill application in other clinical situations. Skill takes intentional practice to develop and to maintain. Where new clinicians are tasked to work in settings where little to no supervision occurs, or workload is too high to seek intentional skill maintenance, or too low to stimulate authentic skill application and practice, skill expertise may not be refined and consolidated. If education organisations could gain insight into whether skill retention differs between two skill teaching methods, this could impact both the patient and the clinical supervisor.

Increased student confidence can be either a benefit or a cost. Student confidence could improve the rapport between the patient and student clinician, and increase already impressive rates of patient permission to student presence and interaction during consultations, procedures and other clinical encounters. Student confidence which is disproportionately higher than their ability becomes a threat to the patient however. This places increased responsibility on the clinician to identify such circumstances and manage them in ways which cares for the emotional wellbeing of the patient and their family members, and promotes insight and learning for their student.

2.6.5.2 Clinician considerations

Student performance impacts on the clinician's role in relation to teaching effort and student contribution to the clinical workload. The clinician has an interest in providing skill training which results in maximal skill retention (or in hosting students who have received such training) as this will result in less corrective training to obtain competency, and decrease the need for supervision. This will save time and other resources. The student who is optimally competent with skill performance may have greater cognitive availability for other areas of learning during a clinical experience or placement. This hypothesis is consistent with Fitts' model of autonomous skill performance with associated multitasking ability, such as walking and talking (Fitts & Posner, 1967a, p. 14). Skill performance and retention is a relevant aspect in considering medical liability issues within clinical practice. Improved skill performance leaves less potential for medical misadventure in the student placement setting, which may provide greater protection from malpractice claims for hosting clinical settings.

Student confidence and insight into their performance is also a consideration for this relationship. The student who has maximal confidence at performing clinical skills may be a potential threat in the workplace. Studies have already demonstrated that student perception of confidence does not correlate to actual performance ability (Dunning, Johnson, Ehrlinger, & Kruger, 2003; Wayne, Butter, et al., 2006), and this student may be less aware of their own limitations. Where student confidence aligns with performance ability, their accurate insight becomes a comfort to the clinician and appropriate learning plans can be more easily put in place. Student confidence should not be considered in isolation from actual capability.

The time that skill training requires impacts directly on the clinician. Between clinical, educational and research commitments, the clinical educator's time is at a premium. The clinician has an interest in saving time on potentially laborious teaching methodologies unless there is an established benefit. The ease of a teaching strategy impacts on this relationship also. If for a moment we consider teaching as more than just a simple transference of knowledge, but as an *art*, we recognise the educator as an artist. When we consider a comparison of two teaching methods, does a restriction to either of these methods restrict the artist's craft? It may be that forcing educators to teach with a 4-stage teaching technique rather than a more traditional 2 stage technique creates stress for the educator who has had their natural inclinations bound. But at the same time, a skilled evidence-based educator may choose to clip their own artistic wings and conform to a tested 4-stage dance if the evidence builds a compelling basis for its benefit to other stakeholders.

2.6.6 Institutional Axis (teaching/research institution - student - health service)

2.6.6.1 Health service provider

The health service has an interest in determining the most effective way to teach clinical skills because it has a duty of care to its patients and other staff who may be affected by poor performance. The health service carries less risk of litigation when students and employees practise with a higher skill level as the likelihood of complications due to poor technique decreases. Students and clinicians who practise with poorer skill performance (for example due to low acquisition or retention) may place increased stress on other members of the clinical team due to fear of an adverse medical event. This in turn carries increased risk of stress leave, sick leave, burnout and fatigue.

Conversely, with improved student performance, receiving a student for clinical placement brings a more reliable team member to the clinical setting which provides additional workforce members. Though, this ought to be measured against the risk of disproportionately high student confidence which could lead to a barrier for learning in the clinical setting. These students may lack insight into their limitations and therefore have a limited capacity to reflect on their performance and accept guidance, and additionally may present to their clinical facilitators as more capable than they are. The Halo effect may be evident where a student performs well in one area (such as patient interaction or communication, for example because they feel confident in their ability), which could lead facilitators to see other practice in a disproportionately positive light (Iramaneerat & Yudkowsky, 2007). The risk to patients is where this biased perception of a student's ability results in greater autonomy than they ought to have.

When clinicians spend more time teaching, this cost is passed on to the health service as time away from patient care and other administrative tasks. The question of cost is also raised in respect to teaching resources. A one-on-one 4SA will logically require twice the disposable resources than 2SA as the skill is performed twice through for 2SA and four times for 4SA (excluding further practice). To date this cost has not been compared in published literature. With the proper planning and curriculum design, 4SA could be split between the two stakeholders in this axis: Stages 1-3 can take place in the skills laboratory in the teaching institution, with Stage 3 being reviewed before progressing to Stage 4 within the health service. This is a truer example of the marriage of two entities of a symbiotic axis, however this particular example may require significant coordination and record control, so any benefit to employing 4SA in this way would first need to be established. Regardless of the skill teaching approach used, this is a strategy to blend both learning environments (the institutional and practical environments) into the student's teaching, allow time to reflect, and place the initial teaching responsibility with the education institution while acknowledging the authentic experiential learning opportunities within the clinical care setting.

2.6.6.2 Education provider

Teaching institutions have an interest in graduating students with the highest competence possible. This bolsters the university's reputation, and generates further income through the prestige gained from widespread knowledge of a university's standard. So, the education provider benefits from being able to provide superior teaching.

Teaching activities carry quantifiable costs in educator time, re-useable and non-reusable resources, and room bookings. These financial resources may be diverted from other areas of need, such as course marketing and educator development. Therefore, a teaching organisation must account for the financial cost of teaching activities in conjunction with its effectiveness.

All teaching organisations (whether they are a formal institution such as a university, or a health service offering clinical placements) have an interest in rapid clinical education where it is efficient; hence both time and effectiveness must be investigated. Financial burdens will vary between different contexts, and likewise the ability to supervise and monitor ongoing performance till vary, so the balance will look different in different contexts. The cost-benefit relationship has not been considered in this way in current literature, but it is of high relevance to the institutional axis.

2.6.7 Social Axis (government - student - community)

This axis addresses expectations placed on the government to provide relevant and appropriate clinical care to the community.

2.6.7.1 Government's duty

Education carries costs, however effective clinical education, manifesting in superior student/clinician performance, will alleviate cost demands on the government by improving the community's health status. Improved skill performance will lead to decreased complications, and reduced mortality and morbidity. (There is the potential that improved performance may conversely result in decreasing the experience of managing such complications, therefore lead to under skilling clinicians in troubleshooting poor performance.) Additionally, time and resource efficient clinical education will alleviate some of the financial support that teaching institutions require from government funding bodies.

2.6.7.2 Community's health needs

Improved skill retention will improve the health status of the community by reducing complications and mortality from poor practice. This will lead to improved quality of life, and less years lost unnecessarily. The improved health of the community also feeds back into the government's gain through decreasing ongoing demand on health care resources, in addition to increasing taxpaying years of its citizens.

2.6.8 Applying the symbiotic model

It is clear then, that there is more to cost-effective skill education than competency and immediate monetary cost of education. The Symbiotic model helps us to understand cost-benefit in a more holistic way by addressing a wider catchment of stakeholders. The axes are not entirely discreet. Even though they are represented in two dimensions, the model is much more multifaceted. Above, I have tried to identify the various considerations of cost-benefit in relation to teaching clinical skills according to the various axes, but it is clear that these axes interact and blend. It is helpful to separate them out as a way of organising knowledge and information, but important to understand that an impact on one stakeholder often overflows to the others.

The monetary (and other) costs of teaching are of high relevance alongside with skill performance and retention. But data comparing these two factors in terms of skill education with 2SA and 4SA has not been published to date. In this thesis, I will compare the costs associated with 2SA and 4SA, in light of data on skill acquisition and retention, and the educator perspective. This thesis will also include a critical analysis of assessment procedures developed and employed.

2.7 Conclusion

When initially comparing 4SA and 2SA it seemed obvious that 4SA would be a more effective teaching method, however there is some conflict and many assumptions made in the literature around the use of 4SA. Many authors promote the adoption of 4SA for the superior acquisition of skills (Barelli & Scapigliati, 2010; Hamdorf & Hall, 2000; Lake & Hamdorf, 2004; Peyton, 1998; Resuscitation Council UK, 2008; Thomas, 2012; Wall, 1999), and with some merit given the *potential* for decreasing the cognitive load required by the learner to organise the learning into schema. Seeing a procedure performed and hearing it described have each been identified as activating neurone necessary for the motor response, and 4SA has a greater repetition of these aspects than a 2SA approach which may lead to the creation of a supportive neuronal framework upon which may be beneficial when the student comes to perform the skill, however delaying the motor function in order for this to occur is not yet established as leading to superior learning. The human body may have a natural inclination to need to involve motor neurons (action output, rather than just action planning) and muscle responses earlier in the learning cycle.

Skill and expertise development are dependent on practice and reflection (or feedback). Kolb calls this "experiential learning" which is more than just performance of the actions of a procedural skill; it also involves intentional reflection and active experimentation (Kolb, 1984, pp. 21-22). Experiential learning, and the context in which it often occurs is the authentic clinical setting. 4SA was originally advocated for this setting, although it has now transitioned to take place predominantly in a simulated training setting. Billett's work on situated workplace learning identifies the importance of context, environment, and the social aspects to learning a set of professional skills and knowledge (Billett, 2009). The 4SA, even in the clinical setting, does not make reference to factors important for developing a professional identity in the clinical workforce through learning. 4SA seeks to move students upward in Miller's pyramid (G. E. Miller, 1990), however evidence of its comparative effectiveness with 2SA is inconsistent. Data relating to the cost-effectiveness of 4SA in comparison to other approaches is absent in current literature.

This thesis opens a new opportunity and paradigm for medical education. It is imperative to gather rich data on the cost-effectiveness of this teaching method, which has come to be used worldwide. Health services who employ clinical graduates, patients, communities, education institutions, clinician educators, and education funding bodies all have an interest in not only effective, but cost-effective education of tomorrow's practitioners. We need to understand the financial and clinical impact of teaching methods in reference to the retention of clinical skills, with the knowledge already that resuscitation skills are known to decay.

This thesis creates its original contribution through interrogating 4SA with reference to Worley's Symbiotic framework for clinical education. This will help organise the relevance and impact of the questions asked and data obtained. This will protect the thesis against education for education's sake; or research for the sake of research. This framework will retain a focus on the underlying *why* behind clinical education: the patient who sits in a context of social, institutional and clinical perspectives, who just wants to receive skilled care.

3 STUDY 1: ACQUISITION OF MANUAL DEFIBRILLATION SKILLS: A COMPARATIVE TRIAL

3.1 Introduction

The literature supporting the use of 4SA is conflictual. On the one hand, some authors cite well accepted educational principles like the ladder of competence (Lake & Hamdorf, 2004) and Miller's pyramid to spur on the use of 4SA(Lake & Hamdorf, 2004), and on the other hand at a closer look, these principles do not always apply as rigorously to the claims made. The comparative studies which have sought to identify whether 4SA is more effective than 2SA have not reached consensus (see section 2.4.1), and variability in reporting measures makes a meta-analysis impossible. Additionally, no studies to date address a baseline skill performance ability. The problem of effective skill education is significant. Health services need competent clinical staff to treat patients effectively, however patient outcomes have not yet been investigated in reference to the 4SA. Teaching institutions need insight into how their chosen andragogical approach and strategies might transfer into clinical practice.

This pilot study asked "Do 4SA and 2SA have the same impact on skill acquisition?" For this purpose, the 2SA used in the study was Stages 2 and 4 of the 4SA. Historically, medical education employed a "*see one, do one*" teaching approach, with an optional "*teach one*" sometimes added. In that regard, *see one, do one* would strictly encompass Stages 1 and 4 of 4SA. The 2SA used in this study (and study series) is, then, a diversion from this tradition. The reason for this is based on current skill teaching practice emphasising the need for explanation during demonstration, hence the 2SA described is a more realistic modern-day comparison. The focus was on skill performance improvement, with the acknowledgment that students will bring previous experience and awareness with them. In order to prevent the influence of previous knowledge and skills on the study, this baseline performance were obtained prior to teaching. Through a cluster controlled trial, half of the teaching groups were allocated to 2SA, and the other half will be allocated to 4SA, and all video performances blindly assessed according to both global scales and skill-specific checklists.

The original contribution of my research lies in the comparison of 2SA and 4SA in terms of performance outcomes according to baseline performance ability. Previous studies to date have not reported baseline performance. Using first year paramedic students, this study helps educators and curriculum designers understand the impact of these two teaching methods on this unique cohort of students. The skill of interest in this study was manual defibrillation. Defibrillation of a patient who is in ventricular fibrillation or ventricular tachycardia is demonstrated to be a significant priority in the management of cardiac arrest (ANZCOR, 2016a) due to its proven positive impact on resuscitation outcomes when performed early.

Understanding the effect that a teaching method may have on paramedic students' ability to perform lifesaving skills such as manual defibrillation has a significant potential to impact on tertiary skills training, professional development strategies, professional emergency courses such as ALS, PALS, and other accredited courses for the provision of specialist emergency skills. Performance of critical skills such as manual defibrillation in an effective and timely way saves lives (Cave et al., 2011; Chan, Krumholz, Nichol, & Nallamothu, 2008; M. Larsen, Eisenberg, Cummins, & Hallstrom, 1993; Spearpoint, McLean, & Zideman, 2000). In light of this impact, educational researchers have an obligation to work *today* to help *tomorrow's* emergency paramedics perform these critical skills more effectively.

3.2 Methodology

The research question reflects the ontological approach that reality is observable, measurable, and absolute (Brink, Van der Walt, & Van Rensburg, 2012, p. 24). This knowledge can be known by direct observation or measurement in a way which is consistent. This epistemology upholds a focus on strict and objective measurement to insure against bias[3], and may be referred to as quantitative methods. Thus, the positivist approach was adopted as consistent with this perspective, and within this approach a comparative trial offers a strategy and foundation which aligns with this paradigm. An intervention study was identified as a suitable approach to discern whether an observable difference in skill performance scoring was evident within the ontological framework stated, as randomised subject recruitment was not feasible.

3.3 Methods

First year paramedic students were invited to learn manual defibrillation in a simulated environment by one of the two methods which was randomly allocated after study enrolment. A pre/post design was employed to measure the participants' skill improvement (as a proxy for learning), as it has been identified that previous studies report on the outcome of skill performance without identifying pre-existing knowledge and ability. Measuring this baseline knowledge with a repeated measures design allowed data to more clearly reflect how much of the post-instruction performance results from the teaching session, rather than assuming all participants enter the study with comparable ability.

[3] Mechanisms employed to minimise various biases in this study are listed in a later discussion on threats to validity, in section 3.3.3.

3.3.1 Variables of interest

The independent variable employed in this study is the teaching method used to teach Manual Defibrillation: either the four-stage approach (4SA) or a more traditional two-stage approach (2SA) as outlined in Section 1.1. The dependent variables of interest include global and skill-specific checklist markers of skill performance both immediately following training, and then at 6 months after.

3.3.2 Study design

A sample size calculation was performed (using α[4] = 0.05 and β[5] = 0.1), and a sample of 21 participants in each group was required[6].

Following recruitment, the first-year paramedic students were allocated to a teaching group and randomly assigned a teaching strategy (2SA or 4SA). Participants' baseline skill performance in manual Defibrillation was assessed with a high-fidelity simulation tool. Skill performance was video recorded for blinded assessment at a later date. Participants then entered the assigned training session and then repeated the skill performance immediately after the teaching session. Participants were planning to return at two months and six months following instruction to demonstrate retention under the same conditions, however external factors impacted this reassessment and the study was adapted to instead consider skill acquisition only. Following recruitment, an unrelated training session was organised by a third party where many of the study participants would have an opportunity to learn and practise the skill. Due to a limited sample size, it would not be possible to isolate the effect of this from any possible teaching effect due to the teaching session. Therefore, this study was amended to a skill acquisition study only, and a new retention study was planned.

3.3.3 Principles of rigour and reliability: how to address threats to validity

In his book *Research Design,* John Creswell (2013, pp. 174-176) outlines possible threats to a study's validity which have been addressed in the study design. Creswell splits these into internal threats, which may impair the researcher's ability to draw correct conclusions about what the data means for the participants in the study, and external threats which impact the researcher's ability to draw accurate conclusions about the wider population from the study. These are addressed in Table 4 and Table 5.

---

[4] Level of significance (*not* Cronbach's alpha)
[5] Risk of failing to detect a difference if one exists
[6] Using μ_0 = 60% and μ_1 = 66% based on findings from Herrmann-Werner et al. (2013), with a paired t-test.

3.3.3.1 Threats to internal validity of the study

Table 4: Threats to Internal Validity

Threat	How the threat was addressed in the study design
History	All data required were gathered in a single session.
Maturation	As all data were gathered in a single session, maturation effects are not applicable.
Regression	Students were not selected on the basis of any previous score or performance. This type of study is likely to appeal more to students who seek out opportunities to learn, and so there may be some natural bias in the student who is likely to choose to participate.
Selection	All participants who met the inclusion criteria were invited to participate, and all those who asked to participate were accepted. The researcher had no choice over who is accepted into the study, and selection bias from the researcher's perspective is null. Students were selected on the basis of inclusion/exclusion criteria, and willingness to participate.
Mortality (or Drop-out)	Involvement requirement in the study was very brief (approximately 60 minutes). No students who registered for the study and completed the initial baseline skill performance withdrew prior to its completion.
Diffusion of treatment	Students would be free to communicate with their peers about their experiences, however the effect of this (for example while passing in the hallway between the teaching session and second performance) was not deemed to have an impact on the study.
Compensatory demoralisation	Both groups were invited to learn the skill of manual defibrillation, which is a second year paramedic skill. Participation in the study would offer both the control and intervention groups this training, just in slightly different ways. Efforts were made not to stress that either skill teaching method is likely to be more effective, just that the study was investigating whether or not this was the case.
Compensatory rivalry	(as above)
Testing	The same test conditions, script and room setup were used for all student performances. The post-education performance may be increased slightly due to the participants' recent practise at the test just prior to teaching, but this is anticipated to be equal (or near equal) for all students.
Instrumentation	The same assessment checklist was used for both the pre-training and post-training assessments. These were developed based on an audit of the teaching sessions to ensure participants would only be assessed on items which they had been taught.

Adapted from "Research design" by John Creswell, pp. 174-176, copyright 2013 by Sage publications.

3.3.3.2 Potential threats to external validity of the study

Table 5: Threats to External Validity

Threat	How the threat was addressed in the study design
Interaction of selection and treatment	Only paramedic students were recruited for this study. Findings will therefore be limited to similar types of students. These students have achieved a high Australian Tertiary Entrance Rank (ATAR) to gain access to the paramedic degree, and have not yet gained employment with the Ambulance Service. Therefore findings may be most generalisable to tertiary students learning psychomotor skills in professional degrees which are competitive (for example an ATAR score above 90), with an inclination towards a career which is autonomous and clinical in nature (for example physiotherapy, medical students or midwives).
Interaction of setting and treatment	The request to perform manual defibrillation occurred in a consistent setting, with consistent access to equipment. The facilitator's script is outlined in Appendix 10.2.3.
Interaction of history and treatment	Not applicable (the study only required around 60 minutes of participant time).

Adapted from "Research design" by John Creswell, pp. 174-176, copyright 2013 by Sage publications.

3.3.3.3 Other considerations of bias

Other sources of possible bias were also considered in the study design. By recruiting independent assessors who were blinded to the participants' group allocation, the impact of the assessors' expectation of the results on the data they provide is minimised. Videos were provided in random order so it would be less obvious which related to the initial performance, and which related to the post-instruction performance (although this was likely to be obvious from the performance demonstrated). Assessors were not advised of the identities of other assessors until marking had been completed, in order to limit any possible discussion which could influence their judgements. Additionally, only enough information was provided on the study design as was necessary. This was kept fairly vague so as to not communicate suspicions that one teaching intervention was superior. I reduced the experimenter-expectancy effect by ensuring I did not access or conduct the teaching or assessment sessions.

3.3.4 Recruitment

3.3.4.1 Ethics

Ethics approval was obtained from the Social and Behavioural Research Ethics Committee (SBREC) at Flinders University.

3.3.4.2 Student participant recruitment

Bachelor of Paramedic Science who were in their first year of studies at Flinders University of South Australia were invited to participate in the study, so long as they did not meet the following exclusion criteria:

- Primary language is not English
- Enrolment in the topic PARA2002 at Flinders University of South Australia (FU) as manual defibrillation is taught in this topic
- Not fit to perform the skill of manual defibrillation
- Under the age of 17 years (as ethics approval was obtained for the recruitment of adults only)

Participants with a primary language other than English were excluded because there was a potential for this to impact upon the way they learn clinical skills. Andrew (2002) identified difficulties among first year nursing students' science and nursing practice course, particularly due to language skills and cultural factors. Further, Zollo (1998) identified that students from a non-English speaking background (NESB) were four times more likely to fail a first year practice course (which included both knowledge and clinical skills), whereas academic performance can be greater for NESB in other courses (Birrell & Khoo, 1995). Salamonson and Andrew (2006) note that students with a NESB are statistically more likely to demonstrate academic underachievement in nursing, and Windle (2004) notes language as an indicator of academic success, albeit within a complex interaction of other ethnicity factors which this study does not have the scope to investigate. Students whose primary language was not English were invited to register their interest for a training session, although their data would not be recorded and reported. No such students registered.

Participants were recruited in three ways:

1. A face-to-face announcement during a first year topic (PARA1000) introduced the study and students were invited to enrol online or contact the researchers.
2. Participants were also invited by email to enrol in the study online. The email and attachments are located in the appendix of this thesis.
3. A follow-up phone call was made to each eligible participant extending an invitation for the study.

The students who chose to participate in the study may be disproportionately enthusiastic in their approach to learning opportunities in comparison to the rest of their peers. This may introduce bias of representing data reflecting only the higher achieving or more enthusiastic students. Participants were blinded to their allocation as much as is possible for a study such as this.

3.3.4.3 Educator selection

The selected educator was a paramedic tutor in 2013 at FU who received instruction in the teaching methods to be used, and agreed to participate in the study as designed. The educator was fully informed of the study purpose and had no acquaintance with any study participants prior to the study.

The educator was briefed in a neutral way which sought not to identify either method as superior. Briefing occurred by verbal, written and practical means, including multiple opportunities for feedback, clarification, and practise.

3.3.4.4 Skill assistants

Assistants were recruited from second-year Paramedic students at Flinders University. They were asked to wear their Paramedic placement uniform to make them appear as unremarkable and unmemorable as possible on the skill performance videos. This was intended to avoid creating a bias in marking by keeping the assessors from recognising variations in dress, as much as possible.

The skill assistants:

- Registered students as they arrive for their sessions
- Performed CPR to assist the teaching session
- Assisted with CPR in skill performance sessions, and
- Facilitated skill performance sessions

3.3.4.5 Assessors

Three paramedic tutors were recruited to mark the student's video performances. Two were asked to mark all performances, and a third was recruited to referee where these two assessors disagreed on binary marking values.

3.3.5 Blinding

Participants were not blinded to the study's intention of investigating possible differences in skill uptake when different methods are used to teach the skills. They were not explicitly informed of the teaching methods, and they were blinded to their allocation. It is recognised that following the teaching session, it could be easily determined by the students which allocation they received through discussion with other students. The educator was not blinded to the study design.

Video assessors were blinded to participant allocation and performance sequence. They were not provided with participant names, student numbers, or identification beyond the video footage. As I performed both the randomisation and analysis, I was *not* blinded to the participants' allocations. I also recorded the skill performance times.

3.3.6 Skill performance

The skills were performed under video recording at Sturt campus of Flinders University. Video (including audio) performance was recorded from two angles, and participants were not provided with prompts or feedback during this session. A skill facilitator was present to welcome and identify the participant, instruct them to perform the skill, and direct them either to the training room (after Performance 1) or to dismiss them (after Performance 2). A skill assistant was present to perform CPR on the manikin as this would ordinarily occur alongside defibrillation attachment and charging.

Both assessment rooms were comparable size, and provided access to the same manikin and equipment. The skill performance area provided the participants ample room to complete the task unhindered. The facilitator’s simulation equipment was set to manual defibrillator mode, and was transmitting heart rhythm of coarse Ventricular Fibrillation (VF) to the receiving unit. Rhythm analysis was not assessed.

The assessment facilitator was asked to follow a pre-determined script for each performance. The performance session ended shortly after chest compressions had been recommenced following the defibrillation, or after a pause of five seconds had indicated that the participant would not recommence CPR following defibrillation. The assistant and participants were briefed that no feedback or advice would be provided during the assessment.

3.3.7 Teaching

3.3.7.1 Teaching session

Participants entered the training room following their initial skill performance. The skill training room was set up with the same equipment as the performance rooms. A single camera was set up to record the training session to monitor teaching practice during these sessions. The skill assistant performed the same role as for performance rooms, and the use of simulation equipment was consistent with the performance room.

3.3.7.2 Teaching format

The control (2SA) teaching protocol was to follow the format:

- The skill was demonstrated slowly by the educator, with a step-by-step explanation of the skill and associated rationale.
- This was followed by an invitation to ask questions.
- Then the participant demonstrated the skill to the educator. During this stage, the educator provided feedback and answers questions if necessary. The students could be asked about rationale during this stage. Each student completed this step in the training session and had the opportunity to witness others' attempts and feedback.

The intervention (4SA) teaching protocol was *intended* to follow the format:

- The skill was to be demonstrated in real time to the student (s), without particular explanation.
- Then, the skill was to be repeated slowly by the educator, with a step-by-step explanation of the skill and associated rationale.
- This was to be followed by an invitation to ask questions.
- The third stage would involve the student verbally prompting the educator to perform the skill.
    - Where errors occurred, the educator would ask other students (if present) for a prompt, before confirming the correct step.
    - Where there was more than 1 student in the group, only one student would be asked to provide prompts to the educator
- This was to be followed by an invitation to ask questions.
- At the fourth stage, the participant would demonstrate the skill to the educator. During this stage, the educator provided feedback and answers questions if necessary. The students could be asked about rationale during this stage. Each student completed this step in the training session and had the opportunity to witness others' attempts and feedback.

Each training session was audited to ensure compliance with teaching content and teaching strategy. This is further discussed in Section 3.4.2.1).

3.3.7.3 Group Allocation

Participants were allocated to a teaching day based on their availability, and randomized into a teaching group using an online randomisation. 48 participants were randomised. Stratified random allocation of teaching interventions was employed which effectively randomised participants in clusters to either an intervention group (IG) or control group (CG). Following randomisation, there were some changes to the groups due to participant availability. Participants were blinded to their allocation (CG or IG) until they experienced their teaching session, therefore these changes were not considered to effect randomisation or have any statistical impact on the study outcomes.

The following schedule was developed:

Table 6: Pilot Study Teaching Schedule

Group	Allocated Teaching method	Allocated day	Allocated start time (initial assessment)	Allocated start time (teaching)
1	2SA	Monday	1:40pm	2:00pm
2	4SA	Monday	2:00pm	2:20pm[7]

[7] Following the execution of this stage of the study, it was noticed that an oversight in the schedule has rostered group two 20 minutes too early, leaving only 20 minutes allocated for group 1 teaching. On the day, this caused group 2 to experience some unexpected delay until the group 1 teaching concluded.

3	4SA	Tuesday	9:00am	9:40am
4	2SA	Tuesday	9:40am	10:00am
5	4SA	Tuesday	10:20am	10:40am
6	2SA	Tuesday	11:40am	12:00pm
7	4SA	Tuesday	12:20am	12:40pm
8	2SA	Tuesday	1:00pm	1:20pm

Note: 2SA= two-stage teaching approach and 4SA= four-stage teaching approach

The training schedule was arranged in this way to disperse training with both methods throughout the day. This helped address the potential impact changes in the way students learn depending on the time of day (Winocur & Hasher, 2004).

3.3.8 Data

Identifiable data are stored on a restricted drive at Flinders University, as per the ethics agreement for the study. Information was collected by video recording and pre-study enrolment forms. Enrolment information was used to:

- correctly identify student performances
- link student performance to demographic and experience data
- ensure participants met inclusion criteria

3.3.8.1 Performance criteria

Two sets of performance criteria were used to gather data relating to student skill performance. These were initially referred to as "objective" and "subjective" marks, however further reflection of assessment validity principles prompted a move to instead refer to a "skill specific checklist", and a "global scale" respectively. This alteration of language coincides with the development of my personal epistemological assumptions, which grew to doubt the possibility of purely objectivity observations, when such observations are obtained by people who interpret even seemingly objective data through their own experiential, theoretical, professional, personal, spiritual and socio-cultural (etcetera) lenses. Such perspectives are therefore always unique, regardless of the attempt to create objective certainty.

3.3.8.1.1 Skill specific checklist

This assessment tool, initially intended to provide "objective" performance data, was developed from an audit of the content presented in the teaching sessions. This was intended to ensure that participants were only assessed on content they had been exposed to, and hence the criteria would be a fair assessment to all. The checklist excluded some items which were prompted by some of the skill performance facilitators. For example, some participants were invited to don gloves or safety glasses, hence assessing this component would not reflect the teaching provided, but rather the individual facilitator. In Table 7 below, Items 1 to 25 were marked as "Yes", "No", "Unable to assess", or "Other" (with explanatory comment from the assessor). An example which could leave an item being "Unable to assess" poor camera angle. Where assessors agreed that the item could not be definitively assessed, a decision was made to award a "no" for that component. One mark was awarded for each "yes", and this series of binary scores was added with equal weighting to produce a score out of 25.

3.3.8.1.2 Global score

Additional subjective parameters were included to allow consideration of factors which could not be captured in a binary style checklist (for example apparent overall confidence/competence). Four items were rated as global items, with a mark of 0 (not at all), 1 (below average), 2 (average), 3 (above average), or 4 (outstanding/highly skilled in the area). These are also listed in Table 7 as items 26 to 29, and provide a score out of 16.

Table 7: Assessment Items for Manual Defibrillation Skill Performance

		Item	Score
Skill specific performance checklist	1	Instruct CPR providers to continue with CPR during pad placement	Yes, No, Unable to assess or Other
	2	State compression: ventilation ratio of 30:2	Yes, No, Unable to assess or Other
	3	Place pads correctly	Yes, No, Unable to assess or Other
	4	Appears confident with pad placement	Yes, No, Unable to assess or Other
	5	Compressions are maintained during charging	Yes, No, Unable to assess or Other
	6	States that defibrillator is about to charge, or that defibrillator is charging	Yes, No, Unable to assess or Other
	7	Charge defibrillator before analysing rhythm	Yes, No, Unable to assess or Other
	8	States charge defibrillator to 200J	Yes, No, Unable to assess or Other
	9	Hands off to analyse rhythm	Yes, No, Unable to assess or Other
	10	Verbally confirm shockable rhythm prior to defibrillator	Yes, No, Unable to assess or Other
	11	Looks at monitor and appears to assess rhythm prior to shock	Yes, No, Unable to assess or Other
	12	Confirms rhythm is "VF" (ventricular fibrillation) prior to defibrillator	Yes, No, Unable to assess or Other
	13	Verbalise "clear" (or similar) for defibrillator	Yes, No, Unable to assess or Other
	14	Oxygen is away from patient prior to defibrillator	Yes, No, Unable to assess or Other
	15	The CPR person is prompted to hold hands up prior to defibrillator (either verbally or queued non verbally)	Yes, No, Unable to assess or Other
	16	Asks the CPR person "are you safe?"	Yes, No, Unable to assess or Other
	17	CPR person's hands are visualised by the person executing the defibrillator, prior to delivery of charge	Yes, No, Unable to assess or Other
	18	Eye contact with other practitioners to ensure clear prior to defibrillator	Yes, No, Unable to assess or Other
	19	Patient's top section is checked as clear prior to shock	Yes, No, Unable to assess or Other
	20	Patient's mid-section is checked as clear prior to shock	Yes, No, Unable to assess or Other
	21	Patient's bottom section is checked as clear prior to shock	Yes, No, Unable to assess or Other
	22	Shock was delivered	Yes, No, Unable to assess or Other
	23	Verbalises the decision to deliver a shock to the patient	Yes, No, Unable to assess or Other

	24	Verbalises when shock is actually being delivered	Yes, No, Unable to assess or Other
	25	Instructs to resume CPR immediately following defibrillator (within 4 seconds)	Yes, No, Unable to assess or Other
Global scale	26	Appears generally confident with iSim use in defibrillator mode	0-4
	27	Provides clear communication throughout	0-4
	28	The skill progresses smoothly without fumbling	0-4
	29	How safe and competent do you consider this person to be at the skill?	0-4

Note: For the skill specific checklist, an item marked "yes" incurred one mark for the student, and an item marked "No", "unable to assess" or "other" did not. Marks were then added together for a total score out of 25 for the checklist. The marks awarded for the four global items were also added together for a total score out of 16.

3.3.8.2 Assessment of performance videos

The assessors were all FU tutors, and were required to sign a confidentiality agreement prior to receiving access to the videos. Two assessors were invited to mark the performance videos according to the revised marking sheet.

If results for the binary items conflicted between the two assessors, a third assessor was employed to provide adjudication. For the global judgement items, an average of the three markers would be used. The time to perform the skill was recorded from the video by the chief researcher. Time was recorded from the point where the participant knelt beside the simulated patient (manikin), provided verbal direction to the skill assistant, or began operation of the training defibrillator (whichever occurred first), until a defibrillation shock was delivered (simulated).

3.4 Results

3.4.1 Participant response

All eligible participants were invited to attend, and approximately 30-35% of those eligible were available and willing to do so. As participation was voluntary, possible sources for sample bias include participant willingness and enthusiasm to participate in extra-curricular teaching, computer literacy levels and convenience (as only limited times were available to participate). These factors are discussed further in study discussion section 3.5.7.1. Participant recruitment and progression through the study is outlined in Figure 14. Two of the eight teaching sessions were excluded from the study due to deviation from the teaching protocol (see Section 3.4.2.1). 28 students remained in the six included teaching sessions: 15 were in the control group (CG) taught with a 2-stage approach (2SA) and 13 were in the intervention group (IG) taught with a 4-stage approach. This approach was not the 4SA intended, but was a modified version ($4SA_m$).

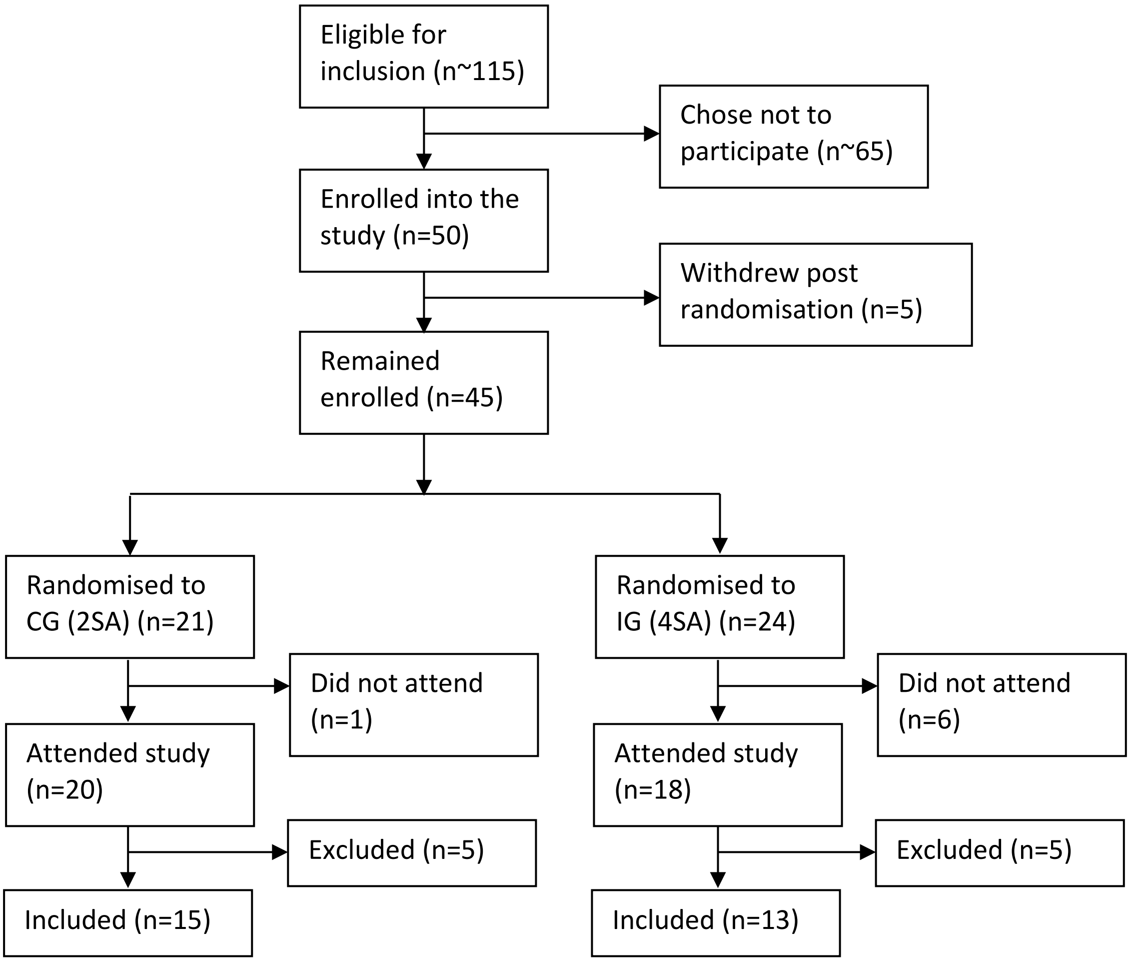


Figure 14: Manual defibrillation study inclusion and teaching allocation. CG = Control group (two-stage approach) and IG = Intervention group (four-stage approach).

The included participants' allocation is outlined in Table 8 below:

Table 8: Participants who Attended and Remained in Manual Defibrillation Study

Participant number	Training group	Intervention	Gender	Both performance videos able to be marked?
1	8	CG (2SA)	Female	Y
2	5	IG ($4SA_m$)	Female	Y
3	5	IG ($4SA_m$)	Female	Y
4	6	CG (2SA)	Female	Y
5	7	IG ($4SA_m$)	Female	Y
6	5	IG ($4SA_m$)	Male	Y
7	4	IG ($4SA_m$)	Female	N (corrupted video)
8	7	IG ($4SA_m$)	Male	Y
9	6	CG (2SA)	Male	N (corrupted video)
10	3	CG (2SA)	Female	Y
11	6	CG (2SA)	Female	Y

12	5	IG ($4SA_m$)	Female	Y
13	4	IG ($4SA_m$)	Female	Y
14	6	CG (2SA)	Female	Y
15	8	CG (2SA)	Female	Y
16	7	IG ($4SA_m$)	Male	Y
17	3	CG (2SA)	Female	Y
18	8	CG (2SA)	Male	Y
19	7	IG ($4SA_m$)	Male	Y
20	3	CG (2SA)	Female	Y
21	6	CG (2SA)	Male	Y
22	4	IG ($4SA_m$)	Female	Y
23	6	CG (2SA)	Male	Y
24	8	CG (2SA)	Male	Y
25	3	CG (2SA)	Female	Y
26	3	CG (2SA)	Male	Y
27	4	IG ($4SA_m$)	Female	Y
28	7	IG ($4SA_m$)	Female	N (corrupted video)

Note: Y = yes; N = no; CG= Control group taught with 2SA; IG = intervention group taught with $4SA_m$

3.4.2 Teaching audit

Each teaching session video was reviewed to confirm compliance to the teaching protocol, and to establish which elements were taught consistently across all sessions to inform development of the marking sheet.

3.4.2.1 Study protocol compliance

Groups 3 and 4 were accidentally taught with the opposite method to the allocation. It was decided that these groups would still be included in the analysis as it was an accidental error made by the educator, and not in response to particular knowledge of the participant cohort, any participant in the group or any stated bias. While the randomisation was not strictly adhered to, it was decided that allowing this data to be included would not introduce a confounder to the sample due to its unplanned nature.

Table 9: Audit of Study Protocol Compliance and Inclusion or Exclusion

	Teaching session							
	1	2	3	4	5	6	7	8
Demonstrate the skill without explanation	n/a	✓#	n/a	✓	✓	n/a	✓	n/a
Demonstrate the skill with explanation at each step	✓##	×**	✓	×**	×**	✓	×**	✓

One student prompts the teacher to perform the skill	n/a	×***	n/a	✓	✓	n/a	✓	n/a
Students perform the skill and receives feedback/prompting if necessary	✓****	✓	✓	✓	✓	✓	✓	✓
Included (I) / excluded (E)	E	E	I	I	I	I	I	I

Note: Shaded cells indicate the two-stage skill teaching sessions, hence Stages 1 and 3 of the four stage approach were not included.
A student was used to do CPR for this teaching session. This blended Stage 1 of the teaching method into a Stage 2 as some explanation was required.
There were some difficulties using the training equipment during this stage.
**The instructor talked through the skill without actually doing it, thus this step became explanation only for those allocated to 4SA.
***The teacher did not actually do it as the student prompted them through it.
****The rest of the student cohort was sent outside the classroom during this step, so they were not able to learn by watching as the other groups were.

Generally it would have been more appropriate for more feedback to be given to the students during the teaching session.

The above review shows:

- Group 1 to be taught with a method different to the other groups
- Group 2 to be taught with a method different to the other groups
- Groups 3, 6 and 8 to be taught with the same method
- Groups 4, 5 and 7 to be taught with the same method

After teaching groups 1 and 2 were excluded[8], the following IG and CG criteria were redefined:

Table 10: Intended Study Protocol

Stage	4SA	2SA
1	Demonstration of skill without explanation	(Stage 1 of 4SA omitted)
2	Demonstration with explanation	Demonstration with explanation
3	Student prompts instructor while the instructor performs the skill	(Stage 3 of 4SA omitted)
4	Student performs the skill	Student performs the skill

Table 11: Actual Study Procedure

Stage	$4SA_m$	2SA
1	Demonstration of skill without explanation	(Stage 1 of 4SA omitted)
2	Explanation only (without demonstration)	Demonstration with explanation
3	One student verbalises skill to instructor (without performance)	(Stage 3 of 4SA omitted)
4	Student performs the skill	Student performs the skill

[8] Skill performances from the two excluded groups were used as a training tool to ensure assessor useability of the marking guide, and validate the marking guide. Following the marking of these videos, the assessors and chief researcher discussed any recommended changes to the marking guide before the videos were marked.

The main deviations from the intended model were Stages 2 and 3 of 4SA. This stage was involves demonstration with explanation, but it was consistently presented as verbal explanation only for those assigned to the 4SA teaching group. It was decided that because the deviation was consistent between the IGs, that the protocol would be adjusted to incorporate it.

3.4.2.2 Consistency of teaching content

The teaching content was audited alongside a checklist of proposed assessment items based on the teaching plan provided to the educator. These items were reviewed to ensure the assessment checklist used to mark performances is a fair reflection of each teaching session.

Table 12: Manual Defibrillation Teaching Session Audit

		Teaching session							
		1	2	3	4	5	6	7	8
	teaching allocation	2SA	4SA	4SA	2SA	4SA	2SA	4SA	2SA
Item	teaching used	2SA	4SA$_m$	2SA	4SA$_m$	4SA$_m$	2SA	4SA$_m$	2SA
1	PPE: safety glasses worn	×	×	×	×	×	×	×	×
2	PPE: gloves worn by both clinicians	×	×	×	×	×	×	×	×
3	Continue with CPR during pad placement	✓	✓	✓	✓	✓	✓	✓	✓
4	30:2 CPR	✓	✓	✓	✓	✓	✓	✓	✓
5	Place pads on patient	✓	✓	✓	✓	✓	✓	✓	✓
6	Maintain compressions while isim is charging	✓	✓	✓	✓	✓	✓	✓*	✓
7	Select defib mode	✓	✓*	✓	✓	✓	✓	✓	✓
8	States when defib is charging	✓	✓	✓	✓	✓	✓	✓	
9	Charge isim prior to analysing rhythm	✓	✓	✓	✓	✓	✓	✓	✓
10	Charge defib (200J)	✓	✓	✓	✓	✓	✓	✓	✓
11	Cease compressions to analyse rhythm	✓	✓	✓	✓	✓	✓	✓	✓*
12	Identify the rhythm as coarse VF	✓	✓	✓	✓	✓	✓	✓	✓
13	Visual check that the rhythm is shockable	✓	✓	✓	✓	✓	✓	✓	✓
14	Move oxygen away from patient	✓	✓	✓	✓	✓	✓	✓	✓
15	State clear top, middle, bottom of patient	✓	✓	✓	✓	✓	×	×	×
16	Identify clear top, middle, bottom of patient	✓	✓	✓	✓	✓	✓*	✓	✓*
17	"Are you safe"	✓	✓	✓	✓	✓	✓	✓	✓
18	Visualize the CPR assistant's hands held up away from patient	✓	✓	✓	✓	✓	✓	✓	✓

19	Make eye contact with CPR assistant on scene	✓	✓	✓	✓	✓	✓	✓	✓
20	Communicate that you have delivered shock	✓	✓	✓	✓	✓	✓	✓	✓
21	Continue CPR straight after defibrillation	✓	✓	✓	✓	✓	✓	✓	✓

Note: shaded columns indicate teaching sessions excluded to maintain consistent comparison between the two teaching strategies.
* indicates an item which was demonstrated but not explicit

Items 1 and 2 were removed from the skill marking checklist as they were not taught in any of the teaching sessions. Items 6, 7, 11 and 16 were taught by demonstration but not explained. These items were included in the marking checklist. Item 15 was not taught to groups 6, 7 or 8 so this item was removed from the marking checklist. Some additional items were added to the assessment checklist presented in Table 7 which was then used by the assessors.

3.4.3 Homogeneity between groups

3.4.3.1 Gender

The number of included participants was proportionate to the eligible population for most groups (11% inclusion from eligible population), with the exception of slightly more males in the IG (16%) which may impact the sample homogeneity.

Table 13: Gender Distributions of Students Included in the First Comparative Study

	Male	Female	Total***
Eligible students	38 (33%***)	79 (69%***)	115# (100%)
Chose to participate in study	16 (42%*)	29 (37%**)	45 (39%)
Included in study	10 (26%*)	18 (23%**)	28 (24%)
Included in IG (4SA)	6 (of 38 = 16%*)	9 (of 79 = 11%**)	15 (of 115 = 13%)
Included in CG (2SA)	4 (of 38 = 11%*)	9 (of 79 = 11%**)	13 (of 115 = 11%)

Note: IG = Intervention group taught with the four-stage approach; CG = Control group taught with the two stage approach
* Percentage of the total eligible male population
** Percentage of the total eligible female population
*** Percentages shows are of the total eligible population
# These 115 subjects were identified from a course enrolment list which omitted information on gender and primary language. The students who did not enrol into the study have gender and inclusion criteria assumed based on conversation content (on phone contact), name (if phone contact unsuccessful), or other knowledge of the participant.

In the absence of evidence that gender significantly impacts a learner's ability to learn, retain or perform a psychomotor skill, the greater number of males in the intervention group is not considered to impact on this study.

3.4.3.2 Age

The intervention group was slightly older than the control group but not significantly different ($p=0.334$).

Table 14: Age Distributions of Students Included in the First Comparative Study

	Total	IG ($4SA_m$) (n=15)	CG (2SA) (n=13)
Age range	17-49	17-49	18-28
Median age	19.5	20	19
Mean (±*SD*)	23.46 (±7.9)	25 (±11)	22 (±4)

SD = Standard deviation; IG = intervention group; CG = control group. All measures are in years.

The sample size was too small to effectively stratify by age, and the four oldest participants were incidentally randomly allocated to the intervention group. Thus, the CG and IG cannot be considered homogenous from an age perspective. Figure 15 demonstrates a similar median, but variable age range between the two randomly selected groups.

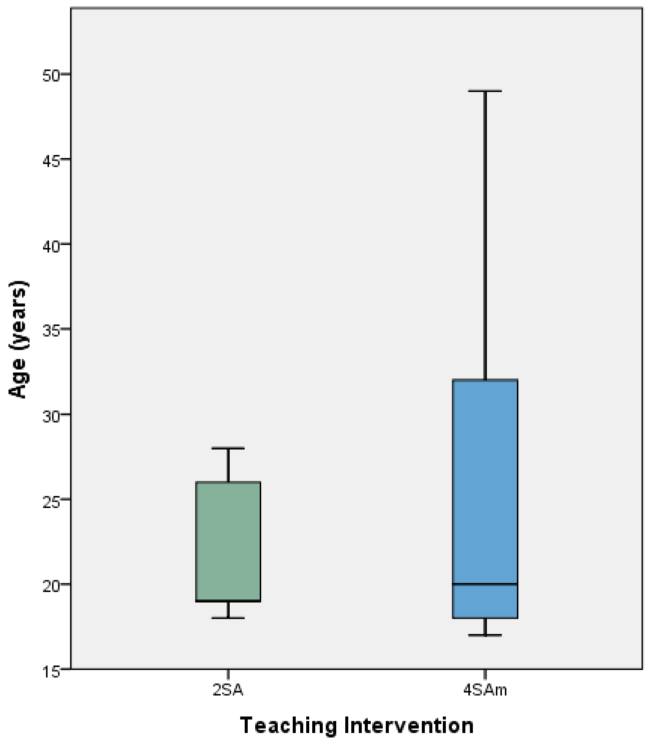


Figure 15: Distribution of participant age for defibrillation for intervention ($4SA_m$) and control (2SA) groups, showing median and interquartile range.

3.4.3.3 Previous exposure

Of the 28 included participants, two stated previous training in the skill of manual defibrillation. These participants were both randomly allocated into the IG, however as baseline performance was measured this is not expected to impact the study outcomes. One participant had no previous experience with the simulation training equipment used. Her initial performance was identified as an outlier (participant 13). Overall baseline performance was identified as statistically comparable between the two groups at attempt 1 (prior to teaching), so the groups are considered equal at the beginning of the study.

3.4.4 Did the teaching session impact performance?

Data were initially explored to compare the performance scores prior to teaching with scores obtained following teaching, to ensure that the skill specific and global assessment strategies identified an effect from the two different time points. If an effect is not measurable, it would raise significant concerns over the mark sheets, or the effectiveness of the teaching session (or both).

3.4.4.1 Skill specific checklist (total score out of 25 binary items)

There was a statistically significant improvement in mean checklist performance scores from baseline (M=7.76±2.437) to following instruction when compared using a paired samples T-test (M= 21.60±2.466), t(24)= -20.656, p< 0.001 (two-tailed). Of the 25 cases, baseline and post-instruction scores were not correlated (Pearson correlation .067, p=.752), so students with greater initial knowledge did not appear to have an advantage in the post-instruction assessment.

3.4.4.2 Global scores (mark out of 16 for the sum of four Likert scales)

Similar results were obtained for the global marks, although marks were out of 16. There was a statistically significant improvement in global performance score from baseline (M=3.94±1.61) to following instruction (M= 10.39±1.48), t(24)= -17.29, p< 0.001 (two-tailed). Of the 25 paired cases, baseline and post-instruction scores were not correlated (correlation .274, p=.185).

3.4.5 Outcome 1: Did the method of instruction alter the skill acquisition?

This question was investigated in a number of ways to ensure thorough interrogation of the data. First, a one-way ANOVA was performed using only the post-instruction performance scores. This analysis did *not* account for baseline performance, and therefore provides some reference to previous studies in the area. Both sets of scores from the skill-specific checklist and the global scales were found to satisfy the assumptions of homogeneity of variance (Levene's statistic 4.056 p=.054 and 0.607, p=.443 respectively indicate). Next, a Repeated Measures ANOVA was performed to understand what change, if any, was evident in performance scores (both skill specific and global) resultant from the teaching method. This analysis included both the pre- and post-instruction scores.

3.4.5.1 One way ANOVA

There was a statistically significant difference between the post-instruction skill specific checklist scores for the students taught with the different teaching methods using a one-way between groups

ANOVA [$F(1, 26)=5.172$, $p=.031$]. The average difference in performance was slightly over two marks ($M(4SA_m)=22.46\pm1.984$), $M(2SA)=20.33\pm2.820$), and the clinical impact of this difference depends on which criteria these two marks refer to.

The comparison of global scores, however, was *not* statistically significant between the two groups [$F(1, 26)=1.100$, $p=.304$]. The actual difference in performance was much more modest ($M(4SA_m)=10.56\pm1.39$), $M(2SA)=9.96\pm1.64$) at just over half a mark. The lack of statistical significance may be the result of a narrower difference between mean scores on this global checklist, in conjunction with low participant numbers.

3.4.5.2 Repeated Measures ANOVA

3.4.5.2.1 Effect of attempt number

This test incorporates the pre-instruction *and* post-instruction scores for each participant, rather than assuming on the basis of low correlation the initial score had no bearing on the post-intervention score. The mean scores for both groups, according to both skill specific and global score data are presented in Table 15 below.

Table 15: Mean Scores for 2SA and 4SA pre- and post- Teaching Session.

	2SA teaching (n=14)		4SA teaching (n=11)	
	Attempt no. 1	Attempt no. 2	Attempt no. 1	Attempt no. 2
Mean Binary score (±*SD*)	7.64 (±2.24)	20.71 (±2.49)	7.91 (±2.77)	22.73 (±2.01)
Mean Global score (±*SD*)	4.12 (±1.33)	10.10 (±1.61)	3.70 (±1.96)	10.76 (±1.27)

Note: Attempt no. 1 refers to pre-instruction teaching, and attempt no.2 refers to post-instruction teaching, immediately after the training session. Binary score refers to the mark out of 25, calculated from the binary items on the skill-specific checklist, and Global score refers to the mark out of 16 for the four Likert items which were more global in nature. Only 25 of the 28 students are represented here, as three videos malfunctioned, leaving only 25 full sets of data.

Tests of within subjects contrasts shows a significant effect for the attempt number (Wilks' Lambda= .043, $F(2,22)= 242.66$, $p<.001$), and a large effect confirmed by partial eta sq = .957. This confirmation of a measurable impact on performance scores from pre- to post- teaching was expected, and supports the validity of the checklist, as it provides a clear distinction between students who have not learnt the skill, and students who have (Cook, Zendejas, Hamstra, Hatala, & Brydges, 2014).

The interaction between the attempt number and the teaching method[9] is not significant (*p*=.153 for global score and *p*=.202 for skill-specific checklist score), indicating that the teaching method did not significantly impact either the skill specific checklist or global scores. However, the graphical representation of these data (in Figure 16 and Figure 17) suggest a possible underpowered interaction as the lines are not parallel. Considering the graphs, the $4SA_m$ *appears* to achieve superior skill acquisition than 2SA.

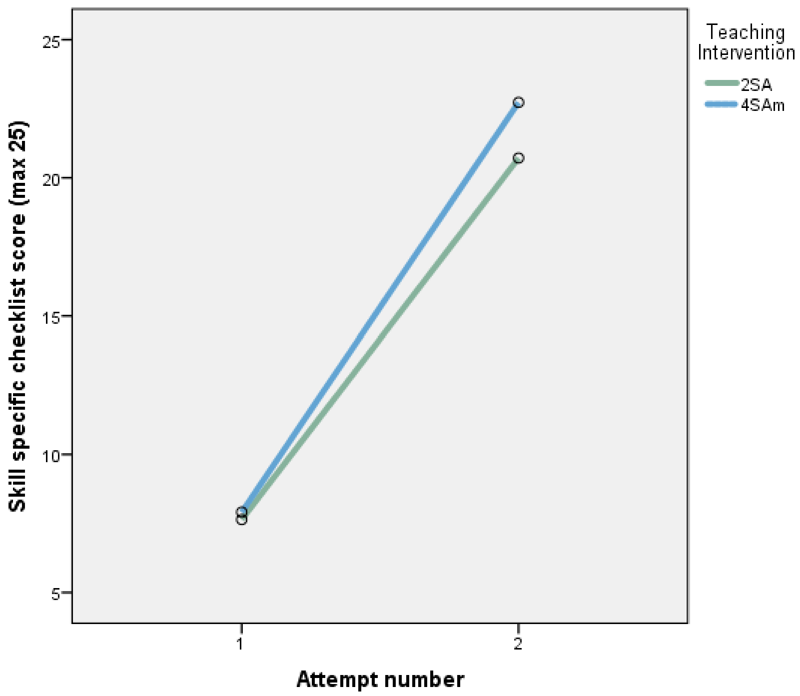


Figure 16: Increase in mean skill-specific checklist score for manual defibrillation following training (max score 25).

[9] (AttemptNo*TchInt) in Univariate Tests

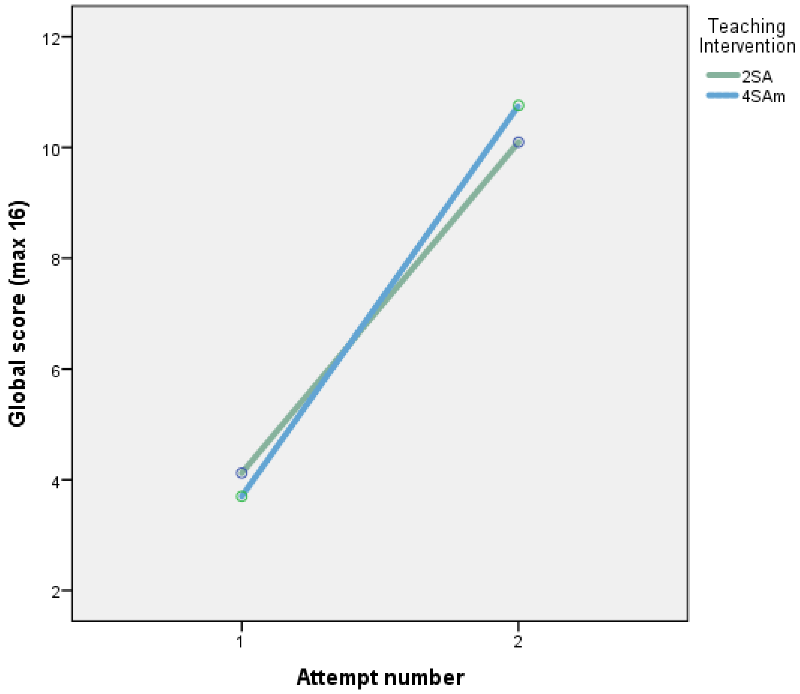


Figure 17: Increase in mean global score for manual defibrillation following training (max score 16).

3.4.6 Outcome 2: Time to teach

With only three groups taught with each method, data were not expected to identify a statistically significant result for a difference in time to teach. With such a small sample of groups, the opportunity to make a type II error is increased.

When the total teaching time is compared between the two different interventions, it appears that $4SA_m$ teaching relates to increased time to teach (see Figure 18).

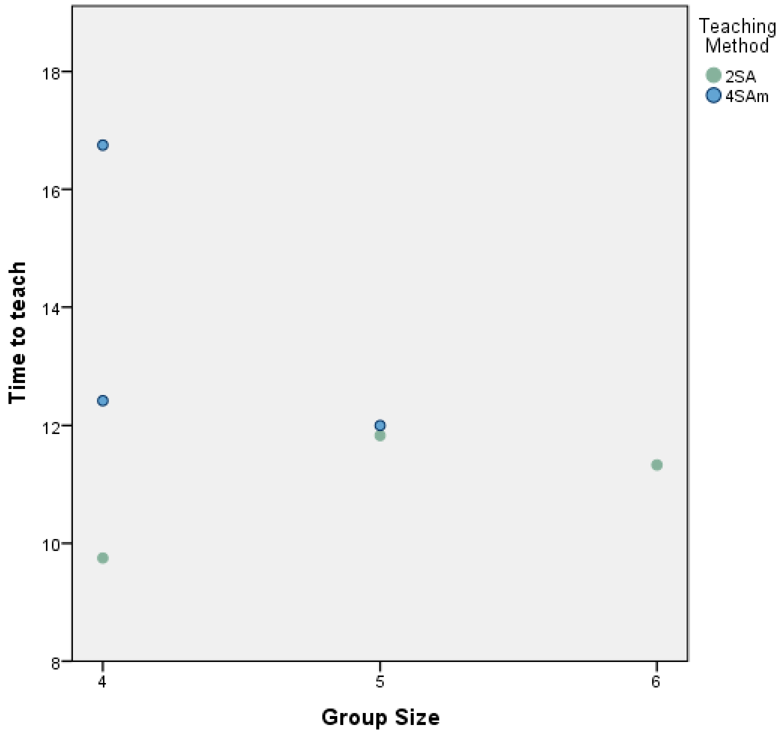


Figure 18: Comparison of 2SA and $4SA_m$ total teaching time by individual group size and intervention

A comparison of the mean teaching time yielded a difference of approximately 2.75 minutes, with $4SA_m$ recorded as taking longer than 2SA (by approximately 25% more time). A one-way ANOVA demonstrated that there was not a statistically significant difference between the two groups [F(1,4)= 2.810, p=.169]. When teaching times are adjusted to include only the first four students' performances (to control in some way for the fluctuating group sizes), the result was still not statistically significant [F(1,4)= 3.198, p=.148], and no further analyses were performed. Mean times are presented in Table 16.

Table 16: Comparison of Teaching Times for Manual Defibrillation Between 2SA and $4SA_m$

	Teaching Method		
	2SA (n=3 groups)	4SAm (n=3 groups)	Difference
Mean time to teach all participants in minutes (±SD)	10.97 (±1.09)	13.72 (±2.63)	25% increase with $4SA_m$
Mean time to teach first four participants only in minutes (±SD)	9.81 (±0.84)	13.22 (±3.20)	35% increase with $4SA_m$

Note: the significance was calculated as p=.169 and p=.148 for the times relating to all participants, and just the first four participants, respectively. However, with such a small number of groups for each teaching method, the risk of a Type II error is extreme.

A graph of the time required to teach the group, including the first four participants' performance of manual defibrillation, indicates that a lack of significance in this finding may be due to a type II error (see Figure 19). The times for the $4SA_m$ sessions (groups 2, 3 and 5 depicted in blue) appear greater than those for 2SA (depicted in green). Larger sample sizes are required to make any further comment on this hypothesis.

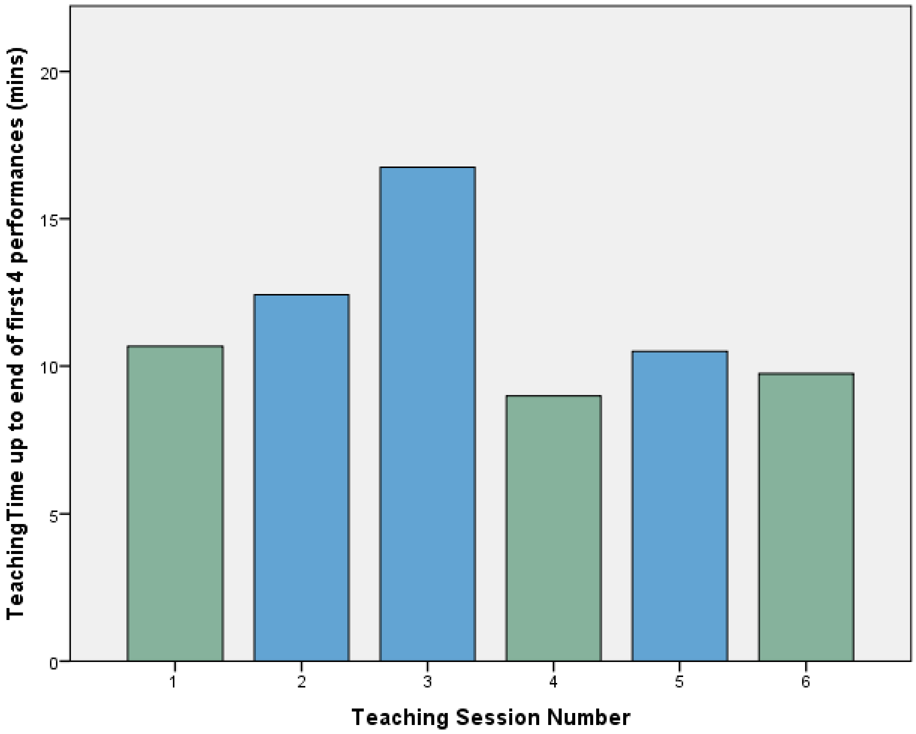


Figure 19: Total time to teach manual defibrillation to end of fourth participant (minutes)

The graph above reflects the total time to teach the skill until the end of the fourth person's performance. This attempted to standardize the effect from having slight variation in group size. The two methods coded in different colours. Again, one-way ANOVA determined that the data did not reflect significant difference. This is not surprising given such a small sample size. The reduction in *p*-value hints that the fluctuation in group size is introducing an additional variable to the time to teach.

3.4.7 Outcome 3: Time to perform skill

3.4.7.1 One way ANOVA

A one-way ANOVA was performed to explore the impact of the teaching method on the time to defibrillate the patient (post-teaching). This was measured from the time of instruction to defibrillate the patient to the time it was provided, in seconds. The mean time to deliver a shock to the manikin was M=43.178 (±12.95) seconds for the control group (2SA) and M=46.077 (±10.41) seconds for the intervention group ($4SA_m$), $F(1,27)=0.417$, $p=.524$. Identifying such subtle differences between groups of limited size is known to be a statistical limitation of small samples.

Time to defibrillate is an undisputed indicator of resuscitation attempt success. Time to defibrillate was not significantly impacted by the teaching method, nor did it correlate to the skill specific checklist score awarded by the markers, or the global score, Pearson's correlation .075 ($p=.705$) and -.073 ($p=.712$) respectively.

The correlations between time to perform the skill, and the performance score are presented in Figure 20 and Figure 21. These figures indicate that although there isn't a statistically significant effect on time from the teaching instruction approach, the post-instruction scores tend to cluster towards the higher end of both assessment score scales, and these performances tend to be faster than pre-instruction times.

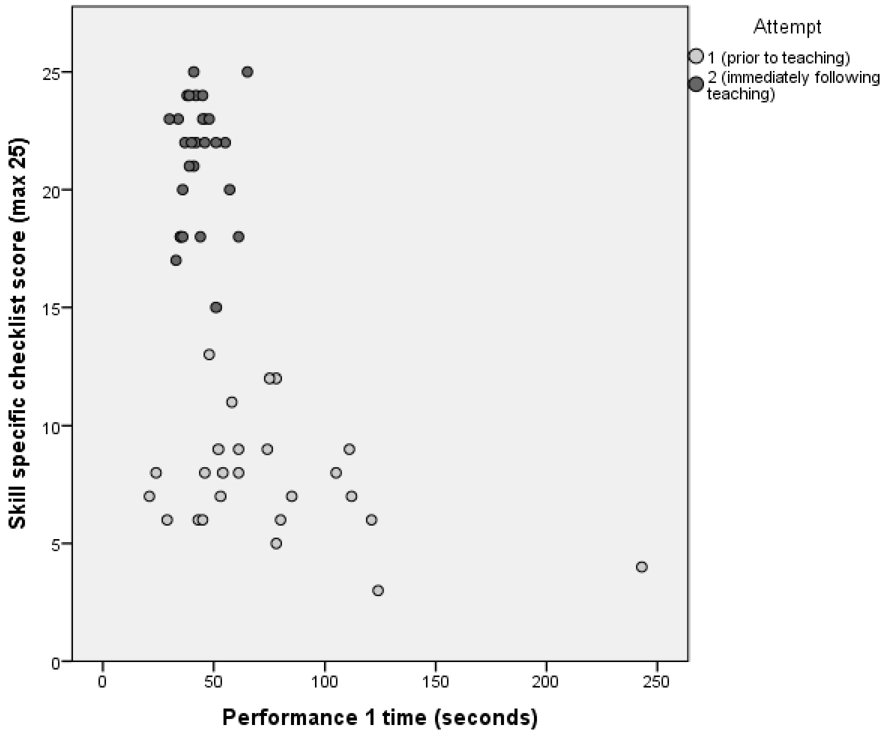


Figure 20: Correlation of time to perform manual defibrillation and skill specific checklist score

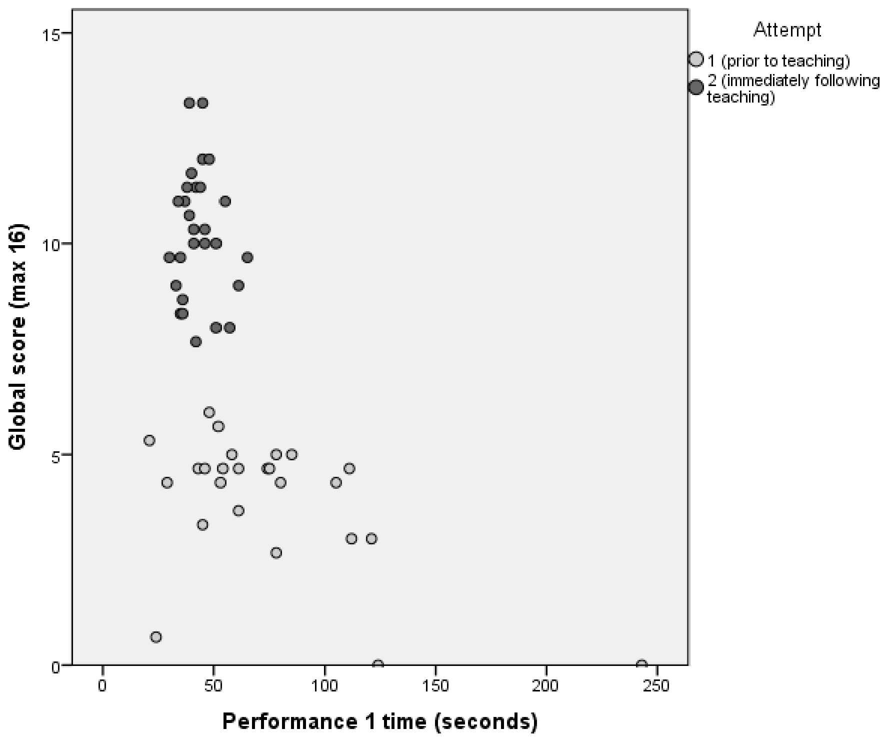


Figure 21: Correlation of time to perform manual defibrillation and global score

Not all of the more rapid applications, however, achieved a high global or skill-specific checklist score. In acknowledgement that the time to perform defibrillation (if appropriate)in cardiac arrest is clinically relevant, the individual components of *correct* skill application may not be as clinically relevant.

3.5 Discussion

In comparing the effectiveness of two different skill teaching methods, this study asks whether 2SA or $4SA_m$ results in superior skill acquisition of safe manual defibrillation skills. While the post-instruction checklist scores for the $4SA_m$ group were greater than the 2SA group, this was not demonstrated for the global scores, or when the mixed design was incorporated (the baseline score). This is an important question to ask wherever students are taught clinical skills, because the extent of the skill acquisition may have an effect on: the workload of clinicians and skills laboratory facilitators who may need to spend time and energy re-training insufficiently learnt skills; the reputation of the teaching institution to the health service who hosts either student clinicians or graduates from that teaching facility; and the level of competent health care provided to the patient consumer. For a skill such as defibrillation in cardiac arrest, these effects are enhanced due to the critical nature of the skill. Effective, timely defibrillation is known to have a significant impact on the return of spontaneous circulation (ROSC) (M. Larsen et al., 1993; Spearpoint et al., 2000) and patient survival to discharge from hospital (Chan et al., 2008). Therefore studies such as this which consider not only the application of the skill, but does so in the context of the dependence on time as a significant clinical predictor, are crucial to understanding effective resuscitation.

Dyson et al. (2015) found that Victorian[10] paramedic staff attend 1.4 resuscitation attempts a year, on average. (As a South Australian paramedic, this number is surprisingly low to me, however my own experience may be impacted by practice within an ageing population which may be more likely to experience such an event.) If these are evenly distributed throughout a year, ambulance staff would attend one resuscitation attempt every eight months, which leaves ample opportunity for skills to deteriorate. Latman and Wooley (1980) argue that significant atrophy of resuscitation skills is measurable in the first 6 months following learning, so without active and intentional practise to maintain expertise, the student's skill performance ability would likely be significantly lower, and slower. This study was initially designed to also measure retention of manual defibrillation skills, however many of the students were invited to participate in an extracurricular training session on the skill, and this was perceived to likely impact the retention data too significantly. A second trial was therefore planned to measure retention of other resuscitation skills, and this is described in Chapter 5.

3.5.1 What do the two checklists actually assess?

The teaching method appeared to significantly impact the skill-specific score following instruction, but not the global score (see section 3.4.5.1). This is the reverse of what Krautter et al. (2011) found. This finding may reflect that the two checklists reflect different aspects of the skill performance. The markers were asked to use their professional judgement when marking according to the global checklists. These judgements are a reflection of the marker's opinions as on-road paramedics, not as first-year university skill tutors. Thus, these marks are intended to reflect authentic practice. The skill-specific checklists, while still potentially requiring some judgement, are designed to be much more objective in nature. They more accurately reflect a student's compliance to the teaching provided which informed the checklist, which may differ from a global performance judgement. The skill-specific checklist may therefore be more a reflection of compliance to the teaching as provided, rather than professional practice ability. The skill-specific checklist, however, still required a level of judgement and even assumption, for example item 11 which calls for a decision on whether the assessor perceived that the student assessed the rhythm when they looked at the monitor, even if

[10]Operating as a pre-hospital ambulance clinician in the state of Victoria in Australia

they didn't state it aloud (item 12). This evidences that efforts to develop an objective assessment tool may still require some interpretation and judgement by the users.

In a repeated measures analysis where the pre-instruction and post-instruction scores are considered, this effect was not observed. The graphs show that if any effect exists, it is small, and this may be why it is not determined to be statistically significant.

3.5.2 Marking sheet validity

The pre-instruction and post-instruction scores obtained for the students were statistically different, with a noticeable improvement in performance rates on both the skill-specific (binary) checklist, and the global rating scales. As these two groups of performances are known to be different, a higher score is expected following the teaching, and confirmation of this in a statistically significant way evidences the ability of the marking sheet to distinguish these.

3.5.3 Baseline performance compared to post-instruction

Correlation between the initial performance score and the post-instruction score for the participants was not found. Therefore, participants who came into the study with more baseline knowledge or ability than others did not necessarily perform the skill better following the instruction. This was somewhat unexpected, with an expectation that the more knowledge a student has prior to the teaching, then if they exert the same cognitive energy to learn new components of the skill as their peers who had little initial performance ability, their follow-up performance would be greater. This relates to Miller's *Magic number seven* (G. A. Miller, 1956), as the new seven pieces of information would be built on a foundation where more knowledge is already assumed.

3.5.4 Teaching audit

A review of the teaching sessions was conducted in order to ensure the skill marking checklist would be fair to all participants, by checking that each item assessed was actually taught. In doing so, an inconsistency in the teaching method was identified. Of the six groups included in the final analysis, three were taught with the 2SA as described, and the other three were taught with a modified 4SA, recorded as $4SA_m$. The various factors that could have caused this unintended outcome, mainly in relation to the cognitive load of the educator are discussed further in Chapter 7.

The skill itself was taught using simulated equipment, which is different to that used in local Ambulance practice. The Paramedic educator had limited experience with the simulated equipment, and this may have contributed to an information overload for the educator. Using equipment which the educator was familiar with may have alleviated this. Additionally, using authentic equipment (in training mode) may have provided more realistic data on the effect of the different teaching methods. All but one of the student participants were familiar with the simulated equipment (in automatic mode, but not for manual mode) for of their course topics. The underlying knowledge they entered the study with was measured in the baseline performance video, however the impact this had on their ability to learn new applications for this equipment is unknown.

A further reason the 4SA may have been adapted may relate to the teacher preparation. While many educators are familiar with 4SA, it is not yet a common teaching practice in paramedic professional education within the local ambulance service. 4SA would therefore be unfamiliar to most paramedics within SAAS, and the educator preparation for this study may have been insufficient, or inadequate practise time allowed for the educator. Finally, it may be that 4SA is simply not a *natural* way for educators to teach with, and this may pose inherent difficulties with educator development.

3.5.5 Time to teach

While there is some indication that 4SA takes longer to teach, more data is required in this area. Such small sample sizes are insufficient to detect even impressive differences in time. The magnitude of the potential difference is of high significance to training organisations who will likely be accountable for the time spent training clinicians, and resources spent employing trainers. Where one teaching method may require 25-35% more time to teach, this may translate to a costly impact on the training organisation, and this needs to be understood in light of the impact this has on student learning (if any).

3.5.6 Time to perform

The time to perform manual defibrillation was slightly greater for the group taught with the 4SA, but this was not statistically significant. The difference of 3 seconds is unlikely to reflect clinical significance. Timely and effective defibrillation of VF and ventricular tachycardia (VT) are known to increase the chances of patient survival following cardiac arrest (Chan et al., 2008; M. Larsen et al., 1993; Spearpoint et al., 2000). The literature has identified a one minute delay in defibrillation to relate to a 7 to 10% decrease in survival (Cave et al., 2011), a 3 second delay would therefore relate to a 0.5% reduction, assuming a linear relationship. If the delay is an outworking of patient safety such as ensuring an appropriate (shockable) Electrogardiograph (ECG) rhythm, or safety to other clinicians (such as ensuring that all are clear prior to defibrillation), this may be acceptable. As Nelson (1989) identifies, regardless of how many steps of a procedure are performed correctly, some skills are of a more time critical nature, therefore the time taken to complete effectively ought to be a significant feature of whether a skill has been performed to an acceptable standard.

3.5.7 Limitations

3.5.7.1 Enrolment bias

Students with a keener interest in extra-curricular development are more likely to respond to the invitation to participate in this study. Aside from getting ethics approval to make involvement in the study a compulsory component of their coursework, this potential bias was unavoidable.

3.5.7.2 Sample size

A number of factors contributed to a small sample size. This was compounded by the exclusion of two of the eight teaching groups. This made a cluster analysis of the data impossible, so data were examined according to overall intervention without clustering. The reliability of an ANOVA output may be impacted by small sample sizes (Field suggests degrees of freedom smaller than 20 and the smallest response category containing less than 20% of the responses). Additionally, accuracy of the F-statistic may be impacted by skew when group sizes are not equal (Field, 2013). Therefore efforts will be taken to maximise participant rates in the next study.

3.5.8 Impact this pilot study had on the design of the following study

This pilot study may be considered as a preliminary study for Chapter 5, with the intention that it identify possible challenges and oversights in the planning of that study. As such, some changes were made to the way in which the next comparative trial was approached. Agreement was obtained with students' semester two topic coordinator that the skills taught in the next study would *not* be taught to the students, as per their previously established skill study plan. This would protect the study from this type of contamination between teaching and retention performance, however students were still free to engage in any other teaching activity of their choice (for example through extracurricular student groups) as this data would be obtained at the retention stage of the study. The equipment used to teach the next round of resuscitation skills would be familiar to the educator to ensure this would not become a distraction from their focus on the teaching methods, especially as it is unlikely to locate an available educator who is experienced in 4SA within the ambulance service. More thorough preparation would also be provided to the educator for the next trial. Additionally, nursing students would also be invited to attend the study, in an effort to boost registration, in addition to providing inter-professional insight into the application of resuscitation skills which may be used by other health professionals.

Finally, as a rigorously validated checklist was not used in the data collection for this study, the benefit of performing statistical analyses on such data (such as inter-rater reliability of individual assessment items) was not substantial. It will be more appropriate to perform a more thorough analysis of the assessment and data collection tools used in Chapter 5.

3.6 Conclusion

This study compared changes in student performance between students taught manual defibrillation with 2SA and those taught with $4SA_m$ according to a skill-specific checklist devised to reflect the content in the training sessions, and a global performance scale. No studies to my knowledge have compared the two teaching methods with reference to initial (pre-instruction) skill performance ability. During the study described in this chapter, compliance to the skill procedure as taught in the instruction session and assessed by the skill-specific checklist was greater in the intervention group ($4SA_m$), however this did not significantly impact global rating scores or the time required to defibrillate the patient. These later two measures are arguably the more genuine measures of practical performance ability.

Unanswered questions remain regarding the time (and associated costs) required to teach with 2SA and 4SA. This is a pertinent question for both initial training and reaccreditation bodies who need to consider budget and time restraints relating to clinical training. Indicators exist in this data that 4SA takes longer than 2SA, however current sample sizes are too small to make definitive claims. The next trial will seek to address this question more definitively.

Other changes will also be implemented in order to maximise retention data (6 months following instruction), as a clearer understanding of the patient and clinical impact of resuscitation skill deterioration is required to address a shortfall in the literature. Assessment tools which more reliably reflect the clinical application of the resuscitation skills compared in the next comparative trial will be developed in order to gather date of high clinical credibility.

4 STUDY 2: DEVELOPMENT OF ASSESSMENT TOOLS USING A MODIFIED DELPHI APPROACH

4.1 Introduction

4.1.1 Resuscitation skills

International Advanced Life Support (ALS) protocols may require a clinician to insert a Laryngeal Mask Airway (LMA) (American Heart Association, 2016; ANZCOR, 2015, 2016b; Maconochie et al., 2015; Soar et al., 2015) or Intraosseous needle (IO) (American Heart Association, 2016; ANZCOR, 2016c; Greif et al., 2015; Maconochie et al., 2015; Soar et al., 2015) during the management of cardiac arrest or the deteriorating patient. Patients who are unconscious may be at risk of airway compromise due to the obstruction caused by anatomical position and loss of muscular tone in the airway. A LMA is a supraglottic device which may be used to maintain an open passageway between the mouth and the hypopharynx. This is aimed at overcoming the occlusion of the upper airways caused by poor muscle tone in deeply unconscious patients, and novice practitioners can successfully insert the device in the resuscitation setting (Hein et al., 2010). An IO device may be used to gain access to the venous circulation in situations where Intravenous (IV) access may be delayed, or unsuccessful. In the resuscitation setting, the device is used for fluid and medication administration (American Heart Association, 2016; ANZCOR, 2016c; Greif et al., 2015; Maconochie et al., 2015; Soar et al., 2015).

4.1.2 Education as a bridge between evidence and practice

Among other procedural skills, LMA and IO are part of the recommended approach to both adult and paediatric ALS. Their use is advocated by international advisors on resuscitation (American Heart Association, 2016; ANZCOR, 2015, 2016b, 2016c; Greif et al., 2015; Maconochie et al., 2015; Soar et al., 2015) however the scientific advancement of such recommendations is always limited by the dispersion of awareness (through education) and translation to clinical practice. The Utstein formula is represented in Figure 22 (Søreide et al., 2013) argues that a focus on education has a prominent role in implementing research outcomes.

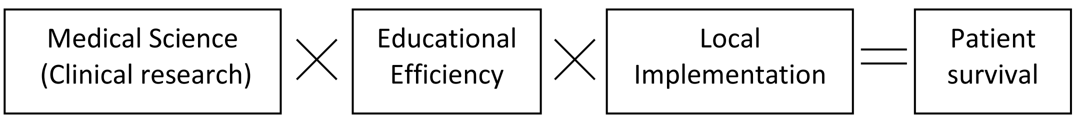


Figure 22: Utstein formula (the formula of survival)

The education and assessment of resuscitation skills in the pre-hospital setting must consider the unique challenges inherent in this environment. In order to appropriately understand the educational bridge between the science and practice of resuscitation skills, the education and assessment strategies must be understood in the appropriate context. I sought to understand the acquisition and retention of LMA and IO insertion and further compare the two teaching methods previously investigated (2SA and 4SA), however in order to do so, I needed to use an appropriate skill assessment tool for the purpose and setting of the study.

4.1.3 The aim of this chapter and how it connects with chapters to come

This chapter outlines the development of an assessment tool for both skills based on expert consensus using a modified Delphi approach. In this chapter, I will use clinical experts' opinions to create a clinically credible checklist, which will be further used in Chapter 5. These tools will aid the rigorous assessment of LMA and manual IO insertion by paramedic students, after no appropriate validated checklists for these skills were located in the published literature. These tools were developed in order to obtain data regarding student performances of LMA and IO, and compare the effectiveness of two skill teaching methods (see Chapter 5). While the primary purpose prompting this assessment tool development was for data collection regarding student performance, the clinical expertise used in developing the tool gives credibility to its use in a wider range of clinical education and assessment settings. This and strategies for further validation of the tools will be addressed in Chapter 6.

Chapters 4, 5 and 6 will therefore form a sub-set of studies aimed at ensuring the core research question (to understand the comparative cost-effectiveness of 4SA) is answered using credible tools (Chapter 4), and with an interrogation of the validity of the tools used and the data produced by them (Chapter 6).

4.1.4 Global rating checklists

Van der Vleuten (2000) argues that accreditation, education and assessment bodies have a social responsibility not only in teaching, but also in accurately assessing graduates. Clinical students will enter the workforce as health professionals who will be trusted by the public to competently meet their needs, and assessment procedures should "provide a guarantee to society that the training programme delivers competent [health care professionals]" (p. 1217). Global assessment scales have been used by expert assessors with much reliability and validity, however these tools depend on an expert assessor who is able to make accurate judgments based on their clinical experience (Doyle, Webber, & Sidhu, 2007). The development of skill-specific checklists would also present an education and assessment tool for use by less experienced assessors, as global ratings tend to be more accurate when used by more experienced assessors (Govaerts, Van der Vleuten, Schuwirth, & Muijtjens, 2007). Huang et al. (2009) argue that their work in developing a procedural assessment tool supports the argument that assessing student clinical performance in a simulated setting against strict criteria based assessment checklists "tend to be highly reliable and valid approaches to technical skills assessment compared with gestalt observations." Clearly the clinical education community is divided on this issue, with other researchers supporting global scales such as the global rating index for technical skills (GRITS) tool (Doyle et al., 2007), and the Integrated Procedural Performance Instrument (IPPI) (Kneebone et al., 2006) demonstrating superior concurrent validity than checklists alone (Brady et al., 2015).

4.1.5 What this study adds

This study will outline the development of two skill-specific checklists to guide the education and assessment of two resuscitation skills: IO insertion and LMA insertion. The study will use manufacturers checklists, other published guides and the opinions of experienced pre-hospital clinicians to develop checklists which reflect authentic clinical practice, and therefore give clinical credibility to the output data. The development of these checklists will be critiqued in Chapter 6 once they have been applied. This chapter, in conjunction with the two which follow, will come together to ask crucial questions about skill assessment, and validation processes, and as such they should be considered as a sub-series within this thesis. Chapter 4 outlines the development. Chapter 5 applies the checklists. Chapter 6 critiques the development process, application, and interpretation of data based on modern approaches to validation procedures. Thus, this study is the first step in a series which approaches skill assessment in a well-rounded way. The assessment tool development and critique are essential bookends for the Chapter 5, which would demand excessive trust from the reader without them.

4.2 Methodology employed in this study

4.2.1 Research aim

The research imperative for this study is to understand what expert clinicians regard as important when they perform these skills in the practical setting. This foundation ensures that the resultant checklists are based on current expert practice in the pre-hospital context, in order to provide an authentic assessment tool reflecting a student's ability to practice. The intention of this study is to build a set of criteria relating to each skill based on clinical practice. To achieve this, the participant sample was targeted, to ensure a panel of experts who had credible and current qualification, experience, and opinions regarding the two skills. I sought to understand the value of various aspects of the skill performance to our participants, and in identifying these values and priorities, to construct a new meaning of these two skills in practice.

4.2.2 The qualitative approach

While the post-positivist approach in the previous chapter was suitable to ask whether a difference exists in the skill acquisition between students taught with two different methods, that approach is limited in its ability to draw knowledge from a range of subjective value statements. This may be one of a number of reasons for the RCT's place as "gold standard", true scientific research, protected from the bias of the researchers is challenged by some as the pinnacle of evidence (Cartwright, 2007; Kaptchuk, 2001; Sullivan, 2011). The ontological basis for this present study upholds that knowledge is formed within the subject on the basis of their experience (Brink et al., 2012, p. 25), rather than uncovered by means of a cause and effect trial. Natural variability in the opinions expressed by participants is expected, and will give strength to the validity of the data. However, the participants' belonging to a professional group will necessarily form a common ground from which the group will construct a meaning of both skills reflective of the common experience, culture of practice, and social expectations of that professional membership (Brink et al., 2012, p. 25). While natural variations will occur, common ground will be identifiable due to the subjects' belonging to a particular culture or role. These values and priorities will have been individually informed by education, training, professional experience and reflection on theirs and others' practice. It can be measured by a qualitative approach which seeks to articulate the meaning and priorities clinicians maintain during the application of the two skills being studied. By examining a variety or perspectives, this study aims to distil a common basis from which to identify expert consensus. In his chapter *Particularly appropriate qualitative applications*, (Patton, 2002a, p. 193) confirms the appropriateness for a qualitative study design in this type of assessment tool development.

4.2.3 Constructivism and social constructionism

The constructivist framework seeks to understand knowledge on the basis of the meaning which objects, processes, concepts or relationships have for the individual (Berger, Luckmann, & Zifonun, 2002; Creswell, 2013, p. 8). These are subjective, varied and numerous, thereby leading the researcher to understand the breadth of such views in order to understand the meaning. This move from objective truth, the presence of which is assumed to be true throughout the study design in Chapter 3, to one of perceived value is a necessary shift to address the question within this chapter. This shift is referred to by Lincoln and Guba as a "move from the objective to the perspective" with true objectivity noted as an *illusion* (Lincoln & Guba, 1985, p. 55 & 72). This sense of a constructed reality is well suited to inform the methods by which I seek to answer the question, however it lies in some tension with the focus of consensus, which is concerned less with the variety of meanings, and more so in the commonalities within this variety. As such, the social constructionist tradition, which may also be referred to as social constructivist (Creswell, 2013, p. 8) is one which values the meaning created not just by individuals, but by what the individuals bring to the meaning developed by the group as a dynamic entity (Crotty, 1998, p. 57). The Delphi approach is a useful technique for obtaining "pooled judgement" (Moore, 1987), as it draws individuals' meanings into a collaborative meaning where, despite anonymity of participants, interaction occurs and individual participants may adapt their responses based on group data.

4.3 Methods

4.3.1 The Delphi method

The Delphi technique is "a method of eliciting and refining group judgments" (Dalkey, Brown, & Cochran, 1969) with the aim of achieving agreement (Keeney, Hasson, & McKenna, 2011). The Delphi method can be useful when current information or direction is contradictory or insufficient (Hasson, Keeney, & McKenna, 2000). It is a multi-stage approach initially developed to identify expert agreement in the predictive military setting (Keeney et al., 2011) . A strength of this process is the distillation of expertise from a panel with identifiable expertise relevant to the topic, based on the notion that "two heads are better than one" (Clayton, 1997).

The Delphi approach has received criticism for lacking methodological rigour, among other critiques (Hasson et al., 2000). One cause of this intellectual concern is the variance between different Delphi studies (Boulkedid et al., 2011). However this challenge can also be viewed as the strength and adaptability of the study type (Keeney et al., 2011). Despite the variability seen, some features are common to all Delphi studies. Key features of a Delphi study are participants' anonymity to the wider group, controlled feedback between multiple iterative stages, and utilising a range of group responses to estimate an aggregated score for each item (Dalkey et al., 1969; Hasson et al., 2000).

4.3.2 The Modified Delphi approach

To create an assessment checklist appropriate for teaching and assessment in the pre-hospital emergency setting, I conducted a modified Delphi study. A conventional Delphi would ask the group to contribute the items to formulate the initial list, then progress in stages where the items were scored, refined, and a final agreed list would remain (Clayton, 1997; Hasson et al., 2000; Keeney et al., 2011). A modified Delphi, however, uses a preliminary list generated outside the panel, either by the research team, or other sources (Keeney et al., 2011). This method was chosen for this study in order to relieve the demand required on the busy clinicians who chose to participate in the study, and therefore to optimise completion. The project design was based on previously published studies with a similar aim (see Table 17).

4.3.2.1 Ethics

Ethics approval was obtained from the Social and Behavioural Research Ethics Committee (SBREC) at Flinders University, in addition to the ethics committees responsible for research in SA Health, and South Australian Ambulance Service (SAAS).

4.3.2.2 Expert panel

The concept of an expert is both a social and a scientific phenomenon (Clayton, 1997).The illusion of expertise is identified as a potential weakness of this type of study (Keeney et al., 2011), and as such, *expert* was clearly defined for this study as: formally trained in LMA and IO insertion; experienced in both LMA and IO insertion (with a stated minimum number of insertions into real patients); and authorised to practice either as a SAAS Intensive Care Paramedic (ICP) or SA Health Medstar (retrieval) Clinician. Due to the relative infrequency of LMA and IO insertion (Carley & Boyd, 2004; Hallas, 2012), I defined "experienced" as a minimum of two LMA and IO insertions into real patients in the last five years.

I used stratified sampling in order to capture views from a representative expert panel of intensive care paramedics (ICPs) and a specialist group of paramedics, nurses and consultants who work as pre-hospital and emergency retrieval specialists known as Medstar clinicians. ICPs and Medstar clinicians have complementary experiences from which to draw an opinion, and the respondent mix was stratified to be almost equally representative from both of these groups. While all participants are pre-hospital clinicians, understanding the wide range of experience and qualification both within and outside of hospital provides assurance that the group is indeed heterogeneous. Nine respondents is considered sufficient for this purpose given the participants were a heterogeneous representation of pre-hospital emergency clinicians who employ these skills (Clayton, 1997). This panel therefore maximises the credibility and clinical relevance of the results obtained (Boulkedid et al., 2011).

4.3.2.3 Recruitment

All potentially appropriate clinicians were invited to participate in the online study by electronic mail which outlined the study purpose and process. As participants registered interest, they were provided with a link to the online survey for the first stage of the study. Registration was closed once nine clinicians had completed the first survey.

4.3.2.4 Anonymity

The anonymity upon which the process is built provided a level ground between all participants, and controlled the influence of more dominant respondents, which may be a cause for bias in a face-to-face group collaboration (Clayton, 1997; Dalkey et al., 1969). As I was aware of the respondents' identities and their scores, the term "quasi anonymity" may be more accurate (Hasson et al., 2000).

4.3.2.5 Preliminary checklist

The preliminary checklists for the first Delphi round contained items collated from manufacturer's instructions (CareFusion™, 2014; "LMA Classic Excel™," ; "LMA Classic™, LMA Flexible™, LMA Flexible™ Single Use & LMA Unique™,"), current teaching practice in our local context, and other published literature (Hein, 2009; Hein et al., 2010; Lammers et al., 2009; Oriot, Darrieux, Boureau-Voultoury, Ragot, & Scépi, 2012).

4.3.2.6 Round 1

Panel members were asked to rank each item on the preliminary checklist from 1 (*not important at all*) to 9 (*mandatory*). The 9-point Likert scale and scoring system described by Huang et al. (2009), was used for this study. In this scale, 1-3 represented *not important*, 4-6 represented *somewhat important,* and 7-9 represented *very important*. The scale is a deviation from a purist Likert approach, where the central point is effectively neutral, and each extreme reflects strong agreement or strong disagreement (Clayton, 1997). The scale used here, as with other previously published studies of this type measures only level of agreement, rather than disagreement also, and this adaptation is consistent with the research question. Berg et al. (2014) employed a scale of 1 to 7, where 1-3 was not *important*, 4-6 was *somewhat important*, and 7 was *mandatory*, resulting in an asymmetrical scoring scale. The mean score at which items were included varied slightly for Berg et. al.s' studies, but tended around 5.5. When translated to a hypothetical inclusion score for a 9-point scale, the mean score which separated inclusion and exclusion was approximately 7.01 (see Table 17).

Participants were also asked to suggest amendments to current items, suggest additional items, and indicate whether they believed that incorrect performance of each item would associate with an increased patient morbidity or mortality risk if performed incorrectly. Items with a mean score of 7.0 or higher were considered to demonstrate consensus and included in the final checklist, based on the mean scores presented in Table 17.

Table 17: Summary of Previous Delphi Studies Determining Expert Consensus

Year	Reference	Scale	Panel size	Statistical feedback to panel members	Point of automatic exclusion	Minimum mean score of items in final list	Minimum mean score of items in final list (adjusted for scale of 1-9) ^	Minimum median score all items in final list	Number of rounds	IRCC	List length
2009	Huang, G. C., Newman, L. R., Schwartzstein, R. M., Clardy, P. F., Feller-Kopman, D., Irish, J. T., & Smith, C. C. (2009). Procedural Competence in Internal Medicine Residents: Validity of a Central Venous Catheter Insertion Assessment Instrument. *Academic Medicine, 84*(8), 1127-1134.	1-9	7	*M*, med *SD*	*M*<3	6.0	6.00	7	2	α= .94	24
2013	Berg, D., Berg, K., Riesenberg, L. A., Weber, D., King, D., Mealey, K., . . . Tinkoff, G. (2013). The Development of a Validated Checklist for Thoracentesis: Preliminary Results. *American Journal of Medical Quality, 28*(3), 220-226.	1-7	8	*M*, med *SD*	*M*<3	5.5	7.07	6.5	2	α= .94	23
2013	Berg, K., Riesenberg, L. A., Berg, D., Mealey, K., Weber, D., King, D., . . . Tinkoff, G. (2013). The Development of a Validated Checklist for Adult Lumbar Puncture Preliminary Results. *American Journal of Medical Quality*, 28(4), 330-334.	1-7	9	*M*, med *SD*	*M*<3	5.8	7.46	6	2	α= .79	20
2013	Riesenberg, L. A., Berg, K., Berg, D., Schaeffer, A., Mealey, K., Davis, J., . . . Tinkoff, G. (2013). The development of a validated checklist for nasogastric tube insertion: preliminary results *American Journal of Medical Quality, 28*(5), 429-433.	1-7	9	*M*, med *SD*	*M*<3	4.8	6.17	5	2	α= .80	19

2013	Riesenberg, L. A., Berg, K., Berg, D., Mealey, K., Weber, D., King, D., . . . Tinkoff, G. (2013). The Development of a Validated Checklist for Paracentesis: Preliminary Results. *American Journal of Medical Quality, 28*(3), 227-231.	1-7	8	*M*, med *SD*	*M*<3	5.5	7.07	6	2	α= .92	24
2014	Berg, K., Riesenberg, L. A., Berg, D., Schaeffer, A., Davis, J., Justice, E. M., . . . Jasper, E. (2014). The Development of a Validated Checklist for Radial Arterial Line Placement: Preliminary Results. *American Journal of Medical Quality, 29*(3), 242-246.	1-7	9	*M*, med *SD*	*M*<3	5.9	7.59	6	2	α= .99	22
2014	Riesenberg, L. A., Berg, K., Berg, D., Davis, J., Schaeffer, A., Justice, E. M., & Tinkoff, G. (2013). The Development of a Validated Checklist for Femoral Venous Catheterization Preliminary Results. *American Journal of Medical Quality, 29*(5), 445-450.	1-7	8	*M*, med *SD*	*M*<3	6.3	8.10	7	2	α= .99	29
2014	Hartman, N., Wittler, M., Askew, K., & Manthey, D. (2014). Delphi method validation of a procedural performance checklist for insertion of an ultrasound-guided internal jugular central line. *American Journal of Medical Quality, 31*(1), 81-85.	1-9	13	*M*, *SD*	*M*<3	6.62	6.62	Not reported	2	α= .94	30

Note: IRCC refers to the inter-rater consistency coefficient used. Data were used to inform the design and decisions made in the current study. ^ the adjusted minimum inclusion mean depicts an equivalent value for a 9-point scale. The average lowest included mean for each study, converted to an equivalent for a 9-point scale, was 7.01. α refers to Cronbach's alpha. *M* = mean, *SD* = standard deviation, med = median.

4.3.2.7 Controlled feedback

The feedback normally provided to respondents between rounds typically includes a central measure of tendency and a measure of dispersion (Keeney, Hasson, & McKenna, 2006). Where comments submitted by individuals are fed back anonymously to the group at each stage, the respondents have the opportunity to refine their opinions based on others' rationale, and hence a more accurate group opinion is sought (Boulkedid et al., 2011; Clayton, 1997). In this study, feedback to participants was limited to mean scores, standard deviation and additional items for scoring. Participant rationale for their score of each item was not invited, as this was seen to pose a risk to completion rates. Participants were all shift-workers and due to the nature of their role, had limited access to a computer during their work shift, and asking for rationale of each item was seen to be an excessive burden which may jeopardise the study. Participants were able to provide rationale if they wished, in addition to suggestion of further items for inclusion.

4.3.2.8 Opportunity for participants to change their scores

Respondents had an opportunity to review their score, based on the group average, in order to improve the accuracy and reliability of study findings (Dalkey et al., 1969). As the same cohort of participants is required to complete all stages of the study, a risk to the study is incompletion of both stages (Hasson et al., 2000). For this reason participants were not asked to review the scores of all items, but rather only those which were not strongly indicated for exclusion or inclusion (based on a mean of 3 or less, or 7 or greater, respectively). Again, this was in order to alleviate the workload of clinicians and maximise chances of completion as this type of study can be seen as tedious or repetitive for participants.

4.3.2.9 Round 2

Items with a mean score between 3.0 and 7.0 were re-distributed for re-scoring along with the new items. Some of these mid-range scores were anticipated to change in light of the inclusion or exclusion of other criteria. All panel members' contributions held equal weight. The final round was identified by a failure to identify new items for consideration.

4.3.3 Agreement

Agreement was defined by a mean score of 7.0 or greater for that item. This ensures that, on average, the panel regards each item as at least "very important". Given this definition of agreement, and the methodological underpinning of this study being steeped in the constructivist approach, it is not appropriate to perform further analytics on these values, for example for the purposes of inter-rater consistency analysis as performed by other studies of this kind. Kane (2006, p. 27) reminds us of our preoccupation with numerical certainty, likely due to the perceived objectivity understood by a numerical value. Furthermore, when we understand that the scores awarded in this study represent a value judgement, rather than *an amount of something*, we see that no certainty is provided by an analysis of these numbers in any case. The qualitative philosophy guiding the study is not consistent with a focus on statistical arguments for or against consistency, as the findings are constructions of a new truth/concept, rather than findings or discovery of a previously hidden truth. Variance among the raters would be expected for LMA insertion in particular, due to the nature of this skill. This equipment may be inserted in a variety of ways, all of which may be correct and appropriate. Factors such as this are not accounted for with consistency analysis, and therefore it is not suitable for this study.

The studies which informed this one all report on the internal consistency of the scale using Cronbach's Alpha (α), with reference to the mean and standard deviation as measures of central tendency and dispersion. Some studies include the median score. It is unclear on what basis items are included or excluded, however a mean value tends to delineate the two. Cronbach himself later doubted "that the [alpha] coefficient was the best way of judging the reliability of an instrument to which it was applied" (Cronbach & Shavelson, 2004) and, in apparent distress of the misunderstandings surrounding the use of α, later argued that α is only a small part of what should be wider analysis in investigating the reliability of a test (Cronbach & Shavelson, 2004)

4.3.4 Validity

4.3.4.1 Addressing validity in the study design

Potential sources of threats to validity were managed in the study design. Guided by Creswell's outline, these are presented in Table 18 below (Creswell, 2013, pp. 174-176). Some aspects did not apply as the design was not experimental or comparative between cohorts.

4.3.4.1.1 Threats to internal validity of the study are addressed

Table 18: Threats to Internal Validity

Potential threat to validity	How the threat was addressed in the study design
History	Some participants delayed the completion of the second round of the Delphi study. This increased their time to reflect and exposure to atypical cases which may impact their practice for the skills in question, but this was not expected to change their views of the procedures substantially as their expertise had been developed over years of clinical practice. There were no changes to practice communicated to the participants from their employers during the time of the study, to our knowledge.
Maturation	Data collection took place over a time period considered too brief to pose a significant risk to validity in this way.
Regression	N/A
Selection	Participants were selected due to their experience and expertise. As a result, they may hold slightly different perspectives than newly qualified

	clinicians who may not have similar experience. This selection bias was intentional, to capture truly expert and experienced clinicians.
Mortality (or Drop-out)	There were no participant deaths prior to the study completion. This risk to validity closely relates to the risk of participant drop-out. The risk of participant failure to complete the second round of the Delphi was addressed to some extent through the modified design (removal of the primary Delphi round) which ordinarily sees the Panel build the preliminary checklist themselves. This lessened the workload with the aim of improving participant completion of both rounds.
Diffusion of treatment	N/A
Compensatory demoralisation	N/A
Compensatory rivalry	N/A
Testing	N/A
Instrumentation	The rating scale (1-9) remained standard between the two rounds of the study. Individual criteria developed and changed between the two rounds as part of the study design.

Note: items which were not considered a significant threat to the study's validity are marked as *not applicable* (N/A)
Adapted from "Research design" by John Creswell, pp. 174-176, copyright 2013 by Sage publications.

4.3.4.1.2 Potential threats to external validity of the study are addressed

As I considered the implementation and reporting of the study results, I made effort to recognise the impact of external threats to the validity of the study outcomes and conclusions, again guided by Creswell in Table 19 below:

Table 19: Threats to External Validity

Potential threat to validity	How the threat was addressed in the study design
Interaction of selection and treatment	This aspect limits our ability to generalise the findings to the wider international pre-hospital care workforce due to participant membership to a local service. As such, I have made efforts to encourage adaption to local practice, and further validation.
Interaction of setting and treatment	The application of these results to the emergency and pre-hospital environment is appropriate, however implementation in other medical settings such as inserting an LMA in the surgical theatre become less valid.
Interaction of history and treatment	The results gained reflect the views and practice of a cohort of clinicians in 2015. Their views have developed from their training and current organisational practice, and therefore as further research becomes available in coming decades, we may see aspects of these practices evolve. The checklists, therefore, are not timelessly prescriptive.

Adapted from "Research design" by John Creswell, pp. 174-176, copyright 2013 by Sage publications.

4.3.4.2 Assessing validity of the study output

The panel's anonymity to each other, their expertise, and their diversity all contribute to the scale's validity. The Delphi approach assumes that a panel of people are less likely to make the error that one person alone might make (Hasson et al., 2000). Non-completion of round 2 posed a potential risk to the study's validity, and diligent follow-up with participants resulted in full completion of both rounds. As outlined in the study's methods, open questions seeking rationale for each participant's rating, and subsequent feedback of this rationale to the group between rounds were limited to alleviate the workload on respondents (see section 4.3.2.7). This deviation from a core aspect of this type of study restricted the panel members' understanding of each other's rationale from their scores.

4.4 Results

4.4.1 Demographic information

Nine clinicians participated in the study: five ICP/ECPs, one Rescue Paramedic, and three Medstar clinicians (such as retrieval physician, anaesthetist, or retrieval nurse). All participants had at least two years' clinical experience (see Table 20).

Table 20: Participant Demographic Data

		Number of participants (total n=9)
Clinical background	Intensive Care Paramedic	5
	Rescue Paramedic	1
	Medstar (non-Paramedic)	3
Age of participant in years	30-39	3
	40-49	3
	50-59	3
Years authorised to insert IO	2 to 4	2
	5 to 10	1
	11 or more	6
Number of IOs inserted in real patients (in past 5 years)	1	0
	2	0
	3 to 4	2
	5 to 10	4
	11 or more	3
Number of (manual) IOs inserted in real patients (in past 5 years)	1	2
	2	1
	3 to 4	1
	5 to 10	4
	11 or more	1
Years authorised to insert LMAs	2 to 4	2
	5 to 10	1
	11 or more	6

Number of LMAs inserted into real patients (5 years)	1	0
	2	0
	3 to 4	3
	5 to 10	2
	11 or more	4

4.4.2 Response rates for all rounds

Data collection commenced prior to completion of registration. More than 5 ICPs volunteered to participate in the study, however as the study was expected to require only 8 to 9 participants, after 5 ICPs had completed the first round of the study the remaining places were allocated for Rescue paramedics and Medstar clinicians. All participants who completed Stage 1 also completed the second stage of the process.

4.4.3 IO

4.4.3.1 Checklist items

Of the 29 preliminary items, 18 were automatically included in the first round (indicated by a mean score of 7.0 or greater). Two additional items were suggested by the panel, and these were distributed for scoring with the eleven items which returned a mean score between 3.0 and 7.0 after round one. Four of these items assessed in the second round were included in the final checklist. There were 22 items in the final checklist. The mean score of included items ranged from 7.0 to 9.0. All items had a median score of 7 or above. These data are presented in Table 21 with information on central tendency (mean and median), and range including Standard deviation (*SD*), and Interquartile range (IQR).

4.4.3.2 Morbidity/mortality

In terms of anticipating risk to the patient during IO insertion, one item was highlighted as being likely to increase the risk of both mortality and morbidity when performed incorrectly, and three additional items were highlighted to likely increase morbidity. This was indicated by seven or more panel members indicating "yes" or "possibly" to this survey question.

Table 21: IO Checklist Item Scores, Inclusion/Exclusion Decisions and Morbidity/Mortality Data.

Item no	Item	Round 1 mean (SD)	Round 1 Median (IQR)	Round 2 mean (SD)	Round 2 Median (IQR)	Decision	Mortality risk[##]	Morbidity risk[**]
1	Clinician wears gloves	9.00 (0.00)	9 (0)			Included at round 1	5	5
2	Clinician wears safety glasses	7.67 (2.50)	9 (1)			Included at round 1	0	1
3	IO inserted into appropriate location (generally)	8.89 (0.33)	9 (0)			Included at round 1	7[#]	7*
7	IO is inserted into flat surface of proximal Tibia	6.78 (3.08)	8 (3)	7.44 (1.24)	8 (1)	Included at round 2	2	4
8	IO is inserted 1-2 cm (1-2 fingers) below tibial tuberosity	8.00 (0.87)	8 (2)			Included at round 1	2	6
8.1	IO is inserted 1-2 cm (1-2 fingers) below tibial tuberosity then in line with medial tibia			7.67 (1.00)	8 (1)	Included at round 2	4	6
9	IO site is prepared (non-specific)	7.56 (1.51)	7 (2)			Included at round 1	6	8*
10.1	IO insertion site swabbed with alcoholic chlorhexidine and allowed to dry (2-3 seconds)			7.11 (2.26)	8 (1)	Included at round 2	5	7*
11	IO insertion site swabbed with antiseptic	7.56 (1.42)	8 (2)			Included at round 1	6	6
13	Holds limb/body part secure during IO insertion	7.56 (2.60)	9 (2)			Included at round 1	2	4
17	IO needle inserted at 90 degrees to bone	8.44 (1.01)	9 (1)			Included at round 1	3	5
18	IO needle rotated back and forth along its axis	7.33 (2.74)	9 (2)			Included at round 1	3	4
19	Apply gentle, constant pressure during insertion	8.33 (0.87)	9 (1)			Included at round 1	2	3
20	Clinician stops applying pressure once the resistance of the outer cortex of the bone subsides (like a "pop")	8.56 (0.53)	9 (1)			Included at round 1	5	8*
21	Depth guard wound down so it is flush with the skin	7.22 (2.05)	8 (4)			Included at round 1	4	4
22	Depth guard not over-tightened	7.44 (1.74)	8 (3)			Included at round 1	4	5

23	Blue cap removed from top of IO	9.00 (0.00)	9 (0)			Included at round 1	3	3
24	Stylet removed from inside IO needle	9.00 (0.00)	9 (0)			Included at round 1	3	3
25	Stylet discarded in sharps container	9.00 (0.00)	9 (0)			Included at round 1	0	0
26	IO visually assessed to stand in the bone without needing additional support	7.67 (1.58)	8 (2)			Included at round 1	3	4
27	IO placement assessed by either aspirating bone marrow or flushing with Normal Saline	6.78 (3.26)	8 (4)	7.67 (1.66)	8 (3)	Included at round 2	5	6
28	IO placement assessed by aspirating bone marrow and flushing with Normal Saline	8.22 (1.99)	9 (0)			Included at round 1	5	6
4	IO site penetrates minimum muscle	6.78 (2.04)	7 (2)	6.11 (2.32)	7 (2)	Excluded at round 2	3	4
5	IO site chosen to avoid damaging superficial blood vessels	6.11 (2.41)	7 (2)	4.33 (2.24)	4 (1)	Excluded at round 2	4	5
6	IO is inserted into medial surface of proximal tibia	5.78 (2.95)	6 (3)	6.67 (1.73)	7 (2)	Excluded at round 2	2	4
10	IO insertion site swabbed with iodine	6.78 (2.30)	7 (3)	6.22 (2.28)	7 (2)	Excluded at round 2	7[#]	7*
12	IO insertion site swabbed in one direction with one side of swab, then in same direction with other side of swab	3.89 (2.72)	5 (4)	3.22 (2.22)	3 (2)	Excluded at round 2	2	3
14	Knee flexed for insertion	4.00 (2.95)	4 (5)	4.00 (2.87)	5 (4)	Excluded at round 2	0	0
15	Place blanket under knee to slightly bend limb for insertion	3.78 (2.32)	4 (4)	2.89 (2.31)	2 (3)	Excluded at round 2	0	0
16	IO needle held with one end in palm of hand, and the index finger resting ~1cm above bevel of needle to avoid inserting too deep	5.56 (3.32)	6 (5)	4.89 (2.71)	6 (5)	Excluded at round 2	2	3
29	IO placement assessed just by flushing with Normal Saline	5.00 (2.72)	6 (3)	5.33 (2.00)	5 (2)	Excluded at round 2	5	6

Note: Items in shaded cells are excluded from the final checklist due to a final mean score of <7.
* indicates an item which is expected to increase morbidity risk if performed incorrectly
** the number of clinicians who indicated that incorrect performance at this part of IO insertion could possibly, or would likely increase the risk of mortality
[#] indicates an item which is expected to increase mortality risk if performed incorrectly. Shaded boxes indicate items which were excluded on the basis of final mean score.
[##] the number of clinicians who indicated that incorrect performance at this part of IO insertion could possibly, or would likely increase the risk of morbidity
SD = standard deviation; IQR = Interquartile range

4.4.4 LMA

4.4.4.1 Checklist items

The initial checklist of 34 items returned mean scores ranging from 3.0 to 8.89 in the first round. One item (item 10) was excluded from the second round of the checklist as the mean score was 3.0. 13 items (1, 2, 3, 4, 6, 9, 17, 21, 25, 28 31, 32 and 34) had a mean score of 7.0 or above, and were therefore included in the final checklist.

Four new items (5.1, 17.1, 25.1 and 25.2) were suggested by the participants for inclusion in the second round. After the second round of the survey, two more items (7, 8) were excluded and three additional items (5.1, 20, and 29) were included as the mean score reached the pre-determined threshold of 7.0. No new items were suggested during this round, so no further rounds were conducted. The 16 items with a final score of 7.0 or greater were included in the final checklist. The included items ranged in mean from 7.00 to 8.89 and had a median of 7 or above.

4.4.4.2 Morbidity/mortality

An increased risk to expected morbidity and mortality was determined with regards to eight items (items 3, 4, 9, 28, 29, 31, 32 and 34). This was indicated by seven or more panel members indicating "yes" or "possibly" for these items.

Table 22: LMA Checklist Item Scores, Inclusion/Exclusion Decisions and Morbidity/Mortality Data.

Item no	Item	Round 1 mean (*SD*)	Round 1 Median (IQR)	Round 2 mean (*SD*)	Round 2 Median (IQR)	Decision	Mortality risk##	Morbidity risk**
1	Clinician wears gloves	8.89 (0.33)	9 (0)			Included at round 1	3	3
2	Clinician wears safety glasses	7.56 (2.35)	9 (3)			Included at round 1	0	0
3	Patient is pre-oxygenated	8.67 (0.71)	9 (1)			Included at round 1	9#	9*
4	Appropriate size LMA is selected	8.67 (0.71)	9 (1)			Included at round 1	9#	9*
5	Remove LMA from packet			8.22 (1.99)	9 (0.5)	Included at round 2	6	6
6	LMA tubing is checked to be free from blockage or debris	7.00 (2.78)	9 (4)			Included at round 1	5	6
9	Integrity of cuff is checked by inserting air into it and ensuring a seal	7.44 (2.65)	9 (3)			Included at round 1	7#	7*
17	Posterior side of cuff lubricated	7.78 (2.28)	9 (1.5)			Included at round 1	2	3
20	Patient's neck is flexed and their head extended	6.89 (2.20)	7 (2.5)	7.00 (1.58)	7 (2)	Included at round 2	4	4
21	LMA inserted with bowl of mask facing anterior	7.78 (1.72)	9 (3)			Included at round 1	5	6
25	LMA inserted in a forward sweeping motion	7.11 (1.54)	7 (3)			Included at round 1	3	4
28	Cuff inflated	7.44 (3.13)	9 (3)			Included at round 1	7#	7*
29	Cuff inflated with no more than maximum stated volume of air	6.89 (3.14)	8 (4.5)	7.00 (2.18)	7 (3.5)	Included at round 2	7#	8*
31	BVM connected while holding LMA secure (before securing device)	7.89 (1.17)	8 (2)			Included at round 1	8#	8*
32	Ensure chest rise and fall with BVM	8.78 (0.67)	9 (0)			Included at round 1	9#	9*
34	LMA secured into place	8.67	9			Included at round 1	7#	7*

		(0.50)	(1)					
5	LMA selected based on patient's weight	6.33 (1.94)	5 (3.5)	6.56 (1.33)	7 (3)	Excluded at round 2	5	5
7	Integrity of LMA plastic tubing is checked by bending it to 90 degrees	4.22 (2.86)	4 (5.5)	2.78 (1.09)	3 (2)	Excluded at round 2	3	3
8	LMA tube is straightened to 180degrees to ensure it does not kink	3.44 (2.40)	5 (4)	2.44 (0.88)	2 (1)	Excluded at round 2	4	4
10	Cuff fully deflated and re-inflated with 150% of the maximum inflation value advised; and examine cuff for leaks, herniation, uneven bulging, and even inflation of the blue pilot balloon	3.00 (3.04)	1 (5)			Excluded at round 1	3	3
11	Cuff fully deflated and reinflated with 100% of the maximum inflation value advised; and examine cuff for leaks.	6.78 (2.44)	7 (4)	5.44 (2.55)	6 (3.5)	Excluded at round 2	6	6
12	Airway connector is securely connected to the LMA tube (as part of LMA device check)	3.67 (3.12)	2 (5.5)	4.44 (2.24)	4 (2.5)	Excluded at round 2	4	4
13	LMA checked for discoloration	6.89 (2.85)	9 (4)	5.78 (2.68)	7 (4)	Excluded at round 2	2	3
14	Inflation line is checked as secure by pulling it gently	4.22 (1.99)	5 (3.5)	3.44 (2.07)	3 (3.5)	Excluded at round 2	5	6
15	Ensure the two aperture bars traversing the mask of the LMA are intact	3.67 (2.65)	3 (4)	3.56 (1.94)	3 (3)	Excluded at round 2	2	2
16	LMA fully deflated prior to inserting	3.89 (2.71)	5 (5)	3.56 (2.24)	4 (4.5)	Excluded at round 2	2	2
17.1	Lubricate Both sides of cuff			5.11 (3.30)	7 (6.5)	Excluded at round 2	3	4
18	Anterior side of cuff is not lubricated	6.22 (3.93)	9 (8)	4.00 (3.32)	3 (6.5)	Excluded at round 2	3	3
19	Assisting clinician is asked to apply Jaw thrust if necessary	6.11 (3.18)	7 (6)	6.22 (2.77)	7 (3.5)	Excluded at round 2	4	4
22	Head extension maintained during insertion by holding under the patient's occiput with the non-insertion hand	5.44 (2.83)	5 (5.5)	5.22 (2.05)	6 (2.5)	Excluded at round 2	4	5
23	Mask held like a pen with index finger anterior to the tube, at the junction of the cuff and the tube	6.56 (2.60)	7 (4)	5.89 (2.80)	7 (4.5)	Excluded at round 2	3	3
24	Tip of LMA is pressed against the hard palate, and pushed along the posterior pharyngeal wall using the index finger, keeping other fingers out of the patient's mouth.	6.89 (2.57)	7 (3)	5.44 (3.17)	7 (6)	Excluded at round 2	4	5

25.1	LMA inserted into pharynx either directly or rotated on insertion until resistance is felt at about 10cm insertion			6.44 (2.55)	7 (3.5)	Excluded at round 2	6	6
25.2	withdraw 1cm to allow for any potential folding to spring back and the push back in 1cm			4.33 (2.65)	4 (5.5)	Excluded at round 2	5	5
26	Once resistance is felt, tube is held stable with non-insertion hand, and withdraw insertion finger from the mouth.	6.89 (2.42)	7 (2)	5.22 (1.92)	5 (2.5)	Excluded at round 2	4	5
27	Clinician checks that the black line on the tube faces the patient's upper lip	6.67 (2.60)	7 (3.5)	6.44 (2.24)	7 (1.5)	Excluded at round 2	6	6
30	Cuff inflated without holding it secure to allow it to settle into place	6.22 (2.86)	7 (5)	6.11 (2.80)	7 (5)	Excluded at round 2	5	6
33	bite-block inserted	6.44 (1.67)	6 (3)	5.78 (1.99)	6 (1.5)	Excluded at round 2	6	6

Note: Items in shaded cells are excluded from the final checklist due to a final mean score of <7.
* indicates an item which is expected to increase morbidity risk if performed incorrectly
** the number of clinicians who indicated that incorrect performance at this part of LMA insertion could possibly, or would likely increase the risk of mortality
[#] indicates an item which is expected to increase mortality risk if performed incorrectly. Shaded boxes indicate items which were excluded on the basis of final mean score.
[##] the number of clinicians who indicated that incorrect performance at this part of LMA insertion could possibly, or would likely increase the risk of morbidity
SD = standard deviation; IQR = Interquartile range

4.5 Discussion

This study establishes a clinically authentic skill performance checklist for both IO and LMA use in the pre-hospital context, using a modified Delphi method. The study aimed to determine which aspects of LMA and IO insertion were common to a panel of experts, with reference to their clinical practice. The clinical credibility and authenticity of the resultant checklists make reasonable the argument that they may be used to make judgements about a person's performance of these skills in a simulated environment, with consideration of the professional expectations for the skill application.

4.5.1 Participant Demographics

The demographic information gathered on study participants reflects a diverse group of individuals. While establishing a minimum criteria to ensure each individual brought definable expertise, the data also reflects the diversity of the key clinical tiers most likely to use both skills (ICP, Rescue paramedic and Medstar clinician). The study captured participants with a range of ages; clinicians who were relatively newly qualified *and* those who had been practising both skills for over 10 years; staff based in both rural *and* metropolitan settings; and clinicians with primarily pre-hospital experience *and* clinicians who consult in hospitals and also retrieve patients from the pre-hospital setting. The diversity within these background characteristics of the study participants brings a representativeness of the wider emergency clinical community to the data.

4.5.2 Reliability and validity

In the assessment of this study's reliability and validity, I must echo the advice from Van der Vleuten and Schuwirth (2005) that "there is no such thing as *the* reliability, or *the* validity, or any other absolute, immanent characteristic of any assessment instrument" (p. 310). Here I will discuss the reliability and validity of the *study* described in this chapter. Discussion of the reliability and validity of the *implementation* of the assessment tools developed (applied in Chapter 5) will be more thoroughly critiqued in Chapter 6.

4.5.2.1 Reliability

Patton (2002b, p. 193) points out that the extent to which tests "are useful, valid and reliable can be a matter of debate and judgement", implying that no single standard exists by which to measure the diversity of possible claims within the wide range of application and study contexts. When we consider reliability in terms of repeatability (Cortina, 1993), there is no real way to know if another similar panel would yield the same final list of items are described here (Keeney et al., 2011). The focus on stratified sampling to ensure a heterogeneous participant panel consisting of a variety of clinical professionals who collectively have a multifaceted experience base for both skills, a sound case may be made for the representativeness of the panel. They bring a diverse and widely representative range of opinions held by clinicians in the local pre-hospital environment, therefore the findings are credible even if a guarantee cannot be made that identical outcomes would emerge if the study were repeated with a different panel.

Considerations for a qualitative study's reliability may include thorough documentation of study procedures, transcript checking, and coding consistency and analysis checking within the research team (Creswell, 2013, p. 203). One of the key aims of this is to promote sufficient transparency for the reader to be able to decide for him or herself whether the study can be relied upon, with reference to his or her context. While this study is heavily influenced with a qualitative philosophy and approach, evidence of an only very recent emergence from quantitative thinking is clearly seen in the data obtained: the majority of data obtained is a numerical representation of the participants' values, rather than a description of the values and meanings themselves. Thus, the prominent approaches to rigour and reliability for qualitative research designs are not well suited to this study.

Instead, considerations to reliability were centred on allowing the data to emerge from the group without influence (bias) from the research team who may have ideas, priorities or assumptions about what is important during IO or LMA insertion. This is evident in the reception of all suggested criteria, and alignment with other literature for decisions regarding inclusion and exclusion. The final checklists presented in this chapter's results are, therefore, representative of the panel's practice and collective opinion. Having presented the data as they emerge in the study, a further discussion will be conducted for adaptation to local clinical practice and reduction to a more concise series of checklists to aid assessors. As the focus has been maintained on presenting the complete data as it emerged, it now becomes important to develop a coherent interpretation. In the following section, I will discuss a suggested refinement to both lists, for the purpose of greater usefulness by educators.

4.5.2.2 Validity

There are multiple levels to understanding validity in this type of study. In the initial stages, the Delphi design was chosen as consistent with the research question and resultant methodology. The published literature identified the Delphi process as a means of producing a validated assessment tool, therefore the method was selected for this study. Steps taken in the study design to ensure valid data collection and interpretation are outlined in section 4.3.4. A critical review of the study prompted separation of validity in the study design (determining whether the study findings are represented accurately for the researcher and participant (Creswell, 2013, p. 201) and validation of the uses of the study outcomes. This first validity, is not wholly derived from following a pre-determined procedure to ensure an appropriate outcome, although some credibility for the study is achieved in using established research methodology. Rather, attention must be given to crafting the study design and processes to ensure the data is sound, and that the claims made are consistent with the process and the findings.

Considerations for validity in a qualitative study design may differ from those discussed in section 4.3.4, which are primarily for quantitative designs. A difficulty in this particular study is the use of quantitative (numerical) expression in understanding a qualitative phenomenon (the importance of a particular item in application of the skill). Threats to internal and external validity were necessarily addressed, however other important factors must also be considered, namely those suggested by Creswell (2013, p. 202), represented in Table 23 below.

Table 23: Factors Considered in Addressing Validity for Qualitative Aspects of the Study

Factor	Strategy to address it
Triangulation	The concept underlying triangulation in building "a coherent justification for themes" is confirmation of findings from several, independent sources. This is achieved through determining an acceptable definition for "agreement" at the outset of the study, with items only included if the average importance of that item is sufficient when considering all participants.
Member-checking	Between the first and second round, participants were invited to alter the scores given to items which had not achieved a polar enough mean score for a definite inclusion or exclusion decision. This was done with reference to the mean score that item achieved from the previous round, and with confirmation of which items were definitive enough to be included in the final checklist. This allowed participants to alter their decisions based on data emerging from the previous round.
Using rich descriptions	This aspect was not favoured, with the intention that the final checklist would be easily utilised by educators and assessors. Long, rich descriptions could impose unnecessarily complexity, so descriptions of each item were intentionally kept as short as practical in conveying their meaning.
Bias	Compliance to the study approach as outlined by previous studies was maintained as closely as practical to ensure minimal researcher bias on data collection and inclusion/exclusion decisions. Deviations from this were explicitly justified.
Present data which counters the themes identified	A cited weakness of the Delphi process is the focus on agreement, rather than disagreement which may hold valid, yet minority views. Perspectives falling outside the majority were not the focus of the study which by definition, is a consensus study.
Spend prolonged time in the field	Embracing the quantitative mindset to bias, my background as a pre-hospital clinician and clinical educator were set aside during the course of the study in order to allow the participants views to be of most importance during the study.
Peer debriefing	(not addressed)
External auditor	(not addressed)

Adapted from "Research design" by John Creswell, p. 202, copyright 2013 by Sage publications.

This Delphi process is concerned with determining a valid expert consensus, rather than revealing an objective truth (Keeney et al., 2011). The employment of experts who meet pre-determined criteria design contributes face validity that the outcomes are reasonable, and provides an argument for authenticity of the assessment to actual practice (Boulkedid et al., 2011). Van der Vleuten and Schuwirth (2005) argue that "competence is highly dependent on context or content", hence there is much value in the participants' intimate understanding of the practice content. This also raises the importance of appropriate use of the marking tool. Additionally, the validation principles presented by Kane (2006) urge us to critically examine the data and the proposed uses and interpretations of the tool which has been developed, arguing that true validation an ever-incomplete, ongoing process (Kane, 2006). For this reason, the study output cannot be considered as a *validated checklist* if it has not undergone some level of critique and justification for the application to which it is intended to be applied. As a result, I have been careful to phrase this study as an assessment tool development study, rather than the development of a validated checklist.

Boulkedid et al. (2011) also warn that resting on the face validity of expert participants may lead us to neglect other important measures as we review the results, namely reliability, sensitivity, specificity, and feasibility. These factors need to consider the tool's purpose too. A tool designed to make pass/fail decisions will necessarily consider sensitivity and specificity, but a tool designed for other purposes may be less concerned with these measures. Kane (2006) guides a modern approach to interrogating an assessment tool's validity. He helps create a clearer delineation between validity as a concept which is addressed in a study's design (for example the validity and reliability of data collection and interpretation) to a focus on the study's implications. Assessing the validity of an assessment tool which is developed is closely connected to the proposed uses and interpretations of the tool. In order to examine whether the use and interpretations are valid, we must make them explicit, along with any assumptions and arguments which support or refute them (Kane, 2006). This is further addressed in Chapter 6. "[A]ssessment instruments are not goals in themselves" (Van der Vleuten & Schuwirth, 2005, p. 310), hence a separate chapter will critique and explore the validity of the checklists developed by this Delphi study, in light of the results generated by it.

4.5.2.3 Review of Inclusion and exclusion parameters

Earlier in this doctoral research, I discussed the alignment to externally identified procedures as a means by which the data emerged without risk of bias or influence from the researcher, however in recognition of some variability among the studies, the decision to include items with a mean score of 7.0 or higher was reviewed. The two studies identified in Table 17 which used a scale of 1-9 included items with a slightly lower mean score (6.0 and 6.62), so I considered the impact of including items with a mean score of 6.0 and above. For IO access, this resulted in the inclusion of two items which were near-duplicates of items already included (items 6 and 10), and one item which was also not agreed to be a critical component of the skill in terms of morbidity and mortality (item 4). For LMA insertion, lowering the inclusion mean to 6.0 saw the inclusion of five additional items (5, 19, 25.1, 27 and 30. These items demonstrated low levels of agreement with relatively large IQR (and *SD*) noted for most. Additionally, item 5 is very closely related to item 4 (included), item 19 may not be possible during a cardiac arrest where the other clinician is performing chest compressions, and item 27 is assumed by the inclusion item 21. The choice to maintain the minimum mean value for item inclusion at 7.0 was therefore upheld, but was necessary to scrutinise.

The checklists reduced much more in length than the studies identified in Table 17. This reduction reinforces that a distillation of core, agreed principles of IO and LMA insertion had occurred through this study. Further suggestions of adapting the checklists to more concise and possibly more useable assessment tools are included in section 4.5.4.

4.5.3 Items pertaining to increased expected morbidity and mortality

It is important to recognise that this is the clinicians' overall expected risk. Clinical competence must reflect safe practice (Oermann, Yarbrough, Saewert, Ard, & Charasika, 2009), and these items identify potential safety issues within a skill performance. I am unable to confirm these aspects as actual high risk by means of an experimental study would be ethically and clinically irresponsible, however a further study will investigate whether these items correspond with global student scores (see Chapter 6).

4.5.3.1 IO morbidity and mortality

Relatively few items for IO insertion were predicted to impact on morbidity (three of the 22 items, or 14%) and even fewer to impact on mortality (one of the 22 items, or 5%). These items indicate the steps during IO insertion where experienced practitioners exercise particular caution, and assessors may consider using these items as mandatory elements of assessment. These items include:

- Item 3: Inserting the IO needle into the appropriate location
- Item 9: IO site is prepared (non-specific)
- Item 10.3: IO insertion site swabbed with alcoholic chlorhexidine and allowed to dry (2-3 seconds)
- Item 20: Clinician stops applying pressure once the resistance of the outer cortex of the bone subsides (like a "pop")

- Item 10: (excluded from final list due to mean score <7.0) IO insertion site swabbed with iodine

The key areas that these items address are IO location (relating to the function of the IO cannula in accessing the venous circulation), and site preparation (including anti-microbial consideration). While there are multiple locations where an IO needle may be effectively placed in order to allow administration of medication to the venous circulation (Pasley et al., 2015), these may not all be appropriate within the choreography of an out of hospital resuscitation attempt. For example, many authors suggest sternal insertion for the administration of medication and fluids (Carefusion Corporation, 2014; Harcke, Crawley, Mabry, & Mazuchowski, 2011; Hartholt, van Lieshout, Thies, Patka, & Schipper, 2009; Lewis & Wright, 2015; Pasley et al., 2015; Phillips et al., 2010a, 2010b; Von Hoff, 1991), however where CPR is required or potentially required this may not be appropriate. Aside from hand placement during CPR, the external cardiac compressions themselves may cause sternal fractures (Krischer, Fine, Davis, & Nagel, 1987; Lederer, Mair, Rabl, & Baubin, 2004), and thus insertion of an IO device into this fractured bone may impair its function and this is therefore a contraindication for site selection (Pasley et al., 2015). This, therefore, ought to be considered in the post-arrest patient where an IO device may be used for post-ROSC administration of fluids or medication. Site location may refer to function (for example selecting a bone which has an intramedullary cavity for infusion), or it may refer to risks such as avoiding the epiphyseal plate, sometimes referred to as the "growth plate". Concern exists that if this bone tissue is disrupted through IO needle insertion into a paediatric patient, the limb may not grow appropriately (Pasley et al., 2015), however this is not supported by the literature (Lewis & Wright, 2015; Paxton, 2012).

Antimicrobial consideration is a consistent concern for experienced pre-hospital clinicians, however the data may suggest that this is disproportionate to the actual risk. Only 0.6% of intraosseous devices result in infection (Rosetti, Thompson, Miller, Mateer, & Aprahamian, 1985; Santos, Carron, Yersin, & Pasquier, 2013), compared to a documented 88 in 875 incidence of local complications[11] due to IV cannula insertion, amounting to complication rates exceeding 10% of intravenous devices inserted (Soifer, Borzak, Edlin, & Weinstein, 1998). This could be a reflection of the greater regard clinicians have for infection risk in intraosseous procedures, or it may indicate that intraosseous devices are not as great an infection risk as commonly thought.

[11] Including Tenderness >4cm from site, warmth, erythema, induration and/or swelling >3cm from site, or cord 3-6 cm from site (see Soifer, 1998)

4.5.3.2 LMA morbidity and mortality

Of the 16 items included in the LMA assessment checklist, eight were indicated as expected to increase both morbidity and mortality if performed incorrectly (50%). This is much higher than for the IO checklist, which may either relate to the relative importance of the airway in resuscitation, or the natural variability in LMA insertion techniques, resulting in a shorter agreed list of critical items. This is supported by the following statement from the Australia and New Zealand resuscitation councils:

> There is insufficient data to support the routine use of any specific approach to airway management during cardiac arrest (ANZCOR, 2016b).

Resuscitation guidelines have historically consistently advocated for early attention to a patent airway, then breathing, then circulation, or *ABC* (Wiles, 2015). Relatively recently, the measurable positive impact of early defibrillation on resuscitation (Greif et al., 2015; Soar et al., 2015) has moved this aspect of resuscitation upwards, however, with relatively less advanced airways advised for early use, and follow up with a more definitive airway following defibrillation (if indicated), and IV access. This may impact how experienced clinicians approach an airway device. Regardless of *how* an airway is secured (chin lift, OPA, ETT, surgical airway), it must be performed effectively in order to progress the patient management. The items flagged for a potential impact on morbidity and mortality include:

Item 3: Patient is pre-oxygenated

Item 4: Appropriate size LMA is selected

Item 9: Integrity of cuff is checked by inserting air into it and ensuring a seal

Item 28: Cuff inflated

Item 29: Cuff inflated with no more than maximum stated volume of air

Item 21: BVM connected while holding LMA secure (before securing device)

Item 32: Ensure chest rise and fall with BVM

Item 34: LMA secured into place

These items reflect the patient's need for oxygenation prior to LMA insertion (given that LMA insertion may prevent the clinicians' ability to ventilate the patient with IPPV for a brief period), adequate seal achieved (including size selection, testing, and cuff inflation), and maintenance of appropriate location and seal (through securing the device). These all directly impact the loss or maintenance of a paramedic's ability to provide intermittent positive pressure ventilation (IPPV) for a patient who is not breathing. Poor airway care can impact patient morbidity and mortality (Owen

& Plummer, 2002), and this list of items provides more specific guidance of this consideration important task in the pre-hospital setting.

4.5.4 Refining the lists further

The final checklists developed by the Delphi process were then reviewed by a small research team and further amendments are suggested in order to provide a more useable assessment tool.

4.5.4.1 IO list refinement

Items 1 and 2 may be combined into a single item of Personal Protective Equipment (PPE). Items 7, 8 and 8.1 all refer to the proximal tibia as the IO insertion site, and so are very similar in nature. Sources collated for the preliminary checklist introduced the proximal tibial as an insertion site, which introduced potential bias. As IO devices can be effectively used in various anatomical sites during resuscitation (including the sternum, femur, distal tibia, humeral head, distal radius, ulna, iliac crest, and clavicle (Harcke et al., 2011; Hartholt et al., 2009; Rosenberg & Cheung, 2013), I expected less agreement on the importance of the proximal tibia specifically. The introduction of this site in the preliminary checklist may have influenced panellists to consider it of greater importance than other sites, however the literature indicates that the proximal tibia is often considered the most clinically appropriate choice in paediatric patients (Boon, Gorry, & Meiring, 2003; Lavis, 2000; Lewis & Wright, 2015; Oksan & Ayfer, 2013; Santos et al., 2013).

The tibia's selection as the most favoured site for IO insertion may relate to convenience of access for the resuscitation team as it allows other treating clinicians more room to perform important tasks such as external cardiac compressions, airway management and defibrillation which require access to the patient's upper body.

Item 9 is non-specific, and was retained from Lammers et al. (2009). This may be considered to refer to antiseptic practice covered by items 10.3 and 11 which are both concerned with insertion site preparation. Condensing these criteria into a single item regarding antimicrobial site preparation will simplify the list, and allow clinicians and educators to adapt to local organisational practice guidelines as appropriate. Items 27 and 28 are also very similar in nature (assessing placement of the IO device). Course and assessment designers would be encouraged to choose the most institutionally appropriate items. I therefore propose use of the 12-item performance checklist presented in Table 24.

Table 24: Suggested IO Insertion Checklist

No.	Item	Original list item (s)	Critical element

1	Clinician wears appropriate Personal Protective Equipment	1, 2	
2	IO is inserted into an appropriate location. (If proximal tibia is selected, the IO is inserted into the flat surface, 1-2 cm below tibial tuberosity and in line with medial tibia(Paxton, 2012))	3 (7, 8, 8.1)	Y
3	IO site is prepared with an antimicrobial swab	9, 10.3, 11	Y
4	Clinician holds limb/body part secure during IO insertion	13	
5	IO needle inserted at 90 degrees to bone	17	
6	IO needle rotated back and forth along its axis, with gentle, constant pressure during insertion	18,19	
7	Clinician stops applying pressure once the resistance of the outer cortex of the bone subsides (a "pop")	20	y
8	Depth guard is wound down so it is just flush with the skin, without being over-tightened	21,22	
9	Blue cap removed from top of IO	23	
10	Stylet removed from inside IO needle and discarded in sharps container	24,25	
11	IO visually assessed to stand in the bone without needing additional support	26	
12	IO placement assessed by aspirating bone marrow and flushing with Normal Saline	27, 28	

Note: This checklist intended to guide education and assessment, showing original items and items expected to adversely impact mortality or morbidity if performed incorrectly.

The guide is intended for use by novice clinicians, educators and assessors, where a global scoring system requires a greater level of expertise to apply. This could be considered parallel to the reasoning applied by relative novices compared to the expert clinician who applies more subtle judgements and reasoning in the management of a patient interaction (Gingerich, Kogan, Yeates, Govaerts, & Holmboe, 2014). Similarly, novice assessors may be more included to employ checklists based on observed criteria whereas expert assessors have a bank of experience upon which to build accurate and informed judgements. Subjective assessments with some structure may be highly consistent and reliable so long as the examiners have expertise in both clinical and examination aspects of the assessment (Schuwirth & Ash, 2013, p. 413), thus reliable observation-based assessment (subjective judgements) require significant examiner training. Item 2 of this abridged checklist still requires some clinical judgement if the IO is inserted into a site other than the proximal tibia.

4.5.4.2 LMA list refinement

I also propose an amalgamation of some items on the LMA insertion checklist for more feasible use in some clinical assessment settings. Items 1 and 2 are combined into a single item regarding PPE. Although it was identified as likely to impact patient morbidity and mortality if improperly performed, item 5.1 was reviewed to be unnecessary for the skill assessment as no part of the skill can proceed without removing the equipment from the packaging. Items 6 and 9 have been combined to address a review for manufacturer's faults, and items 21 and 25 regarding insertion have been pooled in the proposed checklist below (see Table 25).

Some comments from the panel participants speak to the outcome-focussed philosophy of LMA insertion, with the principal focus being placed on timely, effective airway attainment. One participant commented that "The [LMA] procedure is actually very simple, and there are too many criteria here, which over complicates the lma (sic) insertion". Another stated there is "no time to stuff around!", and that "LMA is a rescue device, so messing around adjusting volumes... is time consuming and counter-productive." These exasperations speak to the context specific nature of this data, for example: LMA use in pre-surgical anaesthetised patients occurs regularly and must be performed effectively, however the considerations for a pre-hospital resuscitation clinician are significantly different. Interestingly, the LMA device is contraindicated for use in patients whose recent fasting state cannot be confirmed, are obese, more than 14 weeks pregnant, are involved in an emergency or resuscitation event, or who cannot adequately answer questions regarding their medical history ("LMA Classic™, LMA Flexible™, LMA Flexible™ Single Use & LMA Unique™,"), that is contraindicating use in the patients for whom the apparatus is used in the pre-hospital setting.

Table 25: Suggested LMA Insertion Checklist

No.	Item	Original list item (s)	Critical element
1	Clinician wears appropriate Personal Protective Equipment	1, 2	
2	Pre-oxygenate patient	3	Y
3	Appropriate size LMA is selected	4	Y
4	Briefly check the LMA for manufacturers faults	6 and 9	Y
5	Posterior side of cuff lubricated	17	
6	Patient's neck is flexed and their head extended	20	
7	LMA inserted with bowl of mask facing anterior, in a forward sweeping motion	21 and 25	
8	Once inserted, the cuff is inflated with no more than the maximum stated volume of air	28 and 29	y
9	BVM connected while holding LMA secure (before securing device)	31	Y
10	Ensure chest rise and fall with BVM	32	Y
11	LMA is secured into place	34	y

Note: This checklist intended to guide education and assessment, showing original items and items expected to adversely impact mortality or morbidity if performed incorrectly.

4.5.4.3 Further validation

More rigorous validation of this checklist will be explored in Chapter 6, where I will explore and critique the implications for use, and the underpinning assumptions for the tools' application. This study was conceived in order to develop an assessment tool, based on clinical expertise, in order to distil the core, agreed components of both IO and LMA insertion. The intention is to then apply these criteria to a comparative trial in order to compare the acquisition and retention of these skills when taught with two different teaching strategies.

The checklists, being objective in nature (if true objectivity were possible), will contain a categorical score for each evenly-weighted item (yes or no). The summed scores will be considered a continuous (discrete) outcome variable reflective of the individual performance it has been applied to. Consequently, the assumption of this interpretation is that one performance of, for example 16 marks is the same proficiency as another student's performance of 16 marks, with different scores for different items. This assumption may be reasonably challenged, as it implies that each item is on equal weighting with each other item, and this is not necessarily the case for categorical data. One way to address this assumption is to weight the score of each item based on, for example, the average score given to it by the panel. This would more heavily weight items which the panel considered more important, but if we reconsider the panel's scores as a value statement, rather than principally a numerical statement, we risk over-complicating the numerical calculation of a performance score based on these weights, and the assessment tool becomes unreasonably complicated. Another proposal may be to interpret the score with reference to the items marked as morbidity/mortality flags. These items have been identified by the expert panel as a more significant risk to increase adverse outcomes for the patient, and so these should be addressed with special consideration.

So in response to the question of validity, what is of interest is whether these checklists measure what they are intended to measure, and what they claim to measure. This will be more comprehensively addressed in Chapter 6 where the assessment tools will be applied to real performance data and critiqued more thoroughly. This critique will include:

- Inter-rater consistency of the assessment tools when used by independent assessors
- Comparison of checklist scores with other scoring strategies already well accepted as valid and accurate assessment tools, such as global rating scales (Doyle et al., 2007)
- Whether expected differences are identified between untrained practitioners, recently trained practitioners, and those who have been given opportunity for naturally expected attrition of skill and knowledge.

4.5.5 Limitations

4.5.5.1 Device-specific assessment tools

The assessment tools developed are device-specific, thus with the advent of newer equipment such as automated Intraosseous drills and LMAs which do not have inflatable cuffs, some items on the lists become redundant. The checklists developed in this chapter ought to therefore be adapted to local contexts (including the available equipment) in order to be valid procedural guides. Some items (such as location of IO insertion, IO site preparation, pre-oxygenation for patients receiving a LMA, and assessment of effective ventilation following LMA insertion) will apply to the techniques regardless of the equipment used.

4.5.5.2 Local expertise

The quality of this type of study relies heavily on the participants' expertise, experience and commitment to completion. Participants were all employed by the South Australian health department (SA Health) at the time of the study, and therefore their practice may represent the institutional practices and culture of this organisation. By stratifying the sample of participants across various clinical levels, the study draws together consensus from a range of pre-hospital practitioners and minimises bias from a particular clinical level, but some local influence may still be evident.

4.5.5.3 Pre-hospital setting

The checklist produced in this study is specific to the pre-hospital emergency setting, and may reflect local attitudes, culture of practice and assumptions. This somewhat exclusive clinical relevance was intentional as the study is concerned with resuscitation skills, however it will be a consideration for others wishing to adapt the criteria to their own setting. IO access may be required for purposes other than pre-hospital resuscitation, in which case it will be important to contextualise details such as most appropriate location for the desired therapy, for example.

4.5.5.4 Defining consensus

Ironically, there is little consistency in the literature as to how consensus should be defined (Boulkedid et al., 2011). Many definitions of consensus are proposed by various authors, including criteria based on median score, interpercentile range, proportion of responses in extreme tertiles, proportion of responses within a three-point range (Fitch et al., 2014, p. 57), stability across multiple Delphi rounds (Green, Jones, Hughes, & Williams, 1999), percentage agreement (Green et al., 1999; McKenna, 1994) and various combinations of these. Definitions of consensus may be further classified as strict or relaxed, depending on the spread or outlying responses (Fitch et al., 2014, p. 57). This may be due to the flexibility of the Delphi methodology which is governed by principles rather than strict universal guidelines (Hasson et al., 2000). This raises the very important and yet unresolved question: What is consensus? I defined consensus based on the mean value as established in previously published studies, however this remains debatable (particularly for ordinal data). The definition of consensus with reference to mean score was retained, particularly in light of criticism surrounding percentage agreements (Hallgren, 2012), although other measures may be equally acceptable.

4.5.5.5 Consensus or variation?

A perceived limitation of a consensus study is that it can devalue natural and appropriate variability (Keeney et al., 2011, p. 13). The approach which guided the research question is one which values the breadth of opinion and expertise, in order to identify common ground from a heterogeneous panel who had slightly varying perspectives of the two skills. This is a strength of qualitative research, where a new truth is constructed with each participant contributing uniquely to the final product. A consensus or agreement study could be perceived as being at odds with the constructivist foundation, by excising the value of variability. This tension is offered some resolve in the social constructionist tradition where meaning is constructed not from an individual basis, but from a group basis, where the members are grounded in a particular setting, culture and social space. This social setting is the professional identify to which the participants all subscribe in slightly varying ways, dependent on their membership to the group as a whole, and their individuality within that group (for example, a Medstar retrieval nurse and ICP bring individual experiences, perspectives, practice protocols and subcultures to their wider group membership as pre-hospital clinicians).

4.5.6 What a pre-hospital assessment tool adds

Many items in our list parallel Oriot's scale (2012), with a small proportion of items proposed by us which did not closely mirror the in-hospital checklist . These include items 13, 19, and 23-25, where the latter items related more closely to the specific manual device used in our study. Our list was longer, and more comparable in length to those in Table 17. Where the list is adapted to local practice guidelines, it may reduce in length, creating a more manageable teaching and assessment tool (see section 4.5.4 on Refining the lists further). The most significant benefit of our list in the pre-hospital teaching setting, beyond the development of a context-specific assessment tool from which to gain valuable data for the intended comparative resuscitation skill study (see Chapter 5), is the guidance this may provide not only assessment personnel who may be required to make difficult decisions on student competency, authority to practice, and formative feedback, but also students who are seeking guidance on their practice development (Saewert, 2013). Chapter 6 will address the validity of the tools for these purposes.

4.5.7 Further research

Kane's definition of validation is confirmed as “the process of evaluating the plausibility of proposed interpretations and uses" (Kane, 2006). Based on evidence that supports or refutes these interpretations and uses, validation cannot be considered possible by means of a pre-determined process which provides a study output based on numerical decisions. While examining the validity of the development of these two checklists is required and has been addressed, attention to the validity of a study's processes does not necessarily address and satisfy a validation process or argument. A deeper analysis of the arguments for and against the proposed uses and interpretations of data resultant from the assessment tools, including an examination of reasonable evidence pertaining to such claims is required for these tools to be considered "validated" for a given purpose (Kane, 2006, p. 17).

4.6 Conclusion

This study describes the identification of clinical priorities during LMA and IO insertion from nine pre-hospital experts. The resulting checklist brings rigour and credibility to the education and assessment of new clinicians through the experts' combined clinical experience. Educators may consider adapting these criteria to suit a local context, where appropriate. This will require oversight to understand and retain the sentiment behind the criteria (for example aseptic technique and infection control) rather than an unwavering and uncritical employment of the specific items which reflect the sentiment (for example swabbing the IO site with alcoholic chlorhexidine specifically).

This study has produced two clinical tools to guide the education, assessment and skill maintenance of LMA insertion and IO insertion. No such skill-specific checklists were located in the literature for application of these two skills in the unique, pre-hospital resuscitation setting. Clinical educators now have a tool which will ensure high quality training and a guide to specific assessment foci for these rarely used but critical pre-hospital resuscitation interventions. The importance of such tools will only grow as IO access becomes more widely advocated for use in the emergency setting, in addition to the development of the ambulance paramedic profession.

The credibility of the clinical experts involved in this study, and the minimum exposure requirement stipulated for inclusion with the study ensures that participants not only have training and clinical authority to use the skills, but also current experience in both skills, which are infrequently used in most pre-hospital ambulance settings. A strength of the study is the homogenous backgrounds of participants with a range of pre-hospital experiences, from which the common elements of practice has been distilled. The checklists will be used to gather data from a comparative trial in the following chapter, in order to understand the key research question within this study series: is 2SA or 4SA a more cost effective teaching strategy? The comparative trial which follows will use these skill-specific checklists in order to answer this question, and ultimately guide educators, course designers and teaching organisations in effective skill teaching strategies with reference to organisational costs. Further validation of the checklists will occur in Chapter 6, including a review of the data gained in the comparative trial of Chapter 5, and an analysis of the tools' application and interpretation of outcomes. Considering this study as part of a subset with Chapters 5 and 6 will connect the development, application and critique of the tools, and from which questions surrounding skill assessment will emerge for further discussion.

This Delphi study is part of a subset with Chapters 5 and 6, and contributes a set of rigorous, clinically focussed assessment tools from which to glean performance data, and therefore understand the cost-effectiveness of 4SA compared to 2SA in Chapter 5. This presentation is intentional, ensuring that the form and content of this doctoral research aligns. Therefore the development (Chapter 4), application (Chapter 5) and critique of the tools (Chapter 6), are tethered and carefully develop new knowledge. This structure allows questions surrounding skill assessment to emerge in the use of these assessment tools for further critique and discussion.

5 STUDY 3: ACQUISITION AND RETENTION OF INTRAOSSEOUS AND LARYNGEAL MASK AIRWAY SKILLS: A COMPARATIVE TRIAL

5.1 Introduction

A recent study has identified that in one Australian context, cardiac arrest is encountered by paramedics less often than expected (Dyson et al., 2015). Wisher et al. (1999) argues that procedural skills such as those employed in a pre-hospital resuscitation decline rapidly, "in just a few weeks or months", so it is possible that even the experts of pre-hospital care have reason to be concerned about their skill maintenance. It may be months or even years following training before an ambulance clinician may employ rarely used but critical resuscitation skills on a real patient. In the resuscitation setting, a decline in skill retention could have a marked impact on patient morbidity and mortality through increasing the time required to perform the skill, affecting clinician confidence. Training organisations assume that money invested in training staff will eventually pay off in staff ability to perform the trained tasks (Wisher et al., 1999), however many studies demonstrate that memory and retention are not stable over time (Archer et al., 2014). Special patient groups such as paediatric patients, will have even lower paramedic attendance rates. Lammers et al. (2009) note that "Measured in average days between patient encounters per provider, a paramedic will manage an adult respiratory patient once every 20 days as compared with once every 625 days, 958 days, and 1,087 days for teen/preteen, child, and infant patients, respectively."

Understanding how different clinical skills atrophy for new learners is a critical area for teaching organisations and health care providers to understand, in order to anticipate the level of care given by staff, and tailor skill retention incentives. Understanding the impact of initial training on skill retention could allow teaching organisations to consider the education implications for the workforce and the patient, in light of training costs.

5.1.1 Research question

The retention of infrequently used skills in the pre-hospital setting is therefore of considerable concern, but the power educators hold to effect it remains unclear. In Chapter 3, I investigated whether 4SA results in better skill acquisition than a more traditional 2SA. This chapter will address a similar question on skill acquisition, incorporating the lessons learnt from the previous studies. But this study further aims to compare the costs involved with the two teaching methods, the degree of skill retention, and the use of assessment checklists which reflect expert opinion of skill application (developed in Chapter 4).

The key research question addressed in this chapter is: "Do two common skill teaching methods (2SA and 4SA) have the same impact on skill acquisition, skill retention and teaching costs?" Cost was measured in time to teach, as this directly impacts the teaching organisation's financial outlay, and effectiveness was defined as student performance scores for the skills, both initially after the teaching session, and after a time delay. The two teaching methods are those previously described in Chapter 3.

5.1.2 Hypotheses

The following hypotheses were developed based on current learning theory, literature referring to the 4SA, and the first study:

A. When students are taught with 4SA, they perform better than students taught with 2SA according to:
    a. Skill-specific checklist scores, and
    b. morbidity/mortality risk scores
B. When students are taught with 4SA, they perform better than students taught with 2SA according to a Global rating scale.
C. 4SA takes longer to teach than 2SA.
D. Student confidence of perceived competence is greater for students taught with 4SA during the retention assessment
E. A student is more likely to insert an IO effectively into the bone cavity after 6 months, if they have been taught with a more robust teaching method (4SA).

5.2 How the research question and hypotheses relate to the theoretical framework

5.2.1 Context of this study

5.2.1.1 Theoretical framework

The theoretical framework described in section 2.6.4 grounds the study measures to the complex relationships between different stakeholders within clinical education. These relationships, each of which hold the student in the centre, may either be mutually supportive or at tension. As described in Chapter 2, the model of symbiotic clinical education is employed in this thesis to identify the key aspects of cost-effective resuscitation skill education to the stakeholders within the model.

The institutional axis involves the student's position between the health service and the teaching institution. This chapter is of value to the understanding that clinical skills education is potentially a great cost to both the teaching organisation (universities and professional accreditation bodies) and the health service (in-service training requirements, reaccreditation demands, and up-skilling existing workforces). The cost of initial training manifests in resources required, educator time, and facility hire costs, among other factors. These all place a potential burden on training budgets, so identifying the teaching costs associated with various training methods is of great importance to both sides of the institutional axis. Similarly, the effectiveness of skill performance beyond the training session is an aspect of clinical education often undervalued, despite many studies demonstrating that skill atrophy is common. This chapter will address not only skill acquisition, but also retention six months following teaching, in order to get a more realistic understanding of the performance of infrequently used clinical skills.

The clinical axis involves the student's position between the clinician and the patient. By understanding the output of teaching strategies, educators can understand the potential risk to patients due to skill atrophy, in addition to the potential burden on clinical supervisors in the workplace as they seek to correct and reinforce skill performance in the clinical setting. If one teaching method is demonstrated to achieve superior skill retention, even if it is more costly to teach initially, it may still be considered the most cost effective strategy in light of remediation efforts. Currently there is limited evidence to understand this factor within the clinician-patient relationship, and the impact that a student may have resulting from skill teaching practices.

The personal-professional axis allows the student to grapple with the expectations of their professional identity in light of the student's own worldview, ethical values and personal standards. This axis is not addressed in this chapter. The social axis places the student amongst the community which they serve, and the government which has a responsibility to build medical and healthcare capacity in areas (clinical and geographical) which address this community's need, with the resources available. In rural clinical settings, student clinicians may have greater autonomy than those placed in busy metropolitan settings (Walters et al., 2012; Worley et al., 2006), and junior clinicians may be required to manage patient care for longer periods, with less clinical supervision than in more resource-rich settings. While this environment can contribute to rich learning for these learners, the community needs clinicians who are able to apply the critical skills which are infrequently used, to a higher standard than those who may have a senior supervisor nearby to guide, assist and correct. Communities in less resourced locations depend on clinicians to retain skill excellence, and this study will investigate whether one of the teaching methods investigated may do that.

5.2.2 How this study connects to previous and subsequent chapters

5.2.2.1 Chapter 3

This study follows on from the study described in Chapter 3, which identified a series of improvements which would be incorporated into this second trial. Some of these are described in section 3.5.8, and include the following:

Table 26: List of Ways the Defibrillation Study Impacted this Study's Design

Problem	Strategy to address it
Contamination prior to retention data collection	Agreement sought with first year topic lecturer that the second-year skills taught in this study would not be brought into the first year skill teaching programme (participants were advised that their engagement in other learning activities would not exclude them from the study).
Equipment used to teach	These would be familiar to the educator on order to avoid unnecessary increase in the educator's cognitive load. As the educator teaches two skills, there are a number of additional complexities, so reducing the need to remember unfamiliar equipment may limit distraction from the teaching protocols. The clinical equipment which would be taught to the students has not yet been used in their formal study. While they may have been exposed to the instruments previously, this has not been part of a formal teaching session for their studies, therefore baseline performance is likely to be lower. This will give a more authentic insight into novice acquisition and retention.
Adherence to 4SA teaching protocol	More thorough preparation would be offered to the educator prior to this study. This includes additional written resources, and additional face-to-face preparation.
Low sample size	First- and second-year nursing students enrolled at Flinders University would be invited to participate in the study along with the first -year paramedic students. This will not only allow the opportunity for how different professional students learn and retain resuscitation skills, but will also allow more teaching groups to run. This will provide more sessions from which to gather data on the time requirements for both skills. Additionally a second skill will be incorporated into the design, with all students scheduled to learn two skills, and will learn with both methods.
Low participant attendance	Of the 48 participants who registered for the previous study, only 38 attended to register in person on the day of the study. Of these, the before and after videos of 28 students were recorded. Of these 56 potential videos, three were corrupted and unable to be marked. Full data remained for 25 participants. For this retention study, a power calculation was not performed. Instead, the intention was to invite as many participants as were available so that considerable exclusion or dropout prior to the commencement of the study would be less likely to result in an under-powered study.

5.2.2.2 Assessment tool development chapter

I searched the literature for a validated assessment tool. The search returned no validated tools specific to either skill for use in the pre-hospital setting. Due to the uniqueness of this relatively uncontrolled environment, it was clearly essential to develop an assessment tool for both skills in the pre-hospital emergency setting as significantly different factors impact on the use of such devices in this environment compared to other settings.

In order to gain credible data which reflects student performance of IO and LMA insertion, with reference to authentic, expert pre-hospital clinical practice, and anticipated patient risk, the assessment tools developed in Chapter 4 were applied to this study. Using these tools, this study will not just consider the student's compliance to teaching (though this may be relevant in other studies), but will aim to understand the clinical impact of the two teaching methods, and as such a practical approach to the central research question of comparing the costs and benefits of the two teaching approaches.

5.3 Methodology

The methodological approach adopted for this study was guided by the research question. The stated question implies the existence of a clear, measurable, quantifiable and statistically supported answer. Unlike the previous chapter, the question in this chapter is not value-laden, and dependent on the individual experiences, context and perspective of the participants in determining a credible assessment guide, but rather it is a finite question, to which an answer existed and was waiting to be discovered. In seeking an answer to the question, objective measurement could be applied to provide data from which clinicians and teaching organisations could determine what resource requirements and outcomes could be expected from a chosen teaching method.

This research question is deeply rooted in post-positivism. While it makes use of the assessment resources socially constructed through a study of shared meaning, the post-positivist approach which guided the study presented in Chapter 5 fits the set of ontological and epistemological assumptions which argue that the answer to this question is absolute, and may be directly observed or measured. The quantitative paradigm will focus on measurements of extent and numerical value in determining the comparative participant performance, based on observation according to a checklist which was designed, in part, to guide reliable data collection.

Post-positivism values a study design which allows the data to speak for itself, without impact or bias from the researcher. A randomised controlled trial (RCT) approach is consistent with this intent, however in order to maximise timely subscription to the study, all students were invited to participate, rather than just a randomised selection. Bias in quantitative research may impact the data from various points, many of which have been controlled for in the study design, for example through double-blinded assessment of performance, independent assessment of global and skill-specific performances by different assessors, consistency of skill performance conditions, etc. A well-designed trial will allow observable data to emerge which will allow for the comparison of a change in behaviour between the two teaching methods (a proxy for learning). Time measurements from instruction videos will also allow for objective measurement of the time required to teach with both skills in order to compare the cost relating to educator wages.

5.4 Methods

5.4.1 Recruitment

Following ethics approval by the Social and Behavioural Research Ethics Committee (SBREC) at Flinders University, targeted recruitment took place for educators, student participants and study facilitators. A single educator was used for all of the teaching sessions, in order to minimise variance due to individual factors. The educator was well experienced in teaching, however the 4SA was new to him. He was fully briefed and had the opportunity to ask any clarifying questions prior to the teaching sessions. He had clinical experience and expertise relating to both skills being taught.

Students were eligible for the study if they were:

- enrolled in the first year of the Paramedic Science degree, or in their first *or* second year of the undergraduate Nursing degree at Flinders University
- spoke English as a primary language
- were not undertaking formal studies in a topic in semester two which would teach IO or LMA insertion (though students would not be asked to avoid incidental or opportunistic training)

These criteria were confirmed for each participant at the time of enrolment. Recruitment for Paramedic students included a brief presentation during a lecture, an email invitation, and a follow-up phone call. Nursing student recruitment was restricted to an email invitation only. Students were allocated a teaching session time according to their stated availability. This allocated time would include a video recording of baseline performance, training session and post-training skill performance.

5.4.1.1 Skill facilitators

Second year paramedic students were recruited by email to assist with student registration and skill facilitation. Facilitators were briefed on their roles on the day of the study, and a "cheat card" was explained to them and a hard copy provided (see Appendix 10.4.1) to maximise consistency in participant treatment. After the second day of the study, the performance videos revealed some facilitator inviting the participant to wear gloves for the skill. The verbal briefing then evolved to remind facilitators not to prompt students in this way. This also impacted the assessment of skills to exclude these prompted items.

5.4.2 Participant tasks

5.4.2.1 Baseline performance

Following registration, participants entered a skill performance room by himself/herself, and was asked to insert an LMA into a manikin and an IO into a chicken leg according to the instructions scripted on the facilitator cheat card. Attempts were made to standardise the equipment and setup for each performance. Facilitators were asked to refer to "skill performance" rather than "skill assessment" or "skill examination" in order to minimise the stress associated with examination conditions.

5.4.2.2 Equipment and room setup

The room setup was intended to be consistent for all study participants. Equipment for both skills were placed within easy access for all participant. This was intentional in order to ensure the study measures the participants' skill performance, rather than other skills such as familiarity with the kits and ability to find the appropriate equipment. A key element of a study's validity is whether the data and measurement say what the researcher and interpreter claim they say. Where the paramedic students have ongoing exposure to the kits, the nursing students do not. If a superior baseline performance or skill retention for paramedic students was noted, there would be no way to know whether this is actually measuring kit navigation skills, or ongoing exposure to the kit layout during the interim semester.

5.4.2.2.1 LMA setup

The Portex ™ Soft Seal Laryngeal Mask was used for this study due to availability. This LMA is comparable in application to those produced by Carefusion, the LMA Classic™ and Unique™ produced by Teleflex, and those currently in use by SA Ambulance Service involved in the assessment tool development described in the previous chapter. Figure 23 shows the room setup for the LMA insertion station. The blue "airway kit" was left opened, as pictured, with the ventilation equipment set up and easily accessible. This ensured that the equipment was readily and equally accessible so that the observed performance was reflective of skill ability rather than the student's ability to locate equipment in a kit. In authentic practice, unwrapped LMAs and other equipment would not be placed on an unclean surface due to for example the risk of contamination. A selection of LMA sizes was available, and as these are used repeatedly for training, they had been removed from their packages. Two different syringes were available to inflate the cuff, as was lubrication.

Figure 23: Performance room setup for LMA insertion

5.4.2.2.2 IO setup

Manikins were used to allow the participant to indicate where they would insert the IO device (and why), and then IO needles (Carefusion ™ Intraosseous Infusion Needle) were used to achieve central vascular access using a chicken drumstick to simulate the patient (see Figure 24 and Figure 25).

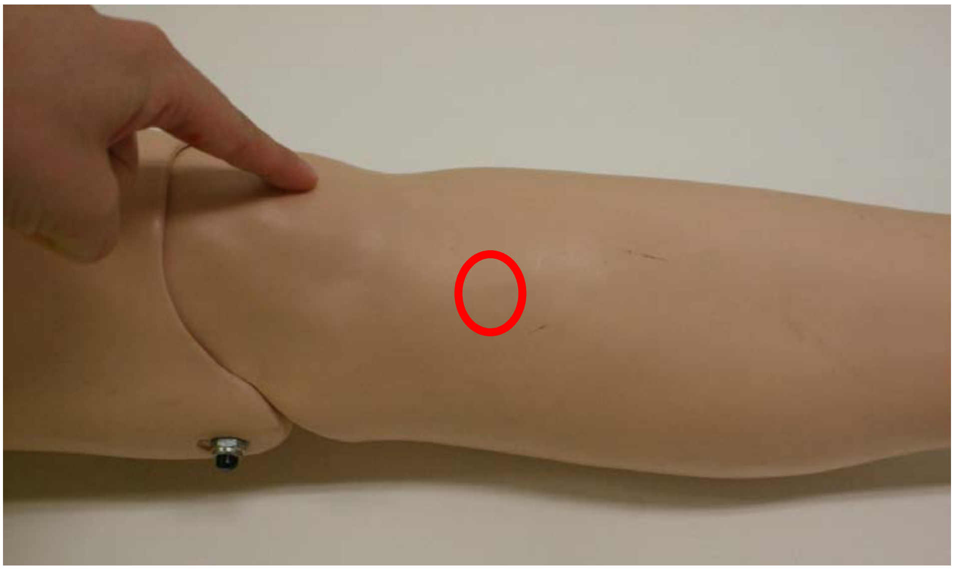


Figure 24: Manikin used for students to indicate IO location, with patella location identified. Taught IO insertion site identified by red circle.

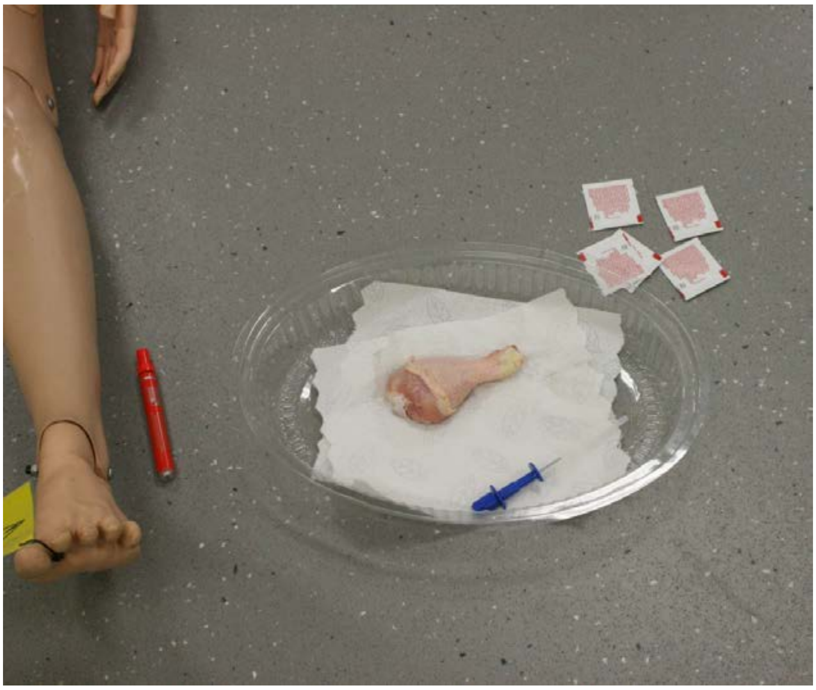


Figure 25: Chicken leg and IO setup

Participants were asked to indicate on a manikin leg where they would insert an IO needle prior to inserting it into the chicken leg. After the skills had been completed, the facilitators were asked to comment on whether the IO needle was secure in the bone before turning the video recording device off. These data were documented during video editing prior to external blinded marking. Video recording of the skill performances were taken from two angles: one from a ceiling mounted camera, and the other from a portable video recording device.

5.4.2.3 Teaching session

The participants then entered the teaching room in their assigned groups and were taught with the randomly selected intervention for that group (see section 5.4.3) before performing both skills under the same conditions, and in the same order, again. The teaching session was also recorded to review consistency and compliance, and allow a comparison of time data between the two skills and two methods.

5.4.3 Intervention allocation

Allocation to an intervention was stratified in clusters of teaching groups. The incorporation of two new skills into this study was necessary due to participant carry over from the initial trial into this one. The new allowed each student to experience both skill teaching methods, and both skills. The two skills were selected as they did not form part of the students' formal study program between initial teaching and reassessment, and they were perceived to be of similar manual complexity. Four possible teaching combinations were prepared which would be applied to each teaching session:

Card 1: taught IO with 4SA then taught LMA insertion with 2SA
Card 2: taught LMA with 4SA then taught IO insertion with 2SA
Card 3: taught LMA with 2SA then taught IO insertion with 4SA
Card 4: taught IO with 2SA then taught LMA insertion with 4SA

Two sets were prepared for the Paramedic student cohort (eight cards in total) and these sealed envelopes were parked with a green sticker. One set was prepared for the nursing student cohort as enrolments were much lower and therefore only four teaching groups were scheduled. This stratification was intended to result in a roughly equal number of students in each allocation. During analysis, it was anticipated to consider only two interventions: IO with 4SA and LMA with 2SA (cards 1 and 3); and IO with 2SA and LMA with 4SA (cards 2 and 4). The inclusion of four distinct cards was intended to evenly distribute any effect of teaching order.

A card was randomly allocated by sealed envelope selection at the beginning of the teaching session, with the cohort colour (paramedic or nursing) the only identifying feature.

5.4.4 Data collection

5.4.4.1 Variables of interest

Variables were carefully selected to reflect both the effectiveness (see Table 27) and cost of 2SA and 4SA.

Table 27: Dependent Variables Measuring Teaching Effectiveness

Variable name	Explanation	Value range
GRS_Total25	Student performance score according to Global Rating Scale made up of the 5 relevant items from Doyle's global rating scale for the evaluation of technical skills (GRITS) (see appendix 10.3.1)	5 to 25
Chlist_total	Student performance score according to skill-specific checklist prepared in previous chapter, excluding items prompted by some facilitators	0 to 22 (IO) 0 to 17 (LMA)
Chlist_taught	Student performance score according to skill-specific checklist prepared in previous chapter, excluding items on the checklist which were not explicit in the training session.	0 to 16 IO) 0 to 11 (LMA)
Morbid_Mortal	Morbidity or mortality rating	0 to 3 (IO) 0 to 8 (LMA)

The key variably measuring cost was the time to teach the group, excluding delays and tangents, but including changeover and re-setting equipment. This is abbreviated as TTT1_4exTD.

5.4.4.2 Teaching audit

I reviewed the teaching videos a minimum of three times: to ensure the teaching protocol had been applied, to code the time taken to teach, and to ensure the content of each teaching session was consistent regardless of teaching method. The time taken for various teaching activities (answering questions, providing correction and instructing) was manually coded according to the definitions in Table 28.

5.4.4.2.1 Time

All teaching time was coded to the closest five second point similar to the strategy used by (Walters, Prideaux, Worley, Greenhill, & Rolfe, 2009). The total session time (TST) was calculated from either the beginning of the demonstration or explanation of rationale (whichever occurred first). The teaching session was considered complete once the final person in the group had performed the skill and any relevant feedback or discussion had finished.

For the analysis, 2SA was considered equivalent in definition to Stages 2 and 4 of 4SA rather than Stages 1 and 4. This definition of 2SA was also used by (Archer et al., 2014; Bitsika et al., 2013; Herrmann-Werner et al., 2013; Jenko et al., 2012; Krautter et al., 2011; Orde et al., 2010). I developed a series of codes (expressed in Table 28) to more precisely describe the teaching actions within the teaching session, rather than the more crude measure of total time.

Table 28: Elements of Teaching Time

			Included in TTT1-4exTD
TST	TT 1-3	Q 1-3	Yes
		T 1-3	No
		D 1-3	No

S 1-3		Yes
R 1-3		Yes
P 4	Q 4	Yes
	C 4	Yes
	T 4	No
	D 4	No
S 4		Yes
R 4		Yes

Note: TST = Total session time; TT = teaching time, S = swap over time, R = resetting equipment; P = student performance; Q = questions (asked and answered), T = tangents (discussion not pertinent to the stated teaching objectives or content); D = delays (such as equipment failure); C = correction of student performance; 1-3 = Stages 1-3 of 4SA (or the first stage of 2SA), 4 = the final student performance stage of 4SA or 2SA; TTT1-4exTD = Trimmed teaching time for all stages of 2SA or 4SA teaching, excluding tangents and delays. This measure is proposed as the most authentic measure of time to teach.

The outcome of measure for this question is TTT 1_4 exTD (Trimmed teaching time for Stages 1 to 4 excluding tangents[12] and delays[13])

Question and correction time were included in this time, but delays and tangents were excluded. This is because questions and teacher correction or prompting are integral parts of teaching, and should be considered part of the standard teaching session. Delays and tangents were measured and excluded, because these were not directly relevant to teaching the skill and may artificially inflate the time to teach the skill through unrelated mechanisms (such as problems with equipment).

I compared the time for ST (the time between the start time and end time as defined above) and the sum of TT 1-3, S 1-3, R 1-3, P4, S4 and R4 to ensure all of the teaching time was accounted for and coded. The data were cleaned until these two values were equal and each five-second portion of each teaching session was captured. This neutralised the risk of counting the same piece of time twice, or omitting other pieces of time.

5.4.4.2.2 Content

The teaching points presented in each session were noted and collated to provide an overview of consistency between the sessions and fairness of the applied assessment tools. Items which were not consistently taught to each group were excluded from the assessment checklist (forming the score for "Chlist_taught" for the respective checklist developed in Chapter 4) in order to assess the features which best reflect learning from that teaching session.

[12] Tangents were defined as questions, comments or discussion which was not considered part of the outlined teaching session.

[13] Delays included time spent waiting for a fresh chicken bone, or new equipment due to failure, and was excluded as this is not directly caused by the teaching method.

5.4.4.3 *Skill Performance*

Participants performed the skills under video recording, which were then edited and assigned a random number for marking by assessors. Assessors were blinded to the student's allocation but not the overall study protocol. Two assessors were each provided with a validated marking tool specific to the skill being assessed (see validation chapter), and two more were provided with an adapted global rating scale containing six items. Items 1-5 are from the relevant criteria from Doyle et al. (2007, p. 552) and Item 6 is included from the final, overall criteria noted by Kneebone et al. (2006, p. 1110). The full scales are listed in the appendices.

5.4.4.3.1 GRS

The Global Rating Scale (GRS) used in this study is explained in Table 29.

Table 29: Explanation of the GRS used in this Trial

Item	Score				
Respect for tissue	1 Frequent unnecessary force on tissues or caused damage by inappropriate use of instruments	2	3 Careful handling of tissue but occasionally caused inadvertent damage	4	5 Consistently handled tissue appropriately with minimal damage to tissues
Time and motion	1 Many unnecessary moves	2	3 Efficient time/motion but some unnecessary moves	4	5 Clear economy of movement. Maximum efficiency
Instrument handling/ knowledge	1 Tentative/awkward moves or inappropriate use	2	3 Competent use of instruments, occasionally awkward	4	5 Fluid moves with instruments. No awkwardness
Flow of operation	1 Frequently stopped, seemed unsure of next move	2	3 Some forward planning, reasonable progression	4	5 Obviously planned course, effortless flow
Knowledge of specific procedure	1 Deficient knowledge. Required specific instruction at most steps	2	3 Knew all important steps of operation	4	5 Demonstrated familiarity with all steps of operation
Overall skill performance	1 Globally incompetent, unsafe, and/or unable to perform skill	2	3 Adequate skill performance	4	5 Skill is performed expertly with every regard (safety, competence, confidence)

Note: the first five items are adapted from "A universal global rating scale for the evaluation of technical skills in the operating room" (2007) by Doyle, J. D., et al., *The American journal of surgery,* **193**(5) pp. 551-555, and are accounted for in the variable GRS_total25. The final item is adapted from "An Integrated Procedural Performance Instrument (IPPI) for learning and assessing procedural skill" (2008) by Kneebone, R., et al. *The clinical teacher* **5**(1) pp. 45-48. This was used to calculate a global score out of 30 in conjunction with GRS_total25, however results were highly comparable to those already obtained therefore this data did not contribute to the research question. The full scales presented by Doyle et al and Kneebone et al are included in Appendix 10.3.1 of this thesis.

5.4.4.4 Assessors

Five assessors were used to gather participant performance data. They were used strategically to ensure that each video was marked by two assessors according to the GRS, and by another two assessors according to the skill-specific checklist.

Table 30: Allocation of Assessors

	LMA	IO
GRS	Assessor 1 (Paramedic) Assessor 4 (Paramedic)	Assessor 5 (Paramedic) Assessor 3 (ICP)
Skill-specific checklist	Assessor 2 (ICP) Assessor 3 (ICP)	Assessor 1 (Paramedic) Assessor 2 (ICP)

5.4.4.5 Confidence

Confidence of both skills was investigated by means of a survey upon the retention performance. Participants were asked to indicate on a scale of 0 to 10 how confident you feel about performing the skills competently today, where 0 is *not at all confident*, and 10 is *completely confident*.

5.4.5 Principles of rigour and reliability: how to address threats to validity

Again using the threats to a study's validity identified by Creswell (2013, pp. 174-176), the tables below outline how these threats have been addressed in this chapter's study design.

5.4.5.1 Threats to internal validity of the study

Table 31: Threats to Internal Validity

Potential threat	How the threat was addressed in the study design
History	It was not possible to have experimental and control groups experience the same events throughout the course of the study, which spanned over 6 months. Skills which the students were not scheduled to learn, and were not likely to encounter on their student placement (as first year students) were selected for this study to minimise confounding experiences outside of the study. Participants were asked to complete a brief survey on returning for the final skill performance, to understand what level of exposure they had had to the skills prior to the final performance. I chose not to ask that they do not participate in additional skill sessions where they may encounter these skills, as I did not want to match their generous and voluntary participation with a demand which they perceive could disadvantage them in other areas.

Maturation	All participants were at the same point in their study (first year paramedic, or nursing student). *Fast track* students, who complete topics in a different order, were excluded from the study population for this reason.
Regression	Similarly to the study reported in Chapter 3, involvement in this study was optional, and students were not selected on the basis of any previous score or performance. This type of study is likely to appeal more to students more likely to seek out opportunities to learn.
Selection	All participants who met the inclusion criteria were invited to participate, and all those who asked to participate were accepted, upon confirmation that they met the inclusion criteria. In this way, the researcher had no choice over who is accepted into the study, and selection bias from the researcher's perspective is limited.
Mortality (or Drop-out)	There were no participant deaths prior to the study completion, however one participant moved interstate prior to the final performance, and three others did not respond to multiple attempts to arrange the final performance.
Diffusion of treatment	Students would be free to communicate with their peers about their experiences. All students participated in a control and an experiment group. This was aimed, in part, at decreasing curiosity between participants in different groups.
Compensatory demoralisation	Teaching all participants both skills using both strategies was used in part to address this threat.
Compensatory rivalry	Teaching all participants both skills using both strategies was used in part to address this threat.
Testing	The same test conditions, script and room setup were used for all three student performances. This is known to potentially impact findings as participants may demonstrate an improvement from the testing which is separate from the intervention (D. Larsen, Butler, & Roediger III, 2009). Any effect is likely to equally manifest across all participants.
Instrumentation	The assessment scales remained standard between the three time periods assessed in the study, for each skill, to ensure direct comparison between all performances.

Adapted from "Research design" by John Creswell, pp. 174-176, copyright 2013 by Sage publications.

5.4.5.2 Potential threats to external validity of the study

Table 32: Threats to External Validity

Potential threat	How the threat was addressed in the study design
Interaction of selection and treatment	Student selection was primarily paramedic students, with a minority of nursing students. This was a natural result of varying recruitment methods between the two cohorts. This aspect will limit my ability to generalise the findings to the wider international pre-hospital care workforce or other health professions due to participant membership to a local professional student group in a single university. That said, the participants used were unlikely to be enculturated into the pre-hospital ambulance profession, and as such an argument exists to understand their results as somewhat representative of the sort of person who will learn resuscitation skills, for example in a first aid course or ALS course in-hospital.
Interaction of setting and treatment	The performance of both IO and LMA insertion occur in a controlled setting, where equipment is readily accessible and reasoning skills to guide the rationale of whether insertion is required (or not). This was arranged to ensure skill performance was measured, not a multifaceted professional performance, however it should be noted that participants' future performance of these skills

	will be in conjunction with other aspects of practice, for example rationale, assessment, handover, extrication planning, and as such it will be beyond the scope of this project to make strong claims about a student's predicted performance in this holistic setting.
Interaction of history and treatment	If the study was held during the participants' second- or third-year paramedic students due to a different level of underlying knowledge and experience. Thus, the results are better understood in terms of new skill acquisition and retention, rather than refresher training. This may still apply to established professionals learning a new skill, with the reservation that they are likely to have a more established schema on which to support new knowledge.

Adapted from "Research design" by John Creswell, pp. 174-176, copyright 2013 by Sage publications.

5.5 Results

The study design and participants numbers at each stage are presented in Figure 26.

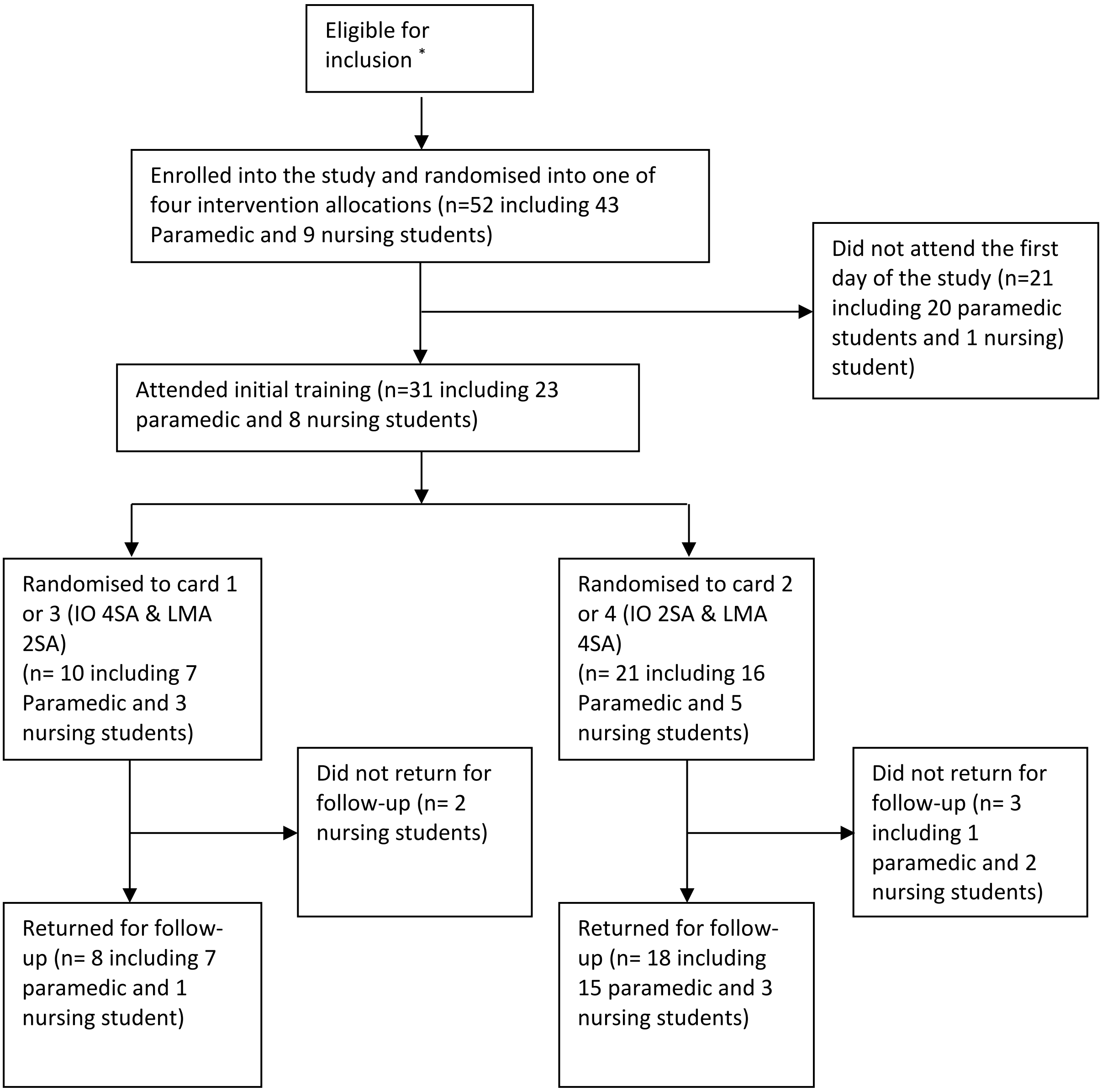


Figure 26: Participant inclusion and teaching allocation for retention study

* The number of total eligible students is unknown as access to nursing student enrolment records was not available[14]. However, approximately 115 paramedic students were identified as eligible for the study.

5.5.1 Demographic data

Participants' professional stream of study is represented in Table 33.

Table 33: Participants' Undergraduate Professional Study Stream.

		Registered	Attended training	Performed retention	Registration attrition	Attendant attrition
Total	Total	52	31	26	26	5
	Paramedic	43	23	22	21	1
	Nurse	9	8	4	5	4
Cards 2 and 4	Total	30	21	18	12	3
	Paramedic	24	16	15	9	1
	Nurse	6	5	3	3	2
Cards 1 and 3	Total	22	10	8	14	2
	Paramedic	19	7	7	12	0
	Nurse	3	3	1	2	2

Note: Groups taught with randomisation cards 1 and 3 were taught IO with 4SA and LMA with 2SA; Groups taught with cards 2 and 4 were taught IO with 2SA and LMA with 4SA. Attended training = students who attended initial teaching session; Performed retention = students who attended initial training and follow-up assessment 6 months later; Registration attrition = Number of students who registered but did not complete final assessment; Attendant attrition = Number of students who registered and attended initial training but did not complete final assessment.

5.5.2 Participant attrition

The nursing students formed a much smaller cohort, and represented much higher attrition than the paramedic students. Students were contacted multiple times both by mobile phone call (with messages left), and email in order to complete the data. The commencement of nursing placements may have been a factor in preventing the availability to return for follow-up, despite a high level of flexibility in scheduling the retention sessions.

5.5.3 Average time delay with retention data

The mean delay between the participants' teaching session (and therefore baseline and initial skill performance) was M=187.77 ±10.1 days (95% CI 183.67-191.86). The delay is positively skewed with

[14] A rough estimate of the number of nursing students enrolled is 300-400, however information on the number of eligible students was unavailable.

only 4 of the 26 completed final performances occurring more than 187 days following the initial teaching. Despite this non-normal distribution, the median delay was very close to the mean at 185 days, so overall this range is unlikely to impact the results significantly. A visual representation of the time delay for all participants who completed the final performance is found in Figure 27. Throughout the remaining study I will refer to the time delay as 6 months, with the acknowledgement that this is an approximation, and some students did not complete the assessment until 7.5 months later.

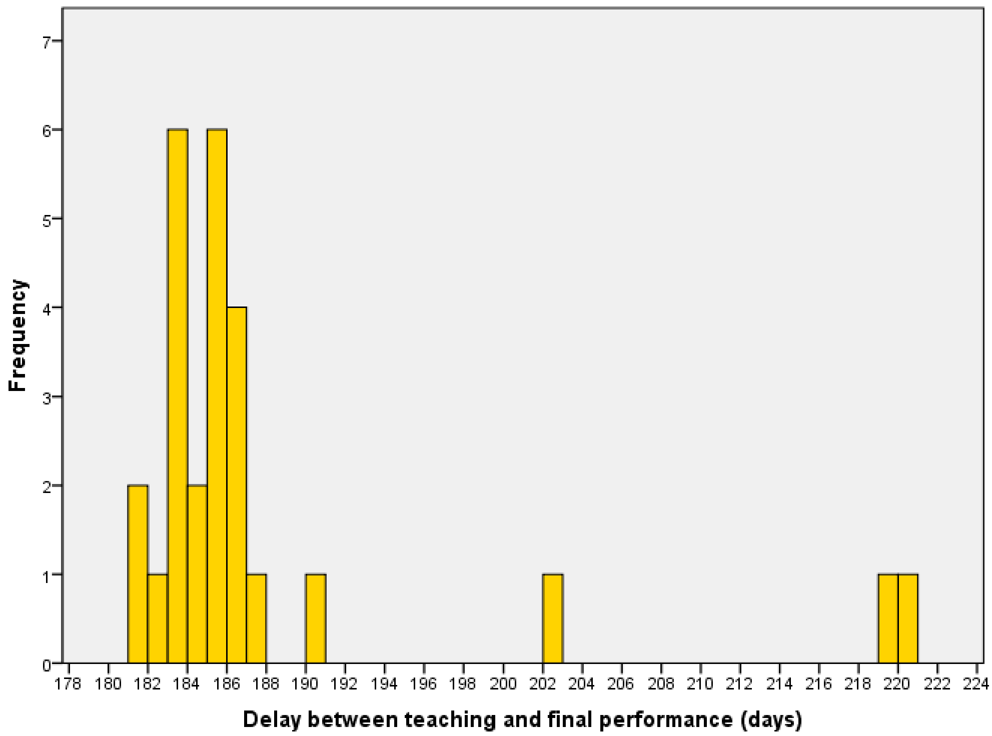


Figure 27: Time delay between skill teaching session, and the final performance of both skills

5.5.4 Exposure to skills between teaching and re-assessment

Between the initial teaching and assessment day and the reassessment, no study participants had inserted an LMA or IO device into a real patient (for example, for students who may have been completed clinical shifts or placements with the ambulance service who may have seen these skills during their shift). No participants had participated in a training session for IO insertion, however some participants had been exposed to skills refreshers in other ways (see explanatory note in Table 34).

Table 34: Participant Exposure to LMA and IO

		Total	Paramedic	Nurse
Participated in LMA training session*	Total	3	3	0

	Allocated to 2SA teaching	1	1	0
	Allocated to 4SA teaching	2	2	0
Witnessed LMA insertion**	Total	7	6	1
	Allocated to 2SA teaching	2	2	0
	Allocated to 4SA teaching	5	4	1
Witnessed IO insertion**	Total	5	5	0
	Allocated to 2SA teaching	4	4	0
	Allocated to 4SA teaching	1	1	0

Note: Exposure is considered witnessing, performing or being trained in either skill between initial teaching and final assessment
* Training sessions included a part of their normal study schedule, external student association training events, ALS courses, informal sessions during a clinical placement, or another source of exposure such as YouTube.
**This includes observing or being present, but not actually being trained, for example in the situations noted above

5.5.5 Participant attendance and teaching content audit

Of the 12 sessions allocated to 2SA, all 12 adhered to the study protocol. Of the 12 sessions allocated to 4SA, nine adhered strictly to the protocol. Three varied in a small way, but were still included in the final analysis as the variation was not expected to significantly impact the results.

Table 35: Audit of Study Protocol Adherence with Reference to Group Demographic and Attrition

Group	Profession	Skill	Method	Strictly adhered to protocol?	Included	Enrolled	Attended Day 1	Completed third attempt	Attrition
1	Paramedic	IO	2SA	Yes	Yes	5	4	4	0
		LMA	4SA	No	Yes- minor variation***				
2	Paramedic	IO	2SA	Yes	Yes	5	6	6	0
		LMA	4SA	Yes	Yes				
3	Nurse	IO	4SA	No	Yes- minor variation***	2	2	1	1
		LMA	2SA	Yes	Yes				
4	Paramedic	IO	4SA	Yes	Yes	6	1	1	0
		LMA	2SA	Yes	Yes				
5	Paramedic	IO	2SA	Yes	Yes	5	2	1	1
		LMA	4SA	Yes	Yes				
6	Paramedic	IO	4SA	No	Yes- minor variation***	4	2	2	0
		LMA	2SA	Yes	Yes				
7	Paramedic	IO	2SA	Yes	Yes	5	4	4	0
		LMA	4SA	Yes	Yes				
8	Paramedic	IO	4SA	Yes	Yes	4	2	2	0
		LMA	2SA	Yes	Yes				
9	Nurse	IO	2SA	Yes	Yes	3	3	3	0
		LMA	4SA	Yes	Yes				
10	Nurse	IO	2SA	Yes	Yes	2	2	0	2
		LMA	4SA	Yes	Yes				
11	Paramedic	IO	4SA	Yes	Yes	5	2	2	0
		LMA	2SA	Yes	Yes				

12	Nurse	IO	4SA	Yes	Yes	1	1	0	1
		LMA	2SA	Yes	Yes				

* Session 1 (LMA) began with Stage 2, which was then interrupted with Stage 1, then Stage 2 was completed and continued with Stage 3 and 4.
** Teaching session 3 (IO) began with Stage 2, which was then interrupted with Stage 1, then Stage 2 completed and continued with Stage 3 and 4
***Teaching session 6 (IO) delivered two demonstrations (Stage 1). In order to calculate the time taken to deliver the training, the average of these two times was used to calculated the TTT1-4exTD.

The stratified random allocation strategy evenly distributed the teaching allocations between the groups, however the different allocations were not evenly distributed amongst all participants as group size fluctuated from group to group. The following table outlines participant retention (including only participants who participated in the study):

Table 36: Participant Allocation and Retention Information

Participant number	Training group	Completed retention study (Y= yes; N=no)
1	3	Y
2	3	Y
3	9	Y
4	9	Y
5	9	Y
6	10	N
7	10	N
8	12	N
9	1	Y
10	1	Y
11	1	Y
12	1	Y
13	2	Y
14	2	Y
15	2	Y
16	2	Y
17	2	Y
18	2	Y
19	4	Y
20	5	N
21	5	Y
22	6	Y
23	6	Y
24	7	Y
25	7	Y
26	7	Y
27	7	Y
28	8	Y
29	8	Y
30	11	Y
31	11	Y

5.5.6 Hypothesis A: Skill specific checklist scores

This hypothesis is investigated with reference both to the skill-specific checklist scores, and morbidity/mortality risk scores according to instruments developed in Chapter 4. These scores were considered prior to teaching, immediately after teaching, and 6 months later.

5.5.6.1 Skill-specific checklist scores (variables "chlist_total" and " chlist_taught")

The data were examined to assess the suitability of parametric testing. Using SPSS, normality was investigated for each variable and teaching allocation. The parametric tests used were considered robust for an isolated violation of normality[15], in light of all other data satisfying this assumption.

In order to compare the acquisition and the retention/skill attrition over time between the two teaching methods, I ran a Repeated Measures ANOVA using performance data from the full validated checklist (Chlist_total, providing a maximum score out of 16 for LMA and 22 for IO performance), and an abridged version (Chlist_taught) which excludes the items which were either not addressed in *every* teaching session, or where participants were prompted to perform that component of the skill by the facilitator (maximum score of 11 for LMA, and 17 for IO). LMA and IO data were analysed separately.

5.5.6.2 LMA data

18 participants were taught LMA with 4SA, and 8 with 2SA.

5.5.6.2.1 The effect of teaching method

Following investigation of data suitability[16], the main effect of the teaching method on the performance measures demonstrated a non-significant effect regardless of whether performance is assessed with the complete validated checklist or the abridged version to account for teaching content and facilitator input (f(1,24)=0.724,*p*=.403 and F(1,24)=0.864, *p*=.362 for Chlist_total and Chlist_taught, respectively), using a Repeated Measures ANOVA. Hence, no further analyses were

[15] The LMA insertion data for the second performance violated the assumption of normality by means of both the Kolmogorov-Smirnov (K-S) and Shapiro-Wilk (S-W) tests (*p*<0.05). As the score for items taught is a subset of the whole checklist, it is unsurprising that both of these measures are non-parametric. As all other items satisfy normality (using K-S and S-W p>0.05), the data was analysed using parametric tests.

[16] A non-significant Mauchly's test confirms that the assumption of sphericity has not been violated by either the full validated checklist data, or the abridged version (*p*=.322 and .432 respectively). Likewise, Levene's test is not significant, F(1, 24)= 1.206, 0.352, 3.782, 0.372, 0.035 and 2.281 for LMA_Chlist_total at attempt 1, LMA_Chlist_total at attempt 2, LMA_Chlist_total at attempt 3, LMA_Chlist_taught at attempt 1, LMA_Chlist_taught at attempt2, LMA_Chlist_taught at attempt3 with *p*=.283, .558, .064, .548, .852 and .144 respectively, so the assumption of homoscedasticity is upheld.

performed. No significant interaction was detected between attempt and teaching method (*p*=.212 and .167 for the Chlist_total and Chlist_taught respectively).

5.5.6.2.2 The effect of attempt number

The attempt number had a significant effect on both outcome measures, F(1,24) =62.746 and 141.705 for Chlist_total or Chlist_taught respectively, p<.001 for both checklists.

The main effect of attempt number demonstrates that the baseline performance, immediately post-teaching performance and 6 month delayed performance are all statistically significant for the total validated checklist (*p*<.001 for all), and for the abridged checklist (*p*<.001 for all except between attempt 2 and 3 where *p*=.002). These are represented in Figure 28 and Figure 29.

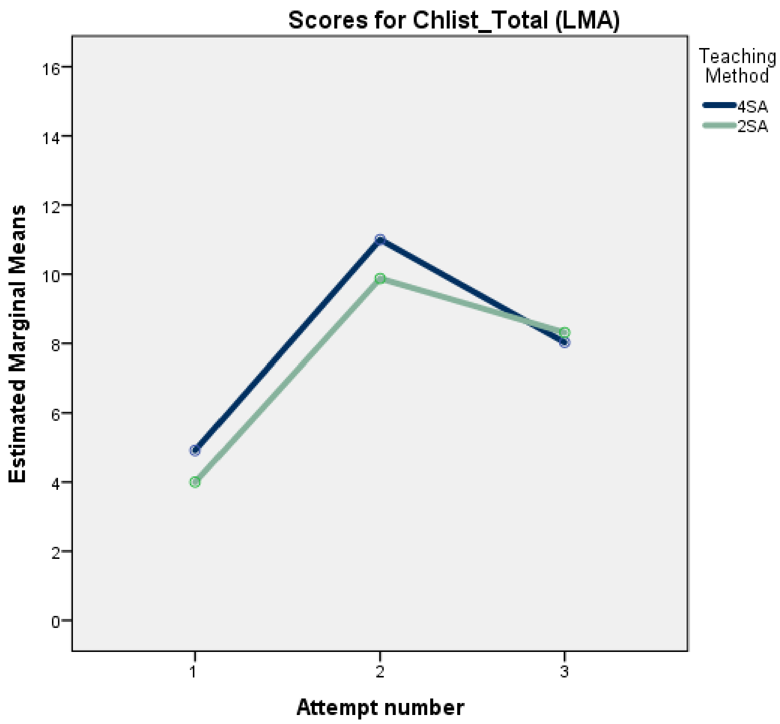


Figure 28: LMA score comparison between teaching methods (according to the Chlist_Total score for LMA). Max score possible = 16, and attempt number 1 = prior to teaching; 2 = immediately following teaching, and 3 = approximately 6 months following teaching)

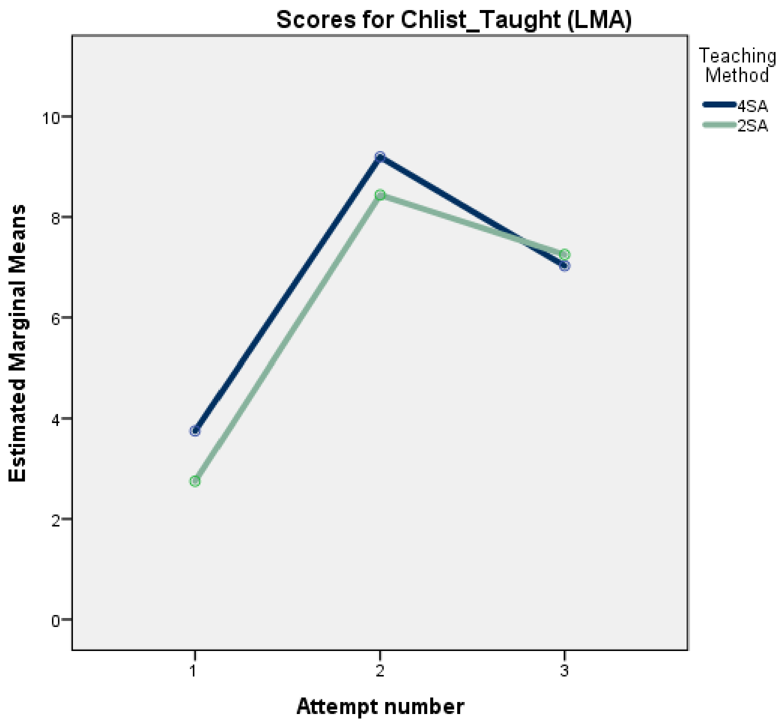


Figure 29: LMA score comparison between teaching methods (according to the Chlist_Taught score for LMA). Max score possible = 11, and attempt number 1 = prior to teaching; 2 = immediately following teaching, and 3 = approximately 6 months following teaching)

5.5.6.3 IO data

5.5.6.3.1 The effect of teaching method

The main effect of the teaching method on the performance measures demonstrates a non-significant effect on performance using both the complete validated checklist and the abridged version to account for teaching content and facilitator input (*p*=.823 and .855 respectively) in a two-way mixed ANOVA, following investigation of data assumptions[17]. No further analysis were performed. No significant interaction was detected between attempt and teaching method (*p*=.339 and .365 for the complete validated checklist and the abridged version respectively).

[17] Mauchly's test is not significant for both the full validated checklist score, and the abridged version (*p*=.677 and .545 respectively). Levene's test is not significant (F(1, 24)= 0.745, 1.133, 0.010, 0.202, 0.071, 0.023) for IO_Chlist_total at attempt 1, IO_Chlist_total at attempt 2, IO_Chlist_total at attempt 3, IO_Chlist_taught at attempt 1, IO_Chlist_taught at attempt2, and IO_Chlist_taught at attempt3, with *p*=0397, .298, .922, .658, .792, .882 respectively). The assumptions of sphericity and homoscedasticity are therefore not violated in the data.

5.5.6.3.2 The effect of attempt number

The attempt number had a significant effect on both outcome measures, $F(1,24)=170.581$ and 163.608 for the complete validated checklist or the abridged version respectively; $p<.001$ for both checklists.

The main effect of attempt number demonstrates that the baseline performance, immediately post-teaching performance and 6 month delayed performance are all statistically significant for the total validated checklist ($p<.001$ for all), and for the abridged checklist ($p<.001$ for all). These are represented in Figure 30 and Figure 31.

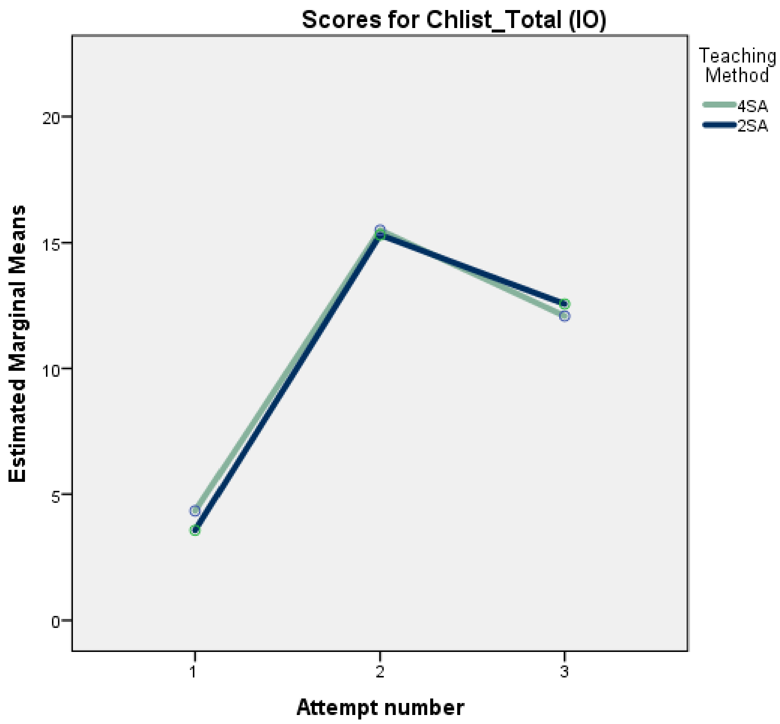


Figure 30: IO score comparison between teaching methods (according to the Chlist_Total score for IO). Max score possible = 22, and attempt number 1 = prior to teaching; 2 = immediately following teaching, and 3 = approximately 6 months following teaching.

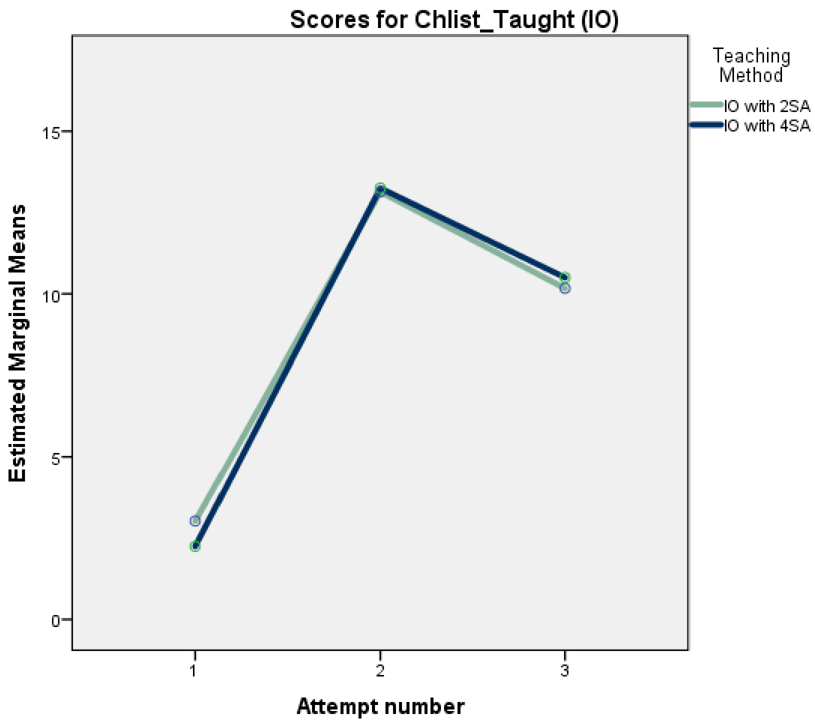


Figure 31: IO score comparison between teaching methods (according to the Chlist_Taught score for IO). Max score possible = 17, and attempt number 1 = prior to teaching; 2 = immediately following teaching, and 3 = approximately 6 months following teaching)

The students taught LMA with 4SA had a sharper decline in LMA insertion ability between the last two attempts than their peers who were taught with the 2SA, visible in Figure 28 and Figure 29. These same students had also had a slightly more prominent decline in IO ability when they were taught with the 2SA than their peers who were taught with the 4SA. As a similar trend is seen with the same group of students with both skills (sharper decline in IO performance ability alluded to in Figure 30 and Figure 31), even though the skills were taught with different teaching methods, I suggest that this is due to participant variation, and not an experimental intervention.

5.5.6.4 Morbidity/mortality risk scores (variable "morbid_mortal")

These scores were obtained from the critical items identified in sections 4.4.3.2 and 4.4.4.2. They reflect the items which the panel generally agreed would likely result in increased patient morbidity or mortality if performed incorrectly.

5.5.6.4.1 LMA morbidity and mortality factors

The main effect of the teaching method on the performance on the morbidity or mortality items was not significant ($F(1,24)=0.680$, $p=.418$). Hence, no further analysis were performed. Repeated Measures ANOVA was performed although the assumption of homoscedasticity was not upheld at every point in the data[18]. Puristically, a different statistical test could be pursued which is does not demand homoscedasticity, however RMANOVA was considered sufficiently robust. These data are depicted in Figure 32.

5.5.6.4.1.1 The effect of teaching method

No significant interaction was detected between attempt and teaching method ($F(1,24)=0.680$, $p=.220$).

5.5.6.4.1.2 The effect of attempt number

The attempt number had a significant effect on the student's morbidity/mortality score, $F(2,48)=98.411$, $p=<.001$. The morbidity/mortality score at each attempt was statistically different from both other time points ($p=.005$ for attempts 2 and 3, and $p<.001$ for all other comparisons).

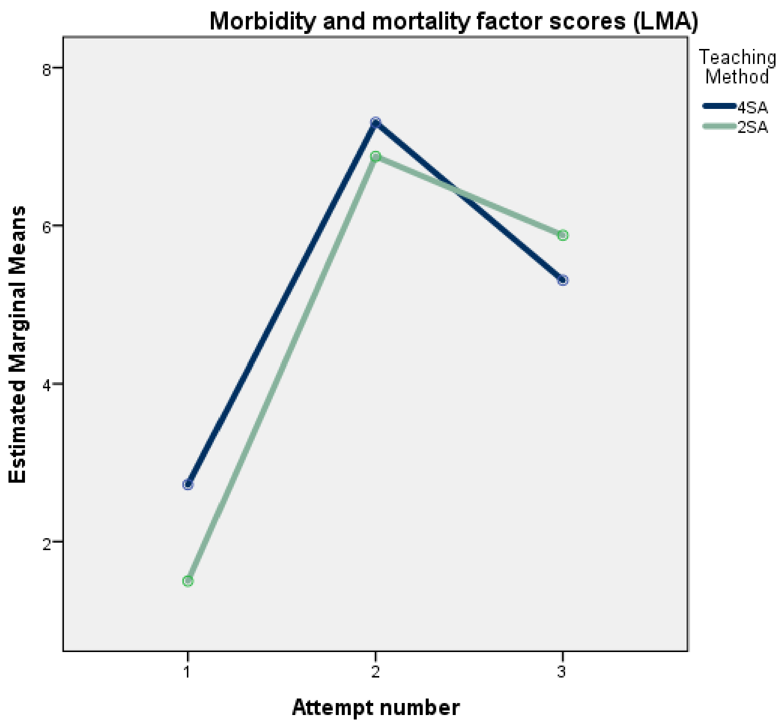


Figure 32: Scores for morbidity and mortality items performed correctly during LMA insertion (where a high number represents lower patient risk with more of these items performed correctly). Max score possible = 8, and attempt number 1 = prior to teaching; 2 = immediately following teaching, and 3 = approximately 6 months following teaching.

[18] A non-significant Mauchly's test confirms that the assumption of sphericity has not been violated by either the full validated checklist data, or the abridged version ($p= .09$). Levene's test is not significant ($F(1, 24)=0.172$, $p=.682$) for attempt 2, however is significant ($F(1,24)=4.290$ and 9.375, $p=.049$ and $.005$) for attempts 1 and 3 respectively.

5.5.6.4.2 IO morbidity and mortality factors

The main effect of the teaching method on the performance measures is seen in the "tests of between-subjects effects", and demonstrates that the teaching method does not have a significant effect on the morbidity/mortality risk associated with the student's IO performance (F(1,24)=1.068,*p*=.312). Hence, no further analysis were performed.

5.5.6.4.2.1 The effect of teaching method

No significant interaction was detected between attempt and teaching method, F(1.765, 24)=1.561, Huynh-feldt corrected *p*=.223[19].

5.5.6.4.2.2 The effect of attempt number

The attempt number (within subjects) had a significant effect on the student's morbidity/mortality score, F (1.765,42.363)=169.735[20], *p*=<.001. The morbidity/mortality score at each attempt was statistically different from each other time point (*p*<.001) except for the comparison between attempts 2 and 3 (*p*=.063).

---

[19] A significant Mauchly's test indicates a violation of the assumption of sphericity (*p*=.036), and as the Greenhouse-Geisser value is greater than 0.75, the Huynh-Feldt correction of 0.883 will be used to correct the degrees of freedom (Field, 2009a, p. 562). Levene's test is not significant for the first attempt, F(1, 24)=0.196, *p*=.662, however scores for attempts 2 and 3 for IO morbidity/mortality are heteroscedastic, indicated by significant Levene's test (F(1,24)=4.454, *p*=.045 and F(1,24)=5.051, *p*=.034 respectively). This indicates that the variance in morbidity/mortality scores following the training (attempts 2 and 3) was different between the two teaching methods, but prior to training it was not. Mean scores for morbidity and mortality items are presented in Figure 33.

[20] Fuymh-feldt corrected

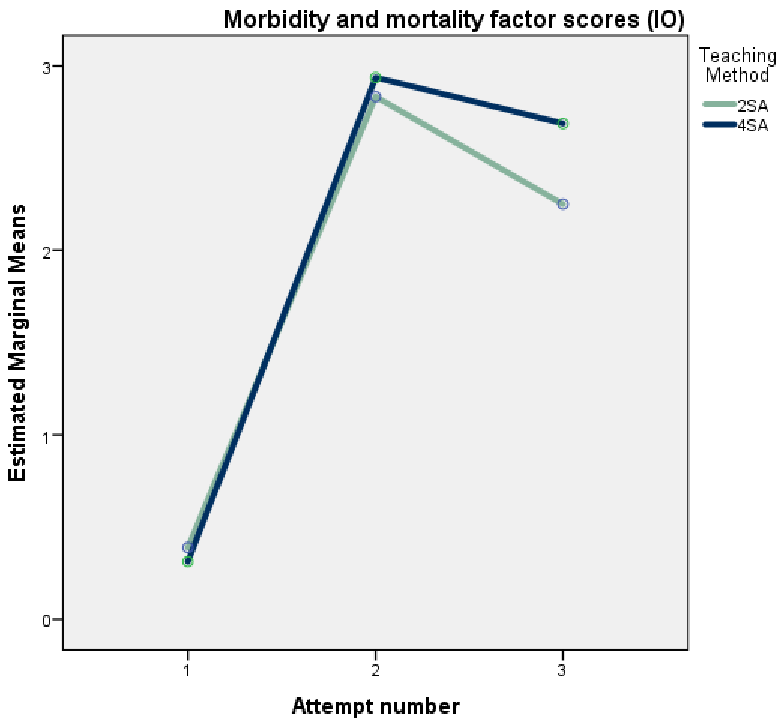


Figure 33: Scores for morbidity and mortality items performed correctly during IO insertion (where a high number represents lower patient risk with more of these items performed correctly). Max score possible = 3, and attempt number 1 = prior to teaching; 2 = immediately following teaching, and 3 = approximately 6 months following teaching.

5.5.7 Hypothesis B: Global scores (GRS_total25)

The GRS data were examined for distribution trends to assess the suitability of parametric testing. Normality was investigated for each variable and teaching allocation[21]. Mixed ANOVA was used to investigate the impact of teaching method on skill acquisition and retention over time.

5.5.7.1 LMA global scores

Levene's test produced a non-significant result for attempts 1 and 3 $F(1, 23)= 0.008$ and 1.564, with corresponding significance of $p=.929$ and .224 respectively. Data for the second attempt violates the assumption of homoscedasticity $F(1, 23)=4.676$, $p=.041$ as identified in Levene's test. This compromises the accuracy of the F-test for that attempt (Field, 2009b). The attempt number was shown to have a significant effect on the GRS, $F(1.758, 23)=94.138$, $p<.001$ (with Huyn-Feldt corrected F-value[22]). These data are presented in Figure 34.

[21] The global rating score for the first attempt of IO insertion violated the assumption of normality by means of both the Kolmogorov-Smirnov and Shapiro-Wilk tests ($p= .045$ and .009 respectively), however as all other items satisfy normality (using K-S and S-W $p>.05$), the data was analysed using parametric tests.
[22] Mauchly's test indicated that the assumption of sphericity had been violated, $\chi^2(2)=6.668$, $p=.035$. Therefore, the Huynh-Feldt correction will be applied to the LMA data F-values (Field, 2009b).

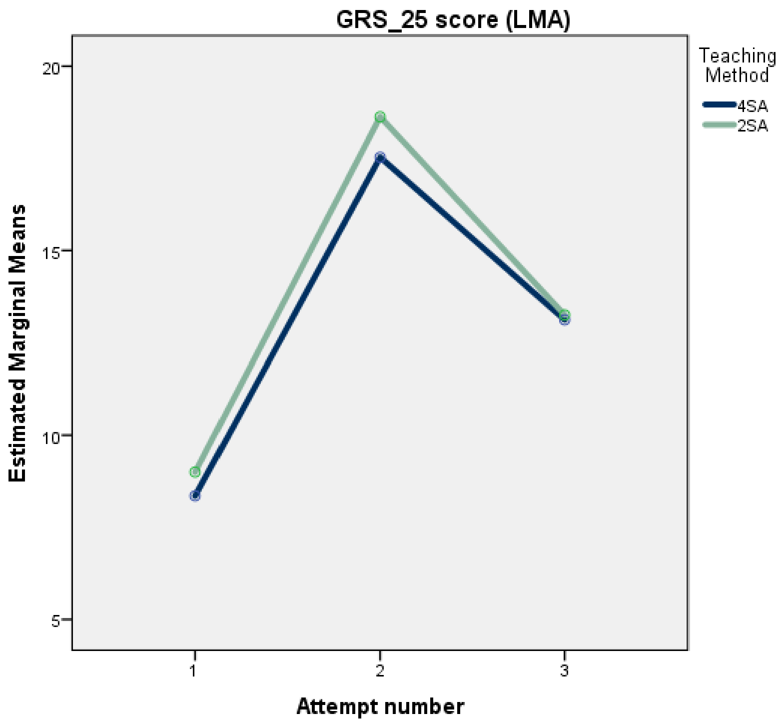


Figure 34: Global scores for LMA insertion, consisting of 5 Likert scales of 1-5. Possible score range 5 to 25, and attempt number 1 = prior to teaching; 2 = immediately following teaching, and 3 = approximately 6 months following teaching.

5.5.7.1.1 The effect of teaching method

The main effect of the teaching method on the participant's global performance score is not significant ($F(1, 23)=.506$, $p=.484$). No significant interaction was detected between attempt and teaching method ($F(1,23)=0.458$, $p=.723$). Hence, no further analysis were performed.

5.5.7.1.2 The effect of attempt number

The attempt number was shown to have a significant effect on the GRS, ($p<.001$).The main effect of attempt number demonstrates that the baseline performance, immediately post-teaching performance and six month delayed performance are all statistically significant for the GRS scores ($p<.01$ for all).

5.5.7.2 IO global scores

Mauchly 's test indicated that the global score data for IO performance did not violate the assumption of sphericity ($p=.299$). Levene's test indicated that the assumption of homoscedasticity is also not violated for each attempt scored by the GRS ($F(1,23)=3.034$, 0.222 and 1.754 for attempts 1, 2 and 3 respectively; $p=.095$, .642 and .198). These data are presented in Figure 35.

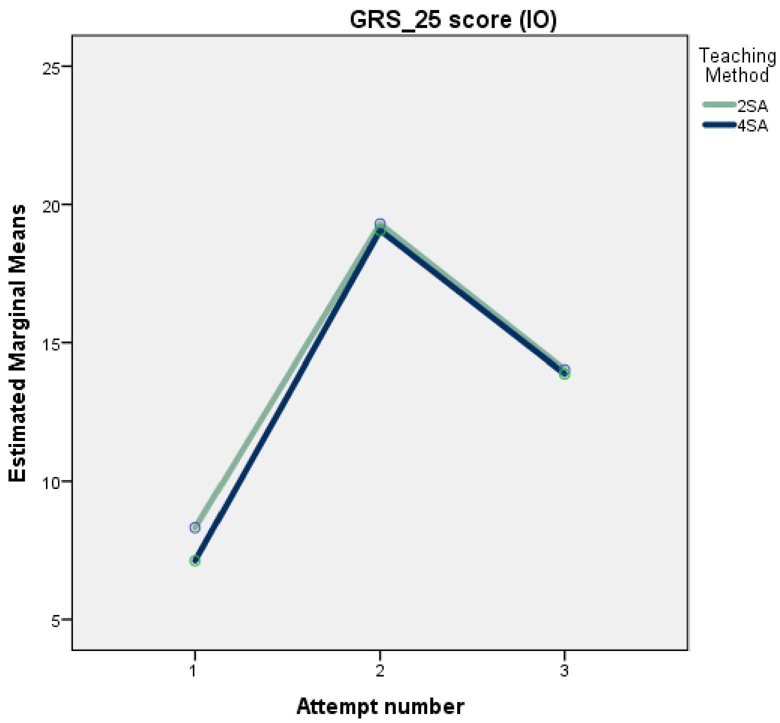


Figure 35: Global scores for IO insertion, consisting of 5 Likert scales of 1-5. Possible score range 5 to 25, and attempt number 1 = prior to teaching; 2 = immediately following teaching, and 3 = approximately 6 months following teaching.

5.5.7.2.1 The effect of teaching method

The main effect of the teaching method on the participant's global performance score is not significant (F(1, 23)=.271, *p*=.607). No significant interaction was detected between attempt and teaching method (F(1, 23)=.058.*p*=.505). Hence, no further analysis were performed.

5.5.7.2.2 The effect of attempt number

The attempt number was shown to have a significant effect on the GRS, *p*<.001.The main effect of attempt number demonstrates that the baseline performance, immediately post-teaching performance and six month delayed performance are all statistically significant for the GRS scores (*p*<.01 for all).

Therefore, with the data obtained I was not able to reject the null hypothesis that the teaching method has any measurable impact on the student performance scores, whether a skill specific checklist with binary itemised coding is applied, or a global rating scale is applied.

5.5.8 Hypothesis C: Time comparison

5.5.8.1 Multiple variable analysis

A visible trend that the time to teach 4SA is greater than that required to teach 2SA is demonstrated in Figure 36. This was therefore entered into the multivariable model last in order to understand how much of the variance is explained by the intervention variable (teaching method) and secondly the type of skill being taught (IO or LMA). More detail of these three models is presented in Table 37.

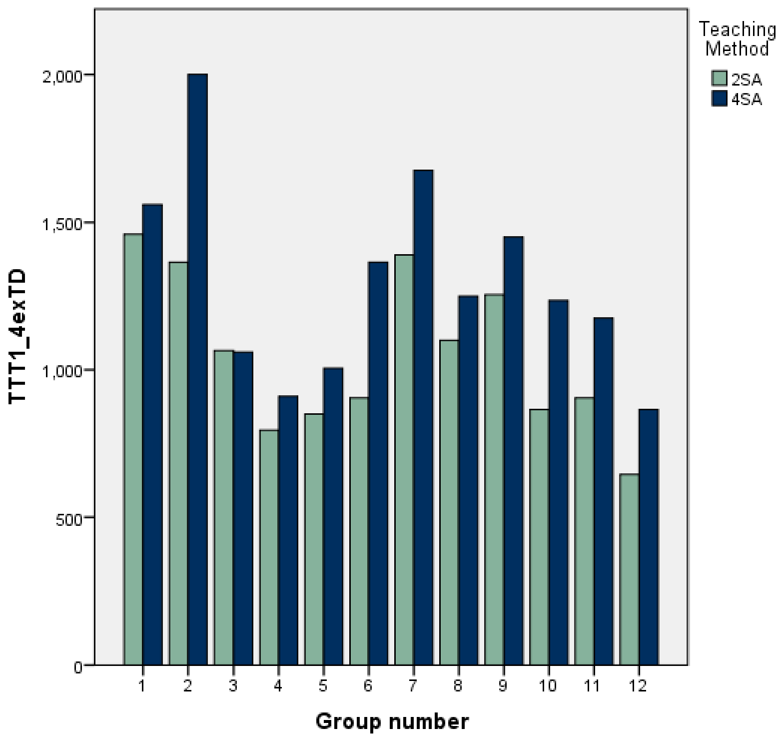


Figure 36: The time to teach both skills for all groups, identified by teaching method. TTT1_4exTD refers to the time to teach, excluding tangents and delays in seconds.

In considering the predictors in the correlations output in SPSS, the highest correlation to the outcome measured (time taken) is the group size at .835 (p<.001). The correlation (standardised coefficient) of the teaching method to the time required to teach is R=.391, and increases to only .398 when the type of skill is also included in the model, indicating that the skill taught had little impact in this study. This is reported in Table 37.

The first model, accounting only for the effect of the teaching method on the time required to teach, is not significant (p=.059) even though the above graph (Figure 36) seems to indicate visually that this is the case. Model two also demonstrates that the teaching method does is not significantly impact the teaching time, however model three which incorporates the teaching method, skill, *and* group size is very significant with a more substantial F-ratio (F=39.34, *p*<.001). A model which incorporates group size, given it demonstrated the highest correlation to teaching time, is important to better understand the individual impact of other variables. This indicates that more of the variation in teaching time is explained by the variables included in the third model. For this model (Model 3 in the table below), it is observable that when group size is taken into account, the teaching method has a significant impact on the teaching time required (*p*<.001).

Table 37: Coefficients Relating to the Impact of Teaching Method, Skill Type and Group Size on the Time to Teach

Model		Unstandardised Coefficients		Standardized Coefficients	t	Sig.
		B	Std. Error	Beta		
1	(Constant)	804.167	194.918		4.126	.000
	Teaching method	122.917	61.639	.391	1.994	.059
2	(Constant)	737.917	274.196		2.691	.014
	Teaching method	122.917	62.905	.391	1.954	.064
	Skill	44.167	125.810	.070	.351	.729
3	(Constant)	247.600	126.828		1.952	.065
	Teaching method	122.917	26.741	.391	4.597	.000 *
	Skill	44.167	53.482	.070	.826	.419
	Group size	189.800	19.350	.835	9.809	.000 *

Note: Time to teach is investigated using the variable TTT1_4exTD, measured in seconds. The sample size is relatively small with only 12 sessions for each teaching strategy
* indicates statistical significance (*p*<.05).

5.5.8.2 Ratio analysis (IO:LMA)

The 12 teaching groups each participated in two skill sessions: one skill by one method, and the other skill by the other method. Using multivariate analysis could be considered a violation of the assumption of independence of cases. I then converted the two times for each group to a single ratio of Time to teach IO to time to teach LMA (calculated as TTT1_4exTD for IO and TTT1_4exTD for LMA). This allows each group to act as its own control, and the impact of fluctuating group sizes becomes, to an extent, accounted for. This new variable, "Ratio_IOtoLMA" is then explored for normality (K-S and S-W both not significant) before an ANOVA is run. The assumption of homogeneity of variances was not violated (Levene's test p >.05). The ratios are compared on the basis of whether IO was taught with the 4SA (and therefore LMA was taught with the 2SA), or vice-versa.

A one-way ANOVA (between groups) identified a significant difference, dependent on which teaching method was employed $F(1,10)= 25.586$, $p<.001$, $\omega=.673$. This indicates a large effect size. The 95%CI confirm significant difference, as they do not cross zero ($M_{4SA/2SA}$ =1.237, 95% CI=1.047-1.428, and $M_{2SA/4SA}$ =0.810, 95% CI=0.706-0.914). These values indicate that the 4SA takes approximately 1.24 times the amount of time that 2SA does (24% longer). These data are represented in Figure 37.

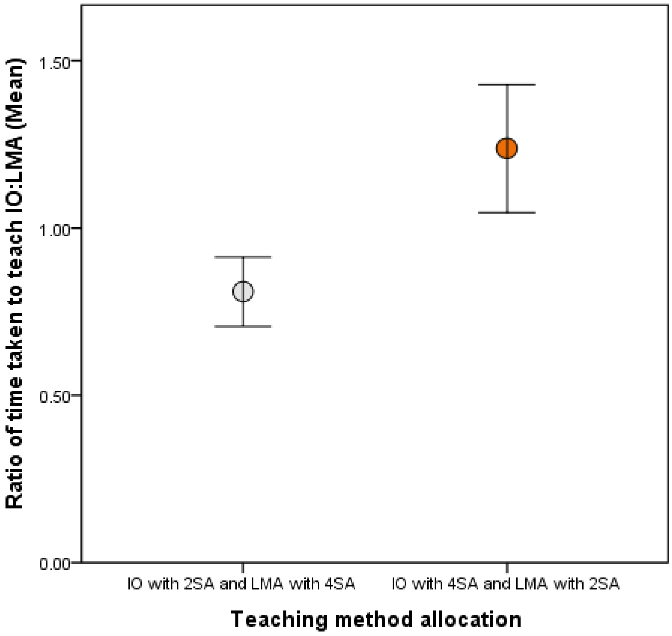


Figure 37: Ratio of time required to teach with the two methods. The box to the left reflects data for the times of groups allocated with cards 1 and 3; and the box on the right reflects data for the times of groups allocated with cards 2 and 4 (Mean ±SD)

Individual group data depicting the time ratio of IO:LMA are shown in Figure 38.

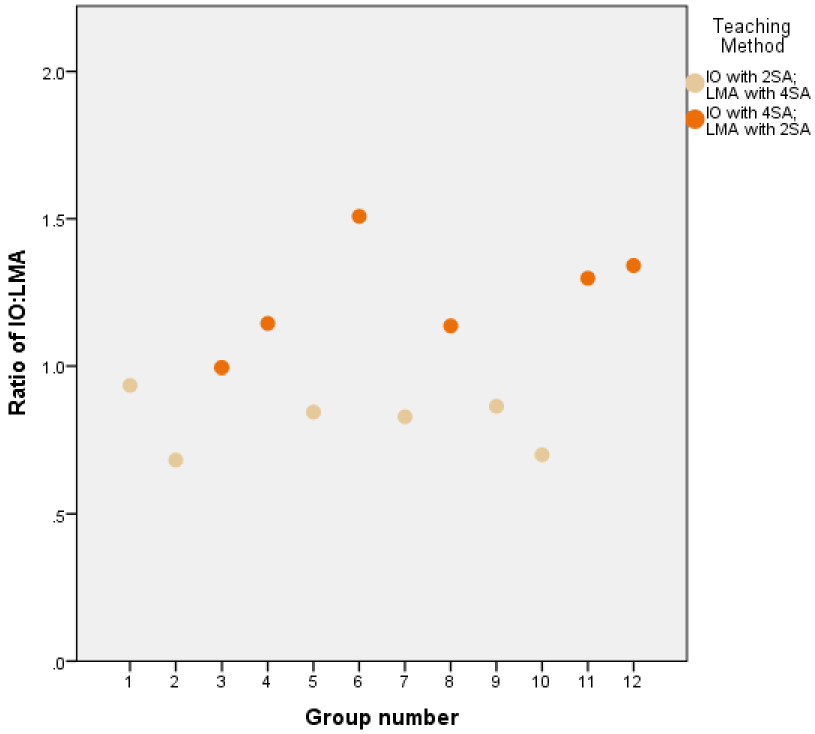


Figure 38: Ratio of time required to teach with the two methods, with each group represented as a separate data point.

The previous finding that the skill taught has no significant impact on the time to teach (see Table 37), it seemed reasonable to compare the time to teach with 2SA and 4SA in this way, thereby controlling for individual group factors as each ratio is calculated using both skill session times for one group.

5.5.9 Hypothesis D: Confidence

An independent samples T-test was performed to compare the mean confidence rating for students taught LMA and IO insertion by the different methods. Students rated their confidence on a scale of 0 to 10, according to the scale noted in Appendix 10.4.4. This test found that students who are taught with 2SA are no more or less confident of their ability to insert an LMA correctly (confidence = 4.38±1.41) than students taught with 4SA (confidence =5.61±1.79), $t(24)= -1.73$, $p=.097$.

The median and interquartile range for these ordinal data are shown in Figure 39 as a complement to that just discussed. It shows a slightly higher confidence for the students taught LMA with 4SA, which is the same group marking a slightly higher confidence with IO insertion in Figure 40. The lack of statistical significance of these data in the independent samples T-tests described is supported by the marked crossover of interquartile ranges between the two teaching interventions.

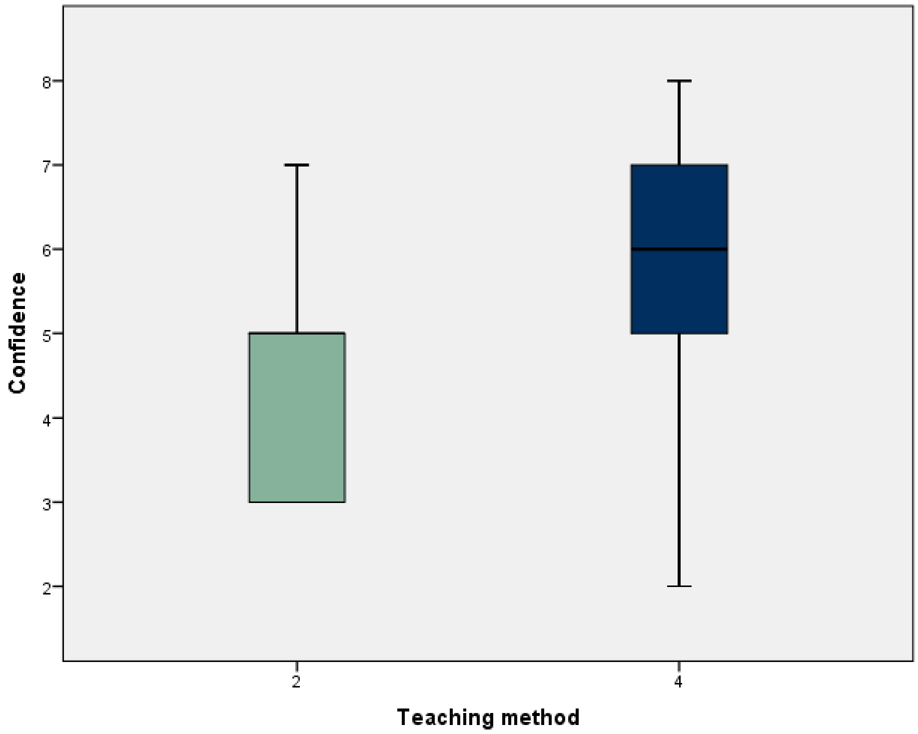


Figure 39: Comparative student confidence (median and IQR) of their LMA insertion ability on a scale of 0 to 10 (where 0 is not confident at all, and 10 is completely confident). Teaching method 2=2SA, and teaching method 4 = 4SA

Students who were taught IO with 2SA are similarly no more or less likely of their ability to perform IO insertion correctly (mean confidence = 5.00±1.680) than students taught with 4SA (4.25±1.488), $t(24)=1.085$, $p=.289$. The median and interquartile range are shown in Figure 40.

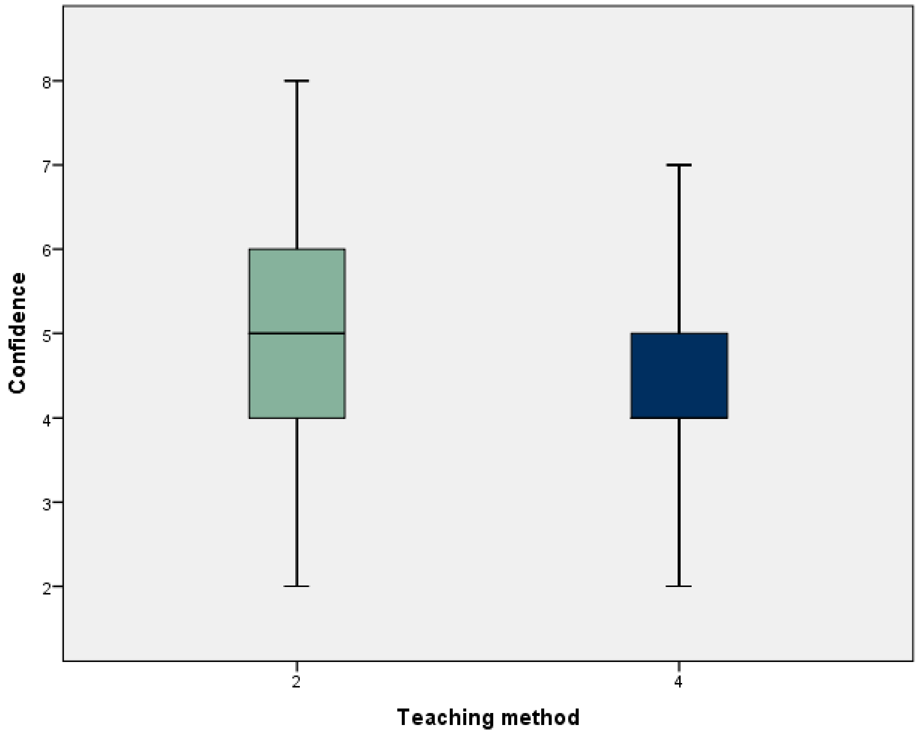


Figure 40: Comparative student confidence (median and IQR) of their IO insertion ability on a scale of 0 to 10 (where 0 is not confident at all, and 10 is completely confident).

The group of students taught IO with 2SA (and therefore LMA with 4SA) appear to have a slightly higher average confidence when both skills are taken into account (average confidence of 5.31±1.58) than those taught IO with 4SA and LMA with 2SA (4.31±1.10) but this is not statistically significant, $t(24)=1.603$, $p=.122$. Hence, no one group is identified as generally more confident than the other. It may be that this study is underpowered to demonstrate such a difference, as indicators may be seen in the slightly higher reading for LMA confidence in the 4SA group, and the IO confidence in the 2SA group.

A paired samples T-test compared whether students are more confident with LMA insertion (mean confidence 5.23±1.751) than IO insertion (4.77±1.632), and this is not statistically significant, $t(25)= -1.513$, $p=.143$. Students who were more confident in one skill tended to me more confident in the other (Spearman's rho .573, $p=.001$[23]), however despite this correlation, there is still reasonable spread noted on the correlation graph, with some indication that greater confidence in LMA is still common despite low confidence in IO. These students (circled in Figure 41 in red) did not fall into a single intervention allocation.

[23] One-tailed

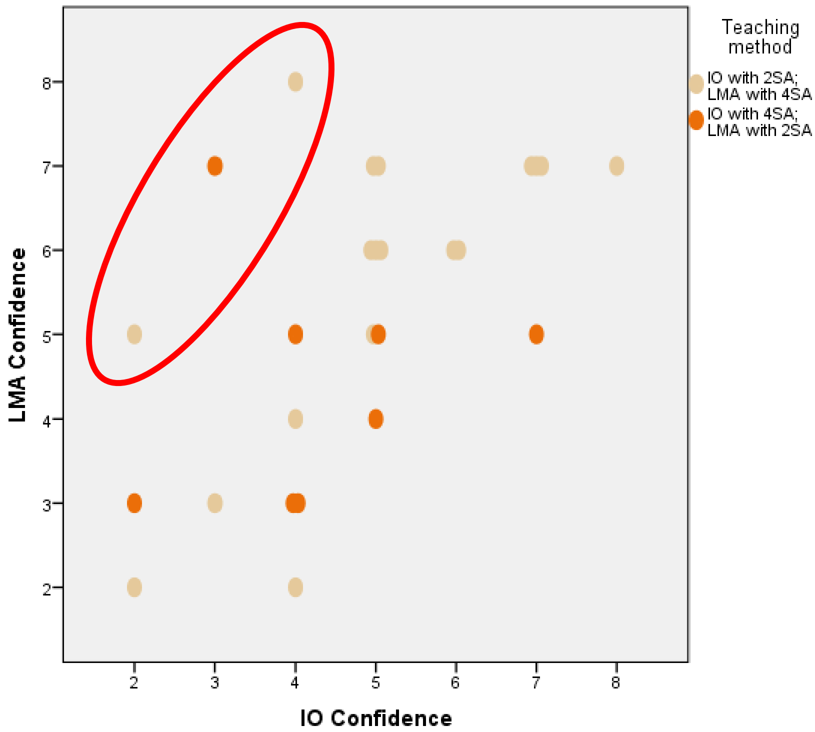


Figure 41: Correlation of IO and LMA confidence with each data point representing a single student.

5.5.10 Hypothesis E: IO placement into medullary cavity

The percentage of students who insert the IO into the bone's medullary cavity is fairly proportional across both groups (see Table 38), indicating that the teaching method does not impact the success of functional IO access.

Table 38: Rates of Successful IO Insertion, by Teaching Method

Teaching method	Successfully placed IO into the bone	Unsuccessful IO placement	Total participants in each group
4SA	57% (n=4)	43% (n=3)	7
2SA	61% (n=11)	39% (n=7)	18
Total	60% (n=15)	40% (n=10)	25

Note: the success or otherwise of IO placement into the bone was recorded by the skill facilitator during the performance video. I edited these data out of the videos distributed for marking, and noted successful placement during video editing based on facilitators' comments following the performance.

This analysis speaks to the idea that a student can perform the skill effectively (that is, placing the IO needle in the bone for fluid and medication administration during resuscitation), even if they choose the incorrect site, neglect infection control processes etc. While this poses some risks to the patient, in the resuscitation setting, in the absence of IV access, this will at least allow access to the circulation for Advanced life support, and other potential problems may then be managed if the patient survives the cardiac arrest. These data do, however, indicate that successful insertion of a manual IO device may only occur around 60% of the time, only 6 months following training.

5.6 Discussion

This study addressed the cost-effectiveness of clinical skill education, in comparing two commonly used teaching strategies. Cost was measured in the use of time and resources, and 4SA was found to be significantly more costly in both of these terms. Effectiveness was measured in student skill performance according to specific checklists, morbidity and mortality estimation, and global rating scales. No clinically or statistically significant difference was identified when I compared students taught with 2SA and 4SA. I also recorded student confidence and perceived competence in both skills, to understand whether, aside from actual competence, if there was a difference between the two methods, and no difference between the two methods was identified. Thus, the research question initially phrased simply as referring to cost-effectiveness, has fantailed into many practical considerations of clinical skill education design.

5.6.1 Demographic/exposure data

Of the students who had witnessed (but not been formally taught) LMA insertion between the initial training and retention data collection, more belonged to the intervention teaching allocation (5 for 4SA compared to 2 for 2SA). The inverse was true for those who had witnessed IO insertion (4 for 2SA and 1 for 4SA). As this was not a teaching session, it was anticipated to have minimal impact, and limited sample sizes did not allow for further analysis with this additional variable.

5.6.2 Findings in reference to the theoretical model

The theoretical model employed to understand key aspects of cost and benefit in clinical education allows the study findings to find a relevant place in the practical setting. Of the eight key stakeholders identified for symbiotic clinical education, the clinicians, patients, government, community, health service and teaching institutions were considered in terms of these measures of cost and effectiveness.

5.6.2.1 Cost of teaching

Training time heavily impacts the demands of a health service provider. Training (both initial, ongoing, and *upskilling*) are essential to the development and ongoing competence of any health professional. It is often required for career progression, and some courses (such as ALS or PALS) are often an organisational and government requirement to maintain national health standards within an organisation (Australian Commission on Safety and Quality in Healthcare, 2010). The time taken to complete such training becomes a burden either to the organisation, or the health professionals who may complete it in their own time.

The cost of resources is likely to be closely related to the specific skill, and what equipment it requires. For example, a skill such as application of a leg splint for a femoral fracture uses equipment which can be used many times over without the use of consumables, whereas other skills may require expensive equipment which can not be re-used. As such, the cost for different skills should be considered on an individual basis.

The finding that 4SA requires significantly more time than 2SA outlines a relatively taxing burden on teaching institutions. Whether the teaching institution is a university, education unit of a health service, or a private provider, available funds are often limited and careful consideration must be given to the commitment of such funds to training. This aspect is better understood with reference to student performance and skill retention, as a greater initial outlay may be acceptable where re-training requirements and risk to patients is lessened.

A Bland-Altman analysis was considered as a more established method to measure the agreement (or difference) between the two times recorded for each group, in order to understand if 4SA was discernibly different to 2SA. Rather than converting the two scores to a ratio, this test considers the two measures, the difference between them (by subtraction), and the mean (Giavarina, 2015). It aims to compare two quantitative measurements of the same phenomenon (for example two different marking checklists both applied to the same skill performance), whereas the two sets of times recorded in this study were two different skill teaching sessions.

5.6.2.2 Effectiveness

Patient morbidity and mortality risk factors were not significantly different between the two groups (2SA and 4SA). Student performance, in terms of both validated skill-specific checklists and GRS were found to be statistically similar at each time point (prior to teaching, immediately after teaching, and approximately six months after teaching). From these findings I argue that the 4SA, while underpinned by and developed with regard to accepted learning theory, does not result in practitioners who will necessarily be able to practice with increased patient safety or proficiency according to the measures used in this study. The measures of expected morbidity and mortality risk include aspects which may be anticipated to relate to infection, de-oxygenation injury to organs and tissue, and may be responsible for considerable cost to the health provider (for example through Intensive Care Unit admissions or infection management), community (for example the cost of disability and death to the community through income loss, and increased welfare through disability support), and the clinical educators who may otherwise need to expend varying levels of time, resources and energy re-training.

5.6.2.2.1 Skill atrophy

An important yet not unexpected finding relates to the quantification of skill acquisition and atrophy. Over the course of six months, the participants' skill diminished significantly following training. When we consider the prevalence of competency based assessment in clinical accreditation and maintenance of accreditation for various professions, this data heralds a warning which is consistent with much other literature: competency immediately following a short course or training session does not imply ongoing competence.

Many critical resuscitation skills are used infrequently, as a result of infrequent exposure to critical cases. It may be well over six months before a clinician trained in resuscitation skills is required to employ resuscitation skills, thus we see the importance of ongoing training and maintenance of expertise move again into the centre stage. Attrition of resuscitation skills is a significant problem (Ali et al., 1996; Amaral & Troncon, 2013; Driscoll et al., 1999; Wayne, Siddall, et al., 2006; Wiles, 2015) with one study noting the clinical setting and subsequent low exposure as adversely effecting ongoing skill ability (Ali et al., 2002).

5.6.2.3 Confidence

Our study found that neither 2SA nor 4SA related to significantly greater in improved confidence. In the literature, confidence tends to be a poor predictor of performance ability (Ehrlinger & Dunning, 2003; Eva & Regehr, 2005; Turner, Lukkassen, Bakker, Draaisma, & ten Cate, 2009; Wayne, Butter, et al., 2006) and should therefore not be relied upon to build an understanding of resuscitation skill ability. Wayne, Butter, et al. (2006) noted that there is "no practical association between observed simulator performance on [Advanced Cardiac Life Support] events and self-assessed clinical confidence", and that residents had "generally expressed high self-confidence" about managing the scenarios. An important tension is therefore revealed: is it better for clinicians to have high self-confidence, even if this is unmatched by competence? Could this help put others in the team, and the patient and patient's family at ease by raising the confidence others have in their ability to manage the problem? Or does this come at a risk to patient care, where further advice or resources may not be perceived as needed even when they are? In my study, as data relating to confidence were only identified during follow-up, I am unable to make comment on trends over time, however Gerard et al. (2011) found that the resuscitation skill ability of residents taught during a PALS course decreased over time where confidence does not. In an investigation of a number of resuscitation skills, Turner et al. (2009) found that some skills (ETT insertion and IO insertion) were associated with greater confidence than other skills. Self-confidence, therefore, is not a reliable function of training strategy nor actual ability. Dunning et al. (2003) further warn their readers that this inability to identify incompetence is a "double curse", in that they are also impaired in identifying others' mistakes.

The implications for clinical practice are multi-layered. On the one hand, where a clinician is unconsciously incompetent due to skill atrophy beyond their recognition, a patient may be at risk. On the other hand, in the resuscitation setting, some infrequently used skills are immediately life-saving, and hesitation due to low confidence may have a detrimental impact. Thus early CPR is a key link in the chain of survival, in addition to defibrillation via an Automated External Defibrillator (AED). The delineating factor between these two situations returns to the original definition of skill, as encompassing expertise beyond performing an action. It entails professional standards, and expectations placed on the practitioner by those who are less-skilled, or un-skilled, and in that perspective, a mismatch between confidence and ability is very different than for the first-aider.

5.6.3 Morbidity and mortality

There was no notable difference in the number of critical items performed adequately between participants from the two teaching interventions. The conclusions drawn from this must be cautioned. First, it is imperative to understand that while these items are developed according to expert clinician consensus, they have not been validated through experimental design. Thus, these values represent the anticipated risk, rather than the actual risk. Williams (2011) note the difficulty in gathering data on actual cardiac arrest survival changes as a function of educational intervention. Reasons for this may include the indirect impact of clinician development on the patient, the multitude of co-variants which may also impact patient outcomes (an increase in education may reflect an increase in funding which may also create an increase in technology which may have an independent yet measurable effect on patient outcomes), and the ethical dilemma of conducting a controlled comparative trial where a group of clinicians is intentionally withheld from training.

5.6.4 Risk of IO insertion

Of the two skills, a similar acquisition-retention profile is observed, however in some ways the two skills are very different. This is first seen in Chapter 4 during the skill assessment checklist development, and will be further discussed in Section 6.3.1.2 with reference to natural variation in correct and acceptable LMA insertion techniques, evident more so than for IO insertion technique. Further, in the local practice setting, IO insertion has been protected as a skill for Intensive Care Paramedics, whereas LMA insertion is a treatment option available to Paramedics and Ambulance

Officers also. This may give the illusion that IO is a more difficult skill, or that it the potential dangers of its use warrant that it be reserved for the clinical elite.

Thus, an argument may be made to health care providers that IO access ought to be available to medics involved in fluid or cardio-pulmonary resuscitation where the timely and effective administration of fluids or medication into the patient's central circulation may be lifesaving, especially where persistent attempts at IV access, or Central Venous Catheter have been identified to decrease success and increase time to achieve central venous access (T. E. Anderson et al., 1994; Paxton, 2012). In 2010 the American Heart Association (AHA) updated their guidelines to promote IO access as a rapid and more straight forward means of vascular access than peripheral IV access (American Heart Association, 2016; Kleinman et al., 2010) with the European Resuscitation Council (ERC) guidelines moving to a recommendation to attempt IO access if IV access is likely to be delayed, or after one minute of unsuccessful attempt (Maconochie et al., 2015). With a movement towards a focus on IO use as a rapid means to achieve vascular access, Paxton (2012) note that it is generally underutilised. Lack of adequate training in the skill was noted as the second most common reason for not attempting IO access when it was perceived as clinically appropriate. The most common reason was, reasonably, a lack of available equipment to do so. In fact, in a study performed by Hallas (2012), only 5.6% of Danish Emergency Doctors surveyed (10 of the 178) had received hands-on training in using an IO device. Low awareness of the benefits and safety IO use may promote a culture where what is familiar to the clinician (IV access) will be a less resistant clinical option, evidenced by a study which found PALS accredited clinicians were more than twice as successful at gaining IO access, and more likely to attempt it (Baker et al., 2009; Paxton, 2012).

The fear of risk associated with IO insertion may also be disproportionate. IO access carries lower risk of infection and complication than is often perceived, even compared to IV access as a relatively routine procedure. Schalk et al. (2011) confirm the use of various IO devices including the manual device studied here, as a rapid, safe and effective form of central venous access, although pain upon immediate infusion in the responsive patient is a key limitation, and ought to be conducted with an anaesthetic agent for these patients.

5.6.5 Generalisability

In considering two different skills, I am able to demonstrate overall consistency in the findings between the two: 4SA had no measurable effect on the acquisition, retention, or self-rated confidence of either skill. Time proportions were not significantly different between IO and LMA teaching. It may be argued, therefore, that other clinical skills of similar manual difficulty will likely require approximately 24% more time to teach with 4SA than with 2SA, and with no discernible differences in skill application or confidence.

In terms of generalising the findings to the clinical setting, assessment of a skill performance in isolation from the complex clinical processes and pressures on a clinician who ought to insert an IO or LMA device may perform differently compared to the controlled setting presented in this study. Van der Vleuten and Schuwirth (2005) argue that more complex areas of clinical practice require a multifaceted collection of assessment methods, prompting improvement in assessment rigour using a programmatic, complementary assessment approach rather than a single perspective on an isolated performance. This is consistent with the definition that assessment is the "process of collecting information about a student's learning and clinical performance over time, generally by using multiple strategies" (Oermann, Saewert, Charasika, & Yarbrough, 2009, p. 274). A randomly selected sample would provide more substantial support for an argument of generalisability. The participants were selected from a specific clinical course, and in addition, voluntary enrolment in the study may have promoted a biased group of participants.

5.6.6 Limitations

5.6.6.1 Participant Attendance

Attendance for this study was again a problem for the collection of data. The relatively small sample left after the initial participant attrition following registration and the project beginning, and the further drop-out of five participants left 26 students with complete data sets for each attempt. This has had two key impacts. Firstly, the group size has fluctuated as a result. Secondly, there is a much increased risk of a type II error due to a small sample size. This unpredictable participant withdrawal prior to the commencement of the study, but following allocation to randomised groups, was unavoidable. Owen and Plummer (2002) note the impact that different group sizes is observed to have on perceived student engagement. Where groups of two students were taught, this provided the students enough of a rest between practices with the facilitator for their arm to recover from the physically demanding act of laryngoscopy during intubation, a skill which requires a very specific coordination and strength in the left (usually non-dominant) arm, which can cause muscle fatigue. They noted that one student per facilitator did not afford the arm muscles enough recovery, and three or more students led to apparent disinterest and distraction while the students awaited their practice. This aspect of group size and the impact it has on student attention and learning, with respect to the individual physical requirements for the skill are not considered in this study, and cannot be with current participant enrolments, but are very practical considerations to the organisation and planning of a teaching session. More research could extend to how personality traits (like introversion and extroversion) may impact the efficiency of various group sizes.

5.6.6.2 Equipment limitations

During IO insertion, the IO needle was re-used and tended to become blocked with fragments of bone from the chicken leg. These fragments would wedge between the inside of the sheath and the stylet, making withdrawal of the stylet difficult, and at times almost impossible. This created disruption in some training sessions and skill performances.

Chicken legs were sourced to simulate a patient's tibia. Alternatives such as turkey thigh/drumstick, pork rib, IO manikins, and even modified/frozen Chocolate bars were considered (Paxton, 2012),and chicken drumstick selected on the basis of availability, affordability and advice in the literature (Ota, Yee, Garcia, Grisham, & Yamamoto, 2003). Chicken drumsticks are a much more affordable and disposable alternative to specially designed part-task trainers such as in Gerard et al.'s study, (2011). In this study, the chicken drumsticks brought some limitations, as they tended to have variable rigidity, and at times the chicken leg would snap during insertion of the IO, even when technique seemed appropriate.

The manikin leg used for the student to indicate IO insertion location was also problematic. While the patella and tibial tuberosity were palpable, being composed of rigid plastic, they were less lifelike and made it difficult to decipher the difference between where the bone would be palpable, and where the muscle would be palpable. Use of the manikins allowed for much better consistency than asking the participant to indicate the insertion location on the facilitator, however. This alternative would have also created logistical problems with camera angle adjustments. The tibial tuberosity was also more subtle on the manikin than it is on many real legs.

5.6.6.3 Assessment method

The assessment method(s) used will inevitably limit the data obtained in the study. Van der Vleuten and Schuwirth (2005) point out that the selection of any method of assessment "inevitably entails compromises and... the type of compromise varies for each specific assessment context." The selection of GRS and skill-specific checklist assessment tools to provide data on simulated, isolated (for example, not in the context of a contextual scenario) skill performances will naturally limit the inferences possible for practice of these skills in a more authentic setting. These limitations were balanced with the benefits to the study in terms of standardisation of assessment to allow more direct comparison between the performances.

5.6.7 Implications

These data suggest that for the two skills used in the study, 4SA does not result in a measurable difference in skill acquisition or retention, but will require significantly more time to teach with (approximately 24% more time than 2SA). Walker and Peyton initially predicted an approximately 30% increase in time required to teach in the surgical laboratory compared to having the student simply observe (Walker & Peyton, 1998), however the difference noted in this study measures the difference between two teaching strategies. Using 4SA in the clinical setting is likely to be greater still compared to not teaching.

These findings challenge the widespread use of 4SA as an everyday teaching method for resuscitation skills. The additional time required to use the four steps will be an additional cost to teaching organisations (such as room bookings, educator wages), and may cause unnecessary student fatigue due to prolonged teaching time. Owen and Plummer (2002) notes that the longer skill teaching sessions he reviewed (>90 minutes) appeared to be too fatiguing for many students, and the data in this study suggests that this would be for no measurable learning benefit.

5.7 Conclusion

This study used a comparative trial to compare the cost-effectiveness of two common teaching methods, with consideration to the time required to teach, and skill performance six months following the teaching event based on skill-specific checklists, global rating scales, and morbidity/mortality subscales. This helps address the underlying problem facing many teaching facilities: How should educators be asked to teach clinical skills? Students were taught LMA and IO insertion in small groups arranged according to their availability. Each group was taught according to a randomly selected card which identified which teaching method would be used for each skill, and which order the skills would be taught in.

Data gathered from blinded assessors who marked the performance videos determined that the teaching method had no significant effect on the acquisition or retention of LMA or IO with reference to the skill specific checklist, the checklist adapted to include only the items taught, or the global rating scale or student confidence at the retention assessment. No impact in morbidity or mortality factors was noted. Neither skill teaching method related to greater likelihood of inserting the IO needle into the medullary cavity of the chicken leg. 4SA was noted to require approximately 24% more teaching time than 2SA, therefore this study concludes that a simpler 2SA to teaching resuscitation skills is more cost-effective. Results obtained in this chapter will be interrogated and critiqued in Chapter 6 in order to either defend or refute the rigor of the claims made about the effectiveness of 4SA in this chapter.

6 A CRITICAL ANALYSIS OF ASSESSMENT TOOLS USED

6.1 Introduction

I began this study by asking: What criteria do I need to satisfy in order to *validate* the checklist developed in Chapter 4 and defend the results obtained in Chapter 5? This question follows the frustrating realisation that following a prescribed and accepted process (the Delphi process) to develop the assessment tools would not, in fact, produce a *validated* checklist. A more critical appraisal of the tools developed, and their application, was required (Kane, 2006). In part, my confusion stemmed from the use of "validity" as a term often referring to appropriate study methods, data collection, and data analysis accounting for sources of bias and aspects of reliability. This chapter's question of validating the assessment tool closely relates to the validity and reliability of the study outcomes, but goes further to consider the application of the study's outcomes.

In the studies upon which the Delphi study in Chapter 4 was modelled (D. Berg et al., 2013; K. Berg et al., 2013; Berg et al., 2014; Nicholas Hartman, Wittler, Askew, & Manthey, 2014; Huang et al., 2009; Lee Ann Riesenberg, Katherine Berg, Dale Berg, Joshua Davis, et al., 2013; Lee Ann Riesenberg, Katherine Berg, Dale Berg, Kathleen Mealey, et al., 2013; L. A. Riesenberg et al., 2013), α is offered as a reflection of the level of agreement (or consistency) between the different raters' scores for each item, although α is used in diverse ways to measure consistency. It has been selected by the above scholars, among others, as a means to argue whether acceptable consensus has been reached. Generally, if $\alpha > .7$, data have historically been accepted as internally consistent.

When I measured α for the two checklists developed in Chapter 4, I found that while the IO checklist inter-rater agreement was considered sufficient (at $\alpha > .7$), the value for LMA was not (at $\alpha < .7$). This prompted much disappointment and months of reading followed in order to better understand inter-rater consistency measures. After investigating the different assumptions behind common coefficients of variation and consistency including Kappa, Fleiss' Kappa, Kendall's coefficient, Scott's Pi, and the Intraclass Consistency Coefficient (ICC), the two-way consistency fixed effects[24]. ICC seemed best suited to the design and data[25], and not α after all. This is consistent with Cronbach's comment that "I no longer regard the α formula as the most appropriate way to examine most

[24] Two-way ICC was used because the same raters were used for both rounds of the Delphi; consistency measures were determined because my interest is in agreement of trends rather than absolute agreement of the number; and fixed effects were used because the sample of raters was not random (intentionally).
[25] Data were ordinal, not normally distributed, and domain referenced. Cronbach's alpha requires , compared continuous, normally distributed norm-referenced data.

data". But after determining the ICC, I again found that it was considered sufficient for IO, but not for LMA (with a value approximating .4, and a non-significant p-value). I was at a fork in the road: I could choose to rely on the Delphi process and the expertise brought to the study by the clinician participants for the face validity of the study, or I could state valid results for the IO checklist only, based on the statistical analysis performed. Building an argument on the basis of the expert participants felt to me to have only limited credibility, hence I could not silence the reasonable arguments that failing to meet an arbitrary standard for consistency analysis rendered the entire LMA checklist invalid. This tension led me to further scrutinise the context, professional practice, and methodological approach underpinning the Delphi study. Through months of confusion, I now better understand where consistency analysis may offer helpful insight, and where it may overstep its appropriate boundaries.

In this chapter, I will review the different ways in which researchers tend to approach validity, in order to critically examine the assessment tools developed in Chapter 4. I will make clear the distinction between addressing valid *data collection and interpretation* in the study design, and addressing the validity *of the study outcomes* (the assessment tools which were developed). I will review the consistency between participants in the development of the tools, the trends in data obtained from the applied tools, and consider an argument for the validation of these tools using Kane's (2006) validity framework. As such, the findings presented in Chapter 5 regarding the comparative cost-effectiveness of 2SA and 4SA will undergo rigorous interrogation. During this chapter, I will use the term "rater" to refer to the clinical specialists who participated in the development of assessment tools in Chapter 4, and "assessor" to refer to those who applied the tools to a student performance in Chapter 5.

6.1.1 Internal and external validity during the assessment tool development

A focus on gathering quality data through rigorous methods is necessary to build a case for the credibility of the claims made (Patton, 2002b, p. 552). Acceptable and valid analysis is considered necessary to build a case for the rigour of the claims made. So it is clear that validity, credibility and rigour are closely related and contribute to one another. Creswell (2013, p. 201) argues that validity is not "a companion of reliability [examining stability] or generalizability [the external validity of applying results to new settings, people or samples]". Validity, Creswell argues, refers to the accuracy of the data obtained, whereas reliability refers to the consistency of the researcher's approach. The internal and external threats to validity have been examined for both the qualitative and quantitative aspects of the study, however these considerations tend to restrict the discussion to the accuracy of the data and their analysis. The danger, therefore, is a tempting and logical assumption that valid (accurate) data interpreted in a valid (trustworthy, transparent and appropriate) way, having been gathered by means of a valid (established and tested) procedure will result in a valid ("validated") study output. Despite appearing explanatory, I suggest that the logic is fractured, by yet unspoken assumptions. Further examination must be performed in order to present a rigorous and acceptable validity argument, including both a psychometric analysis of the data obtained from the assessment tools, and an interrogation of the assumptions behind the interpretation of such data.

6.1.2 Psychometrics

The psychometric framework is one way to examine the reliability (precision) of an assessment tool, and implies that a true score exists. This may be in the assumption that a true value exists to describe how important each criteria is (as in Chapter 4), or the true score for each student's performance (as in Chapter 5). Hence, measuring the precision of an assessment tool is deeply important to this view, and is often achieved through measuring consistency, either of multiple student attempts, or multiple assessor ratings of one student. Where data clusters closely, precision is accepted, and the assessment tool is argued to be more reliable (Govaerts et al., 2007; Van der Vleuten, 2000). Govaerts et al. (2007) clearly describe further aspects of a psychometric approach to validation, namely that variation is considered error or bias, and is unwanted; raters are interchangeable, trainable, and therefore can be considered highly consistent instruments of measurements in the application of the assessment tool; the ability of students under examination is fixed; and low correlation (for example between GRS marks and those derived from a skill-specific checklist) implies low validity (Govaerts et al., 2007, p. 241). Psychometrical statistical review forms justification for higher stakes decisions such as whether a student should pass or fail (p. 245). This data dependency "will tend to promote data-driven or bottom-up processing", which takes into account the facts and evidence pertaining to an assessment or decision. Thus, measures such as the skill-specific checklists developed in Chapter 4 can argue the constitution of a given score, and be a means to improve assessment standardisation or fairness. An inherent risk noted by Vnuk (2013) is that students end up "practising the checklist and reciting the findings" (p. 50). This assessment may then become a theatrical performance where a student's lines are prompted by a memorised marking sheet without underpinning clinical understanding.

Assessment methods which use a more "top-down" processing, however, tend to demonstrate weaknesses in psychometric analysis. Such assessments use global judgements about a student's ability, based on more holistic impressions steeped in assessor experience and expertise. Such holistic judgements tend to be provoked by "situational cues", assessor's past experience, and therefore tend to exhibit higher variability between raters (Govaerts et al., 2007). Global judgements (such as the GRS used in Chapter 5) tend to overlook the vast detail, and therefore are often easier, faster scoring system. Such variation may be derived from bias or stereotyping (p. 243). Despite variation in how sale terms such as "unsatisfactory" and "satisfactory" are interpreted (Gingerich et al., 2014), however such ratings have been demonstrated to be more accurate reflections of performance, increasingly so when used by more experienced assessors (Govaerts et al., 2007, p. 244).

6.1.3 Kane's "validity as argument"

While a measure of reliability (repeatability, consistency, and data precision) may be reported in a single coefficient, the same cannot be said for validity as it is conceptual (Van der Vleuten, 2000). Furthermore, Kane (2006, p. 17) defines validation as the "process of evaluating the plausibility of proposed interpretations and uses", based on the evidence which supports or refutes them. Thus, validation occurs once these underpinning assumptions have been exposed and examined, and an application intention of the study outcome made explicit, rather than from a defined procedure such as the Delphi model.

Validity procedures are employed to support the data and claims made within the research, however when we turn our attention to validating an educational tool, the focus is on the extent to which the measurement tool is accurate and appropriate in its resultant claims (Schuwirth & Ash, 2013, p. 413), and whether it measures what it claims to (Van der Vleuten, 2000). This depends on valid research processes during the tool's development, but looks beyond the development itself to the application of the tool. So what qualifies an education or assessment tool *validated*? The notion of assessing the validity of assessment tools has evolved significantly over the last century. From early descriptions of the criterion model of validity in the 1920-1950's, to the prominence of content validity and soon thereafter construct validity in 1955 (Cronbach & Meehl, 1955), finally a unified model of validity, with a focus on interpretation, placed criterion and content validity nested within the construct being examined. Kane presents this new age to validation practice with "validity as argument" (Kane, 2006, p. 23). The *interpretive argument* makes explicit the assumptions inherent in the assessment tool's proposed interpretations and uses, and the *validity argument* critiques this (Kane, 2006, p. 23). Validation of an educational assessment tool must therefore address two related facets: firstly, the evidence in support of the interpretation of the tool's measurement, and secondly, the plausibility of such claims. This includes understanding the processes and assumptions behind an observation which impact the scoring decision (Gingerich et al., 2014; Schuwirth & van der Vleuten, 2012), inferences made when extracting universal scores from these data, claims about the target domain, and claims about the wider construct. This particular study focuses on the first two stages. Kane's definition of validation is adopted for this discussion chapter as:

> the process of evaluating the plausibility of proposed interpretations and uses (p. 17).

Validity, in relation to an educational measurement tool will be defined as:

> the extent to which the evidence supports or refutes the proposed interpretations and uses (Kane, 2006, p. 17).

The studies upon which Chapter 4 was modelled all contained the title phrases: "the development of a validated checklist for *(skill)* "; "validity of a (*skill)* checklist"; or "validation of a performance assessment scale for *(skill)* ". All used the Delphi process, and all report inter-rater consistency with α above .79, although most were .94 or above. This statistical figure was reported but not expounded, however the inclusion of such a statistic seemed to intend to argue the findings as consistent and therefore valid, with sufficient agreement between raters on the final checklist. It is clear that there is no such thing as a single validation strategy, so this multi-faceted approach, all the while seeking to retain connection to the outcomes and context of the skills applied, aims to address a range of arguments imperative to critiquing such a tool.

This chapter will present and critique the validity arguments in reference to the assessment tools as developed in Chapter 4, before exploring the application of these tools with reference to the four stages of Kane's argument validity as identified by Schuwirth and van der Vleuten (2012) as outlined in Table 39, and with reference to psychometric data as appropriate.

Table 39: Overview of the Inferences Necessary to Move From Observed Data to Conclusions About a Construct

	Medical example	Example from the checklists developed in Chapter 4	Inferences necessary to move to this stage from the previous
Observation	Acoustic sounds through a stethoscope when a clinician is measuring a patient's blood pressure	An action which is witnessed in the student skill performance	
Test score	The visual reading on a sphygmomanometer relating to the onset and disappearance of the noted sounds	The score, out of 16 (for LMA) or 22 (for IO) resulting from the addition of individual skill components performed	The observations have been observed and interpreted correctly and reliably
Universe score	A diagnosis of hypertension may be made if criteria are satisfied	A statement about a student's IO or LMA insertion ability (for example a competence statement)	That the measurement taken is sufficient upon which to argue a generalisable score

Target domain	The patient's cardiovascular health	A statement about a student's psychomotor skill proficiency	Data from other sources must be obtained and triangulated to form a holistic picture of the target domain
Construct	The patient's general health status	The student's clinical practice ability	Likewise, additional data must be obtained and to form a holistic picture of the construct

Note: The medical example used and the inferences mentioned are adapted directly from "Programmatic assessment and Kane's validity perspective" (2012) by Schuwirth, L., & van der Vleuten, C., *Medical education, 46*(1), 38-48. Copyright 2011 by Blackwell Publishing.

6.1.4 A brief note about consistency

Inter rater consistency coefficients (IRCCs) measure the level of agreement between different raters' scores on a scale. There are two ways I will refer to IRCCs in this chapter: firstly, in reference to the development of the IO and LMA assessment tools in Chapter 4, and secondly in reference to the data generated by multiple markers applying these tools in Chapter 5. It is important to distinguish between these two uses for consistency assessment, because correct interpretation hinges on examining the question underlying the consistency analysis. The questions in the two studies mentioned are very different, and assumptions about the underlying construct vary.

6.2 Assessment tool development

6.2.1 Internal consistency measures of raters' scores during the checklist development

Where $\alpha > .7$, data have historically been accepted as internally consistent, however this reflects a narrow understanding of the statistic (Cortina, 1993). For example, when a set of scores from a student's performance are analysed for consistency and a large α (approaching 1) is obtained, this may indicate a highly repetitive set of items, and therefore item redundancy rather than consistency across a wider cross-section of the construct being studied (Streiner, 2003). In the Delphi studies mentioned earlier, a high α may indicate that a relatively homogenous group has been recruited, and therefore minimal variability in practice or recommendation is evident. This may be as intended, to understand practice within a niche clinical setting, but will limit the generalisability to a wider field of clinical practice. This prompts the need to establish a diverse, heterogeneous participant panel as outlined in section 4.3.2.2. Thus α approaching 1 may reflect consistency due to reliable agreement, or due to homogeneity. The studies in Chapter 4 sought a heterogeneous panel or pre-hospital clinicians, so α approaching 1 is not expected.

A very high α may also be a simple function of the list length, or item list dimensionality (Cortina, 1993). Cortina goes on to state:

> One reason for the misuse of alpha in applied psychology is that there seems to be no real metric for judging the adequacy of the statistic. Experience with the literature gives one some general idea of what an acceptable alpha is, but there is usually little else to go on. Note, however, that those who make decisions about the adequacy of a scale on the basis of nothing more than the level of alpha are missing the point of empirically estimating reliability. The level of reliability that is adequate depends on the decision that is made with the scale. The finer the distinction that needs to be made, the better the reliability must be... Thus, any judgment of adequacy, even in research, needs to consider context (Cortina, 1993, p. 101).

These comments provide an important background to α, and indeed other measures of consistency to an extent. Cortina is referring here to the consistency of different raters' scores on something which is assumed to have a *true score*, therefore precise clustering of different raters' scores around a single value will provide some argument that a) the tool is reliable, and b) the true score likely within or near the cluster. Granted, both of these arguments may be challenged.

This approach in pursuit of the true score of a given item, by means of a consistency rating between different participants is reflective of the psychometric framework. This perspective identifies a "specified level of consistency that is assumed to be conditional on technically sound measurement and the assumption of error when repeated measurements fail to yield consistent result" (Govaerts et al., 2007, p. 241). Govaerts continues on to highlight the lack of regard this approach has for the assessors, context and environmental variation. She argues that "raters are considered to be interchangeable, 'measurement instruments', and ratees' ability is assumed to be a fixed, permanent and acontextual attribute" (Govaerts et al., 2007, p. 241). This approach may be appropriate for comparing assessors' scores using this checklist, as applied to the data in Chapter 5, but for the assessment tool itself, the context to which Cortina (1993) urges us to attend is quite different: there is no *objective truth* to uncover in the assessment tool development. This assumption is key to the driving methodology of Chapter 4 as a qualitative construction based on expert practice. Cronbach himself later reflects:

> In 1951, I published an article entitled, “Coefficient Alpha and the Internal Structure of Tests.” The article was a great success and was cited frequently [no less than 5,590 times].1 Even in recent years, there have been approximately 325 social science citations per year.2 The numerous citations to my article by no means indicate that the person who cited it had read it, and does not even demonstrate that he or she had looked at it. I envision the typical activity leading to the typical citation as beginning with a student laying out his research plans for a professor or submitting a draft report, and it would be the professor’s routine practice to say, wherever a measuring instrument was used, that the student ought to check the reliability of the instrument. To the question, “How do I do that?” the professor would suggest using the alpha formula because the computations are well within the reach of almost all students undertaking research and because the calculation can be performed on data the student will routinely collect (Cronbach & Shavelson, 2004).

A tone of frustration emerges from Cronbach's reflections, echoing a modern academic fascination, almost an obsession with α, although its existence in a research paper tends not to be well understood. There appears to be an expectation in scholarly literature to report a sufficiently impressive α, although Cronbach urges us to first consider if it is appropriate. Cronbach states "I no longer regard the α formula as the most appropriate way to examine most data", and in reference to understanding the consensus formation of Chapter 4, I agree. Having necessarily mentioned consideration for a consistency measure such as α, attention will now rightly be brought to the study's context. As already mentioned, the intention to construct a meaningful description of Intraosseous (IO) and Laryngeal mask airway (LMA) insertion (highly qualitative) is inconsistent with the inherent assumption that precise and consistent scores will necessarily reflect the truth (a highly quantitative construct). Additionally, the context of pre-hospital paramedical practice is one of high variability and uncertainty, therefore the nature of skill application will arguably be more adaptable (and therefore less consistent between practitioners and situations) than those applied in a controlled and consistent environment.

The decision for inclusion or exclusion of criteria in Chapter 4 was based on mean scores, so the purpose of an additional statistical representation of agreement becomes questionable. Kane (2006, p. 27) suggests that we have a preoccupation with numerical certainty as numbers provide us with perceived objectivity and certainty. Surely, if the initial α was sufficiently close to 1.0 to be considered internally consistent, I likely would have been prompt to report it in Chapter 4 without further investigation of its meaning. In order to understand the ramifications of this preoccupation and the risks it may incur, we must seek to understand some level of what consistency coefficients reveal.

6.2.2 Skill type and context: Expected variability

When we consider IO and LMA insertion, variation in consistency is expected. Schuwirth and Ash (2013, p. 415) argue that where variance is assumed, a high internal consistency ought to promote suspicion of *low* reliability. LMA insertion has a much wider variability in practice than the IO device investigated in the previous study, and it is likely that the organisational context has an impact. The use of the LMA as a time-critical rescue device in the pre-hospital setting separates it from another common setting of anaesthetic medicine where its use is more routine and controlled. This is echoed in some participant's comments that:

> I consider LMA an emergency 'end of life' airway rescue device and think that any routine should keep this in mind – no time to stuff around! ... There are often many "right" ways of doing things
>
> (participant 13)

This data was obtained in free-text fields during the conduct of Chapter 4 and confirms the need for further review and validation of both checklists. It brings our attention back to the context of application, and the patient focus of the clinical skill.

When we consider the patients who may receive an LMA device, particularly in the local context where this study was conducted, they may be more likely to be an adult patient compared to those receiving an IO device. There are more options available in the event of a failure of LMA insertion, including an Oropharyngeal airway (OPA), nasopharyngeal airway (NPA), Endotracheal tube (ETT), or a surgical airway whereas an IO is more typically used if intravenous access is not possible or likely to be delayed (Calkins, Fitzgerald, Bentley, & Burris, 2000; James Cheung, Rosenberg, & Vaillancourt, 2014). This may increase the perceived *stakes* and pressure on the clinician who is performing IO insertion. These factors, in addition to the devices themselves, may create a more rigid practice for IO insertion and allow natural and acceptable variability in LMA insertion techniques, which will likely be reflected in any internal consistency rating such as α. Therefore, the value of using α to report on the checklists is further threatened.

6.2.3 Other risks of chasing internal consistency values

Some further dangers remain if we continue to focus on the importance of a reporting statistic. Firstly, consistency measures does not necessarily reflect *value*, meaning that if we are distracted by an impressive statistic, we can overlook what it really means. This may occur in the same way as chasing an acceptable p-value can sometimes distract from the real piece of interest, such as the effect size. As a further example, I will refer back to the assessment tool development in Chapter 4. If, of the 9 raters, 3 scored an item 7, 3 raters scored the item 8 and 3 raters scored the item 9 (mandatory), overall the value of that item would be greater than if all raters scored it 7. But if all scored it a 7, it would be more consistent, and agreement would be more demonstrable. So while I could delete the items which have greater variability from the list in order to retain only the items which have higher consistency, and in doing so we may boost the overall consistency of the list, but that would not make it a more valid list, because we may be retaining less valuable items, and removing more valuable ones. This may be why Cronbach and Meehl (1955) warn that "High internal consistency may *lower* validity" (emphasis in original).

High internal consistency may also be a reflection of excessive homogeneity. This was guarded against in Chapter 3 by recruiting a heterogeneous panel (albeit having met strict common inclusion criteria). The group were diverse in age, qualification, practice setting, and role within SA Health. A minimum level of experience benchmark was set in order to be considered "expert", however a diverse range of experience was still achieved (see section 4.4.1). Where some variation is expected (for example in a heterogeneous panel where there is no single correct answer), observing little variation could create concern.

Kane's definitions of validity and validation does not allow us to report a standard procedure (such as the Delphi approach) and a widely accepted measure of consistency (such as α) as a basis for an educational measurement tool's validity. These may be helpful starting points, but the pertinent questions remain: Are the proposed uses and interpretations of scores resultant from these assessment tools valid? Are the assessment tools, developed in Chapter 4, fit for purpose? If we want to measure a student's skill performance, do these tools do that? Caution must be exercised by conclusions on how well the participants have learnt something or how well they can apply it, as the tool becomes a proxy then, for this un-seeable phenomenon (we can only see the output of it). Such conclusions relate to the target domain, which requires additional data (see Table 39).

The aim of this present chapter is to critically analyse and examine the arguments for and against the validity of the IO and LMA skill assessment checklists. I hope to contribute to research on what validation is and how we might examine it, by critically reviewing the findings and implications of the checklists developed in Chapter 4. I will challenge the suitability of a strict procedure (for example the Delphi model) for the development of a "validated checklist", and also challenge the appropriateness for a consistency statistic in presenting a validity argument. This discussion will reflect on the second study, and using data from Chapter 5 and principles in the wider literature, argue for the valid use and interpretation of the checklists developed. Thus, the original checklists developed in Chapter 4, and the original findings in Chapter 5, will be both challenged and underpinned by critically appraising the valid employment of the assessment tools. This is a significant contribution as it will make explicit the uses and valid interpretations of these two checklists, for further adaptation to and use in wider resuscitation and clinical skill education and assessment.

6.2.3.1 Delphi scores from Chapter 4

The use of numbers in this study was only intended as a reflection of a *quality*, not *quantity*. The quality, or extent, of an item's importance was rated on an ordinal Likert scale, which cannot carry the assumption of absolute consistency between raters, or even between two ends of the same scale. The numerical values are a reflection of the clinical importance of each item, however the number alone has no real meaning. For example, an importance of 6 means nothing apart from the value statements which describe each extreme and section of the scale. Only when we know the scale range (1 to 9), extreme values (1 reflects an item which is "not important at all" and 9 is "mandatory"), and sub-scale values (sections 1-3 representing "not important", 4-6 "somewhat important", and 7-9 "very important"), can we begin to understand the meaning of a number.

For this reason, the statistical purist in me now considers it a breach to some extent, to calculate, report and include or exclude items based on a mean score which is appropriate for continuous, normally distributed quantitative data. A pragmatic decision was required, however, to progress the second stage of study in a timely fashion. This use of ordinal data was condoned by the practice published in the guiding studies. In hindsight using criteria such as the RAND UCLA criteria for agreement, rather than the mean score, would have been more consistent with this philosophy and data type. This criteria requires that for a 9-point Likert scale, agreement is obtained if at least 7 of the 9 participants enter a score within the region (1-3, 4-6 or 7-9) containing the median (Fitch, Bernstein, Aguilar, Burnand, & LaCalle, 2001). It is noteworthy that this identifies agreement, but further attention then needs to be paid to the value of the region upon which the participants agree. Thus, the resultant checklist is argued as a distillation of the generally agreed upon items, on the basis of a mean score, by a panel of expert clinicians.

6.2.4 A warning about consistency measures

Assessment (or educational measurement) tools developed using a Delphi approach often report consistency between the expert panellists for the items on the checklist with an "acceptable" consistency coefficient such as Cronbach's alpha (D. Berg et al., 2013; K. Berg et al., 2013; Berg et al., 2014; Nicholas Hartman et al., 2014; Huang et al., 2009; Lee Ann Riesenberg, Katherine Berg, Dale Berg, Joshua Davis, et al., 2013; Lee Ann Riesenberg, Katherine Berg, Dale Berg, Kathleen Mealey, et al., 2013; L. A. Riesenberg et al., 2013), Kappa (Herrmann, Kirchberger, Stucki, & Cieza, 2011a, 2011b; Kors, Sittig, & van Bemmel, 1990; Lemberg, Kirchberger, Stucki, & Cieza, 2010; Rapp & Queri, 2016), correlation R (Sermeus et al., 2009), or percentage agreement measure between participants (Chang, Gardner, Duffield, & Ramis, 2010; Dimitrow et al., 2014; Hagen et al., 2008; Jeon et al., 2015; Logue & Effken, 2013; Noblat, Oliveira, Santos-Jesus, Noblat, & Badaro, 2006; Parratt et al., 2016).

However, where a statistical representation of a phenomenon (such as α representing consistency or agreement) is presented, it must also be interpreted. In the study presented in Chapter 4, this is potentially dangerous. Firstly, agreement on how to insert an LMA is not expected. There are many safe and appropriate ways to insert it, with the exception of some more critical items, as identified by the morbidity and mortality items. This is reinforced by comments from some of the experts recruited urging for the lists to be simplified, and to remember the intention of the device in the pre-hospital setting, which is as a time-critical rescue airway device. The checklists may be a guide to new assessors, but are limited to LMA devices which require cuff inflation rather than the newer generation I-gel LMA which does not require cuff inflation (Kim, Oh, Min, Lee, & Lee, 2014).

Secondly, consistency analysis has a more legitimate place in application to an expected "true" score or value. The philosophy guiding the assessment tool development was qualitative and value-laden, with appreciation for the heterogeneous experience of the participants. These values were understood on a numerical scale, but the use of numbers does not make the study quantitative in a quest for truth discovery; rather it is still qualitative, with a desire to build meaning from the various participants. Analysing these meaning statements statistically beyond a mean or median calculation was not perceived as consistent with the intention of the study.

6.2.5 Context

The context of application for these checklists is for pre-hospital, or emergency resuscitation. Some application may be identified for the in-hospital emergency environment, but the primary focus in developing these education and assessment tools is the pre-hospital (such as ambulance and retrieval) clinical training. The recruitment of clinicians from SA Health may result in some local practices emerging, however local service (ambulance service or hospital) protocols were not employed during the study, therefore there is no apparent reason why the tools cannot be appropriately employed in other pre-hospital contexts.

These checklists have been developed to guide the detailed education and assessment of IO and LMA application. These will find much more reasonable and useable application during isolated skill stations such as an Observed Structured Clinical Examination (OSCE), rather than integration within scenario assessment. A case or scenario assessment will involve many aspects of assessment, decision making, communication, scene logistics and skill application, and to specify a checklist to the detail noted for these two skills will be overly cumbersome and threaten the valid application of the tools.

6.2.6 What do the Delphi scores claim to mean?

I have previously mentioned that the philosophy of the study in Chapter 4 was aimed at building knowledge rather than discovering it. The use of numbers in this study was intended as a reflection of a *quality*, not *quantity*. The quality, or extent, of an item's importance was rated on an ordinal Likert scale, which cannot carry the assumption of absolute consistency between raters, or even between two ends of the same scale. The numerical values are a reflection of the clinical importance of each item, however the number alone has no real meaning. For example, an importance of 6 means nothing apart from the value statements which describe each extreme and section of the scale. Only when we know the scale range (1 to 9), extreme values (1 reflects an item which is "not important at all" and 9 is "mandatory"), and sub-scale values (sections 1-3 representing "not important", 4-6 "somewhat important", and 7-9 "very important"), can we begin to understand the meaning of a number.

Thus, for a study such as that reported in Chapter 4, the checklists produced identify the items which are generally agreed by a heterogeneous, though specialist, panel to be important in the application of both skills. The decision for inclusion or exclusion of criteria was based on mean scores, so the purpose of an additional statistical representation of agreement, such as α or kappa, is questionable, just as the statistical analysis of qualitative data (even quantifies representations of qualitative data) is questionable. Kane (2006, p. 27) suggests that we have a preoccupation with numerical certainty as numbers provide us with perceived objectivity and certainty. Surely, if the initial α was sufficiently close to 1.0 to be considered internally consistent, it likely would have been reported in Chapter 4 and this entire discussion and investigation may have never been considered. In order to understand the ramifications of this preoccupation with statistical measures and the risks it may incur, we must seek to understand some level of what consistency coefficients reveal.

6.3 Assessment tool validation

6.3.1 How does the observed data lead to conclusions about a score?

In order to interpret the data gained in an assessment, it is usually converted to a score on a scale, be it quantitative (for example the number of times a student compresses the chest during CPR, per minute) or qualitative (for example the perceived standard of communication demonstrated in a clinical interaction). Observed data (in this case viewing a student's IO and LMA insertion performance) is used to determine a test score (according to the checklists developed in Chapter 4), from which an argument may be made on the universal score (which reflects the student's ability to insert an IO device or LMA), then subsequently the target domain (the student's skill performance), and finally the construct (the student's clinical performance). These four transitions from observation to construct will be examined using the argument thread outlined by Schuwirth and van der Vleuten (2012).

6.3.1.1 Inter-assessor consistency when applying the checklists

Having considered the IRCCs for the items on the assessment tool in Chapter 4, a different perspective is now adopted to understand consistency of data generated by applying the tools. IRCC is often use to argue the reliability of a Delphi study's findings by measuring consensus, for example using Kappa (Herrmann et al., 2011a, 2011b; Kors et al., 1990; Lemberg et al., 2010; Rapp & Queri,

2016) or Cronbach's α (N. Hartman, Wittler, Askew, & Manthey, 2016; Palter, Graafland, Schijven, & Grantcharov, 2012). Such uses of a consistency coefficient approaching 1.0 implies high agreement between different assessors, and it often used to argue that consensus has been achieved on either decisions (to include or exclude) or numerical scores (rating on a Likert scale) for a list of items. The consistency of the assessment tool will now be investigated to understand how reliably a score may be obtained when the tool is used by different examiners. Beginning with Chapter 5, the student performance exists as fact. It was witnessed by multiple angles, and was recorded to enable replaying in order to allow for accurate blinded assessor marking. The assumption here is that a single truth exists for each of the items on the list: either the student did or did not perform that item. The items were generally very specific, and as such the skill-specific checklists were initially referred to as "objective marking sheets" with the GRS scoresheets considered "subjective". Homogeneity of the assessors also supports the expectation of high consistency due to similar clinical perspectives, though this approach has been challenged during the course of this research series, and a series of possible explanations for variability between assessors using the "objective" marking guide is presented below, in section 6.3.1.2.

6.3.1.1.1 LMA performance inter-assessor consistency

The consistency in LMA insertion scores given by the various checklists is acceptable with Intra-class coefficient (ICC) measures consistently above .78 (ICC exceeding .7 is accepted as consistent), and significance values well below *p*=.05.

Table 40: Consistency Features of Various Measures from LMA Performance Data.

	Assessors	Score	Kappa (κ)	Cronbach's alpha (α) and ICC
Global scores	1 and 4	GRS_25 scores	κ =.057; *p*=.016	α= .863 Single measures ICC = .759 *p*<.001
		GRS_30 scores	κ =.066; *p*=.002	α= .977 Single measures ICC = .781 *p*<.001
Skill-specific checklist scores	2 and 3	Scores from checklist developed in Chapter 4	κ =.060; *p*=.38	α=.949 Single measures ICC=.903 *p*<.001
		Scores from checklist in Chapter 4 (excluding items not taught or those prompted by facilitators)	κ =.205; *p*<.001	α= .946 Single measures ICC = .897 *p*<.001
		Morbidity and mortality items	κ =.407; *p*<.001	α= .966 Single measures ICC = .935 *p*<.001

Note: GRS_30 scores are GRS_25 with the addition of an overall skill performance item, which was imported from Kneebone's IPPI .

6.3.1.1.2 IO performance inter-assessor consistency

The consistency in LMA insertion scores given by the various checklists is demonstrated with α consistently exceeding .9, ICC measures consistently above .89 (ICC exceeding .7 is accepted as consistent) and ICC significance values well below *p*=.05.

Table 41: Consistency Features of Various Measures from IO Performance Data.

	Assessors	Score	Kappa (κ)	Cronbach's alpha and ICC
Global scores	Assessor 3 Assessor 5	GRS_25 scores	κ =.403, *p*<001	α = .995 Single measures ICC = .990 *P*<.001
		GRS_30 scores	κ =.197; *P*<.001	α = .993 Single measures ICC = .986 *P*<.001
Skill-specific checklist scores	Assessor 1 Assessor 2	Scores from checklist developed in Chapter 4	κ =.187; *P*<.001	α = .947 Single measures ICC = .899 *P*<.001
		Scores from checklist in number 4 (excluding items not taught or those prompted by facilitators)	κ =.121; *p*<.001	α = .946 Single measures ICC = .897 *P*<.001
		Morbidity and mortality items	κ =.422; *p*<.001	α = .904 Single measures ICC = .904 *P*<.001

Note: GRS_30 scores are GRS_25 with the addition of an overall skill performance item, which was imported from Kneebone's IPPI.

It seems that in the efforts to create an "objective" assessment tool, and reduce the assessor to a "well-tuned analytical machine" (Gingerich et al., 2014, p. 1058) have been somewhat successful, given the agreement between different assessors on the skill-specific assessment tools' application. The specific skill checklist reveals a quantitative researcher's desire for a reliable piece of "laboratory hardware" such as a highly calibrated microscope or chemical testing solution. Roberts (2005) aligns this inclination with that of the chemist whose trade turned from emphasis on sensory input (the smell, feel or taste of chemical concoctions in the eighteenth century to an objective, "exact science" supported by apparatus which are now considered superior to human senses and perception (pp. 106-107). But this abandonment of perception in the striving for precision assumes an improvement in accuracy which promotes an evolution of scientific epistemology. This leaves the value of qualitative, sensory or otherwise perceived data somewhat in the past, and often undervalued. She comments "the deployment of the chemists' bodily senses was subordinately tied, almost to the point of invisibility, to laboratory apparatus that yielded evidence in the form of quantitative measurements" (Roberts, 2005, pp. 108-109). This eventually removed the scientist's bodily senses from the subject, with sensory evidence playing "less and less of a public role in the scientific determination of knowledge" (Roberts, 2005, p. 109). This is a significant deviation from Jussieu's time (1700's) where chemistry students were advised to employ sensory means to determine readings of higher reliability, such as time and temperature (Roberts, 2005, p. 111). The eventual paradigm shift within this science was multifaceted, including physical danger of chemical exposure, precision, and the mechanical development and development of instrumentation and techniques previously inaccessible. This raises the question: is the superiority of quantitative data, obtained by a skill-specific "objective" checklist with sufficient ICC, more precise and accurate than global judgements based on sensory and cultural contexts, in clinical education and assessment?

6.3.1.2 Possible explanations for variance

Even a specific checklist designed to illicit a series of binary decisions of whether or not an action was performed, may not result in complete inter-assessor agreement. The psychometrician is likely to accept a reduction in variability as clustering around a true score, whereas those who have a more contextual approach to assessment may question the very existence of "objective" scores (Govaerts et al., 2007). The recruitment of assessors assumed to be able to apply "objective" scores which are not expected to vary according to their experience, context or personal values is in some ways naive. A more homogenous cohort of assessors will likely agree on performance scores, however a heterogeneous group of assessors who have greater variability in their scores does not necessarily imply a faulty or imprecise checklist. Some explanations for variance may include differences in judgement, error, or perception.

6.3.1.2.1 Variance in judgement

The development of the assessment checklists was an attempt to obtain a set of objective, binary items for both skills, however the assessing clinician must still apply judgement in marking the item as achieved or not. This is feasible in some items which demand interpretation, for example the use of appropriate personal protective equipment. One clinician may demand the student to wear safety glasses and gloves. Another may only expect such protection in the presence of potentially infectious body fluids or contaminants. Another may only expect such protection to be worn if the potential for such exposure exists, even if not yet realised. These underlying judgements may be driven by professional practice culture, or ingrained through a standard that failing to enter the assessment room with gloves and glasses worn results in an instant failure of that practical assessment.

6.3.1.2.2 Variance due to error

Other sources of variation where there is an objective observation which requires no judgement or perception may simply be error, such as a coding error. If we assume that an actual level of performance occurred in each of the video performances marked in Chapter 5, then it is reasonable to assume that if marked by multiple assessors, the scores for each item will cluster around the "true" score. The tighter the clustering, the more agreement exists for that item between the assessors, and therefore the stronger the argument of accuracy and reliability in the assessment tool

We may use an inter-rater consistency coefficient (IRCC) to determine how consistent the scores for each item on the checklist are. A higher consistency coefficient reflects lower variance between different markers, and while some minor variation is expected through judgement, perception or other factors, part of the validity argument is borne from the reliability of the tool between different markers, demonstrated through IRCC approaching 1.0.

6.3.1.2.3 Variance in perception

Even for a binary truth, perception can divide a nation. A recent photograph of a duochromatic dress (Lafer-Sousa, Hermann, & Conway, 2015; Winkler, Spillmann, Werner, & Webster, 2015) recently created a media frenzy in Australia, with the nation divided over whether the dress was blue and white, or black and gold. With the objective nature of colour being the result of light waves of a specific frequency, if that frequency were to be measured, surely an undisputable measurement of colour would result. So why then did such division occur? It seems that even where some things are objectively measurable (such as the frequency of a light wave or cleaning the site of a medical procedure), perception still occurs on an individual level. Sacks ponders this, by asking, "to what extent are we - our experiences, our reactions - shaped, predetermined by our brains, and to what extent do we shape our own brains?" (Sacks, 2005, p. 25). Indeed, even with the intention of an objective, indisputable binary checklist, perception plays an important role. We must understand that the individual assessor brings with him or herself a unique biology, impacted by their previous experiences, which impacts the way they experience. Gingerich et al. (2014) asserts that:

> [S]ocial judgements and impressions made by raters are typically viewed as sources of... construct-irrelevant variability in performance ratings contributing to the 'noise' in the instrument. However...multiple 'signals' do exist within the 'noise' of interrater variability in performance-based assessment. It may be valuable to understand and exploit these multiple signals rather than try to eliminate them (p. 1519).

The variability seen between multiple assessors may not, therefore, be a reflection of error, bias or poor tool reliability. Within this variation, however, consistent divergence between assessors may bring attention to a reason for variation, and discarding this as "error" may cause assessment reviewers to miss important features of the data, assessor behaviour and student performance.

If we assume that an actual level of performance occurred in each of the video performances marked in Chapter 5, then it is reasonable to assume that if marked by multiple assessors, the scores for each item will cluster around the "true" score. The tighter the clustering, the more agreement exists for that item between the assessors, and therefore the stronger the argument of accuracy and reliability in the assessment tool. We may use an inter-rater consistency coefficient (IRCC) to determine how consistent the scores for each item on the checklist are.

6.3.1.3 What do the test scores imply about the observed performance?

Once this checklist is applied to student performances, it allows a score to be applied to that performance. Where each item is subsequently equally weighted, a performance score out of 16 (for LMA insertion) and 22 (for IO insertion) will be attributed to that performance. Assumptions springing forward from this include that: two students with the same score have the same ability; students with the same difference in score from performance one to performance two demonstrate the same amount of learning; and a student with a higher score has greater ability. These inferences, however, can be challenged.

Two students who achieve the same score may *not* have performed the skill to the same level of proficiency. With most clinical activities, some actions, omissions or errors are more critical than others, and some are *perceived* as more critical than others, for example the items noted as likely to impact on morbidity and mortality. The items of the lists developed vary in ease and clinical importance, so an omission of one item may reflect superior clinical ability than omission of a different item. I have endeavoured to identify these critical items through the investigation of clinicians' perception of processes which may increase patient morbidity and mortality risk.

The same increase in score for two different students may *not* necessarily reflect the same amount of learning, especially if learning is understood as a conceptual cognitive, behavioural, social, and professional identity shift, rather than a quantitatively measurable event. However, within this limitation the observable presence or absence of an action may be quantitatively measured as a proxy for the learning which we may not be able to define.

In addressing Kane's key concerns of what the assessment tools claim to assess, and what they actually assess, I return to my initial intention to develop a tool to compare students' learning resulting from two different teaching strategies in Chapter 5. However, learning and the cognitive processes associated with it are not directly measurable currently, and certainly not with these assessment tools. These checklists guide assessment through identifying certain actions, the sum of which result in a number which may be interpreted as a level of ability or performance. But this type of assessment only measures the actions (motor output) and not the processes (understanding and reasoning), interpretation (sensory input) and behaviour or change thereof for the developing health professional student, which may result from the development of their professional identity and role (tis aspect will be discussed further in Chapter 7. The checklists therefore are unable to make great claims relating to learning, however their ability to identify change in student's action and recall and application of these aspects of the skill performance are supported by a change in score from before, immediately after and then six months following the teaching session.

6.3.2 How does a test score lead to conclusions about a universal score?

The score determined by the checklist exists in this series of studies to make a statement about the ability of the student in performing either IO or LMA insertion. Interrogation of this inference will consider the logic of the tool's outputs (whether students perform the skills better after training than before it), whether data generated using the tools agree with other assessment tools which are accepted as a reliable measure (convergent validity), and a discussion around standard setting based on the application of the tools.

6.3.2.1 Are the tool's outputs reasonable?

If the assessment checklists are able to detect differences in student performance ability, however, as discussed, these differences are not uniform being ordinal in nature, then applying the checklist to untrained students would yield significantly different results compared to students who have received training. In sections 5.5.6.2.2, 5.5.6.3.2, 5.5.6.4.1.2, and 5.5.6.4.2.2, it is clear that statistically different scores are obtained using the checklists for IO and LMA insertion based on whether the attempt was prior to training, immediately following training, or six-months following training. This very expected result supports the intention that the checklists can at the very least delineate between untrained practitioners, those who have just received training, and those who have had the opportunity for their skills to atrophy. These clusters of scores from attempts 1, 2 and 3 are visually represented in Figure 42 and Figure 43.

6.3.2.2 Convergent validity

The analysis will include the immediate intention for the tools (for use in Chapter 5), and also for wider application (education and in-service assessment). Additionally, this discussion will go on to address whether the outcomes from the assessment tools are reasonable and expected in the context of the data gathered. I expect the scores obtained from the performance of students prior to training to be statistically different (lower) compared with their performance following training. This strategy will investigate to what extent the scores obtained from the checklists correlate to those from GRS. Where evidence for convergent validity exists, it is reasonable to argue that on the basis of the global checklist's accepted validity, the skill specific checklists developed in Chapter 4 produce similarly valid scores. This will help clarify whether differences in checklist scores can reasonably be understood to relate to a difference in actual performance, based on the established validity and reliability of global assessment tools when used by experts. However, the opposite is not necessarily true. For example, where a global scale and skill specific checklist assess two different constructs (for example the former may inherently assess fluency, confidence, and professional application, where the latter may assess compliance to a specific procedure), a lack of coherence between the two may identify different perspectives rather than one valid and one invalid assessment.

To answer this question, I again considered the scores from a modified version of the IO and LMA checklists (omitting items which were not addressed in the teaching session, and items which were prompted by the facilitators in some of the assessment sessions). This produces a checklist mark for each student out of 11 for LMA insertion and out of 17 for IO insertion. When these grades were measured against the sum of validated GRS items (maximum mark of 25), a correlation was measured. For both skills, a high Pearson correlation of .803 (*p*<.001) and .844 (*p*<.001) was demonstrated for LMA and IO respectively. Figure 42 and Figure 43 demonstrate the relationship between the GRS and checklist scores, with further reference to the student's attempt number.

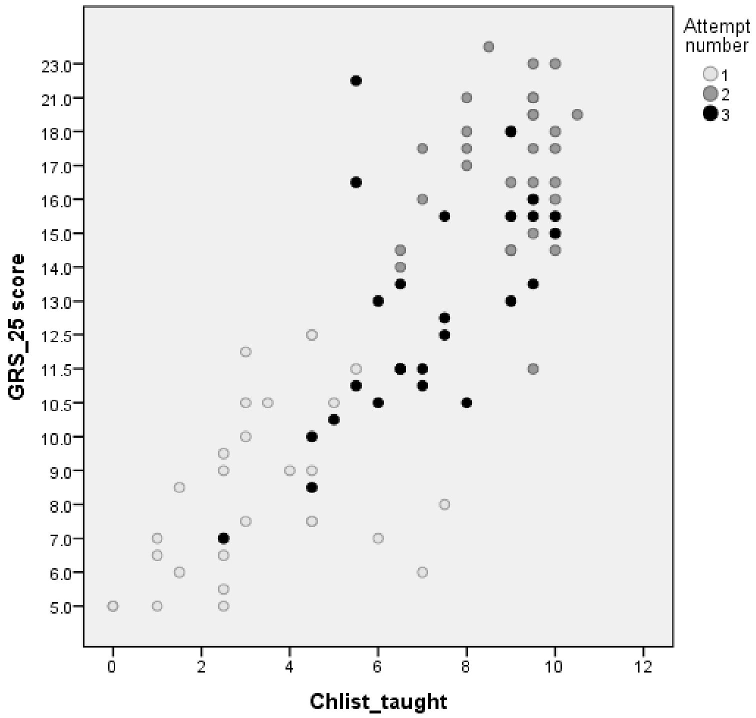


Figure 42: Correlation of GRS and checklist Scores (LMA), identified by attempt number.

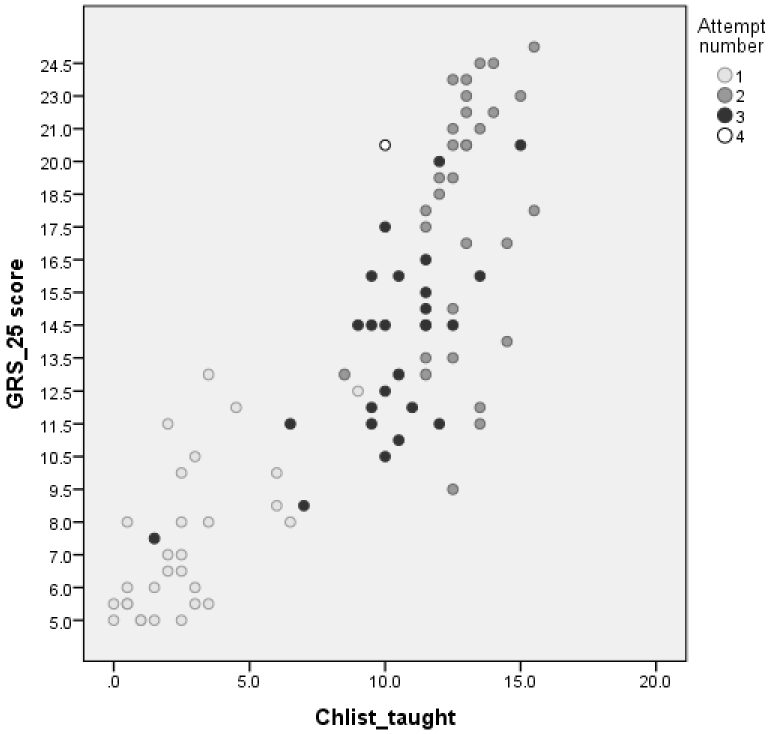


Figure 43: Correlation of GRS and checklist Scores (IO), identified by attempt number.
Note: The data point for Attempt 4 in Figure 43 relates to a student who completed the initial Attempt 3, and wished to re-perform it. For the purposes of all other analysis, the fourth attempt was excluded to remain consistent with other performances, but was included here for interest.

Similar correlations are obtained when the total checklist score is compared to the GRS_25 score, with Pearson's correlation of .801 (p<.001) and .857 (p<.001) for LMA and IO respectively.

6.3.2.3 Decisions on pass or fail standards

Finally, using the well-established and validated global scores, and conducting an analysis of contrasting group means (Norcini, 2003), the score which best delineates those who are competent from those who are not will be established. Where acceptable convergent validity exists, this may be translated to a checklist score, and further scrutinised on the basis of sensitivity and specificity.

A variety of options exist from which to identify a suitable "pass mark" for the checklists. Norcini (2003) suggests that the first step is in identifying the purpose of the standard. If using these checklists as a means to compare the effectiveness of different teaching methods, an absolute standard is appropriate as the focus is on the individual students' performance. A relative standard would be most appropriate for analysis of a set proportion of students, or where a fixed number of students is required (for example, to identify the top 10%). Norcini (2003) outlines four strategies which may be used to establish a minimum standard for the assessment, including the fixed percentage method, Angoff's method, contrasting groups method, and the Hofstee method.

A modified contrasting groups method was chosen for this purpose. In the method, a selected panel of assessors will review a selection of performances, and judge whether the performance is a pass or fail. Those who have passed are clustered, as are those who have failed, and all are graphed on a single set of axes. A decision is then made about how best to separate the two groups, and this is often the point of intersection, or of least overlap between the two groups. Thus, the assessors identify the contrasting groups (pass or fail), and a point (score) is sought which reflects minimal false positives, and minimal false negatives. This allows a score to be identified for consideration of future pass/fail decisions. A weakness of this method exists where assessment tools discriminate between the different groups poorly, thus the point of intersection entails many false positives and many false negatives.

This method will be modified slightly, by using a combination of GRS and morbidity/mortality factors to discriminate between performances which will be considered a pass and a fail. A pass score will be defined as: GRS of 3, 4 or 5 from *both assessors* for each of the GRS_25 component in addition to at least one assessor measuring satisfactory completion of each critical component of the checklist (items 3, 9 and 20 for IO, and items 3, 4, 9, 28, 29, 31, 32 and 34 for LMA). A fail will be defined as a GRS of 14 or less, which implies that at least one item on the GRS_25 scale has been deemed to be performed below acceptable standard by both assessors. Global scores were made up of 5 items all scored on a scale from 1 to 5, hence a score of 3 is *just* adequate, so a score of 15 will be *just* adequate. Anything less than that will be inadequate in at least one criteria.

Once the performance had been awarded a pass or fail resultant from the GRS and morbidity/mortality criteria, they were separated into groups, and data from the relative checklist scores analysed. These data are presented in Table 42.

Table 42: Mean checklist (total) scores for IO and LMA

	Mean pass score±*SD*	Mean fail score±*SD*
LMA	11.09±0.664 (n=11)	5.53±2.82 (n=49)
IO	15.33±1.44 (n=23)	7.224±5.12 (n=49)

Note: not all cases fell clearly into a "pass" or "fail" decision, and these were excluded from this analysis.

6.3.2.3.1 LMA pass mark

When the performances of the students who had been deemed to have passed were graphed alongside those who were deemed to have failed, the intersection of this graph sets a pass mark above 10, but below 10.5 (see Figure 44).

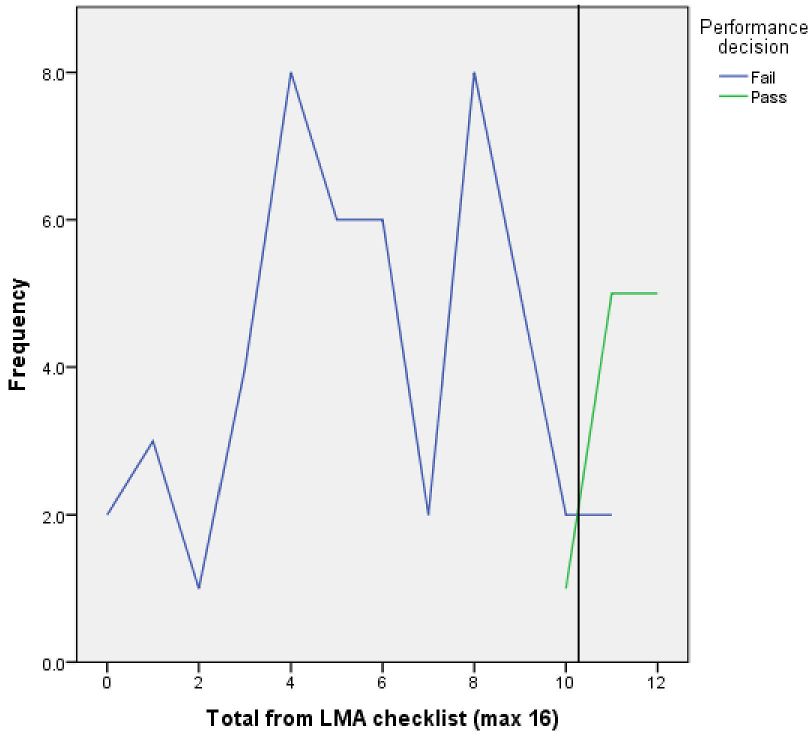


Figure 44: Distribution of LMA skill-specific checklist scores once performances had been separated into a pass or fail according to GRS and mortality/morbidity criteria, including crossover at approximately 10.2 of these contrasting groups (LMA).

Setting the pass mark at 10/16 will, for this data, result in:

- Two performances deemed to have failed by the criteria above to be granted a pass mark (false positive or FP of 2).
- 47 performances will be identified as "fail" on both scales (true negative, or TN of 47), and 11 performances will be identified as "pass" on both scales (true positive or TP of 11).
- No performances which have passed according to the global items and safety items have been awarded a fail grade with the pass mark set at 10 (no false negatives or FN).

Therefore, if we use the global expert assessor judgements (which are established in the literature as a valid measure of competence), in conjunction with the morbidity and mortality markers of safe practice as the criteria for whether a student ought to pass the skill or not, setting the minimum pass score of 10/16 carries a sensitivity of 100%, thereby identifying all the students who ought to pass using expert global rating scales, and specificity of 96%, thereby appropriately "failing" 96% of the students who ought to fail is assessed by an expert using a global scale.

Where sensitivity = TP/ (TP+FN), and specificity = TN/ (TN+FP):

- Sensitivity = 11/ (11+0) = 1.0 (100%)

- Specificity = 47/ (47+2) = .959 (95.9%)

Video 66 met all of the mortality and morbidity items satisfactorily, however the GRS was below 3/5 for two items. Video performance 74 was determined by one of the two assessors (not both) to unsatisfactorily meet one of the morbidity and mortality items, and in one GRS component received <3/5. Increasing the pass mark to 11 will increase the specificity to 100%, but decrease the sensitivity to 92.5% with the 4 new false negatives. (Sensitivity = 49/(49+4) = 92.5%).

6.3.2.3.2 IO pass mark

The frequency graph of scores obtained on the checklist developed in Chapter 4 contains two intersections between the performances judged to "pass" and "fail" according to a GRS of 3, 4 or 5 from *both assessors* for each of the GRS_25 components, in addition to at least one assessor measuring satisfactory completion of each critical component of the checklist. Intersections occur at a checklist score of approximately 12 and 14, according to the Figure 45.

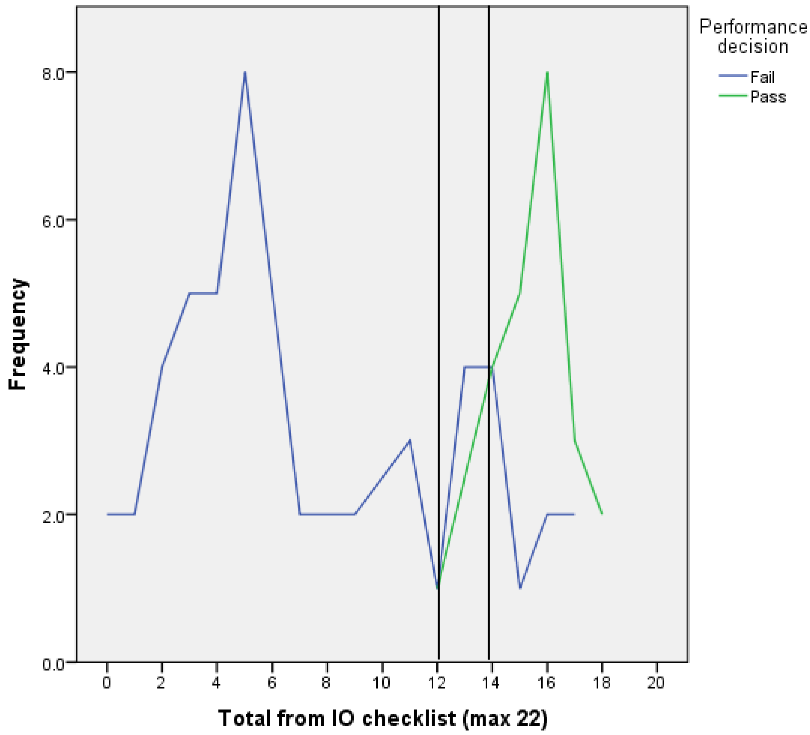


Figure 45: Distribution of IO skill-specific checklist scores once performances had been separated into a pass or fail according to GRS and mortality/morbidity criteria, including two crossover points. These occurred at approximately 12 and 14 of these contrasting groups (IO), thus the point of deliniation was taken as a score of 13.

If the point of 13 is accepted as the pass mark for this checklist:

- 11 performances deemed to have failed by the criteria above to be granted a pass mark (false positive or FP of 11).
- 48 performances are identified as "fail" on both scales (true negative, or TN of 48), and 22 performances are identified as "pass" on both scales (true positive or TP of 22).
- One performance which passed according to the global items and safety items has been awarded a fail grade with a checklist score below the set point of 13 (video 7 obtained a checklist score of 12.0, however this score was excluded from the analysis in Chapter 5 as it was a fourth attempt (false negative or FN=1).

Therefore, if we use the global expert assessor decision in conjunction with the morbidity and mortality items to determine whether a performance is satisfactory to "pass" or not, setting the minimum pass score of 13/22 carries a sensitivity of 96%, thereby identifying a very high number of the students who ought to pass using expert global rating scales, and a specificity of 81%, thereby appropriately "fail" most of the students who do not meet the expert criteria, based on the checklist scores alone.

Where sensitivity = TP/ (TP+FN), and specificity = TN/ (TN+FP):

- Sensitivity = 22/ (22+1) = .956 (95.6%)
- Specificity = 48/ (48+11) = .814 (81.4%)

Depending on the stakes of the assessment may alter what decisions are made on these data. For example, where practitioners will be practising independently and without supervision either from senior staff or peers, this specificity may be considered unacceptably low.

6.3.2.4 Other considerations for standard setting

These calculations include only the videos for which a clear fail/pass decision could be extrapolated from the global scores and safety items. Some video performances did not fall easily into either category, based on the criteria used to separate the two groups. An alternative calculation could include the use of the overall skill performance (OSP) mark, determining an OSP of 3 or greater as a pass mark, and less than 3 as a fail. While this would include all of the videos and possibly give a more complete picture of sensitivity and specificity, this method would exclude consideration of mortality and morbidity factors, and therefore offers a more limited perspective.

Regardless of the pass mark that is identified from the data available and the contrasting groups strategy used, Kane (2006) urges us to examine what the data and implications really mean, and what we understand them to mean. In order to establish a minimum standard, all the videos were pooled, and a strict criteria of what should be considered a pass or a fail were applied. The scores used for this purpose included those awarded to pre-instruction skill performance. This could be criticised as an unrealistic pool of data from which to make conclusions, and real-world data from students performing these tests in an OSCE or similar assessment would likely include fewer marks at the lower end of the scale. The risk of standard setting will be further discussed in Chapter 8.

During my own tertiary education, I typically needed a mark of 50% or greater in order to pass each of my assessments and subjects. The minimum mark of 10/16 and 13/22 are 62.5% and 59% for LMA and IO respectively. Thus, the standard identified in this chapter for a student to pass IO and LMA insertion are slightly greater than what I have experienced to be the arbitrary level of implied competence. This may result from the criteria used to determine what a "pass" performance entails, as it included a minimum GRS in addition to all critical criteria being addressed satisfactorily, which likely boosted a "pass" standard.

Overall, having demonstrated the consistency with expected difference between un-trained and trained students, inter-assessor consistency when using the skill-specific and global rating scales, and the consistency between the skill-specific outcomes and the GRS outcomes, various arguments for the valid and acceptable use of these skill assessment tools within the pre-hospital setting have been established. However, this is not proposed to replace the value of integrated case-management scenario assessments, or the value of programmatic assessments (Schuwirth & Van der Vleuten, 2011). Van der Vleuten and Schuwirth (2005) argue that the assessment tools "are not goals in themselves" (p. 310), but instead must be contextually appropriate in order to be considered valid. Conducting ongoing, carefully selected assessments within the intentionally planned range of evidence strategies is key in achieving reliable and valid educational assessments (Van der Vleuten et al., 2012). This allows assessors to obtain a whole picture of applied competence in the varying situations of clinical medicine remains an important component of authentic and reliable clinical assessment. These tools are offered to complement and not replace such a strategy.

6.3.3 How does a universal score lead to conclusions about a target domain?

The universal score provides a statement on how well a student has inserted the LMA or IO device. In order to make statements about the target domain (the student's skill performance standard generally), a wider picture of the domain is required. For example, performance of a series of other skills in varying contexts (authentic and simulated) will assist in making such statements.

Extrapolating conclusions about a student's skill application ability from a single recording of an IO or LMA insertion in a simulated, controlled setting is beyond the scope of these checklists. A solid argument cannot be made on this domain without data on other skills assessed in a programmatic fashion to give indications of skill application in the clinical context.

6.3.4 How does information on the target domain lead to conclusions about the construct?

The target domain may then become a basis for making judgements relating to a construct. Schuwirth and van der Vleuten (2012) argue the need for data to be obtained "from other sources and triangulated to support a more general conclusion". The student's ability to perform clinical skills tells part of the study of their clinical practice ability, but without further information on their clinical reasoning ability, patient communication skills, and other elements of clinical practice, judgements about this construct would be speculative.

6.4 Further discussion

This series of studies has addressed a range of strategies in order to validate the checklists developed in Chapter 4, by critiquing the assumptions behind the proposed uses and implications of results obtained, correlating scores with a previously validated assessment tool, measuring the assessment tools' reliability when used by two different assessors, and further analysing performance data to propose a possible standard for pass or fail decisions. This was important on the basis that while valid research and data collection methods were applied during the creation of these checklists, application validity must critique the context, assumptions, and results of applying the checklists. Development methods alone, and the face validity of recruiting an appropriately qualified panel are not sufficient for demonstrating valid application of the assessment tools.

Given that much of the validation argument depends on the intended uses of assessment tools, it is importance to be clear that the original intent in the Chapter 4 was to develop tools to apply in Chapter 5, and obtain data which measured the extent of learning during the trial described. This is an example of assessment of learning, rather than assessment for learning, which as a philosophy has impacted modern medical education to see assessment as intertwined within the process of learning (Schuwirth & Van der Vleuten, 2011). Of course, this purpose is challenged by the development of my definition of learning, as outlined in Section 1.2, which has evolved over the course of this PhD.

6.5 Conclusion

By reviewing the concept of validity, this chapter has critiqued the assessment tools developed in Chapter 4 beyond the standard limitations of internal and external threats to gathering and reporting valid and acceptable data. Additionally, a resistance to rest on an accepted and established procedure has been demonstrated. This was not in response to criticisms of the Delphi approach, but rather in response to pressing and previously unaddressed questions around the modern approach to validity: What does the tool claim to measure, and what evidence supports this? The Delphi method is not a complete strategy for developing validated clinical, education, or assessment checklists. Validation cannot be achieved by applying a protocol. It requires analysis and interpretation such as is found in this chapter.

A key contribution to knowledge is the explicit critique of inter-assessor consistency measures as a largely unquestioned strategy to argue the reliability or validity of an assessment tool. This has not been reported to such depth in the literature, to my knowledge, in reference to the Delphi technique where consistency ratings (especially α) are common. A defence is mounted for qualitative studies such as the Delphi approach, and a challenge to the relevance of consistency statistics for data which is ordinal and representative of a value. Thus, α or ICC approaching 1.0 ought not to be the apex in this context of studies. An "insufficient" consistency statistic may instead reflect valid and appropriate variation depending on the skill itself. Such measures may be appropriate, however, when applied to the implementation of such tools in the skills assessment setting. Both the LMA and IO checklists demonstrate acceptable Intra-class correlation to argue for reliable application by independent assessors. Additionally, high correlation to a previously validated and accepted marking system (GRS) and the demonstrated discrimination between students prior to, immediately following, and 6 months following their education session further supports the reliability of these skill-specific skill assessment tools.

It is important to be mindful, however, that setting out on the task "to validate a checklist" or "use a validated assessment tool" is incomplete, and is therefore not practically possible. A tool is only ever validated *for a purpose*. This chapter grapples with validation arguments for (and sometimes against) the use of a skill-specific checklist for the measurement of a student's ability to perform IO and LMA insertion, however contextual adaptation for the tools' implementation must be considered to uphold their valid use. The use of global judgement scales have been well accepted for valid use by expert clinical assessors who may exercise well-rounded top-down processing based on the observed data. For less experienced assessors, however, these tools offer a means to perform bottom-up processing in order to determine a level of student performance proficiency.

This is an important contribution to resuscitation medicine. It will allow for the responsibility for assessment of such skills to be less dependent on the small number of assessors in this environment who are adept at the use of GRS styles: namely those who are both experienced clinicians and skilled at judgement-based assessment. This benefit is particularly relevant for locations and contexts which are under-resourced. This may include remote settings, developing communities and settings where these skills are new to a service provider. Contexts such as these may experience significant difficulty in accessing adequately skilled and available clinical educators and assessors. For these situations, checklists such as these two may provide a structured approach, which may be adaptable to the local practice guidelines and culture, in order to maximise staff education and therefore also patient care. This chapter is an important addition to the enquiry into the central research question comparing 2SA and 4SA, as it provides a basis from which to understand the credibility and validity of the data gained by use of the assessment tools, and therefore supports the rigor of the data used to report results.

7 STUDY 4: DO-ING OR WHO-ING? A QUALITATIVE INVESTIGATION

There is no single reason why any educational system succeeds or fails.

Pasi Sahlberg (2012, p. 6)

7.1 Introduction

In previous chapters, I have described two trials aimed at comparing the cost effectiveness of two common skill teaching methods: the four-stage teaching approach (4SA) and a traditional two-stage approach (2SA). In these studies, I noted that 4SA seemed to be applied less consistently and less correctly than 2SA for each teaching session in the Chapter 3, and 25% of the sessions (though with educator correction) during the second trial (see Chapter 5). The 4SA is used widely to teach clinical skills, notably in advanced life support (ALS) courses internationally. As an example, the Australian Commission on Safety and Quality in Health Care (ACSQHC) have set guidelines requiring health care facilities to identify a "Rapid Response Team" able to recognise and respond to the deteriorating patient, with access to ALS trained clinicians (Australian Commission on Safety and Quality in Healthcare, 2010). ALS training in Australia follows the model set in the United Kingdom, with new instructors provided with “A Pocket Guide to Teaching for Medical Instructors”. This guide (Resuscitation Council UK, 2008) outlines 4SA, adapted slightly for a group setting, the final stage simplified to a skill performance rather than verbalisation before each step as initially described by Walker and Peyton (1998), and is encouraged by European and Australian based ALS instructor development material (Australian Resuscitation Council, 2015, p. 6; Resuscitation Council UK, 2011). Clearly, the 4SA has reached far and wide within the Australian and international medical education community, through the consistent approach to ALS instruction, and its influence on educators within local health care providers.

7.1.1 Hypotheses and aim of this study

I developed a series of hypotheses based on these findings and educator interaction. These include:

- The 4SA appears to be straight-forward to learn, so educators expect to find it easier to apply than it actually occurs in practice, and
- Although 4SA may alleviate some of the extraneous and intrinsic cognitive load (Sweller et al., 1998) on students, it may do so by transferring that load onto the educator, and this may impact on the educator's ability to perform the teaching process.

The aim of this study is to investigate the educator perspective of 4SA, specifically: How do educators experience 4SA during their own training and implementation? And, why might adherence to 4SA be lower than 2SA? In terms of health care education, these questions are important for educator development programmes such as ALS, EMST and PALS instructor development courses, and other health professions education courses which advocate the use of 4SA based on its theoretical value. In reference to the central research question, such questions address non-quantifiable measures of monetary cost, or effectiveness in performance scores, but allow the investigation of teaching philosophy, meaning, experience and confidence as markers of cost and benefit.

This chapter outlines a multifaceted qualitative study aimed at exploring these important questions. Through it I hope to stimulate thought on the educator response to innovations in teaching, as a necessary consideration for curriculum developers and education organisations. Thematic analysis of the data will result in five key ideas: subscription to 4SA; consistency and control of information dissemination; cognitive load; complexity and adaptability; and finally identity. Some of these may illuminate the phenomena seen in Chapters 3 and 5, and some will take this chapter well beyond the initial hypothesis to a place where differences emerge between the clinician who teaches, the clinical educator, and the educator craftsman. Where 4SA offers benefit to one, it may be an obstacle to the other.

These findings are significant to the development of clinical education as a clinical specialty. Historically, clinicians who have experience or expertise are expected to teach the next generation of health professionals, however this chapter forces the clinical community to consider the effect of developing true expertise in educators similar to any other specialist skill set. Supporting the autonomy and expertise of educators has been shown in other settings (Sahlberg, 2011b) to improve education in multifaceted ways, and health professional education has much to gain from these lessons.

7.2 Methodology and methods

7.2.1 Method of inquiry

The influence of my inherently quantitative inclination is evident in this qualitative study. As an example, the statement of initial hypotheses is made in order to be explicit, and transparent with the intention that stated hypotheses may not bias the thematic analysis like an unstated assumed hypothesis can, and the data may be allowed to speak for itself without being forced to conform artificially to a pre-determined expectation. In quantitative research this is considered bias, however in qualitative research this is often accepted as part of a researcher's perspective.

The shift to qualitative inquiry is driven by a different question. Rather than asking which teaching method should be used based on measurable performance and cost data, this study will consider priorities, values and experiences, in order to understand the educator's perspective of 4SA.

7.2.2 Constructionism as epistemology

The core epistemological approach employed is constructionism. The meaning of the educators' experience with 4SA is therefore constructed rather than discovered as an external truth. Such construction relies on multiple levels of interpretation. This epistemological perspective recognises that meaning is derived for the individual on the basis of his or her experiences and inner reflection of the teaching approach. Within this constructionist approach (Crotty, 1998, p. 8), I sought educators' interpretations of their experiences to understand 4SA from their perspective. To seek a holistic understanding of the topic, I used strategies consistent with both a constructivist and social constructionist perspective.

7.2.2.1 Social constructionist and Constructivist perspectives

The key aim is "constructing *knowledge about* reality, not constructing reality itself" (Shadish, 1995, p. 68, emphasis in original text) and "any notion of truth, then, becomes a matter of consensus...". The social constructionist perspective sought to understand the "shared, social constructions of meaning and knowledge" (Schwandt, 1994 quoted in Crotty, 1998). Through a series of group data collection activities (focus group and debrief session), participants built meaning of the topic collaboratively. This is dependent on ontological relativity where two people can hold very different worldviews (background, history, culture, assumptions or language) in the same world.

Conversely, constructivism focuses on the "meaning-making activity of the individual mind" (Schwandt, 1994, p. 127 quoted in Crotty, 1998), and in order to explore these, a series of surveys and individual interviews sought to understand the individual cognitive processes at work as educators employ 4SA. These two perspectives were woven together to address the study question to provide a sturdy foundation from which to gather a wider range of dimensions. This is in slight tension with Patton's definition (Patton, 2002b), that the constructivist perspective is concerned with multiple realities, and the implications these have on social interactions (p. 96), which Schwandt and Creswell would consider social constructionism.

While the initially developed hypotheses concerned the two educators used in the initial comparative trials, the focus in this study is educators' experiences of 4SA generally, and what might be perceived barriers or benefits of the method, in order to better understand how 4SA compares to 2SA. As such, a phenomenological approach (Crotty, 1998) influenced the data collection and analysis, as I sought to understand the educators' experience.

7.2.3 The sample

As I was interested in understanding educators' perspective of 4SA, I used purposive sampling of clinical educators who had learnt this teaching method. The two groups immediately accessible for the study included a cohort of post-graduate clinical education students, and ALS instructors. The two groups served slightly different purposes. The ALS instructors had all received tuition in 4SA to use in future ALS courses, and had the option to use it during their other clinical teaching, and the post-graduate students were taught 4SA as an option to consider in their various education roles.

Using this purposive sampling, I was able to gather data from educators who had received training in the 4SA, had an opportunity to practise it, received feedback on their teaching, watched others perform the 4SA and had experience in health professions education. This group also included educators who facilitated the teaching sessions, and nursing/medical students who were available to be taught the medical skills. I acknowledge that this may restrict generalisability through relatively homogenous sampling, however the purpose of this study was to obtain rich, contextual qualitative data rather than generate a probability value (*p-value*), so a representative sample of all clinical educators was not appropriate (Guest, Namey, & Mitchell, 2012).

Ethics approval was granted by the Social and Behavioural Research Ethics Committee (SBREC) at Flinders University prior to recruitment for this study.

7.2.4 Methods

Over a 14 month period, 117 surveys, 8 semi-structured interviews, a focus group and an educator debrief were conducted, with multiple stages of textual analysis to identify emerging themes which guided later stages of data collection as it unfolded.

7.2.4.1 Survey

I developed a survey to: collect background information on participants' experience with 4SA (questions 1-3); inquire about educator experiences learning 4SA, expectations about using 4SA for the first time prior to practice, during practice, following practice (prior to feedback), and following feedback (questions 4-7); and understand the educator's perception of 4SA (see appendix 10.5.1).

Free text comments were invited through optional comments at each question. Survey participants were invited to provide their contact details if they consented to a phone interview to discuss further.

7.2.4.2 Interviews

Semi-structured interviews were conducted to further investigate the ideas raised in the survey and other data. The first three were more highly structured than those following. As themes emerged from these interviews adapted to further investigate these themes. An overview of the development to less structured interviews (albeit not unstructured interviews) is outlined in appendix 10.5.3. The surveys and interviews both addressed a constructivist approach by understanding the educator's *individual* meaning of the common experience.

7.2.4.3 Focus group

A focus group (FG) was heterogeneous in clinical background, but homogenous in that all participants had recently been taught and encouraged to use 4SA in a small group setting. Some chose not to use 4SA, but each small group contained 4-5 educators, so each group had at least one who chose to use the method. The focus group was guided by Associate Professor Anna Vnuk who as a supervisor of the PhD, was briefed in the main themes which had emerged to that point[26]. The benefit of the focus group was the social construction of shared knowledge, allowed by the group dynamic which is distinct from individual data collection (such as social constructionism compared to the constructivist focus during individual interviews). Of the 22 participants there were 15 females (referred to as *female participant 1* to *15*), and 7 males (referred to as *male participant 1* to *7*).

7.2.4.4 Educator Debrief

Finally, an educator debrief was conducted following the focus group mentioned above. This is considered a separate data collection strategy as it was a compulsory component of an educator development intensive, unlike the focus group which was optional. I conducted the debrief as a part of a tertiary topic I was teaching at the time, with the consent of all participants. It had a different intention to the focus group, which was more investigative in nature. The debrief had a similar social constructionist value, with the inclusion of other facilitators and students (who had been taught by the cohort of educators). The main idea of the debrief was to discuss the small group teaching sessions with minimal agenda other than reflection. This activity is scheduled into each intensive for

[26] The submission for ethics approval at Flinders University included the clause that I would not be able to access the transcripts for this FG until after the semester grades had been finalised. This ensured participants could be more open about perceived difficulties or benefits of 4SA without fear that their opinion could impact their grade.

the topic, but this particular one was recorded for this study. Some thoughts may have been altered, developed or challenged during the focus group, however the relatively unstructured and unplanned structure to this session may have allowed other ideas to emerge.

7.2.5 Data collection and analysis

The aim of this study is to explore the educator's perspective of 4SA. The data is drawn from a multifaceted qualitative study, involving a series of (117) surveys, eight semi-structured interviews, a focus group and an educator debrief. The surveys and interviews were conducted with ALS instructors who had just received training in 4SA, and the focus group and skill-teaching debrief were conducted with medical educators completing graduate studies in clinical education.

I collected a series of surveys, conducted interviews, arranged the FG, and facilitated a debrief session to ask educators about their experiences were using 4SA. Thematic analysis occurred during data collection, at six main points, as indicated in Table 43.

Table 43: The Stages of Thematic Analysis.

	Surveys	Interviews	Skill teaching debrief	Focus group
Stage 1	1-30	1-3	---	---
Stage 1.2	---	---	22 participants	(conducted)
Stage 2	31-70	4-5	---	---
Stage 3	71-90	6	---	---
Stage 3.2	---	---	---	Data available
Stage 4	91-106; 115-125	7-8	---	---

Note: Surveys and interviews were conducted on ALS instructors in training. Surveys were numbered 1 to 106, and then 115 to 125. Upon collection of the last batch of surveys, I did not have convenient access to the earlier forms, so re-initiated numbering at 115, expecting that this may appear like surveys 107 to 114 were missing even though they never existed. Re-numbering the surveys following analysis was seen as increasing the risk of erroneous numbering, so this was avoided. The survey numbers therefore reach 125, however only 117 surveys were collected.
The debrief and focus group were conducted with students completing a post-graduate course in clinical education which addressed teaching clinical skills. These data gathering sessions were aimed at investigating these students' experience as clinical educators.
--- indicates that data from that source were not gathered at that particular stage

7.2.5.1 Surveys and interviews

The first stage of data collection occurred through optional surveys. These were developed to investigate instructors' perception of how easy or difficult the teaching method was to learn at three stages: as they learnt it, while they practised it, and during reflection after they had received feedback. These were distributed among ALS instructors during their instructor training course.

A series of survey questions was crafted to capture multiple perspectives, including data relating to the educator's introspection/self-reflection and also eliciting the educator's reflection of external feedback from the trainer. Answers to the first three questions give insight to the respondents' previous experience and exposure to 4SA. The next five questions aimed to understand educator implementation of 4SA from five perspectives: two questions look at anticipation, followed by actual implementation and reflection. Specifically, the survey investigated the perceived level of difficulty of 4SA during their training session (question 4), anticipated level of difficulty to use 4SA (prior to implementation, question 5), actual level of comfort using 4SA (during implementation, question 6), and then reflection on performance, subject to trainer feedback (questions 7 and 8). The final question gives some clues as to whether the 4SA is as easy/difficult as it first appears, once the educator has considered their own performance, then their performance in light of additional external feedback (see Appendix 10.5.1). The survey allowed participants to provide contact details if they agree to be contacted for a phone interview, and otherwise they were anonymous. In the results of this study, "interview" and "interviewee" refer to data obtained during these individual phone interviews.

7.2.5.2 Skill-teaching debrief and focus group

After I had received 30 surveys and conducted three semi-structured interviews some themes began to emerge. I then conducted an educator debrief for a group of (22) students enrolled in a core topic of the postgraduate clinical education course at Flinders University. This group of students had just completed a skill teaching session in small groups, with a medical/nursing student. They received feedback in their small groups from a facilitator and their peers, then participated in a debrief with the larger group to discuss different aspects of the teaching session.

On the same day, participants were invited to join a focus group to further discuss the 4SA. De-identified transcription was not available for inclusion in the third level of analysis in order to assure participants that no statements made during the focus group could impact students' progression through the topic. The interviews and group sessions were recorded for transcription with participant permission.

7.2.5.3 Analysis

Data were analysed using and open and axial coding in NVivo (NVivo 11, QSR International). Using the constant comparison (Kennedy & Lingard, 2006), emerging themes were identified as they took form and were used to inform subsequent data collection during the focus group and interviews. Each layer of thematic analysis reviewed, refined, and contributed to the arrangement of codes in a staged fashion. The stages of data collection are identified in Table 43. As analysis progressed, previously coded data was re-coded in light of new themes. This technique is often connected with a Grounded Theory (GT) approach (Kennedy & Lingard, 2006) to qualitative research. While this study is not Grounded Theory, the strategy was used to allow themes to develop and infuse the study as it developed, in a way which was congruent with the study's purpose. The data within each theme were then reviewed to ensure codes and themes were representative of the data they contained. This study, however, is closer to phenomenological research than GT. The qualitative approach is used to understand the phenomenon witnessed in the comparative trials conducted in Chapters 3 and 5 by engaging with those who have experienced the teaching strategy of interest. Without an explicit statement of the meaning imposed on the object of the research by the subject of the research (Crotty, 1998), this study and its analysis is influenced by a phenomenological flavour, but is not a purist phenomenological study. The data collection was informed initially by the hypotheses which informed the survey, and then developed as data were further collected and analysed.

7.2.6 Validity

Creswell (2013, p. 202) identifies a number of considerations relating to the validity of a qualitative research design. In order to produce reliable and credible themes, the data were reviewed independently by myself, and PhD supervisor Associate Professor Anna Vnuk.

Table 44: Factors Considered in Addressing Validity for Qualitative Aspects of the Study

Factor	Strategy used to address it
Triangulation	Data were collected in a variety of means, from a variety of people. In the results, attention has been given to representing data from the surveys, interviews, focus group and debrief session, in order to provide a representative picture of the range of input.
Member-checking	This occurred through the interactive nature of the interviews, focus group and debrief session, however was not possible during the written survey as this was not an interactive data collection method which would allow checking facts and interpretation of language. Due to the use of constant comparison during data collection, themes were checked and refined by subsequent participants in the research.
Using rich descriptions	Data are presented in section 7.3 with reference to the context of conversation, and explanation of implications inherent in terminology used by participants and researchers.
Bias	The potential bias brought to the study and its analysis is alluded to with the explicit identification of the hypotheses identified in section 7.1.1.
Present data which counters the themes identified	Perspectives of agreement and disagreement are presented where appropriate.

Spend prolonged time in the field	My background in clinical resuscitation, education of students, and education of educators allows me to understand terminology, nuances and assumptions inherent in the data obtained. Where this may lead me to a biased and pre-determined conclusion, the use of peer debriefing is used as a guard to allow the themes to emerge as intended by the participants.
Peer debriefing	This peer debriefing between Associate Professor Anna Vnuk and myself included discussion of the research aim, methodology, and emergent themes, we refined key themes and groupings, explored possible connections between themes, and ensured the data were well represented by the themes. As themes emerged, we considered which existing theories may explain, challenge or be impacted by the findings. This strategy was employed to ensure themes and findings were not limited by a single researcher's perspective.
External auditor	(not addressed)

Adapted from "Research design" by John Creswell, p. 202, copyright 2013 by Sage publications.

7.3 Results and Discussion

The study commenced exploring educators' perspectives and experiences of 4SA in order to understand why it may not be used when intended for staff by an education organisation. However, the concepts of teaching habit and individual style, and practical considerations emerged very early in the data collection and analysis process. By the third level of analysis, the deeper notions of cognitive load and teaching craft had emerged, and the final analysis showed a much more integrated sense of expertise and identity informing clinical education.

The survey gives a glimpse into how easy or difficult ALS instructors (in training) find 4SA to learn and use. While there is not a marked trend to support the initial notion that it seems easy to learn, or feels easy to implement but is actually much harder in practice to get right, we can see that there is also not compelling evidence to support the notion that educators' expectations and self-reflection is always accurate. The primary function of the survey was to provide direction for the interviews as themes continued to develop, to the findings are presented in appendix 10.5.2 rather than featuring in this chapter with great prominence.

The final theme map resultant from all data generated in the study is presented in Figure 46.

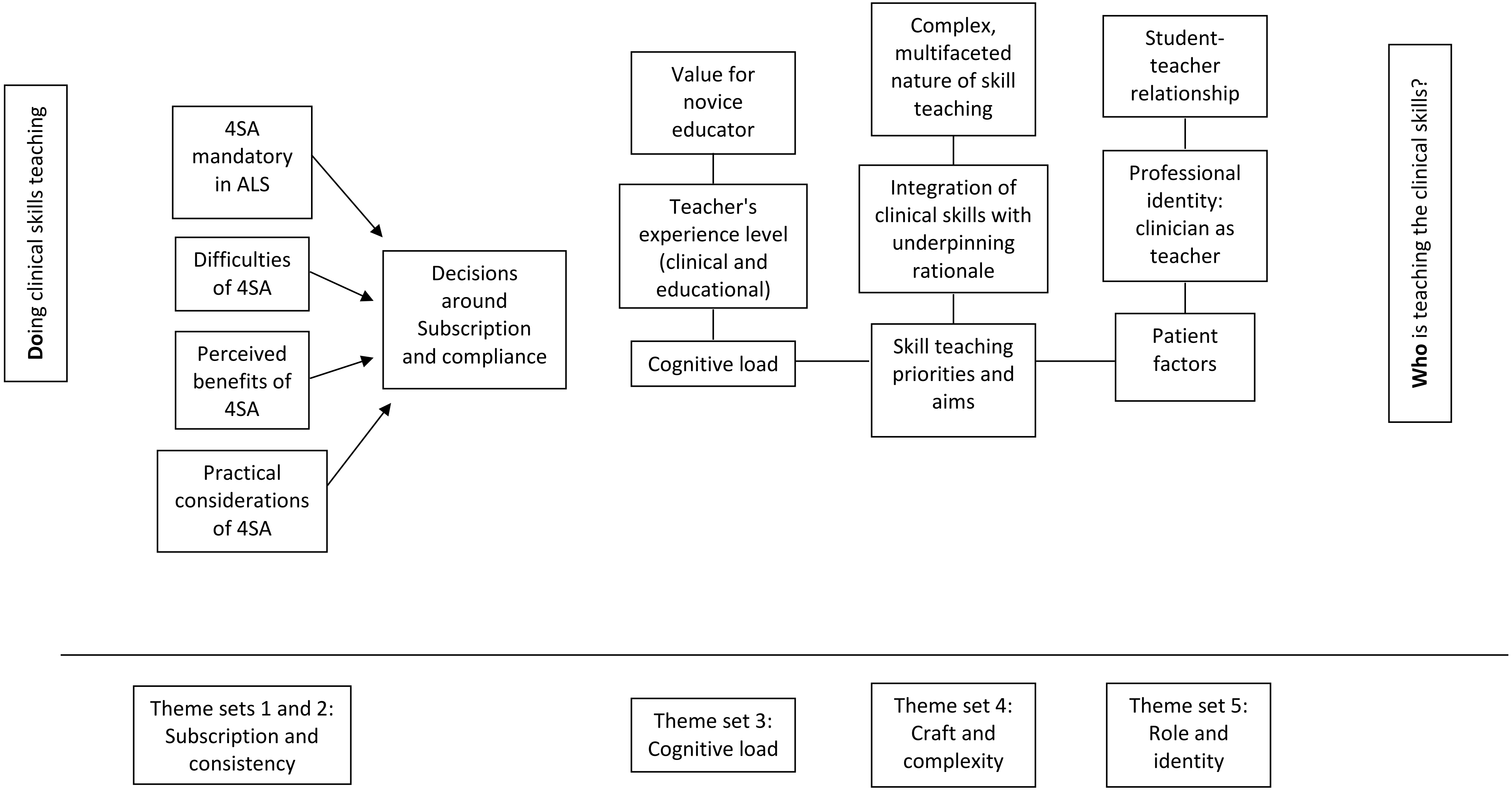


Figure 46: Final theme map

In this theme map, the "do" and the "who" are inextricably linked, albeit at different sides of the schematic. Where a *clinician* may teach in an ad-hoc way, a *clinician who educates* may find comfort, structure and support in 4SA, whereas the *craftsman educator* may find 4SA restrictive to all they aim to achieve in their teaching. If education organisations and educator development programs focus on developing the "doing" part of clinical education, we risk missing the deeper aspects of the person behind the process, both for the student and for the educator. After all, the *educator* is the education tool. They can be seen as an education tool which can be programmed to use particular strategies, or they can be invested in to develop expertise, resulting in skilled practice of their teaching.

7.3.1.1 Theme set 1: Subscription to 4SA

Educators expressed a variety of practical ways 4SA may be of benefit to the clinical education context in which they function, and also examples where 4SA may bring difficulty. Some practical considerations raised include setting, available time, skill complexity, and educator teaching experience.

7.3.1.1.1 Time and clinical situations

There was overwhelming agreement that 4SA is expected to take more time than a simplified 2SA. This was linked closely with the additional consideration of teaching context, with survey responses, Interviewees 2, 5 and the focus group indicating that time pressures in the clinical environment may put further pressure on teaching time.

> Quote index 19
>
> I think a lot of people would find it too cumbersome, too time consuming, I think it would interfere with teaching in clinical situations and they would not do it I think
>
> Interview 2

The time pressure 4SA is perceived to place on the clinical educator may prompt him or her to seek a less time consuming approach. The consistency between the various health professionals who participated in the study provides a strong argument as to the institutional and clinical burden clinical educators believe 4SA may bring. Interviewee 2 above connects this pressure with an interference on the clinical responsibility the educator has, and the potential dissonance which may develop. This tension between clinical and educational demands will be addressed further in Theme set 5.

7.3.1.1.2 Skill complexity

Skill complexity also contributed to the educators' expectation of 4SA's useability in clinical education, however arguments were raised on both sides. But while some saw that 4SA held real benefit for simple skills (Interview 1, 3, 5, FG), others felt 4SA it was too repetitive in this setting, and was better suited for more complicated techniques (Interview 2 and 6).

> Quote index 52
>
> For relatively simple skills things like the jaw-thrust it may be a little bit of overkill and I would maybe feel a bit gumby[27] "right now tell me how to do a jaw thrust… now tell me how to do a chin lift" I don't know if it's as applicable in those simple skills. It's certainly applicable in say putting in an art[erial] line or say a femoral lines if the candidate talked you through it, I think that'd be really valuable because that would get them to think through some more complicated procedure in terms of getting out all the equipment and everything so that would be the limitation in terms of the context that are too simple but otherwise it's good.
>
> Interview 6

The above quote goes beyond the statement that 4SA is unnecessary in simple skills, but that such use would have an impact on the educator. There is implied embarrassment in this participant's admission that to use 4SA would make him feel "gumby". This comment raises a further question: what is a simple skill? What is simple for a skilled practitioner may, after all, be difficult for a novice practitioner. So we are beginning to see that a clinician's practice level may impact how they might teach, through their perception of how simple of difficult a skill is. The skills mentioned above (jaw thrust and chin lift) may be simple to perform procedurally, however the type of patient who requires such interventions is potentially seriously unwell.

7.3.1.1.3 Novice educator

General consensus emerged that 4SA is expected to be most helpful for novice learners, which may support the earlier data that it is more aptly suited for skills laboratory settings than clinical ones. An unanticipated finding was the value 4SA held for novice educators. The value was perceived to hinge on the structure afforded by the method, which may be as true for other teaching templates.

> Quote index 50
>
> I'm relatively young and novice um if I, you know, I'm more likely to want to have a crutch that I can lean on in my session, you know if I get stuck halfway through I know what my strategy is going to be; I've got my four steps.
>
> Interview 6

Novice educators may benefit both from the use of a structure, as stated above, but being relatively inexperienced in teaching, they may not have ingrained teaching habits which may be in tension

[27] A colloquial term for "ridiculous, silly, uncoordinated"

with 4SA. The use of the word "crutch" above implies reliance on and support from the 4SA, and potentially also infers a perceived weakness for which 4SA might compensate. Additionally, a crutch is not something one intends to use forever, so this terminology may imply only temporary use of such a structure, until their expertise develops.

The term novice does not fully encompass the type of educator which participants in this study commonly referred to with the word. For example, the focus group identified a strong notion that experienced clinicians who excel in their clinical practice, are often expected to perform in a teaching position due to their clinical expertise (Cangelosi, Crocker, & Sorrell, 2009), are therefore not novice in clinical practice even though they are novice at teaching. These clinicians may have teaching *experience*, but this may have given rise to teaching habits which are not necessarily flavoured with effective andragogy. For example:

> Quote index 56-57
>
> It’s probably good for the expert clinicians because they’re - knowing how to do it really well, often they don’t know what the steps are so actually doing it with a student and doing that first part where you just demonstrate... I’d like to go back and teach my expert clinicians to teach this way because they often forget the... steps [to the skill] because of that passive knowledge that they’ve developed as an expert. So then when they – so then if they use this approach [4SA], it would actually help them, I think, unpack their - and to be able to think out loud. So just when you were talking about [the benefit 4SA holds for] early teachers, the student teachers, I think that should be the expert clinicians because it helps them to realise what steps they’ve taken.
>
> Focus group, female participant 6

The sentiment conveyed here is not infrequent in the data gathered. It relates to the ladder of competence, with acknowledgement that the expert clinicians may be unconsciously competent in the skill, and this can create difficulty in passing that knowledge on to others. It is not made clear, though, how Stages 1 and 3 of 4SA (the real-time demonstration without explanation, and the student prompting the educator how to perform the skill) actually help the educator make this tacit knowledge explicit in a way that Stage 2 (which is common to both teaching methods) does not. However, the underlying idea tends to be that a novice educator is not always a novice clinician, and an expert clinician is not always an expert educator. Indeed, a novice educator may be experienced in teaching, but without the development of expertise, and thus be considered by novice. This becomes difficult, as the experienced, yet novice, educator may not perceive their own novice.

The suggestion that 4SA may help a novice educator (an unconsciously competent clinician) become more skilled at teaching supported by another focus group participant's comment:

Quote index 61

You can see how… the four step helps what we were talking about earlier on today, the, unconsciously competent to their next – how it can help the teacher become that next state…..what's it called?... mindful, yeah.

Female participant 15

These comments tended to emerge in the focus group, which may reflect a challenge to the identity for the participants in this group. The interviews were conducted with ALS instructors who have very recently completed their training, however the focus group was conducted as part of a postgraduate tertiary topic which takes place over a semester and these participants have a more diverse background, usually already having held a clinical education role in their organisation for some time which is not limited to a single teaching context such as ALS instruction.

7.3.1.1.4 Superior skill retention

4SA was expected by many participants to be expected to increase the consolidation and skill recall through the structure, repetitive strategy and diverse teaching range within the one teaching method. This all provides a reasonable argument for improved skill uptake, and to learn *the precise* way to perform a skill.

Quote index 11

Well I want them to learn the absolutely correct way to deliver the procedure, um so by going through the 4 steps it ensures they have seen, verbalised and can carry out the procedure with all the technical details that is required, so I want them to learn the procedure in a structured fashion so that they are clinically competent... four step um makes absolutely sure that you can.

Interview 2

This comment also brings the patient into the equation. In theme set 4, the complexity of clinical education will be discussed, with some reference to the patient outcomes. In this quote, even though this participant saw 4SA as appropriate for the skills laboratory setting, they were teaching with the future, hypothetical patient (s) in mind.

7.3.1.1.5 Perceived mandatory subscription for specific courses

When asked whether there were aspects of the 4SA which were difficult to teach or easy to teach, one interviewee stated:

Quote index 13

I don't, I actually um [*IOR*[28]] failed my station when I did it to start with because... I'm teaching stuff all the time and I don't necessarily follow a four step thing... and I try to I guess

[28] IOR refers to a section which was "Inaudible on recording"

> turn it into a 2 step situation so they failed me and I redid it and I went home and I really thought about it a lot and I thought "oh I'm not really sure... why they failed me is correct" but when I went back and did it the second time and really thought about it overnight I think [*IOR*] you have different learners who learn differently, you different people with different skill sets but this forces everyone to learn using the 4 step thing, very clear, exactly what they wanted by enforcing the four steps so I do think it's a good way for the acquisition of the skill.
>
> Interview 2

This interviewee began by externalising the responsibility, stating "they failed me", rather than "I failed". It seemed that this interviewee perceived that the participant's skill teaching was effective, but if they did not perform it according to the template of 4SA, they were not considered competent. Shortly thereafter, the interviewee expressed subscription to 4SA, which may be in order to successfully pass the ALS instructor course. She later stated its value for consistency in teaching, remarking that it's what "they[29] wanted". This raises suspicion that she will believe in it not because she was convinced of it as an educational tool, but because she would not otherwise pass the skill teaching component of the ALS instructor course. This could reflect subscription to 4SA from external pressure or motivation, rather than intrinsic agreement. This is subtly echoed in the later comment from the same participant:

> Quote index 16
>
> And yeah look I've used it since then because I've done some ALS teaching since then and I have um I do think it's good. It doesn't matter if I've got a medical student or an intern or an anaesthetist
>
> Interview 2

This quote occurred following that described in section 7.3.1.1.5 and reflects agreement that 4SA is helpful (albeit "cumbersome"), with the implicit caveat that it has only been adopted for ALS teaching, where (as earlier noted), this educator may feel a necessity to adopt the strategy in order to meet the set expectation. This is echoed by other interviewees. The notion of external pressure leading to educator subscription is further indicated later in the interview:

> Quote index 21
>
> I think the people who do ALS are probably preaching to the converted almost, it will be really interesting to extend it to older clinicians like I am in the workplace trying to teach with it. That would be really interesting.
>
> Interview 2

These practical considerations give some argument as to why some educators subscribe to the use of 4SA, and others do not. However it is clear that this does not begin to adequately answer the

[29] The instructor instructors

question raised by Chapters 3 and 5: Why might the educators not have complied with 4SA as the intervention study protocol? On one hand, the participants recruited (for the survey, interviews, FG and debrief) and the direction of the discussion do not specifically address relating to those educators, who were not part of this study. The impact of subscription on whether an educator will use 4SA or not cannot be easily applied to the initial phenomenon which inspired this study. This phenomenon, of comparatively lower compliance to 4SA was noted during the teaching audits performed in Chapters 3 and 5, thus in this context the educators were not given a choice over what teaching strategy they were to use. The final three themes may begin to more helpfully understand this phenomenon.

7.3.1.2 Theme set 2: Control, consistency or rigidity?

The data in this theme indicates that the consistent approach of 4SA was sometimes expected to create consistent educational outcomes. Some educators suspected that it was the structure and organisation of the approach which would achieve this by preventing the educator from missing important information. Also, as mentioned in section 7.3.1.1.4 and 7.3.1.1.5, 4SA is perceived to address different learning styles and communicate the absolute right way to perform a skill. The implication is that by ensuring a common approach, a standard minimum output is ensured:

> Quote index 42
>
> I think by having this, I mean it is quite a rigid [IOR] that they're teaching, and again it's a bit like trying to make sure you're singing from the same song sheet is not it? That's what it all about. Trying to control the information and the way it's delivered they believe that's going to give them a consistent outcome regardless of the individual and I suppose if you're trying to teach a course that's wide and in fact probably worldwide, and you want to produce an apple every time then that's what you need to do.
>
> Interview 4

The language used in this quote suggests, though, that this consistency is not believed by this educator, but that it is taught this way by their instructors. 4SA is upheld by other participants, however, as a strategy which achieves a dependable, reliable result. Interestingly, interviewees 2 and 3 associate this aspect of consistency with the term "force". 4SA *forces* the educator to be more thorough; it *forces* everyone to learn with the 4SA. This may explain why the idea of *rigidity* was identified in reference to the 4SA. This was sometimes in a way which suggests 4SA should not be used in all situations:

> Quote index 130
>
> the way I reflect on some of my you know teachers that I've had a really good relationship or whoever really admired I think it's really probably effect them a little bit negatively maybe it can be a little bit I guess stifling in terms of if it's formally adhered to kind of, maybe not necessarily, but the particular educators and I'm thinking about probably a different generation

to myself some of them I don't imagine would really adhere to a formal teaching strategy that doesn't necessarily mean that it's not useful but I guess... I don't think that some of the teachings I would admire would probably be using it that well really.

Interview 7

This participant also commented in the security offered by the consistent approach in her comment "if everyone's teaching it in the same way you don't stand out as much". This suggests 4SA offers a level of protection for the novice educator, and that if the protocol fails (for instance, the students do not learn properly, or they do not like me as an educator), then it was the process, not the person using it.

While many participants include caveats on the use of 4SA (for example, for the admirable teacher mentioned by interviewee 7 in quote 130 above), some argued for the perceived universality of 4SA. For example, interviewee 2 mentioned its appropriateness in addressing all types of learners:

Quote index 14

you have different learners who learn differently, you different people with different skill sets but this forces us to everyone to learn using the four step thing, very clear, exactly what they wanted by enforcing the four steps so I do think it's a good way for the acquisition of the skill

Interview 2

This educator came to subscribe to 4SA. This comment echoes a *one size fits all* approach which is also communicated. And later:

Quote index 17

It doesn't matter if I've got a medical student or an intern or an anaesthetist, but whatever the skill is, you go back to core steps and you know, they can see you they can hear you talk through it, they can talk through it, yeah, each step. So a bit cumbersome but I think [*IOR*] way of doing it.

Interview 2

This comment speaks to the learner's level and the skill being taught *both* being sufficiently addressed in 4SA. A comment made by interviewee 6, that "[4SA] pretty much hits the nail on the head" (Quote index 51) now has fuller context, if 4SA is understood in a way which gives it credibility for all learners, in all situations, for a consistent output, regardless of the experience level of the educator.

Interestingly, interviewee 6 later considered themselves "relatively novice" at clinical education, and they feel that 4SA, despite being incomplete ("the thing it doesn't do is repetition, but obviously that can come later if it doesn't apply to a context that can come later as well..."), is fundamentally appropriate in all skill teaching situations. Just as availability bias may cause a clinician to overestimate the prevalence of a diagnosis due to previous exposure to such cases (Zwaan et al., 2016), educators who have been taught of the usefulness of 4SA and the importance of performing it correctly may overestimate its value in the wider clinical education setting, and be resistant to respond to indicators that a different approach would be helpful. This temptation warns that "when you have a hammer, everything looks like a nail".

While some focus has been placed on the consistent release of educational information and direction in achieving a consistent outcome regardless of student, skill and educator factors, the focus group later identified that this consistency is not seen in the clinical context, and may be a pitfall for the student:

> Quote index 88
>
> I think you can teach someone to do a skill and that's fine in that – on that particular day on that particular patient but, unfortunately, in medicine there are no hard and fast rules and everything's different, the situation is different each time, the patient's different each time and keeping the patient happy….. So that's where doing it for me and having ongoing reviews, that's probably a good thing. But once they [the students] do one [cricothyroidotomy] they think they are going to be able to do it, you know… I think, yeah, there's more to it than just, like, one, um, process that you do. I think it's got to be an ongoing review and assessment
>
> Focus group, female participant 8

This comment speaks to one aspect of the complexity of clinical skill teaching, namely the student expectation of being considered competent after a successful skill application. The educator is aware that the variability of patient presentations and other factors leads him or her to want to expose and assess the student's skill performance in a range of settings to be confident of their skill. This draws on the potential confusion created when we strive for unnatural consistency in teaching, as no two patients are exactly the same. The risk here is that a consistent approach to teaching, which attempts to communicate a consistent performance of clinical skills may hinder a developing student's adaptability of practice.

This thread is identified elsewhere in the focus group, in the context that teaching a manual skill using 4SA separates it from the underpinning knowledge, rationale, and information on when not to use the skill and other aspects of clinical expertise. In that regard, the educator who uses 4SA need not be a clinician:

> Quote index 86
>
> You can get a gardener in, and teach them how to put in a cricothyroidotomy, you know, someone without any previous understanding because it is very much - as I said before – a monkey see, monkey do. And as long as you've got that capacity to memorise the steps, you can pass the skill.

Focus group, female participant 15

However, even with a lack of rationale and clinical reasoning to know when (and when not) to use the skill, and how to adapt it, others saw that 4SA was still beneficial for teaching the manual components of the skill:

Quote index 92

We had a demonstration of the skill where it was identified that I'm just going to teach you the skill and not the background behind it …or - um, some of the background information. And that was a more effective way, I think, a student came out... of that from no knowledge to being competent in that skill but without the background knowledge.

Focus group, male participant 3

This participant saw that 4SA was able to be used to effectively teach a student the manual skill (the action) without the background or understanding. This links to the discussion of skill assessment, and what level of competence we assume when a student can correctly perform a manual skill. There is a risk that smooth skill performance may be misunderstood as overall skill competence where the supporting knowledge for safe use of that skill may be lacking. After all, with reference again to Miller's pyramid (G. E. Miller, 1990), an ability to *show how* does not imply an ability to *do* in practice. Additionally, following the prescribed steps correctly, does not necessarily reflect performance which is underpinned by expertise.

In this respect, we see the importance of the authentic clinical setting as it relates to skill teaching. When students are taught an isolated skill in a laboratory setting, with a method such as 4SA which may omit relevant and foundational information on the safe use of that skill, the clinical context, rationale, and troubleshooting becomes all the more important. The following focus group exchange discusses this tension further:

Quote index 99

Focus group, male participant 6: Yeah, you can take (participant 15's) gardener and teach them to teach that skill. They might not get all of the thing [rationale, etc] but that person, they're teaching definite skills and then they can get filled in later on directly.

Response from female participant 5: Well, our students have questions. They – they ask a lot of questions and, um, we're able to answer them because we did have that clinical background whereas if we just taught someone how to teach a skill, [the students] wouldn't be able to ask those questions.

A non-clinician *can* teach the manual components of a clinical skill using 4SA (or some other method). The students can learn where to place their hands on the instrument, and they can learn to do the action. This is sometimes understood as the pinnacle of Miller's pyramid (G. E. Miller, 1990), however this is actually *showing how*, according to Miller's description, not *doing*. The apex of the pyramid, "action", is concerned with "what a graduate *does* when functioning independently in a clinical practice"(p. S63), and is dependent on aspects of professional practice and expectation. However, aiming clinical skills education at *showing* how may contribute to more affordable clinical education, as it can be achieved using "the gardener" described in quotes 86 and 99 above, rather than paying specialist clinical educator wages, as noted by a focus group participant. However, the students are restricted in some of the questions they have answered by the educator. 4SA runs the risk, that when used as a strategy to teach application of a skill or insertion of a piece of medical equipment (a finite task), rather than facilitate transformation in the student as he or she learns (a complex task), a large portion of the teaching potential will be lost. This idea will be further expounded in section 7.3.1.4 .

7.3.1.3 Theme set 3: Cognitive Load

The next set of quotes all hinge together at a critical junction: the cognitive demand on the educator when teaching with 4SA. These factors still speak to compliance to 4SA, but this occurs indirectly through a cognitive overloading of the educator when using a strategy they are not familiar with.

The ALS instructors varied in their thoughts of how 4SA easy was to learn and implement, from "very straightforward [to learn]" (survey 98) and "It makes sense" (survey 25) to "I never think it will be easy!" (survey 44). Most responses hovered around the notion that 4SA is not natural but with practice it will become easier to implement, and this was expressed by participants in interviews 1, 4 and 8.

Interviewee 1 notes the difficulty experienced in remembering the individual steps as related to the unnatural process. They commented that:

> Quote index 4
>
> it takes a while to get used to it, I remember the first time I started doing it I kept forgetting some steps because it’s not something that comes naturally
>
> Interview 1

This difficulty was present when the instructor was learning 4SA, not just when they went to use it. When asked about how easy (or otherwise) 4SA was to learn, prior to an attempt at using it, she responded that the method was "a little bit awkward". This awkwardness seemed to be a function of the educator's unfamiliarity with 4SA, stating that "I've never heard of it before and it was a little bit odd, and a little bit unnecessary... It felt awkward the first two times and then we kind of got the hang of it" (Quote index 6). So there are hints of educator subscription (or a lack thereof) impacting the smoothness with which the method is used, and this may indicate that a person's perceived worth of 4SA may be in tension with the mental processes needed to perform it.

Once the difficulty of 4SA being awkward in the initial phases had been explored, the interviewee moved to acknowledge that the difficulty involved with this "complicated" teaching method was expected to be beneficial in teaching. This appears to be resultant from the repetition observed with 4SA:

> Quote index 8
>
> The benefit [with 4SA] would be again coming back to the same thing, it's a more complicated process [so] it would have good consolidation.
>
> Interview 1

This interviewee noted earlier that she anticipates her student would just want to do the skill, and this could cause frustration. But with the repetition granted by 4SA, and with the more complicated process for the educator, the outcome would be better consolidation for the student. This is the first hint we gain of a shift in cognitive load from the student to the teacher in order to maximise learning for the student. This is further supported by Interviewee 5:

> Quote index 48
>
> [the 4SA] probably makes you be, it makes you more conscious of what you're doing. It makes you put in more effort to planning exactly how you're going to structure the session rather than "oh yeah, I'm just going to teach them how to put in a gauedel[30], and I do that every day so I'm not worried about it" It probably makes you a bit more conscious of how you going to teach it.
>
> Interview 5

This quote conveys the educator's increased intrinsic cognitive load in order to decrease the extraneous cognitive load required by the learner (this is further discussed in Section 8.2). By presenting the information in a way which may be more accessible for the student, they have more capacity for learning. However when it comes to decreasing the cognitive load for the student (Sweller et al., 1998), this is more cognitively taxing for the educator who must then put additional effort into the way the information is communicated. Thus, there is a possible transfer of cognitive load from the student to the educator.

A counter argument for this cognitive load transfer is that some educators consider that it may therefore be patronising for the student. For example:

[30] This is a type of airway support device

Quote index 37

I think [when an educator uses 4SA] the student they sort of stand there looking at you a bit strange like "what are you doing" the first time you just talk about it and the second time you explain it and then you do it again, and I think they're looking at you thinking "do they think I'm stupid?"

Interview 4

This stated discomfort is echoed in a quote mentioned earlier, where Interviewee 6 mentioned feeling "gumby" when using 4SA for simple skills (see section 7.3.1.1.2). Another participant noted their inability to accept the additional cognitive load (in using 4SA) into their practice during the small group session:

Quote index 27

I think I completely missed out the silent step because there were so many things going on. It's just I didn't I couldn't I don't know, it's a time factor and naturally [IOR] I'm just talking as I'm doing stuff anyway.

Debrief, participant 4

Related to this idea that 4SA is not *natural*, many participants commented that it requires practice. Some educators felt that the increased demand the 4SA places on the educator is helpful to make the implicit explicit, whereas others felt that this was a source of the additional difficulty experienced in 4SA implementation.

Quote index 35

You know you are more focusing as much on that [the 4SA] as on the information, and sometimes you forget things because of that and then you go "oh I need to tell you that" or you know a bit disjointed at the end but again I think it's something that with more practice the actual four steps will become sort of second nature and then the content will you know become the primary focus but I think in the beginning you do tend to focus on the technique quite a lot and probably you forget things that if you were teaching in another method you would have covered.

Interview 4

This participant takes the idea of increased cognitive load one step further, and expresses that the cognition dedicated to recalling and applying the 4SA required a lot of focus, at the expense of the content she was asked to teach. And again, in a later interview:

Quote index 47

It's obviously easier to do what comes naturally which is you know talk through something and then show it and not really have a structure… to your teaching and then you know hope

> that the candidate[31] gets what they need so yeah I mean some structure is always going to be harder to implement than no structure but I didn't think that it was particularly taxing.
>
> Interview 5

This comment moderates the argument that 4SA increases the load on the teacher, indicating that it was not a significant increase for this participant. This educator felt that the difficulty came from imposing a structure onto the teaching session, built on the assumption that for them, formal structure was not habitually used in skills teaching, so to change practice to accommodate was difficult. What comes naturally for this educator was to have no structure, although the habitual skill teaching style still tended to have a format, albeit an easier one for them than 4SA. Later, this participant prompted me to consider the role of habit in educator development:

> Quote index 49
>
> I think for people who have been teaching airways skills for 20 years and they like the way they do it and it works for them then it would be hard to change to something that's harder.
>
> Interview 5

This statement prompts the question: when is it better to abandon habit in order to do something which is more difficult, and how do we know when the experience an educator has actually makes them more of a craftsman in education and therefore they may be more effective teaching with that art? This is discussed more in section 7.3.1.4.4. The distress one particular educator felt at trying to use a teaching style which was in conflict with their own style is evident in the following comment made during the focus group:

> Quote index 65
>
> I can't speak on anyone else's behalf but, for me, in my own workplace, prior to taking on this course, I had already developed my own teaching skills, so that's what I utilised to teach this skill. So I found it – I couldn't necessarily – I tried – I tried to remember what – what they [the four steps] were but it – and I - I had prepared but it just – it didn't come through when I actually went to teach the skill.
>
> Focus group, female participant 13

Whether 4SA feels like a natural teaching process, and how it aligns with the educator's habitual style, are both factors which compete with the educator's ability to focus on the content of the session. Survey respondent 41 summed many of these ideas up concisely, commenting that:

> Quote index 2

[31] Clinical student

> Initially it is moderately difficult to unlearn previous bad habits & implement a new one whilst keeping mindful of content.
>
> Survey 41

Existing practice, unfamiliarity with the teaching strategy and competing cognitive demand for the teaching content itself may all contribute to poorer compliance to 4SA. It is unclear at what point the educator becomes so familiar with 4SA that attention to the content is not sacrificed, and it likely varies from person to person. This also raises the question: if the teaching strategy requires deliberate practice, and doesn't appear to positively influence student performance, is it worth asking educators to use it?

7.3.1.4 Theme set 4: The complex nature of clinical skill education

The final two themes identified speak to the complex, variable and dynamic nature of teaching clinical skills. As the skills are of a clinical nature, the complexities of the context in which they will ultimately be applied impacts the educators' teaching considerations.

7.3.1.4.1 Variability in clinical practice

The following focus group participant describes an adaptive teaching technique used when the patient's response to management was unexpected:

> Quote index 60
>
> We had a situation in our department where we had a patient with a dislocated hip. So I quickly put on some videos from YouTube and we looked at three different methods of reducing the hip and then we went in and we used all three of them because we couldn't reduce it
>
> Focus group, female participant 5:

Improvisation was required not only in this clinical situation, but also in the education setting to which it was tethered. This demanded dynamic and adaptive education strategies for the students present, rather than persisting with a single "absolutely correct" approach (see quote 11 from Interviewee 4) which may be detrimental to this particular patient. Just as clinical care needs to be adaptive and flexible to the needs of the patient presented to the clinician, the educator needs to be able to respond to the complexity of the education situation with expertise and flexibility so that tomorrow's clinicians can operate in uncertainty. The focus group later returned to this idea, with a discussion on the importance of a student developing rationale and theoretical support for the manual skill while they learn the physical components.

> Quote index 88-91

> Focus group, female participant 5: I think clinical reasoning is involved as well. I have the skill now and when to use it and when not to. I think Peyton's, the main purpose is just to teach and I think from what I have seen today, it does it very effectively.
>
> Female participant 9: I think it teaches the skills but not when or where to use it, that's all, or how to modify the skill.

In this exchange, the participants are grappling with the limitations of 4SA ("Peyton's") in the clinical setting. The clinical reasoning is inextricably linked to skill application in the clinical context, but 4SA has no room for it, yet we see in the first comment that students may develop confidence in the skill when actually they lack the appropriate direction in that skill, which hinges on understanding of rationale.

Therefore, skills ought to be taught in a mini-spiral curriculum. This concept is widely accepted in medical education, but is yet to formally filter in to skills teaching, even though this is just what these educators are grappling with. Concurrently learning rationale, anatomy, skill application, troubleshooting and other principles like an aseptic technique and communication allows the otherwise separate component of skill performance to develop alongside the development of aspects which will support it. Then, the student will have cognitive scaffolding in place which is connected to their knowledge of the skill, and will allow them to consider ways to safely adapt a process to suit the individual patient.

7.3.1.4.2 How can we teach perception?

The following FG discussion thread builds on the idea of clinical variation, and considers not only how educators teach a student to perform a manual skill, but also the role of perception within the student's learning. This perception is often based on an understanding of clinical rationale or anatomical awareness, as the following excerpt identifies:

> Quote index 84
>
> Focus group, male participant 2: I wonder, too, if – not an extra step but you – you might need to sort of define one of the steps when you're showing students in that they – in some situations it's necessary for them to be involved. Ah, in the case of my procedure, I was teaching how to put an EEG electrode on and – but when you clean the site, I wanted to be able to show the student, um, this is about how hard you need to – to rub when you're cleaning the site. And I'm thinking back to when I was helping deliver a baby the other day, the GP was pulling out the placenta and she got me to put her hand – my hand on hers so that I could feel the tension in which she was pulling. Now I guess that's not officially in the four-step Peyton model but when you've got that fine dexterity at work, you can't always see how hard a doctor's pushing or – you know, I've never seen an epidural put in but I imagine there – there must be, um-----
>
> Facilitator: There's a give, there's a start-----
>
> Male participant 2: Yeah, there's a give.

Facilitator: Telling you to go and then-----

Male participant 2: But you hear about them.

Facilitator: that stops-----

Male participant 2: When you feel that give.

Facilitator: when you feel that give. But how can you teach that?

Male participant 2: How do you explain that, yeah?

This exchange between one participant and the facilitator builds to what seems to be an unanswerable question. How does an educator teach or explain something which is perceived, rather than performed? The educator using the student's hands to allow them to perceive something is offered as a means by which to communicate this sensation, as is verbal explanation. At this point, the group acknowledged that skill teaching has multiple levels beyond recall of the individual steps. As educators, they now understand that they do not just teach an action, they are trying to facilitate the student learning an approach, a set of principles to be applied, a perception, a responsiveness to feedback, and more!

So how do we teach how hard to rub to place an electrode on clean skin? Or the tension to place on the umbilical cord during placental delivery? Or understanding the feedback of different tissue densities during an epidural insertion to know what space the needle is in without visual affirmation? The complexity does not begin with how these aspects are taught, but what they are. They depend on muscular and perceptive feedback which tells the clinician what is happening when they can not see it. This depends on understanding well beyond what can be covered in 4SA, or indeed any stand-alone skill teaching session. It must be intertwined with resources and knowledge which builds a much more integrated cognitive map of that skill.

From this analysis, the question is clear: What is the aim of the skill teaching session? The answer is less clear: Is it to teach an action? Is it to teach an adaptable approach? Is it to inspire the development of a masterful clinician? Or is the aim to achieve all of these aims and more?

7.3.1.4.3 Patient risk

An additional complexity to teaching clinical skills is the clinical implication for the patient. Many educators in the skills laboratory setting have a patient in mind when teaching, therefore are mindful of the potential risks of errors becoming ingrained in their students' practice. In the authentic setting ("on the ward", "on the floor", "bedside", etc), the patient needs are much more paramount. In the context of interrupting clinical skills teaching in order to swap hand placement between the educator and student, the following participant mentioned:

Quote index 83

Focus group, Female participant 14: Yeah, so, like, um, supporting breathing, well, you're going to have to correct the student while they're doing it there and then not on a – you know, as opposed to chest auscultation, letting a student finish and then correcting.

Facilitator: Okay, that's right. *'Now I put the stethoscope here and now you hear how much better that is*?' So we're doing that at a level of risk to the patient that can have a big influence on whether Peyton's will work or not.

It is clear that different approaches to correction during teaching are required for different types of skills which impact the patient differently. Some skills carry much higher risk to the patient if performed incorrectly, and some level of intervention or correction is essential to mitigate this risk. Therefore, the complexity around patient management brings an additional layer of decision making and complexity to a teaching strategy, which further prompts the need for a responsive and adaptable approach when teaching in the clinical setting. The following quote takes this idea a step further:

Quote index 46

again the more complex tasks there's also a sense of more potential for harm so letting someone loose having watched you do it twice is you know more talking you through and then jumping in this often a lot more interference and hands on to set boundaries and parameters to keep a patient safe

Interview 5

This quote indicates a tension between the clinical educator's roles: that of a clinician, and that of an educator. As an educator, the agenda is to guide and facilitate learning, however as a clinician the responsibility is to provide health support to alleviate suffering and disease. This comment highlights an instance of possible tension between these two, and speaks to the educator's identity which will be further explored in section 7.3.1.5.

7.3.1.4.4 The art of teaching

The adaptability demanded of clinical educators due to the complexity of the clinical practice, patient variability, risk considerations, and the demands of teaching something more than just an action are heavily cognitive processes: there is much on the educator's mind at this point. An aspect of teaching which these do not address is the craft of engaging students in meaningful teaching. When this participant was asked what she thought about 4SA, her response focussed on the educator's ability to engage, not a prescribed teaching approach:

Quote index 37

I think [4SA is] a bit long, especially when you know time is short, because you really you're repeating the same thing 4 times, you know like you're doing the task 4 times, but then I do think that when you look at something like BLS[32] I think what makes a difference in BLS is that yes the individual educator makes a difference because you know some people are just better at getting people engaged, and I think when you're engaged you take more away with you.

Interview 4

Later, she commented that:

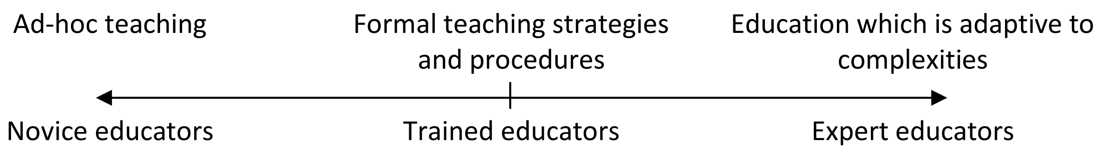
Quote index 39

I do think there are people who are more gifted than others. I think if you're not particularly gifted, then having a very structured… set[33] or you know dialogue can help you.

Interview 4

This participant identifies that whether or not you follow a teaching strategy, different people have different gifts. How one person is able to engage learners may be different to another's ability, otherwise a foolproof strategy on how to obtain engagement would exist. Some suggestions for these exist (such as "hooks"), but it is clear in quote 39 above that these strategies or structures are not perceived as required for "*more gifted"* educators. Further, these quotes suggest that a gift for student engagement appears to be more valuable than an educator who lacks the "gift" but is compliant to a teaching structure. Whether this participant means "gift" as an intrinsic quality, or an art form which has been developed through much practice and reflection over time, is unclear.

If this principle is used to develop a scale, we may have unstructured, uninformed, ad-hoc (even if well-intentioned) teaching on one end, and gifted, engaging and masterful teaching on the other. This gift may an art, developed from a philosophical conviction of what teaching is, or build on years of refining educational practice based on insight, feedback and grappling with the latest educational advances. Either way, it appears that 4SA (and possibly other similar strategies) sit somewhere in the middle. This is depicted on the scale below, with teaching style on the top, and educator stage below:



[32] Basic Life Support
[33] "set" and "dialogue" are terms used in ALS instructor training to refer to the structure of the session.

7.3.1.5 Theme set 5: Identity

The final theme of identity arose in two key ways. It includes the educator's intention to use skill teaching as a part of developing the student's professional identity, and also the educator's perception of their own role in skill teaching and how their clinical identity impacts this. Returning again to an idea which developed in the FG, the *gardener* is offered as a s a symbol of someone who is not a clinician, but may be trained to teach clinical skills using 4SA:

> Quote index 87
>
> You can get a gardener in, and teach them how to put in a cricothyroidotomy, you know, someone without any previous understanding because it is very much - as I said before – a monkey see, monkey do. And as long as you've got that capacity to memorise the steps, you can pass the skill. So how do we make that better? How do we make it, you know, um, medicine nursing specific.
>
> Focus group, female participant 15

This participant saw that one of the aspects to consider with 4SA is what we are intending for our skill session. If we are aiming for the student to be able to reproduce the skill, then they expect that 4SA may be adequate because reproduction does not depend on reasoning, understanding and adaptation. But there is an air of dissatisfaction with this aim and outcome: *how do we make it better*? This implies that teaching a student to perform a skill is not all of what educators hope for a student to do.

Teaching a clinical skill without reference to the reasoning behind it to guide safe and appropriate practice has already been identified as incomplete teaching. The idea of bringing the student beyond simple manual ability challenges 4SA with the foundation of Miller's pyramid, however the four stages of achievement begin with knowledge. The first stage of 4SA gives some overview of what the skill is, but the knowledge transfer is extremely limited, and underlying rationale is likely missed in this performance which occurs without any particular commentary. Hence, additional layers of this pyramid may be built on an inadequate foundation.

But this theme goes further to challenge Miller's pyramid itself with the inclusion of an additional level. Each step up Miller's pyramid reflects a more profound level of achievement, and the data in this study proposes an additional aspiration of becoming. With this additional layer building on what a student does, and reflecting one who *is*, the model then speaks to what many educators see their role as: facilitating the professional development of a clinical student. This is echoed by Cruess, Cruess, and Steinert (2016) with reference to the identify formation resultant from a professional calling which runs much deeper than behaviour or action, however this is also captured in Miller's original description (G. E. Miller, 1990). The development of a professional identity, I believe, can go further still, and will be discussed in Section 8.4.5.

Interviewee 2 states, "I want [the student] to learn the procedure in a structured fashion so that they are clinically competent." This raises two ideas. Firstly, the development of clinical competence may vary in perception between educators. A focus group participant comments that "the student needs to be able to become confident in a skill rather than just teaching it [the procedural steps] to them". This helps us further understand what clinical competence might look like, and it is wider than recall or action. It impacts the student's approach, and may be connected to their developing identity as a transition from student to clinician, as their manual ability, perception, understanding of rationale, and confidence in approach grow. I is also important to consider the educator's identity. The next excerpt speaks to the two roles a clinical educator has: one as a clinician with a patient focus, and the other as an educator with a learning facilitation focus. There is an implicit grappling with the potentially conflicting demands placed on a clinical educator in this situation.

> Quote index 94-95
>
> as a teacher, you need to have that sound clinical background so you couldn't just be an educator and have no clinical knowledge. You need to have those two things to be able to use a bit of judgment as to when it's appropriate
>
> Focus group, female participant

This participant noted that in order to be a successful educator, the person must have clinical expertise, however this is challenged by another group participant, who promptly argued:

> But if you were working in a limited resource environment and you just need someone who can do that skill, do you need to take a clinician off the floor[34] just to teach the skill?
>
> Focus group, male participant 6

These statements speak to the tension between clinical demands and teaching demands, and are steeped in factors such as perceived priorities, cost and risk. Ideally, the person teaching a clinical skill is both a clinician (with clinical credibility and experience) and educator (working under sound andragogical principles), but this is clearly mediated by resources. In a resource poor setting, a non-clinician who educates may be a more cost effective option if using a clinician results in diminished workforce available for patient care. This implies that a non-clinician educator provides less credible education than a clinician-educator, but also assumes that the education occurs in tension with the clinical need. The symbiotic model would argue that integrating clinical education into the health

[34] "off the floor" here refers to bringing a clinician away from their clinical responsibilities with patients

delivery system would retain clinicians in this environment, and meets the education and clinical needs present. This is reinforced in a further comment:

> Quote index 99
>
> Our students have questions. They – they ask a lot of questions and, um, we're able to answer them because we did have that clinical background whereas if we just taught someone how to teach a skill, they wouldn't be able to ask those questions.
>
> Focus group female participant

The clinical and educator roles held in clinical teaching may present competing priorities which need to achieve resolution in the context, but in other ways the educator's clinical role aids their teaching with relevant knowledge and credibility. Whether this credibility is achieved by years of previous clinical experience, or ongoing and current clinical practice was not made clear. This quote goes beyond the "doing" of clinical education, and to the responsiveness, problem solving, perception and adaptability required.

7.3.1.5.1 Proving yourself as an educator

Finally, the idea of an educator's identity is noted not just as something which guides their own behaviour, attitudes, priorities and action, but also something which is seen to be presented to others. In this way there is an intrinsic (perceived or otherwise) identity, and one which the clinical educator may seek to present to others. This idea was raised in the following ALS instructor interview, but was not represented in the focus group or debrief:

> Quote index 44
>
> There is often scrutiny of yourself as an educator, often trying to establish a particular standard and justify your position as an educator so often that there's as you know a little bit of stress and arousal regarding remembering all of the things that you've been taught and applying them appropriately to demonstrate to those instructing the instructors that you're at a suitable level and perform sufficiently to be approved as an educator.
>
> Interview 5

The clinical educator, then, is trying to manage their public identity to "those instructing the instructors", and to others within their work setting where this participant comments he needs to "justify his position", in addition to their clinical identity and recognition of patient needs.

7.4 Discussion

These findings show that underlying an educator's decision to use or abandon a particular skill teaching strategy, many practical considerations are considered, in addition to the flexibility or reliability of a given strategy. Cognitive load factors between the educator and student are present, as well as the need for adaptability in approach due to the complexity of the task at hand.

7.4.1 Are "simple" skills actually simple?

The complexity of a clinical skill was cited as a factor influencing the decision of using 4SA or not. Interviewee 6 mentioned reservations about using such an approach for simple skills such as a chin lift or jaw thrust. This prompts some reflection on the meaning of "simple", and its subjective nature. A patient may require *simple* airway interventions such as a chin lift, jaw thrust or insertion of an oropharyngeal airway when they are unable to protect their own airway. This may occur due to cerebral injury, inadequate cerebral oxygenation (for example due to a lack of blood flow from the heart), or chemical sedation. Depending on the cause for this, what paramedics refer to as, "loss of an airway", the clinician providing these manoeuvres may have a significant amount of urgency, pressure, complex diagnostic reasoning, and scene management to perform. So in context, these skills are not necessarily *simple* when applied. Thus there is significant complexity in teaching even the skills which appear simple, and this appears to be underestimated by some educators. Considering what is meant by the word *skill* may be helpful as educators grapple with the assumptions inherent in teaching the complex expert practice they apply. Is a skill an action to be performed? Or reflective of an instrument to be used? Or does it connote expertise?

7.4.2 Skills are more than doing the right thing with your hands

A recurrent idea in this study is that clinical skills are more than just learning how to hold and what to do with a piece of medical equipment or tissue. A strong and recurrent expression that clinical skills ought to be taught alongside context, rationale, anatomy and practice considerations was evident throughout the data, and the steps in 4SA in and of themselves, do not achieve this. The educators seemed to be crying out for the application of a spiral curriculum to skills teaching (see Figure 6). 4SA may allow a non-clinician to affectively teach the manual components of the skill performance, but this will not always leave the student with the ability to perform the skill competently in clinical practice. Indeed, knowing when not to perform a procedure may reflect superior skill and expertise than performing the procedure (physically) flawlessly.

The integration of teaching clinical skills, with the rationale of applying them appropriately, and appropriate reasoning and decision making is what allows the student to learn to become a skilled practitioner, rather than simply to do procedures well. The alignment of the educator's role to the identity of a clinician and *also* an educator is critical in modelling, explaining and troubleshooting this co-construction of multiple facets required in becoming skilled.

7.4.3 Complexity and flexibility

Clinical skills may be more complex than many clinicians first think. Add to this the variability in patient presentation, and all of a sudden the developing clinician must be armed with a skill approach which is adaptable to that variability. An example of this is the patient who "hasn't read the textbook". This phrase is commonly used by health professionals to describe the variability of patient presentations beyond a single presentation. One might almost imagine a clinician wondering "I gave the patient Salbutamol for this bronchospasm but it is not resolving. Haven't they read the textbook? Salbutamol is supposed to fix bronchospasm!" The phrase connotes the obvious reality that clinical care is not fixed, and applying rigid procedures in clinical medicine will not always achieve the best for the patient. The "textbook" presentations describe how clinicians may expect to see various states of disease manifest in the patient, and while this may be true, presentations vary drastically in authentic clinical practice. If a single approach to the skill application is taught, how will a student be able to adapt it? Alternatively, if understanding of key principles are taught, a student may be much more likely to apply and adapt the process to a variety of patients.

The desire from interviewee 2 to teach their student "the absolutely correct way to deliver the procedure", and the resultant adherence to 4SA is bound up in this assumption that a skill performance has a single correct process, a single rationale, and this will inform the benchmark for s "competent". This is supported in the original documentation of 4SA, with reference to the *workmanship of certainty* and the *workmanship of risk*. The *workmanship of certainty* refers to the standard procedures of clinical procedures, which tend to have a standard outcome. Granted, this is not always possible, but over time many surgical procedures have become a "reproducible operation with predictably good results" (Walker & Peyton, 1998, p. 172). This "rigorous step by step process", Walker and Peyton comment, is what has allowed some more advanced medical procedures to have "flourished". It is from this practice of what is (or appears) *certain* (routine and predictable), a more advanced practice of *risk* can be built.

4SA was perceived by many participants to be expected to increase the consolidation and skill recall through the structure, repetitive strategy and diverse teaching range within the one teaching method. This all provides a reasonable argument for improved skill uptake, and to learn *the precise* way to perform a skill. However, there often isn't one single correct way to perform a skill. Clinical practice should be adaptive, as no two patients every present identically and hence they require a flexible approach, based on sound principles rather than rigid protocols. Section 7.4.4 will expand on this further.

This may relate to why some educators felt that 4SA was too "rigid". Many who commented that 4SA was flexible and adaptable are not referring to 4SA as it is, by definition, set. Removing, repeating or adding steps will change it from 4SA to something else, and the desire to do that reveals something of the educator's desire to be adaptive in his or her teaching, in reference to the context of that patient, on that day, with that skill. As this complex process is responsive, based somewhat on developing experience, constant reflection of the situation, and relatively un-definable, it is closer in description to an art, or a performance craft. It is impossible to pin down, difficult to explain, but it is salty, spicy, and has an unstatable quality much like a highly refined art. Ultimately it is recognisable, but not definable in words.

7.4.4 Identity formation

Miller (1990) is clear that doing (the apex of his pyramid) is action, laden with context and professional practice. Identity formation should not end with contextual action based on professional expectations, but ought to incorporate the clinician as a person, with values, motivation and experiences. Educators have grappled with the language required to teach perception and responsiveness during clinical procedures, and this is consistent with the preoccupation we have with the observable outputs of a body's behaviour, while the internal processes remain somewhat hidden, though still felt and understood. This is a building of the inner person which drives the outward action, rather than the outward action alone. So are words and demonstration the only tools available to help a student develop this?

In his introduction to *Empire of the Senses*, David Howes (2005b) distinguishes the speaking mouth from the tasting mouth, the latter of which tends to be accepted as subordinate (p. 2). His point is the "elevation of language" relating to the speaking mouth is a mirage in comparison the depths at which the things we have no adequate descriptors for (the sensation of taste) can be known and experienced. Later, Howe articulates the Peruvian Cashinahua teaching that:

> a wise man, *huni unaya*, has knowledge throughout his whole body. '*Hawen yuda dasibi unaia*, his whole body knows', they say. When I asked them where specifically a wise man has knowledge, they listed his skin, his hands, his ears, his genitals, his liver, and his eyes. 'does his brain have knowledge?' I asked. *'Hamaki* (it doesn't),' they responded.
>
> (quoted from Kensinger, 1995:280 in Howes, 2005b, p. 6)

Likewise, if our focus is on the hands which do a task, rather than on the hands which perceive and interpret information far beyond the limitations of our words, we are restricted. Thus, a cultural understanding of knowledge may be helpful. Knowing how to perform a skill, knowing when to perform a skill (or not), and being able to appropriately apply and adapt that skill are all components of skilled and professional clinical practice. This application is an outworking of the physical body, and how the different aspects of the physical body *know* what to do. In western medicine we understand the mechanics of movement (dexterity, muscle memory, coordination etcetera) and decisions around appropriate the application of clinical skills to be driven by cerebral processes which are connected to and impacted by the self: intuition, emotion, and identity.

To separate who one is from what they do is artificial. The impact of self-identity on behaviour was measured to exceed that of social identity and an agreed set of cultural group norms, so why is our clinical education not targeted at this depth? The caring professions (education, medicine) demand the worker's self to be brought to the role, but just as assessment becomes more difficult with the higher levels of Miller's pyramid, how might we assess (and teach) the *becoming* proposed by Cruess et al. (2016)?

7.4.5 Educator experience

Another resistance to 4SA uptake was noted as educators' pre-existing habit, or experience. Fullan, Galluzzo, Morris, and Watson (1998) investigates why teacher reforms may fail, with a review of a decade of teacher reform attempts through USA from 1986. Fullan et al. suggest that "many leaders believe that teaching is not all that difficult" (p. 20), and while his reference is to school education, perhaps this explains an expected graduation from clinical expert to educator, which has often occurred without the requirement for formal training and expertise development in education methods (Hagler, Kastenbaum, Brooks, Morris, & Saewert, 2013). The findings in this chapter raise a number of questions around how an educator's experience in teaching may impact their chosen style. It was identified by participants that some educators have a special ability to engage and inspire students, and this has been referred to as a *craft*. How this develops remains unknown. It may be that some teachers have recruited learning principles into their practice and that this craft is informed by theory and refined by practice and reflection over time. Further, should the focus of clinical education in terms of faculty development be to teach procedures, teach strategies, or to invite educators to grapple with concepts of adult education in the clinical context, with thought to the various stakeholders, and make an informed decision on what might contribute to their *craft*? An assumption with this suggestion is that as educators gain experience, they will have the insight to know what will be an appropriate addition to their *craft*, however experience does not necessarily equal expertise.

Educators who have developed their *craft*, as identified in some of the data in this chapter, may be negatively impacted by the introduction of a rigid 4SA into their practice. But additionally, a novice educator who has not developed a *craft* based on expertise (regardless of how experienced they are) is expected to benefit from the use of 4SA. This is hypothesised to relate to the perceived role of the educator.

Where the person teaching identifies as a *clinician*, they may have a more clinical focus and teaching will be more focussed on the patient's needs than andragogy. The teaching format may be "ad-hoc", unplanned, and dependent on time, resources, and the setting as perceived by the clinician. The person who identifies as a "*clinical educator*" or "*clinical teacher*", may put effort into using the strategies presented to be expected of them, with an explicit desire to teach according to recommended strategies, with some focus on both the student's needs and the patient. This teacher may be trained in teaching strategies and principles, but is likely to use such recommendations without significant adaptation either due to the inflexibility of their own training as a teacher, a lack in confidence or existence of their own expertise, or a lack of critical analysis of the complex needs of the situation at hand. Finally, the *craftsman educator* whose attention falls to the needs of the current and future patients, the students as developing professionals, and the many individual factors which must be addressed to contribute to a focussed, helpful, inspiring education experience. The teaching will be adapted as needed. These are represented below in Figure 47, where *craft* is an expert synthesis of andragogy and the individual needs of the situation; *strategies* includes teaching templates such as 4SA, and *what seems right* may be interesting but lacks fundamental principles or reference to the needs of the group.

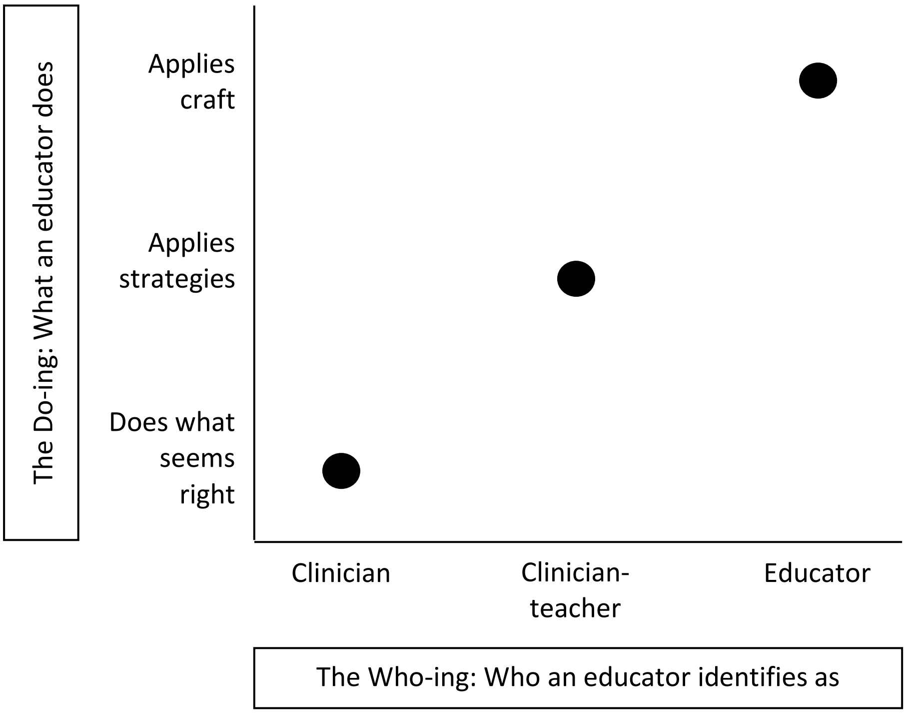


Figure 47: The relationship between a teacher's identity and their teaching. Doing what seems right is very ad-hoc teaching. This clinician may employ strategies they experienced while they were themselves a student, as driven by clinical culture. A Clinician-teacher may be more interested in learning strategies with which to teach, and may use these as a way to structure or support their teaching. A craftsman educator has developed a craft by which they adapt to the complex and unique clinical education situation which they are in. It is infused with insight, educational expertise, and flexibility to the requirements of the situation at hand.

This image is not completely representative of all educators. For example, an educator not represented in the figure and is worth special mention is the one who has experience, and has developed habits, but these habits are not truly adaptive to the complex clinical and educational needs of the situation. This educator may *feel* that they have developed a craft, and may be *experienced* in applying it, but still be compelled more by an individual style than insightful expertise and sound andragogy. This educator is sometimes difficult to spot!

7.4.6 Applying findings to the symbiotic framework

7.4.6.1 Clinical axis

A key strength behind Worley's (2003) clinical axis is the engagement of students in the clinical setting, and the consequent participation in patient care, development of illness scripts, and relationships between students and their patients, in addition to the relationship between student and teacher. Educators tended to see the value 4SA for use in the skills laboratory setting rather than in authentic clinical teaching, indicating that its use as a tool in symbiotic clinical education may be limited.

One of the perceived benefits 4SA may offer is for the clinical teacher who finds supportive structure in such a strategy. Although this may increase the cognitive load for some educators, for others it may provide a useful tool to implement, and then reflect and adapt according to increasing awareness of andragogical principles as the clinician educator develops their expert craft.

7.4.6.2 Institutional Axis

The standardised approach of 4SA is tempting to accept when a consistent output is desired. While some educators rested in 4SA as a means of achieving this consistency, others were tentative of this standardised approach. The expectation of consistency implies that each student, skill and situation remains consistent, in order for a standardised approach to achieve consistent outcomes for students, the health professionals who graduate from such a course, and the reputation of the teaching institution.

While strategies such as this are sometimes seen as a means to alleviate educators' concerns (by resting on an external teaching philosophy or reform strategy), Pasi Sahlberg warns against this (Sahlberg, 2011a). Sahlberg argues that by investing in educators and allowing them to develop andragogically sound practice and expertise, and by allowing them to perform their professional role as educators as they see fit according to this expertise, they will be supported as a profession, and feel safe to continually test and refine strategies for the benefit of the individual learners they teach. This principle may be applied to medical education: if medical educators are invested in as specialists within their clinical field, they may flourish in a way that promotes innovation in learning, and may achieve far greater outcomes for the student, patients, health service and teaching institution. Fullan et al. (1998) state that "the problem [of education reform failure] begins with teacher education programs" (p. 20). Thus, investing in educator development and building a solid case for respect of this among other clinical disciplines is a key step. Prescribing actions for teachers to follow is antithetical to partnering with educators in reform and educational research (Fullan et al., 1998, p. 39). Thus, the value of educators as a resource is implied. The authenticity of Worley's model is brought into this argument by addressing the educator, as a clinician, and encouraging their focus on the future patients who will be managed by the developing professionals who they teach, be it in the clinical setting or in the skills laboratory. This perspective integrates their identities as clinician and educator.

7.4.6.3 Social Axis

Worley's social axis highlights the relationship between government policy, budget considerations, and the demographic, cultural and epidemiological needs of the community. Factors such as these may influence not only what skills are taught in a program, but also how the skills are taught. Skill teaching ought to include relevant patient sensitivities (relating to culture), which may be integrated into 4SA. However as identified in the data, some educators express concern over the potential to teach an inflexible approach to skill performance with 4SA, and the resultant inability students may have to adapt skills to different types of patients (and different patient presentations) during subsequent practice. Thus, skill education will be most integrated with the community's needs when it is conducted with the patient, and not just the process, in mind. Therefore, as Fullan argues (1998), "it is pointless to work on school reform without prior community building efforts". Medical education expertise must be invested in, supported, and valued by the wider medical community.

7.4.6.4 Personal Axis

Educators expressed a desire to see students learn to grapple, perceive, adapt and rationalise their approach as a developing health professional, but 4SA was not considered a useful strategy to that end. Indeed, this was a point of non-resolution within the focus group: how do clinical educators do that? When it comes to professional identity formation for the student is not a natural outworking of a manual skill session, especially when divorced from important aspects of implementation such as patient communication and empathy, rationale, and when it is best to *not* apply the skill. Thus, a spiral curriculum which addressed these components during initial training and intentional practice, in a modelled (even if simulated) environment, will allow for more integrated identity development. The student will learn how to be, rather than just how to do.

7.5 Conclusion

This research question was prompted by incidental findings of lower educator compliance to 4SA than 2SA in Chapters 3 and 5. As the intention was to understand the educator experience of the 4SA teaching strategy, a phenomenological influence was appropriate during this qualitative study. Educators have different reasons, caveats and contexts in which they choose to subscribe (or not) to 4SA. This may be impacted by their understanding of the likely benefits and limitations of 4SA, and the expectations they perceive to be placed on them by others.

Some educators perceive that 4SA is likely to decrease the extraneous cognitive load on the student, through a structured approach and repetition inherent in the teaching method. Thus, the student is expected to have more ability to address the intrinsic and germane cognitive load of learning a new skill. This may in effect transfer some of the load from the learner to the educator who may have an increased demand through the implementation of an unfamiliar teaching approach such as 4SA.

4SA is understood by some as a teaching method which achieves consistent and reliable results, which is important for globally marketed training programs such as ALS, and PALS courses. While this consistency provides a sense of structure for novice educators, it is seen as rigid and inflexible by others who expect it would stifle the expertise some admirable educators have developed. The variation seen in authentic clinical practice demands variability in an approach to practice, based on clinical principles rather than rigid protocols, and the same is found to be true for medical education: individual factors relating to the teaching context, stakes, skill complexity, and student variation demands a more flexible and adaptable approach to learning clinical skills.

A clinician who is a novice educator may benefit from 4SA in order to become a more confident clinician-educator, however the expert educator will work from principles, insight, and adaptability in reference to the complex task which is before them, and 4SA may hinder this.

Medical and clinical education is a complex task, however the manual task of inserting an OPA into a manikin (or teaching a learner how to do so) is not. If medical educators intend to teach a task, their game is finite, and a finite tool such as 4SA is appropriate. If medical educators seek to inspire learning and nurture a developing clinician's skill understanding in order to aptly manage the variation they will see in their clinical context, the educators' game is infinite, and a finite tool such as 4SA is much too inadequate (these ideas will be further developed in the context of the thesis, in Section 8.4). For this task, adaptive expertise in response to the complexity before them is required. With the emergence of medical education as a specialist clinical role, we need to step closer to recognising the specialist, adaptive approach required to masterfully engage in the education of a clinical topic which one is already recognised as clinical expert in.

This qualitative review provides a body of evidence to suggest that the using 4SA or any rigid teaching strategy may come at a cost if it encourages educators to overlook the individual student needs, or if it overshadows an individual's teaching expertise. However, for educators who do not have a deeply established and reasoned andragogical approach, 4SA may offer considerable benefit in providing structure and confidence for the more novice educator. Thus, the question of cost-effectiveness about any teaching method is much more multifaceted than first understood.

8 DISCUSSION

8.1 Comparative cost-effectiveness

This series of studies has investigated the cost-effectiveness of 4SA compared to 2SA using a pragmatic mixed methods approach. The first key finding demonstrates that while the modified 4SA approach used in Chapter 3 appeared to relate to superior manual defibrillation performance on a specific checklist than 2SA teaching (immediately post-teaching), a significant difference was not detectable when incorporating initial performance scores in a repeated measures design, or when using a global rating scale.

This study raises a series of questions about how clinical skills are taught and assessed. The two assessment methods (a skill specific checklist and a global scale) were initially considered different as one was "subjective" and the other was designed to be "objective", however some disagreement between the two assessors was noted. This raised suspicion that true objectivity is not possible when human assessors are asked to provide even very binary judgements to student clinical performance. This was the first challenge to the positivist approach which permeated the beginning of this PhD. This finding also challenges assumptions about assessment tools and their interpretation. As further discussed in Chapter 6, an assessment tool is a proxy for something which, in education measurement, is unseen and unseeable. The implicit intention was to obtain quantitative data to compare what is increasingly understood through the thesis to be a qualitative phenomenon: learning. In Chapter 3, the bottom-up, skill-specific checklist was developed following an audit of the teaching sessions and the teaching script provided to the educator by the research team. This assessment tool, then, generates a quantitative indicator of coherence and compliance to a procedure as taught to the students, rather than actual ability to perform the skill. Educators would intend that these two measures overlap as much as possible, but that is not always the case. In this study, the top-down global scores are more reflective of the assessor's opinion of the student's actual ability to perform safe manual defibrillation, with a benchmark of paramedic clinical practice on-road. As the global scores do not demonstrate a difference between the two teaching methods, it may be that 4SA may achieve more powerful *compliance*, but this does not reflect strongly in *practice*.

This finding may be applied to other education settings. In clinical practice, the clinician may be driven by practice principles, without always being wedded to a single strict "right way" to do something, however in some settings this may not be the case. If the intention of a teaching session is to maximise replication of a pre-determined technique, 4SA and other specific teaching strategies may call for further research.

The second comparative trial sought to clarify some of this confusion by applying skill-specific checklists which were developed according to professional skill practice, rather than based on recall of a teaching session. In this trial, no observable difference was evident between students taught with the different methods, but earlier concerns were confirmed that 4SA requires significantly more time (25 to 35% more) to teach. The monetary cost for education organisations for educator wages and site bookings for no measurable practice or patient benefit is potentially significant for large teaching bodies. Given the difficulty some educators have in complying with 4SA, one of the questions implied by this study series is "is it worth asking educators to push through this difficulty in order to find 4SA easier?", then the data from both trials would indicate that from a cost perspective, from a skill acquisition and retention perspective, and from an anticipated patient morbidity and mortality perspective, the answer is no.

These two comparative trials raised an unanticipated research question: How useable is 4SA? Is there a reason why the educators used in these trials performed 2SA more consistently than 4SA? Is there a hidden educator cost to using 4SA? Should we therefore ask this of educators? These sorts of questions were addressed in Chapter 7. The initial hypothesis was that 4SA is more difficult to teach with, and may increase the cognitive load for the educator. Through surveys, semi-structured interviews and group data collection sessions, educators with some experience in 4SA and other methods of clinical teaching were invited to share their experiences. Unsurprisingly, cognitive load was an early theme to arise in the data, and this was especially interrogated through the analysis to ensure it was not artificially introduced. Further themes to arise warrant discussion around the benefits and issues of standardised education and assessment in clinical practice, dealing with complexity in a variable and unpredictable environment, and the role of identity and perceived role in prompting educational practice.

A limitation of the two trials is the assumption that all students have an equal capacity to learn cognitive information, and apply and practice it with equal dexterity. While students who did not speak English as a primary language were excluded from the studies, other potential differences (such as dyslexia or other learning difficulties) were not addressed in this study. While this may be considered a limitation within the comparative trials, it further challenges the notion that a standard input (a regimented teaching approach or template) will result in a standard output (a predictable performance ability), as the students are individual, and have individual strengths and needs.

8.2 Cognitive load

Participants in the final study (Chapter 7) tended to agree that 4SA required a lot of practice, and as a result its demand on the educator sometimes detracts from the teaching content. These comments speak to the notion of habitual practice in clinical education. Such habits may require breaking, but given that doing so places a cognitive demand on the educator, when the task is already, complex, it should be done so where a clear and established benefit is identified. It appears that in this thesis, a clear benefit for the use of 4SA has not yet been established.

When teaching a clinical skill, the educator must be mindful of the practical limitations (for example time, resources, group size, teaching context, stage of learner, skill complexity), in addition to assessment tasks if applicable, the patient implications and the skill itself. Cognitive load theory argues that as more cognitive demands placed on the educator, they are pushed closer to their natural cognitive limit. For an expert clinician who is unconsciously competent, the skill itself takes little to no cognitive load during clinical practice. Sweller et al. (1998) refer to this as "automation". However, placing them in a teaching situation may on one hand increase their cognitive load because they are not used to explaining the clinical process thoroughly and they are forced to recall information and rationale which they are not used to in daily practice, but an argument can be made that they may have increased cognitive capacity for teaching because the skill itself demands so little from them.

Aside from the clinical fluidity with which a clinical educator practises, the data gathered direct attention to the educational fluidity or habit with which a person teaches. When educators are asked to teach using a strategy they are not used to, they find it "difficult" (survey 41), "harder" (interviewee 5), forcing the educator to “focus on the technique quite a lot and probably you forget things that if you were teaching" (interviewee 4), "awkward" (interviewee 1) and "cumbersome" (interviewee 2). These comments were raised even by educators who subscribed to the technique, often on the assumption that it would become easier with practice.

This hypothesis, that a well-crafted skill teaching strategy will alleviate the extraneous cognitive load for the learner, and allow them to prioritise their available working memory towards schema acquisition, allowing for more sophisticated storage for long term memory, is difficult to test (Whelan, 2007). One could hypothesise that if this was the case, an increase may be seen in the recall of students taught with 4SA compared to 2SA, but skill performance is a very crude substitution for assessing a cognitive process such as schemata construction. Whelan (2007) warns that this may be less helpful than when we may expect:

> recent studies show that when learning materials are too easy, collecting reliable measures of cognitive load is much more difficult. When learners are less challenged, their reports of the load reveal an under-load effect, where the greater range of variance in the dependent measures makes it less easy to reliably address questions about the predictive validity of load measurement instruments for design evaluation (Whelan, 2007, p. 2).

Clinical educators are warned here that when the challenge on students decreases too much, it is difficult to accurately measure the cognitive demands of the learning as the learners perceive too light a cognitive load. This effect is noted by Tuovinen and Sweller (1999), who saw a measurable, positive impact of highly structured learning for students who had little or no existing schemas within which to organise their learning, however this result was not the same for students who already have experience in the topic area (Kalyuga, Ayres, Chandler, & Sweller, 2003). The extra effort required by educators to help the student create this schemata is lost on those who have previous experience. Therefore, learners must be sufficiently challenged in order for these aspects of cognitive load theory to reliably apply. Stepping aside from assessing cognitive load briefly, and turning attention to the learner's challenge, recent theories on the optimal challenge point are brought to the fore. Guadagnoli and Lee (2004) argue that this optimal challenge point exists where the material or learning challenge is not too difficult, nor too easy, and this exists at different levels for novices, experts and those in between. By attempting to reduce the extraneous cognitive load on the learner (by potentially increasing the intrinsic cognitive load for the educator), it is possible that the challenge is too low to motivate learning, especially if the complexity of the skill is overestimated for the learner.

However, cognitive load, does not explain the whole question of learning. As Whelan (2007, p. 2) explains, "Learners in today's world are adept at highly complex multiple task performance that blurs the lines between intrinsic, germane, and extraneous cognitive load in practice. Cognitive load theory does not translate readily into a 'full-spectrum' interpretative device." This admission helps explain the beginning of a movement within this discussion to an understanding of learning clinical skills which is beyond the current routine limits of clinical education discourse. The discussion which follows will move towards recent lessons about teacher development within the Finnish education system, and beyond towards complex adaptive systems and how consideration of complexity challenges and enlightens how clinical educators consider teaching resuscitation and other clinical skills.

8.3 Standardising a clinical skill teaching strategy and the Global Education Reform Movement (GERM)

Key themes arising from Chapter 7 show that 4SA may perceived as beneficial for some educators,, particularly novice educators and others who may benefit from its structured approach or who may not be confident in their own andragogical expertise and approach. However for others it may be too rigid, cumbersome, and inflexible in meeting students' individual educational needs in a variety of contexts, settings and skill complexity levels. It appears that using a standardised approach to teaching clinical skills, for example in ALS, PALS, and EMST courses, may be influenced by the key driving factors of Global Education Reform Movement (GERM). The balance between consistency and control is a key discussion point in Sahlberg's GERM. GERM is explained by Pasi Sahlberg as the "unofficial educational agenda that relies on a certain set of assumptions to improve education systems". This international movement is based on an economic, standardised approach to education with a principal impact on pre-primary, primary and secondary school systems (Sahlberg, 2011a). Sahlerg describes the impact unfavourably, with the acronym easily lending itself to connotations of infection, disease and harm (Sahlberg, 2012). Sahlberg states that:

> The Global Education Reform Movement has had significant consequences for teachers' work and students' learning in schools. Because this agenda promises significant gains in efficiency and quality of education, it has been widely accepted as a basic ideology of change, both politically and professionally (Sahlberg, 2012).

The success of GERM in infiltrating education organisations and the teachers who subscribe to it can be seen as related to the gains it implies. Improvement in efficient and effective education, which drives teachers' meaning in their role, makes such a strategy appealing.

8.3.1 GERM's five identifiable features

The first feature is the standardisation of curricula and how they are implemented (Sahlberg, 2011a). This standardisation addresses an imperative (perceived or actual) to meet minimum learning objectives and benchmarks. In turn, this breeds competition and comparison between schools which become ranked on outcomes of standardised testing against these curricula, which contributes to a higher stakes environment within which teachers are rewarded (or otherwise) based on ranked student scores (p. 100).Secondly, an increased focus on core subjects (Sahlberg, 2011a) is closely connected with the previous feature, as it seeks to establish a consistent, minimum expected competency of content for each student. Education becomes a means by which educators produce a predictable product, led by a global market demand, in order to meet a minimum expectation (p. 100). Third, a prescribed curriculum evolves where the interpretation and implications of test results are higher stakes (Sahlberg, 2011a), for example education institution funding, enrolments, project grants. This promotes low-risk education design which is a natural antidote to "freedom and experimentation" in the classroom, which prevents innovation in reaching predetermined learning outcomes (p. 101). Fourth, education models incorporate practices from other industries (Sahlberg, 2011a), for example a focus on economics to build human capital through production-line teaching. This may be aimed at increasing efficiency and output of the educational system, but Sahlberg argues it undervalues the education sector's ability and freedom to learn from its own practices, failures and successes and improve using strategies congruent with the unique context of education (p. 101). And finally, high-stakes accountability policies emerge (Sahlberg, 2011a), where facility and faculty income, accreditation and management is closely linked to student performance (p. 101).

Sahlberg presents the Finnish education system as a remarkable and successful antithesis to GERM. Key strategies in resisting GERM, and thereby achieving undeniable educational success for pupils and professional autonomy and national respect for teachers, include adaptive, customised learning to classes where students learn alongside peers of varying ability; creative learning focussed on the student's development as a whole rather than an isolated learning objective; risk-taking by implementing innovative approaches to teaching and learning; using reflections on the past to plan innovative teaching built on the teacher's wider professional role, and the relationship with the students; and finally by building a national culture of trust, respect and autonomy for teachers, in a system which values education and targets resources appropriately (which is unlikely to be equally), utilising sample-based assessments rather than an over-saturation of regular testing (Sahlberg, 2011a, p. 103). An immutable feature of education in Finland is the extreme popularity of education as a profession, even more so than Medicine or Law (p. 8), despite a salary just above the national average.

8.3.2 Is there a place for GERM in medical education?

Education systems influenced by GERM will promote teaching strategies which may be taught to inexperienced or untrained educators, potentially indicating a lack of trust in their educational ability and expertise. In the health professions, many clinical educators find themselves in a position where they are invited or expected to teach because they are recognised as proficient or expert in the content material, but have little to no formal training, qualification, or even interest in teaching. Newton argues that:

> One cannot assume that all nurses are able to teach students, as this underestimates the importance of the preparation that is required to be an effective preceptor (Newton et al., 2009, p. 10).

Newton is by no means alone in her sentiment (Ash, 2010). It was this environment into which 4SA was developed and took hold. As a result, in the data arising from this study, it is clear that 4SA is perceived as a helpful structure for novice teachers, and indeed for unconsciously competent clinicians who may make a high number of assumptions when teaching. As one ALS instructor explains in Quote 50 (previously discussed on page 213):

> Quote index 50
>
> I'm relatively young and novice um if I, you know, I'm more likely to want to have a crutch that I can lean on in my session, you know if I get stuck halfway through I know what my strategy is going to be; I've got my four steps.
>
> Interview 6

This was spoken by an experienced and skilled clinician, who perceived herself as a novice educator. She found security in the 4 steps, as her expertise as an educator was still under construction. Her response to 4SA in this way may be an assumption of its worth as a teaching strategy steeped in learning theory, as she may have received it during her instructor course as an elite method which if it is done properly, the learning will be done properly. Or it may be valued as a temporary crutch until her education expertise develops (I did not follow up this line of questions). It appears that the benefit here is in the structure afforded by 4SA, rather than the individual steps themselves.

This standardised structure to teaching clinical skills was also reported as a way to achieve consistent teaching despite a range of educators recruited as ALS instructors.

> It's a bit like trying to make sure you're singing from the same song sheet is not it? ... they believe that's going to give them a consistent outcome regardless of the individual and I suppose if you're trying to teach a course that's wide and in fact probably worldwide, and you want to produce an apple every time then that's what you need to do.
>
> Interview 4 (ALS instructor)

Courses which are established to accredit to a standard criteria will logically employ such an approach for instructor development. Educators may be recruited on the basis of clinical expertise and interpersonal ability which is perceived as consistent with that required to educate, but an assumed safety net may exist when all are required to *song from the same song sheet*. This may imply working together in cohesion with the wider purpose and teaching team as with for a literal choir which will disintegrate if some members start singing the song in a different key; and it may also reflect the assumption that consistent input (4SA) will achieve a consistent output (minimum competence), despite variable subjects (clinical students who have different background, assumptions, cognitive capacity) and application in an uncontrolled environment (patients with variable comorbidities, anatomy, and considerations. Teachers' and education organisations' subscription to 4SA appears logical at first glance, but becomes flimsy when more critically reviewed. The loss of adaptability for set teaching strategies such as 4SA is a key criticism of GERM theory. Addressing learners' individual needs is at odds with a marketised philosophy of education courses which encourage (inherently or explicitly) a conveyor-belt learning program. When educators are forced to teach in a pre-determined way, the adaptability of their message may be compromised, their experience and skills restricted, learning can be impacted, and in the context of clinical education, this may have a profound effect on practice and patient care.

8.3.3 What can we learn from this?

In 2011, the Finnish education system invested six times more in teacher development (30 million U*SD* annually) than standardised testing (5 million U*SD* annually), with expectations to double teacher development over the following 5 years. Such an investment, in direct opposition to the core philosophy of GERM, emphasises the flexibility, creativity, adaptability, problem solving and cooperation to which the success of the Finnish education system is attributed (Sahlberg, 2011a, p. 98).

The warnings attached to GERM do fuel a valid argument that all standardised approaches should be abandoned. The context of educational reform in Finland was steeped in highly trained, respected and trusted teaching staff. The minimum qualification required to teach in schools is a Masters degree, competition to enter the career is fierce, and structured mentoring occurs throughout the early stages of a teacher's career (Sahlberg, 2011a). The investment placed in educator development lends itself to high degrees of trust in educators, autonomy, reliable judgement, teaching confidence and educational insight. The context of clinical education has been evolving towards this end, but still has a long way to go. Through progressive investment in teacher development, thereby supporting the status of educators in the health professions, it will be less likely to be perceived as a safety net for those who do not want to (or can not) practice. Instead, Sahlberg's work argues, it will be lifted up as a specialty within medical practice, and the trust, autonomy and andragogical expertise will naturally effect tomorrow's clinicians.

8.4 Finite and infinite problems: complexity in clinical skill education

A key restriction GERM imposes is a lack of adaptability and acceptance of variability. James Carse's (2012) work helps us understand the imperative for expertise and adaptability within the complexity of medical education. The notion of finite and infinite problems was initiated in 1986 by Carse, and at its core is the premise that these two types of problems which have fundamental differences and therefore require fundamentally different approaches and strategies. Carse refers to the two *games* that these problems give rise to: finite and infinite. In clinical practice we, our patients, the public, the health service, and society in general expect us to predict, recognise, control, and generally cure disease processes. In clinical education, students, clinicians, education organisations and health service providers, expect that educators can teach student clinicians to do these things. However, some problems exist which cannot be solved within our current frameworks and social constraints, for example the deeply complex and interdisciplinary dynamic of poverty, the ecological impact of over-fishing, conflict in the Middle-East, or the current refugee crisis. These issues cannot be helpfully addressed with a single approach. This is a key difference between finite and infinite problems.

8.4.1 Finite problems

Finite problems are predictable. They have a known beginning, middle and end (Carse, 2012, p. 4). They are limited, and containable. The finite game is *winn-able*, and the outcome of a finite task is associated with establish-able competency because it involves a limited number of variables. The solution is knowable, for example the administration of adrenaline and antihistamine in anaphylaxis offers a relatively predictable resolution of symptoms and disease. In the context of medical education, the task of teaching a student to insert an oropharyngeal airway (OPA) into a training manikin could be reasonably understood as a finite task. The context is defined, the outcome is measurable and there is a line at which the educator, assessor and student recognised the task as having been achieved. The achievement of this task is the point at which the game has been won (Carse, 2012, p. 19).

8.4.2 Infinite problems

Finite problems have finite solutions, addressed by playing a finite game, but infinite problems do not. An infinite game can not be played like a finite game, although it may include a number of finite strategies. Infinite problems, are *un-winnable*. Therefore the point can not be to win, but rather to keep the game moving towards a more acceptable space (Carse, 2012, p. 3). There is no single measure of success, and the individual's role is to identify where in the confusion and insecurity an option for constructive action exists, with constant reflection and re-evaluation to assess the impact of that move on the rest of the system. The aim should not be to solve the problem once and for all, because this is not possible. These problems generally have a high dimension of variables and differences, and multiple relationships and unpredictable complexities. For the clinical educator, the task of teaching a clinical student to confidently and competently manage a patient's airway in the pre-hospital setting is closer to an infinite problem than a finite one. It does not have a measurable or definable end-point, because so many variables exist: either a soiled airway through emesis or trauma with consideration to cervical stabilisation dependent on mechanism of injury or illness could dramatically alter the strategy used by the pre-hospital clinician. The physical environment the student clinician is preparing for during his or her training is unpredictable, ranging from an upside-down car in a creek bed, to under the desk of a corporate office, or a beachside walking trail, including everywhere in between. There is no one way to approach the goal, or ensure that it has been completed. It involves perception, judgement, and sometimes "just enough pressure" which cannot be communicated wholly in words, but instead develops through practice, reflection, and experience (at least in part). This clinical educator's task is an infinite one, which will include proficiency in OPA insertion, but is well beyond the limits of that single airway strategy.

The key is understanding that a finite game, with its finite end-point, predictable series of outcomes and definable and containable aims is approached very differently to an infinite game, where a determinable resolution for the game is not possible. Infinite games are dynamic, involve unpredictability, and require the application of insight and principles, rather than a pre-determined inflexible plan. If the educator sees his or her task as ensuring a clinical student can replicate the dexterity required to insert a medical device correctly, the game is perceived as a finite one. This finite problem may be appropriately solved with a finite solution. The 4SA to teaching clinical skills may be one such solution.

However, if the educator sees his or her role as an infinite (complex, dynamic, evolving and multifactorial) one, the game is very different. The intention may be that the student to become competent in the application of a skill, which relies on reasoning, problem solving, memory, application in the context of pressing emotional circumstances and adaptability to different anatomy in addition to other factors. In this situation, a finite, protocol-like strategy won't satisfy the problem or the educator who perceives this problem. For infinite problems, Glenda Eoyang (2013) suggests players use adaptive action. This approach is based on Complex Adaptive Systems (CAS), within which the interactions between individual parts, which cannot be predicted, and give rise to a system-wide adaptation over time. The system-wide changes then often reinforce the individual interactions collectively in achieving a sustained change (Dooley, 1997; G. H. Eoyang, 2006). With this description, a CAS approach to teaching and learning clinical skills cannot be convincingly argued in the context of this thesis, however the use of adaptive action as described by Eoyang (2013) is still of particular relevance.

8.4.3 Adaptive action

Adaptive action is described as a reflective cycle of assessing the current state, analysing the key factors, and taking decisive action based on your analysis, then review of the problem in light of action, giving rise to secondary analysis, secondary responsive action, and the cycle continues (G. Eoyang & Holladay, 2013). A prelude to adaptive action is clear in the data arising in this theme: limitations of the 4SA within a complex clinical context, and a desire to either apply 4SA flexibly or use a different strategy altogether. In the 4-stage approach, the action is prescribed, and not specific to the complexity of the evolving and dynamic situation before us. Hence, it may not allow the educator to navigate out of the confusion of an infinite problem.

8.4.4 Will an educator perceive a finite or infinite problem?

The choice as to whether 4SA is appropriate relates in part to whether the educator perceived this strategy to be sufficient for the task. For a finite task, such as an isolated skill in the simulation laboratory, this appears to be much more prominent as the variables can be controlled. For an infinite task, such as seeking to teach a paramedic student how to apply their skill set flexibly in order to manage the undifferentiated patient in the pre-hospital setting with a complex medical history, and external stressors such as violence at the scene, and communication barriers such as an intellectual disability, employing 4SA even in the simulated or skills laboratory setting may be considered insufficient for the task. The complexity is raised by the educators in this study seems to arise through a potential mismatch between the task the educators perceives they have, and the means by which they may achieve it.

The perceived task (or game, as Carse would call it) is not always the explicit task. Where a learning objective may be stated as "proficient insertion of an OPA", the clinician's learning objective may be far more complex, and connote a raft of previous patients' faces to inform what they want their student to be able to do, should they encounter similar variety. This implicit or hidden task may be motivated by experience, reflection on previous mistakes, and their perceived role as an educator, as infused by their identity.

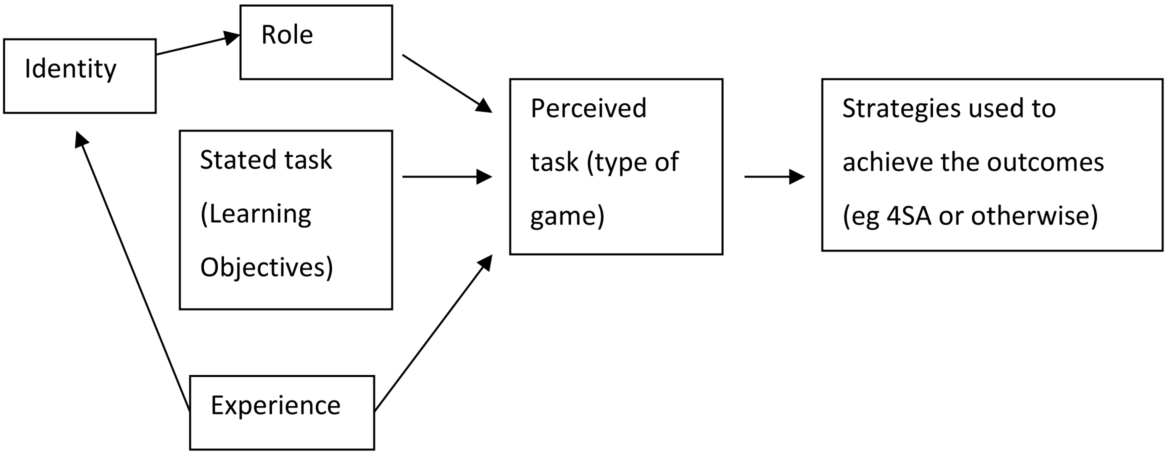


Figure 48: Suggestion of how educators select a teaching strategy resultant from a) their perceived role, task and identity, and b) the type of game they are engaging in.

8.4.5 Identity

The theme of identify emerged in two key ways in Chapter 7: the identity of the educator, and the development of the student's identity.

Firstly, the educator's identity was one which was seen to impact the credibility, depth, currency and approach of the skill teaching session. A hypothetical character who emerged repeatedly through the focus group was "the gardener" who was taught to use 4SA, and could teach the steps of a clinical skill competently: "as long as [the student has] the capacity to memorise the steps, [he or she] can pass the skill" (focus group female participant 15). The clinical identity of the educator was perceived as deeply important in order to accurately and confidently address student questions, with perceived credibility. In the first comparative trial, the educator used was an established and experienced clinician, and also had experience in teaching psychomotor (though had not recently taught clinical) skills. It could be that his identify as an educator with habitual, comfortable, and arguably proven skills in teaching manual motor skills, that to align to a strict teaching procedure was too unnatural. The second comparative trial used a different clinical educator, who was also experienced and well respected as a clinician and educator, but whose reputation as a *clinical* educator was well established. He demonstrated a clear desire to comply to the study protocol, and did so with 2SA even though this may not align to how he would ordinarily teach, and sought to do similar with 4SA because of his high regard for research protocols in a truly comparative trial. Each deviation (3 in total) from 4SA was self-identified, and remedied to the best of his ability, but this phenomenon prompted further investigation in Chapter 7. Findings suggest that for the educators who have developed expertise, such as the two used on Chapters 3 and 5, 4SA is counter to their established practices as an educator. Novice clinicians who perceive themselves as clinicians who are learning to educate value the helpful structure of 4SA, but those who have established expertise in teaching are expected to be "stifled" by it (Interviewee 7). This may be related to the competing priorities in the mind of someone who is a clinician (with patient considerations), and an educator (with student learning outcome considerations).

Some of these ideas are echoed in Ash's findings (2010). In her interrogation of identity in medical education, Ash explains the surgical and medical educators' roles as including both educational and clinical roles, which were sometimes in conflict. Ash found that "all clinicians are role-models to students whether they like it or not" (p. 129) and mindfulness of this impacted how the clinical educator performed their duties. The role of inspiring, guiding and encouraging learners was identified by her participants, in order to develop self-learners rather than students dependent on a teacher's input. Thus, for Ash, identity is intrinsically linked with perceived role, and the resultant approach. The description of these roles, and the resultant tasks of education are infinite in nature: How does one measure the end-point of inspiring a learner? That game cannot be finally and completely *won*, as it is an ongoing relational process, hence a finite approach will not suffice. The perceived role, likely impacted by professional experience and expectations will influence the perceived task, and recognition of a finite or infinite game, hence informing the strategy used to teach a skill.

8.5 Fullan's educational change theory

The educator's perception of their role, task and management strategy is likely to be driven in part by habit, the expectation on clinical educators to adopt 4SA is sometimes met with resistance which may be explained by Fullan's educational change theory (Fullan, 2002, 2006). Fullan's transition to consider educational change rather than applying organisational change theory to the education sector offered a new perspective for its time (Ash, 2010, p. 11). In understanding why some educational innovations are successful and others are not, Fullan notes that educators must have meaning in their role and tasks for change to be sustainable (Fullan, 1993). He writes:

> Teacher education programs must help teaching candidates to link the moral purpose that influences them with the tools that will prepare them to engage in productive change

This morality of teaching connects the educator to the tasks they perform, and the meaning derived for the educator, Fullan argues, promotes the uptake of educational change, but is also protective against burnout. He found that teachers teach due to their personal vision which gives meaning to the work. This personal vision comes from within, but ought to be explicit in order to guide the educator like a compass in his or her teaching role. Fullan argues that "people *behave* their way into new visions and ideas, not just think their way into them" (Fullan, 1993) (emphasis original). If this moral purpose and personal vision conflict with the nature of proposed educational change, educators are unlikely to subscribe and the proposed change will be less successful.

These aspects are seen emerging in the final study of this series. Having identified some possible reasons why 4SA may not have been used as consistently as 2SA in the two trials described, other themes like educator subscription, and the educator's underlying aim of the teaching session hinted that an educator's actions are driven by a perceived role, which reveals an inner identity. It is this inner identity which promotes action. Introducing teaching innovations without reference to or consideration of the social processes which support such reforms within a professional operation will threaten the adoption of such proposed change (Ash, 2010, p. 12), and this is evident in the context of educator development sessions which present a strategy such as 4SA. 4SA is a structured approach to teaching clinical skills, however the focus of the innovation is learning theory (such as Miller's pyramid), and an assumed improvement in skill performance (compared to less structured teaching strategies). Where educators start considering the patient interaction at the bedside, and building rapport with the student to promote engagement and a longer term educational partnership with them, the appropriateness of 4SA to all skill teaching contexts becomes threatened. This response is echoed by Fullan's challenge of Rita Kramer's argument that a "summer of well-planned instruction" sufficient for a person to learn to teach effectively (Fullan, 1993). This philosophy of teacher development seems more aligned to an indoctrination of robotic teaching approaches, rather than an inspiration of the moral value and identity development Fullan argues infuses effective educators. This is supported by examination of the 10 suggestions he makes for superior faculty development. These suggestions use verbs such as *commit to, value, engage in, model and develop, be visible and valued, work collaboratively...* The subject of these suggestions can be considered secondary to the verbs in revealing the key in educator development. The suggestions are not simple, finite tasks. They are ongoing behaviours, reflective of a value which infuses an approach, rather than a simple action. Thus, teacher development ought to extend more deeply than a superficial teaching task such as 4SA.

8.6 Defining and understanding the term skill

A key issue in addressing the way skills are taught is inherent in what is implied and understood by the term *skill*. Skills are referred to in the context of an action which is performed, and in skills laboratories, students are *taught skills*. Using the noun form of the word, rather than the adjectival form may be the root of some of this discussion.

8.6.1 A skill requires knowledge (cognition) performance (doing), and a socio-cultural context of application

Skilled performance of clinical skills involve a level of ability and expertise which separates the practitioner from what the average person could do, within the professional clinical setting. Key learning theory in the area tends to focus on the physical aspects of performance, although Billett brings a perspective which more closely aligns with the contextual nature of clinical skills. The study design and assessment was initially influenced by the cognitive and behavioural aspects of skill learning noted in sections 2.2.1 and 2.2.2. These perspectives focussed on the action of the clinical skill enabled by neural and motor processes. The action of a skill application is relatively straightforward to judge, as it is an objectively seen behaviour (if true objectivity were even possible), from which sub-items of the skill can be identified as performed or not performed. However, this view of objective skill performance is challenged by Billett's movement towards a contextual approach to skill application. Chapter 6 reflects on these assumptions with reference to assessment assumptions behind the measurement tools and decisions or implications driven by them, and through this critique and the data gathered in Chapter 7, Billett's focus on action within the context of a social practice, in a dynamic professional environment prompts the re-evaluation of these assumptions about learning. In his marriage of two theoretical perspectives, Billett (1996, p. 3), confirms that:

> Cognitive psychology posits a pathway to expertise through the acquisition of procedural and conceptual knowledge (cognitive structures), organised and richly indexed to facilitate complex thinking activities, such as adaptability, transfer and non-routine problem-solving (Evans, 1991; Gott, 1989; Royer, 1979; Stevenson, 1986a, 1991). Increasingly, within this discipline, the nature of expertise is viewed as being domain-specific or situational (Alexander & Judy, 1988; Glaser, 1990; Perkins & Salomon, 1989; Sweller, 1989). Therefore, rather than complex performance being associated with the universal application of cognitive structures, a more specific view, involving situationally dependent understanding and procedures, is now being advanced within cognitive psychology.[whereas] A socio-cultural pathway to expertise is associated with immersion in a particular social situation over time, and acquiring not only skilful knowledge, but also the facility to engage successfully in the discourse, norms and practices of the particular community of practice (Fuhrer, 1993; Goodnow, 1990; Lave, 1990; Lave & Wenger, 1991).... Therefore, domains of knowledge are not formal fields of study, as they are often conceptualised (Alexander & Judy, 1988), but are rather a set of rule-based concepts and procedures which are patterned by the social factors within a particular community of practice. Becoming expert is thereby premised on access to the particular social practice and what that practice privileges (Goodnow, 1990; Lave, 1990)

Skill performance is tightly linked to its application. When seeking to understand skill performance, we must address procedural knowledge, conceptual knowledge, cognitive schemata and situational application in addition to the social factors such as culture and community of practice. Billett argues that knowledge and ability is dependent on a practical context, which will vary with the socio-cultural factors in the context. The action (psychomotor skill application) that a health practitioner makes, then, is a function not only of their recall of the steps of a skill, and their motor ability to physically perform it, but also the indoctrination of a professional identity, with explicit (and often also implicit) boundaries of practice, expectation and value.

If this approach to learning clinical skills was upheld at the beginning, a series of quantitative comparative trials which gathered data from the actions observed would *not* have been accepted as consistent. However, through the course of the study series, the underlying approach to measuring knowledge and understanding what is actually represented by what we see, has evolved. Further impacts of this are discussed in section 8.8.

8.6.2 Current methods of teaching are insufficient to teach tactile sensory material

A further aspect of keen relevance to understanding psychomotor skills is that presented by Howes (Howes, 2005b). His monograph approaches manual skill performance with a perspective not yet well represented in the clinical sciences, however the blending of his cultural studies research on the senses and clinical skill education approaches enlightens a key obstacle to the latter. When clinical educators approach a skill, their intent and focus is on performing an action, with varying reference to the socio-cultural context of professional practice. Typically, methods of instruction will involve recruitment of cognitive and motor strategies such as demonstration (visual input), explanation (auditory input), and practice (tactile exploration). But educators tend to start with the earlier two strategies, and likewise when practice is erroneous, correctional advice is conveyed through auditory and visual means: the student is told where their hand should be placed to achieve a better grip, or they are shown by the demonstrator. Howes comments about "Doctor P" from Oliver Sacks' "The man who mistook his wife for a hat" that "'our mental process... involve not just classifying and categorising,' of which Dr. P. was still capable, 'but continual judging and feeling also.'" (Howes, 2005a, p. 22). This conveys the warning that when science becomes too objective and computational, we risk abandoning sensual perception, and becoming no more than calculators, thus losing some of the richness of perception and experience.

Stewart argues that the distinction of the five senses as we know them are "an historical, that is human, accomplishment" (Stewart, 2005, p. 59). She goes on further to battle the potential limitations of sensual input. Instead arguing that not only may we "apprehend the world by means of our senses, but the senses themselves are shaped and modified by experience and the body bears a somatic memory of its encounters" (Stewart, 2005, p. 61). This describes the lasting impact that sensory information has on our bodies as it continues to shape further perception. Thus, the richness of intentional practice in developing skill is further supported.

Senses of sight and hearing are often portrayed as most elevated through modern history, with taste and touch ranking lowest, and smell in the middle. The value of senses which convey information from a distance, and carry the ability of contemplation and abstraction (sight and hearing) hold more socio-cultural value than those who do not (touch and taste) (Stewart, 2005). These later, corporeal senses (Classen, 2005, p. 70) are not more prominent in 4SA than 2SA. Maybe this sensual pecking order explains our preoccupation with verbal and visual instruction for motor skills?

When it comes to communicating and understanding touch, we simply cannot pin it down linguistically. How can touch and the multidirectional transfer of input (sensation) and output (motor function) in exquisite and continual balance be adequately conveyed in words? Indeed, the word "touch" has been identified as one of the longest entries in the Oxford dictionary, likely due to its elusiveness when being constrained to mere words (Mazzio, 2005, p. 86). In much the same way as students' technological use will be limited by the availability of words to enter into a search engine (Ghezzi, Chumber, & Brabazon, 2014), the educator's description of a tactile quality will be limited by the availability of words known to describe such a thing. When physical touch occurs, a plethora of information is processed, consciously and subconsciously. The toucher is engaging directly with the object, whereas the senses of distance (sight and hearing) operate through a medium, and it is this medium which is interpreted by the seer or hearer (Mazzio, 2005, p. 92). We become so focussed on the representation of a thing (the reflective light we see, or the sound waves we hear) that we forget that direct touch is medium-less, and thus the interpretation is internal (reliant on nerves, memory, cognition) rather than external. So how do educators help students understand the touch input when language is insufficient to do so? After all, are words and symbols our thoughts, or the representation of our thoughts (Sacks, 2005, p. 40)?

8.7 What does all this mean for symbiotic clinical education?

Considering aspects of cost, acquisition, retention over time and the educator perspective has allowed this study to be significantly more holistic in its approach than many other comparative studies have been able to. This is reflected and encouraged in the theoretical framework used for the study. Using Worley's model to identify the key stakeholders in clinical education, the skill teaching implications on each agent has been considered, with some reference to each of the axial relationships. This strategy provides a structure to ground findings in the authentic clinical education context that is with relevance to the range of people and groups with varying perspectives and interests in effective clinical education.

8.7.1 Clinical

For the clinical axis, cost-effective skill teaching approaches benefit the clinician-patient relationship. Firstly, by minimising the time away from patient care responsibilities due to teaching responsibilities, and secondly by allowing the student to play a more authentic role in the clinical team, through more sound understanding and retention of the clinical skills taught. Ineffective teaching, conversely, will impair the student's contribution to the team. This relationship may also be impacted by the setting of skill education activities, student confidence, and aspects of skill performance affecting patient safety.

In Chapter 7, the educators tended to express that 4SA was most appropriate for use in the skills laboratory setting rather than in the context of bedside teaching, so the impact of the skill teaching method on the clinician-student-patient relationship at the time of teaching is not prominent. It is reasonable to expect confidence to impact the development of a student-patient relationship, and impact the patient's sense of comfort in the student's ability. Participants' confidence in their ability to perform the skill correctly was not found to be affected by the skill teaching method. I expect, on the basis of expertise development requiring intentional practice competence will be more greatly impacted by time, with guided practice and constructive feedback, however this may not impact confidence.

The clinician allows an increasingly closer relationship between the student and patient as the student's clinical ability increases, however confidence may play a significant role in this. Where a student perceives his or her ability to be greater, or exudes a confidence that this is the case, this may cloud the clinician's judgement and mask possible insufficiencies (Iramaneerat & Yudkowsky, 2007).

A key part of this study series is the extension beyond performance scores to consider the patient impact. In Chapter 5 this is addressed theoretically, with a measure for expected risk on morbidity and mortality. While a difference was not identified between the two teaching methods, a significant decline was noted in the number of critical items performed correctly. This was mirrored in a significant decline of skill performance scores generally, and may reflect significant risk to the patient as skills atrophy, over as little as 6 months. Ali et al. (2002) note maximal attrition of emergency skill decay to occur over 4 years, suggesting further attrition of skills in Chapter 5 is likely.

The impact of this on the patient and the supervising clinician is of great relevance to pre-hospital care. In most Australian ambulance services, students engage in an in-service training program lasting approximately one year. During this time, competencies are measured, and then they are authorised to practise independently (albeit generally alongside a partner who may either be qualified or another student). Where skills are not maintained either in the simulated setting, or in clinical practice due to appropriate caseload, this atrophy is likely, and may not be identified until a patient requires said skills, with significant consequences. This concern is echoed by Smith and Greenwood (2012) who argue that ongoing practice is needed to maintain expertise.

Reflection is aimed at identifying gaps in practice. However, this is dependent on the insight to recognise where aspects of sub-competence exist, and this may be difficult for independent practitioners who receive little external constructive feedback on their practice. While feedback is considered imperative to education (Milan, Parish, & Reichgott, 2006; Oestergaard et al., 2012; Van de Walle, 2004), ongoing peer feedback may not be a cultural norm within many health system delivery models.

8.7.2 Institutional

In the institutional axis, the question of cost-effective skill education is directly related to the quality of health care provided by student and graduate clinicians within the health service. The training received through formal training courses, where focussed on ongoing skill maintenance, will contribute to quality of care. This contextualises the investigation into skill retention in Chapter 5, rather than solely skill acquisition which gives no insight into skill maintenance.

8.7.2.1 Time to teach

Time to teach clinical skills is an important commitment for teaching institutions, and while 4SA has received criticism for taking too long, until now it has not been measured. This also means that until now, education institutions, curriculum designers and educator development initiatives have had insufficient information to decide whether the expected or assumed benefits from 4SA are worth the additional cost.

It is a very reasonable assumption that 4SA requires additional time to teach. Indeed, Walker and Peyton expected that teaching step-by-step in the theatre will require around 30% more time than performing the surgical task without instruction. In Chapter 5, 4SA was found to require around 25% more time to teach than 2SA, which is likely much more than 30% more time than to simply perform the skill, as originally predicted (Walker & Peyton, 1998).

8.7.2.2 Cost of resources

The comparative cost of resources was not formally calculated in this study series, although it can be theoretically determined with the cost of training equipment and estimated extent of equipment re-use. For 4SA, each piece of equipment is used four times (plus once for each additional student in the group), and for 2SA it is used twice (plus once for each additional student in the group). The cost can then be calculated as:

Cost of resources for 4SA = $(\mathrm{C_{unlim}}) + \left(\left(\frac{\mathrm{C_{lim}}}{uses}\right) + \left(\mathrm{C_{sing}}\right)\right) * \left(4 + (n-1)\right)$

Cost of resources for 2SA= $(\mathrm{C_{unlim}}) + \left(\left(\frac{\mathrm{C_{lim}}}{uses}\right) + \left(\mathrm{C_{sing}}\right)\right) * \left(2 + (n-1)\right)$

Where:

C_{unlim}= cost of unlimited use equipment (items which do not wear out)
C_{lim}= cost of limited use equipment (items which can be re-used a limited number of times)
C_{sing} = cost of single use equipment (items which can only be used once)
Uses = estimated number of uses each "limited use" piece of equipment will provide
n= number of students in each teaching group

The above is based on the assumption that additional teacher demonstrations are not necessary. In the adaptive teaching context, this is a difficult assumption to make.

The difference in cost of resources and consumables is therefore potentially significant, depending on the skill. For manual defibrillation in Chapter 3, the initial cost of resources (simulated defibrillator and manikin) are significant, however there were no consumables other than relatively inexpensive gloves. For LMA insertion in Chapter 5, there would be some potential increase in consumables, with the LMA seals eventually likely to perish. These single use items withstood the study remarkably well with no such incidents occurring. For IO insertion, the manual IO devices tended to wear, warp and eventually create a slight opening between the removable stylet and inside of the metal lumen, which would occasionally become jammed with small chicken bone fragments. The chicken legs had a limited and unpredictable stamina, and as the number of insertions in a single leg increased, the rate at which the bone would snap during insertion tended to increase. The manual IO needles tended to change over every 6-10 insertions, and the chicken legs tended to last approximately 5 insertions.

With more consumables, the IO insertion session became more expensive than the other skills from an ongoing perspective although the initial layout of unlimited use equipment was lower. For this skill, therefore, 4SA would undisputedly be more expensive than 2SA from a resource perspective assuming initial unlimited use items were already acquired. It is clear then, that the type of skill must be considered when comparing the cost of resources, but when consumables are required, the cost of 4SA increases the institution's costs again.

8.7.2.3 Educator development

Time to teach impacts casual educator employment costs and potentially room booking costs, and consumable resources potentially increase the cost of teaching with 4SA more so than 2SA, but an accidental finding urges the consideration now of the potential non-financial cost to the educator. Lower teacher compliance to 4SA than 2SA in the two comparative studies prompted the study presented in Chapter 7 where it was found that educators do not agree on the value of 4SA in theory clinical teaching. Learning a skill teaching strategy tended to be well accepted by novice educators engaged in the study, and was expected to be helpful for other novice educators identified by them. The key benefit stated was the structured approach, and the need to break the skill down into smaller steps (neither of which are factors unique to 4SA). However, many educators noted 4SA was cumbersome, required intentional practice to perform well, and required a level of concentration which detracted and distracted from the content of what they were planning to teach. So while 4SA may help some (novice) educators approach a skill teaching session with greater confidence because they are *armed with a strategy* in a sense, it can be stifling and distracting for other educators.

It is not possible to survey the timetable commitments and tutor costs for all clinical skill development programs within health services, private providers and tertiary institutions, in order to calculate the precise cost of teaching using 4SA compared to a simpler, traditional 2SA. It is, however, reasonable to argue based on these data that 2SA is a much more cost-effective teaching approach for teaching institutions, without disparate skill application ability in the students.

8.7.2.4 Perceptual skill education

Howes' *Empire of the senses* contains a series of arguments from beyond the limitations of clinical education and assessment to allow consideration of manual skill application from a variety of perspectives. Stephen Gosson, in "The schoole of abuse" (1579) writes:

> The height of heaven is taken by the staffe: the bottom of the sea sounded with lead: the farthest coast discovered by the compasse: the secrets of nature searched by wit: the anatomy of man set out by experience, but the abuses of Pliaes cannot be showen, because they passe the degrees of the instrument, reach of the plummet, sight of the mind, and for try all are never brought to the touchstone (Mazzio, 2005, p. 92)[35].

This describes, among other things, the inadequacy of some instruments and measures. If we cannot sum up an action clearly using words, or describe just how much pressure is just enough, or just how

[35] Mazzio retained Gasson's original expression and spelling from the 1579 text.

delicate the tissue is, using our easiest medium (words), then the measure (words) is insufficient for the task. Educators need to be using a sufficient measure if the student is to learn to perform, rather than learn to replicate an action. The power and complexity of touch is often overlooked. Consider you are at a theatre, and the light is low. You search your bag for your mobile phone to ensure it is switched to silent, but your search is limited to the sensation provided by your fingertips. Chances are you know your phone as soon as you touch any surface of it, but how do you describe, teach, or otherwise convey that familiarity? It seems that "locating the many possibilities and powers of touch simultaneously is in and of itself a seemingly impossible task" (Mazzio, 2005, p. 92).

8.7.2.5 Assessment

The assessment tools developed in Chapter 4, and applied in Chapter 5 were further critiqued in Chapter 6. This sub-series of studies takes the focus away from 4SA briefly, and brings the thesis to consider what learning is, how it is applied in clinical education, and how it can possibly be assessed. A global-style scoring system was used in conjunction with a list of items developed from the teaching sessions in Chapter 3, and an adapted global system was used in Chapter 5 alongside the performance tools developed from a basis of authentic, experienced clinical practice and awareness in Chapter 4. These studies suggested that while manual defibrillation was performed with greater compliance to the 4SA (adapted) teaching session, the participants' overall ability was perceived as comparable between the two interventions. Likewise, the IO and LMA insertion performance videos scored comparably between the two teaching methods when using the clinically relevant skill-specific assessment tool, and the validated GRS checklists.

This may be because there is in fact no difference in the learning resultant from these two methods, or because the assessment tools which measure a student's observable action are not able to measure learning. In assessment, educators are often required to use a proxy measure for what they are actually measuring, and this is the case in these studies, fuelled by assumptions on what learning is. Learning is a change in behaviour, and can be assessed by movement upwards on Miller's pyramid (G. E. Miller, 1990). Miller sees the pinnacle not as simply doing, but it is a doing infused by the contextual practice of the developing professional. We gain glimpses of this by observing what someone does, but this is only the tip of the iceberg which is visible. Short of in-depth interviews, programmatic assessment to understand patterns of practice, and functional MRI scans to gain insight to cerebral function during skill performance in an authentic setting and other such tools, a student's observable action is the most convenient measure for a comparative trial such as this thesis contains.

Finally, skill education, practice and assessment centres could be ideally located within the health service to maximise the educational benefit to the health service, and the benefit of the clinical context to the educational organisation. Joint ownership, responsibility, resourcing and staffing may promote a symbiotic approach for both agencies to partake in the clinical student's development. This institutional partnership may generate improvements for both institutions through sharing the cost and obtaining more sustainable education strategies for both.

8.7.3 Social

This axis of the Symbiotic clinical education model is concerned with the health needs of the community, and the government's responsibility to resource the health system with the appropriate staff and other incentives to meet these needs. In Worley's model (Worley et al., 2006), the allocation of students to a community placement may be facilitated to strengthen a relationship between the government and the community. When it comes specifically to teaching clinical skills, a symbiotic model will ensure students learn and apply their skills appropriately to the community who needs their care. This may include training in the simulated skills laboratory environment, but training to apply the skill or procedure in a single way will not allow for flexibility required in most resuscitation skills due to the variability in patient needs, co-morbidities, physical space limitations and other aspects of the uncontrolled environment. While some procedures may be highly standardised, many of the resuscitation skills applied in the pre-hospital context rely on the clinician's ability to adapt to the unique needs of the context before him or her, thus a skill teaching approach ought not to aim for the single correct way to perform, but rather build competence and awareness of where flexibility may be correctly and responsibly applied.

The government response to support a symbiotic and integrated approach to learning clinical skills may therefore consider providing resources for increasingly authentic learning environments which may safely simulate some of the variety encountered in practice. Additionally, clinical placements and facilitated learning programs where students have an opportunity to practice skills initially learnt in a simulated setting, with appropriate caseload, time to reflect, and supervised support in skill development needs to be considered in effective skills education.

A potential barrier to this is obvious when we consider the relative infrequency of resuscitation skill application in the pre-hospital clinical setting. Inter-professional skill development and maintenance programs such as assisting with in-hospital airway establishment and management in both adult and paediatric departments, assisting with phlebotomy clinics, or bone marrow testing services may offer pre-hospital clinicians with valuable exposure to these skills so that routine experience with these infrequently used skills may allow a basis for expert adaptation when required in the clinical setting.

This social axis also refers to government policy, and factors influencing quality and safety. Australian Commission on Safety and Quality in Health Care (ACSQHC) advise the existence of a "Rapid Response Team" within health care facilities, with access to ALS trained clinicians (Australian Commission on Safety and Quality in Healthcare, 2010). Obtaining such a qualification to satisfy the guideline is one part of the problem, but the intention is focussed on delivery of patient care, which can only be addressed with skill *maintenance*. Education must be accountable to the intentions of such government guidelines, rather than an isolated educational exercise.

8.7.4 Personal

Having originally not intended to address this aspect of the model, the final study aimed at understanding the educator's perspective of 4SA and skill teaching addressed the personal and professional connection for the clinical student and the clinical educator.

Neither students nor teachers are standardised, programmable machines. Students learn, in part through the authentic application of clinical skills, how to *be* a health professional. Likewise, educators bring their authentic experience from the clinical setting to their teaching, and personal values and expectations which must be reconciled with their education roles. Using non-clinicians to teach integral components of clinical practice may adequately teach the "what", but not the development and reconciliation between the clinical and personal identities which the clinical student must address. In this way, the clinical and educational identities of the clinical teacher will impact the clinical and personal reconciliation for the clinical student.

For the clinical educator, their approach to teaching appeared to be impacted by their perceived role, informed by an underlying identity. Some sought to teach the clinical student the absolutely correct way to perform a skill, confident in 4SA being sufficient to do so (although Chapter 5 and global scores in Chapter 3 suggest that this is not the case). Whereas other educators grapple with the complexity of intending their student to understand the perception involved in a clinical skill, and the principles to be applied and adapted on the basis of natural and expected variability which they will encounter in their future roles. This latter intention is infused with experience and patient-centeredness, and is intensely difficult to describe or achieve within the limitations of our five senses. This much more complex, infinite game will demand an adaptive approach requiring educational and interpersonal insight and expertise, rather than a template teaching strategy. The graduation of one (the template) to the other (adaptability) is a reflection of perceived role and identity: Am I a clinician? Am I a clinician who educates? Or am I an education specialist in a clinical field?

For the clinical student, the climax of 4SA is doing the clinical skill, under the guidance of a supervisor. While some scholars have argued this step to align with the pinnacle of Miller's pyramid (performance), Miller intended that this step relates to action in practice (G. E. Miller, 1990), indeed relating more to behavioural performance, rather than action. 4SA alone, nor I would argue any other isolated skill teaching session will move a student from unconsciously competent, to knowing the skill, to knowing how, showing how and performing it in authentic daily practice. Performance as Miller intended it, is steeped in professional expectations, and applying the skill in the clinical context is part of practice. This is clearly an unreasonable expectation of a once-off teaching session, separate to guided practice, feedback, reflection and development. The student's professional identity is impacted by their application of clinical skills in the professional setting, but 4SA (taken out of the wider educational context within the original text by Peyton) stops short of that end.

8.8 The challenge to andragogical approaches in resuscitation and clinical skill education

This thesis, while initially only intending to compare the financial cost-effectiveness of two skill teaching methods, has expanded to investigate a much more holistic approach to teaching resuscitation and other clinical skills. The key lessons influence how we teach paramedics and other clinical first responders, how we understand clinical students to learn and develop expertise in new skills, how we ought to consider assessment, and a critical response to common educator development strategies.

8.8.1.1 How we ought to teach resuscitation skills: 4SA or 2SA?

These rarely employed but critically important skills may be called upon unexpectedly to be applied promptly in the pre-hospital ambulance setting. Skill atrophy of such skills is prominent, as found in Chapter 5 and the wider literature. A focus on initial skill education is not enough, with a more time-consuming and arguably more thorough 4SA making no measurable difference to the actions performed during student application than those taught with 2SA. I argue that educators must encourage students to strive for a culture of expertise development and maintenance through intentional practice, reflection, and adaption of skills in a wide setting to ensure that when the *real* assessment comes, the patient will receive the best from their clinician. This connects the student's educational activities with their future professional activities.

Educators who have been exposed to 4SA are somewhat divided over its use. Some hold fast to its subscription, and others see it as an obstacle to their purpose. If a structured approach assists the developing educator in building a series of experiences from which they can grow educational expertise, the benefit is tangible. However, the data obtained in these studies urges us to stop pretending that employing 4SA is of any measurable benefit to the student's ability to develop skilful practice, beyond a simpler education strategy.

8.8.1.2 How students learn clinical skills

Howes (2005b, p. 5) warns us that "Science cannot provide a touchstone of truth or a higher authority for cultural analyses." The application of a clinical skill does, indeed, involve a cultural analysis. It is steeped in a profession, with reference to the patient's needs, professional guidelines and expectations, and a plethora of wonderfully unpredictable factors. Sacks (2005, p. 36) speaks of the limitations of visualisation in understanding something which stimulates more senses than just our sight. The lesson here is pertinent to understanding the learning of a psychomotor skill:

> 'A blind person has a better sense of feeling, of taste, of touch,' [Lusseyran] writes, and speaks of these as 'the gifts of the blind.' And Lusseyran feels, blend into a single fundamental sense, a deep attentiveness, a slow, almost prehensile attention, a sensuous intimate being at one with the world which sight, with its quick, flicking facile quality, continually distracts us from."
>
> (Sacks, 2005, p. 36)

When we assume that a student will learn through observing a demonstration and listening to an explanation, we rob the sense most prominently recruited in the application of skills: touch and sensory feedback. Observation of and receiving auditory input for skill instruction have been shown to stimulate neurons required for motor function, which could be seen as the brain's natural inclination to recruit motor neurons in learning a clinical skills. Separating the observed and the heard may be an artificial approach based on the Greek dualism of body and mind, which underestimates the connectedness of one to the other. This poses the question: do we need to get over our reliance on visual and auditory aides in teaching, and just get students experiencing, sensing, feeling the pressure, responding to the weight?

8.8.1.3 Skill assessment

Oliver Sacks (2005, p. 36) prompts us to consider the limitations of observation in understanding. When we assess what we see, rather than what the *seen thing* represents, we are distracted from what is truly before us. The notion of programmatic assessment has been a part of medical education for some time, however when it comes to short courses such as ALS, and PALS, it is not possible. Resuscitation skills taught in other contexts, such as paramedic and other health professional courses, the data in this study series argues that skill education ought to take place early in the program, and be re-applied to increasingly complex cases with increasing anatomical and physiological depth, with regular lower stakes assessments which reflect a wide context of skill application, rather than a limited number of assessments close to the time of initial training. A strategy like this will encourage practice and expertise development and maintenance, with a movement away from a once-off competency-based assessment.

8.8.1.4 Educator development

Where educators are taught a strategy like 4SA, expertise for the able educator may be stifled, while the novice is given a helpful structure. Clearly for the novice this has a benefit. However, if we are to learn from the marked successes of the Finnish education reform, by investing in educators through formal training and qualification, the specialty of clinical education will be recognised for what it should be, rather than something which anyone can do so long as they have clinical ability. We need to invest in our educators' expertise, andragogical principles (not just template strategies), and in doing so allow them to focus not on the standardised tests which are otherwise a marker of success, but rather to focus on the development of tomorrow's clinicians. This mindset will shift an educator from doing the teaching tasks according to the objectives, to considering the patients who their students will treat, and working with their future colleagues for the good of those patients.

8.8.1.5 Patient care and educator identity

Where the patient is the clinical educator's focus, his/her identity will be split. The complexity that ensues ought not be fought, but instead embraced so that the infinite game of helping students understand the principles to apply them in situations beyond which we test and practice in can be identified and played.

8.9 4SA as a skill teaching approach: a summary

The comparison of two skill teaching approaches indicates a series of "what" focussed questions. What is the difference? Is there a difference? What is the measurable impact? These questions were driven by a framework founded in the practical task of teaching tomorrow's clinicians with acknowledgement to the connectedness this task has to the wider health service, community, and government. The expansion of the initial questions warranted a mixed methods approach which allowed for a more critical consideration of the literature surrounding assessment and education. While 4SA may be dogmatically enforced by its subscribers, this may promote the same stifling of educational expertise which is common in strategies prescribed by the GERM. And at the same time, for novice educators 4SA may provide a structure and confidence upon which to build experience and confidence as a clinical educator. It appears there is a time and a place.

The impact of identity in resuscitation skill education is largely under explored. The possible tension between personal values and professional expectations was initially not expected to feature prominently, however this feature may explain different educators' drives to teach in particular ways, depending on their perceived task, role or "self". This tension with identity, habit, and perceived value in certain teaching styles may war with competing cognitive load: patient responsibility, time pressures, resources, and teaching prescription, which may detract from an educator's ability or willingness to teach with a particular teaching method. A shift in approach to teaching technical skills to consider corporeal aspects of learning: perception, an "all body" knowing, and synthesis of such an approach to recall of rarely used skills during critical medical events calls for the reconsideration of teaching philosophy surrounding clinical skill education.

9 CONCLUSION

This thesis has explored resuscitation and clinical skill education as a key part of health professions education. An exploration of the literature revealed many scholars' assumptions that 4SA is a more effective skill teaching strategy compared to a traditional approach (2SA). The arguments behind this are compelling, until a more critical review of the literature is performed. Skilled practice demands expertise, however in clinical education it seems to be typically understood as the performance of a series of actions or movements relating to patient assessment or treatment. This understanding is imperative to the educator's perceived role: Is their role to teach a student to perform a skill, or to perform the task skilfully? This will infuse not only the teaching session, but also the assessment tools and strategies used to examine whether or to what extent a student or clinician is satisfactorily skilled.

9.1 Key findings

The variability in the implied meaning of the term *skill* in different approaches to learning theory is evident. From the cognitive perspective, skills are seen s as something which are learnt, organised within either new or pre-existing schemata, recalled and remembered. An educator with this view of learning will limit new information to no more than seven items (G. A. Miller, 1956), and consider the impact of cognitive load on learning. In this approach, the difficulty of the task itself (intrinsic cognitive load) is finite, but by presenting the knowledge in a particular way (extraneous load), the student's ability to organise the new knowledge for convenient retrieval (germane load) may be impacted (Sweller et al., 1998).

The studies within this thesis did not find superior skill acquisition or retention when students were taught with 4SA (or $4SA_m$) compared to 2SA. The first comparative study (Chapter 3) identified greater post-instruction manual defibrillation scores for $4SA_m$ compared to 2SA, however this was not the case when global scoring was applied, or when a repeated measures statistical design was included to account for baseline performance scores. The skill-specific checklist in this case was likely more a reflection of skill reproduction, as this assessment tool was devised from the teaching session rather than from everyday clinical practice expectations. The loss of significance of this difference when the initial skill performance scores were considered indicates that any difference is subtle at best. Further investigation into this *possible* superiority of $4SA_m$ may be of interest to sectors and professions which require skill reproduction of a taught approach, rather than global and adaptable skill performance.

Two different scoring strategies were employed for the second comparative trial (Chapter 5). In order to ensure a marking scale which would more closely reflect clinical practice rather than skill replication, a panel of clinical experts was invited to participate in a study to determine what aspects are important in the pre-hospital setting during insertion of an IO or LMA. These checklists were then applied to the second trial, alongside previously validated global rating scales. This study prompted much critical reflection of assessment methods used in clinical skill education. The initial search for *just the right* assessment tool led to the conclusion that such a tool does not exist. Any assessment tool when used in isolation from other judgements has inherent limitations, and the question becomes less about choosing the correct method, and more about understanding what the purpose of the assessment is, and therefore what assessment schedule will provide the most valid, useable and defensible data for such a purpose.

The study design revealed a measurement of action as a proxy to assess learning, which is further critiqued in Chapter 6. However, the pragmatic decision was made to apply the assessment tools already developed to compare skill acquisition and retention in this way, with the philosophical critique of that series of assumptions following.

Skill acquisition and retention were found to be unaffected by the teaching method in the trial described in Chapter 5, using both global and skill-specific checklists based on clinical practice. 4SA was found to require 25 to 35% more time to teach. Using the data gathered in this study, the validity of the assessment tools developed for this purpose is critiqued in Chapter 6 using psychometric arguments within Kane's argument and plausibility model of validity.

There were two unanticipated findings from these two trials. Firstly, participant attrition from enrolment to completion was 48 to 38 in the first trial (only 28 of which were included), and 52 to 26 for the second trial. This potentially overlooked detail becomes key during planning phases of further studies. If data becomes available for only around 50% of registrants, either due to technological limitations, poor attendance or study protocol misapplication, more studies may be insufficiently powered, and this ought to be addressed by maximising initial registration wherever possible. Secondly, educator compliance to 2SA was greater than 4SA for both trials.

This difference in trial protocol application between the two teaching methods fuelled the final study in this series. The educator perspective was investigated in Chapter 7 through a multifaceted qualitative investigation, which allowed educators to grapple with the use of 4SA in clinical skill education. 4SA was perceived to advocate the "absolutely correct" way to perform a clinical skill, whereas the application of clinical skills was agreed to be a complex pursuit, with the need to ensure teaching which would allow adaptable practice. The structure inherent in 4SA was perceived to be of benefit to novice educators who were developing their practice, however the masterful educator may be hindered by this potentially rigid practice, as he or she seeks to perform a task which is infinite in nature. The identity development for the student affects the skill teaching session, as does the perceived task, role and identity of the educator.

In the development of an educator's masterful craft, the perception of task and role are key. Expertise and experience are not always synonymous, therefore a focus needs to be places on clinicians who are expert in their clinical field, and who may have developed habitual teaching through years of experience, to develop andragogically sound expertise in teaching. Just as special training programs are expected for clinicians who seek to work in intensive care, paediatrics or cardiology for example, the recognition of education as a specialty area in clinical practice is essential to the ongoing informed development of the educator's craft. Like any other specialty, this professional stream will require specialty training, qualification, assessment and rigorous expertise development. I argue that the dual identity of the educator as both teacher for the student and clinician for the patient will more greatly impact a skill teaching philosophy which is appropriate to the unique complexities of the immediate context, than the application of a single approach for all situations.

4SA is widely accepted as a teaching method which puts the individual components of the skill into the wider procedural, but not clinical, context. However, this first stage, even if it occurs in the clinical environment, in itself does not satisfy an authentic, situated learning context (Billett, 2001, 2008). Situated learning occurs within a setting of professional expectations, workplace culture, and clinical practice. This practice is a culmination of being a health professional within a wider team, and is not fully appreciated by conducting the teaching session within the clinical environment rather than the simulation laboratory. The peak of Miller's pyramid of assessment, similarly, is steeped in this contextual practice, as the student learns to act within their professional context, and not in isolated performance (G. E. Miller, 1990). Harden's spiral curriculum (Harden, 1999) further urges clinical educators not to remove procedural skill education from the practice context of knowledge and attitude. Skill development feeds and is fed by wider aspects of clinical practice as it promotes problem solving, application, adaptability, complexity, and integration of different types of knowledge. Teaching how to perform a skill as a part of clinical practice requires an education approach which is not necessary for teaching a student how to use a piece of medical equipment in a particular way. The former demands a social learning theory, whereas the latter will be satisfied by a cognitive or behavioural approach where learning is based on recall and motor output.

9.2 Doing or who-ing

In investigating the comparative cost-effectiveness of 2SA and 4SA, this thesis has become more of a juxtaposition and comparison of the quantified description of complex, indescribable phenomena. The 4SA is one possible strategy to teach clinical skills, yet teaching clinical skill education is so much more than this one strategy. The bottom-up processing used in the determination of a score from a skill-specific checklist is set alongside the expert's top-down global judgement of a clinical skill ability. A numerical score on an ordinal, interval or ratio scale is just one measurement of a student's performance, but performance of a skill can never be fully described by a single number. An educator's intention to teach a student to perform a skill in “the absolutely correct way to deliver the procedure" (interviewee 2, quote index 11), is later compared to the variability and complexity in clinical practice, and this may impact on the skill application. This challenges the notion that in truly competent, responsive practice there is no one correct way that a skill should always be performed, but instead adaptability is the key. Some educators saw their task as teaching the student to correctly perform a skill (to *do*), and others saw their task as helping their students apply appropriate practice in a clinical setting (to *be*). This leads into the climax of this thesis: with the separation of the tools and strategies used in the act of *doing* teaching, and the art of *being* an educator.

In each of these examples, neither is wrong, but one is defined and limited by its description or its example, and the other is broad and complex. One is a finite representation of an infinite concept, quality or thing. A more limited scope or template may be very appropriate in some settings, especially where the education focus is consistency of practice, or a protocol-fuelled approach. In settings such as accreditation courses where limited time is available for student learning and professional identity development, or where clinician-educators are used to teach, a template-driven teaching may be acceptable. But I would argue that an educator-clinician (note the identity transition from a clinician who is engaged in teaching), who is engaged in teaching students how to be clinical professionals, their expert craft must burst beyond the limits potentially imposed by such teaching protocols. Rather than restricting educators' practice, educator development programs could instead help educators navigate the complexities of their role.

The 4SA offers some supportive structure to novice educators, and it has a place in clinical education, arguably for short accreditation courses run by clinicians who educate, but who do not have a significant education role through which to develop their adaptive education practice. Novice educators may also choose to lean on the 4SA while their confidence to adapt their practice develops. However, expert educators ought not to be restricted by its bounds.

Expert educator-clinicians will grapple with the tensions between their dual identities: the educator's role and the clinician's role. This requires consideration for a myriad of factors, including patient factors, student factors, clinical outcomes and educational benchmarks. An expert educator is more than an educator who has teaching or clinical experience, but is one whose practice is informed and challenged by intentionally applied andragogy. This is made more possible when the educator is *called* to education, rather than expected to teach because of their clinical competence.

9.3 Implications for the professions

9.3.1 For clinical education

Just as medical doctors may choose to specialise in cardiology, and nurses may specialise in burns therapy, and physiotherapists may specialise in paediatric rehabilitation, what if clinical educators were *interested* in, *qualified* in, and needed to demonstrate *expertise* in the specialty of education within clinical practice? What would this specialty look like? Instead of having learnt a strategy to employ to teach clinical skills, they would understand, through critical reflection of their own practice in light of rigorous teaching and learning theory, the principles of educational practice. Developing this specialty is key to expanding our practice beyond the limitations of tools and strategies which may be helpful for the novice educator, or the clinician seeking to educate, but should be a spring board for the educator's specialist expertise development, rather than the goal. This development will not occur as long as educators are taught tricks or strategies which, if performed correctly, imply correct educational practice. Education is complex, just as clinical practice is. Balancing these two roles demands both scholarly and interpersonal expertise.

9.3.2 For clinical *skills* education

Teaching clinical skills is complex. There is a lot to process. When discussing an educator's desire to help the student understand complex perception, responsiveness and the necessary adaptability to the variability of practice that an obvious struggle for words was evident as the facilitator posed, almost rhetorically "but how can you teach that?" A participant in the focus group replied "how do you explain that, yeah?" (Male participant 2). Howes (2005b) approaches the limitations of our words in explaining something so much more dimensional than words can convey. There are many things we simply cannot define or explain, they must be experienced. Through experience, the learner understands what that *give* feels like as they are inserting an intraosseous needle; they understand how an intravenous cannula feels as they advance it into the patient's vein; and through experience they begin to learn how to manoeuvre a cool, lifeless body into a better position for effective and accessible resuscitation. Words simply cannot convey the personal, bodily sensory understanding to such stimuli. Additionally, words carry connotations which vary for different people, yet as an educator we expect that our words are understood the same by our listener as they are for us. In this regard, both 2SA and 4SA are equally limited.

The 4SA is employed internationally, for standardised courses such as ALS, PALS and EMST courses. The factors influencing the Global Education Reform Movement (GERM) are evident in the use of a standardised teaching procedure to (assumedly) obtain a standardised level of competency following training. Such as assumption only stands when each student is standard, which is well known to be flawed. While GERM may pose a risk to inspiring learning in a school setting, with noted benefits in the Finnish context where educators are specialised, trusted, and fairly autonomous (Sahlberg, 2011a), in a setting where educators are clinical specialists but may not have yet developed educational specialty, this approach to teaching may be helpful in providing a support for the clinical educator. However, dogmatic subscription to 4SA as the most effective way to teach is not supported by the studies in this thesis.

9.3.3 For clinical skills assessment

Two skill-specific checklists were developed in this thesis, and they were found to be consistent with global rating scores noted by expert assessors. These checklists may be of great benefit for guiding student practice and education (Saewert, 2013), but will also be of particular use in the pre-hospital resuscitation assessment setting. The tools have undergone rigorous validation for use in such a setting, although transference to other settings with different priorities was not investigated in this thesis. In institutional and health settings which do not have access to expert assessors for new trainees, these tools may guide clinically credible application and assessment of the skills, for the emergency environment. Additionally, the process followed to develop and critique the assessment tools, as presented in this thesis, could serve as a means to develop and critique further practice and assessment tools for the pre-hospital setting. Such tools provide a bottom-up assessment approach, and are particularly helpful in the absence of experienced assessors who have sufficient expertise in top-down (global judgement) clinical assessments.

The strategy for identifying a credible pass mark for each of the two checklists was determined as 10/16 (62.5%) for IO insertion, and 13/22 (59%) for LMA insertion. These are both above the often arbitrarily applied 50% pass mark for may assessments, and may prompt assessment coordinators to consider more representative standards to ensure those who pass the assessment actually demonstrate *competence* rather than a *minimum score*.

9.4 Future research

There is still so much to learn about effective skill education. Currently, it is clear that many clinical educators are seeking to do the impossible: Teach a responsive, adaptive, critical approach to a complex clinical situation, when the ability to do so relies on a whole body knowledge, rather than one which is limited by our ability to describe using linguistic, visual, or cognitive appeals. After all, anyone can be taught to pull, how do you teach a novice practitioner *how much pressure is just enough* for the situation at hand?

I believe that deeper understanding of clinical skill education relies upon recognition of the limits of our words and demonstrations (visual and linguistic communication). Situated learning will allow students to navigate the professional requirements of their skill development and application, and practice will allow their perception to learn to collaborate what feels *normal*, and where adaptability is required. However, further understanding is imperative on the identity development of the clinical student, and how this impacts their professional approach to the motor tasks they perform. Who they perceive themselves to be may be key to understanding what they do, because it is driven by a *why*. Maybe, we do what we do *because* of who we perceive ourselves to be.

A similar notion emerged from the final project presented in Chapter 7. The educator applied their teaching choices based on their perceived roles and tasks. Where they wanted to teach the single correct way to perform a skill, a template was often accepted as suitable. But where the clinical identity (with complex patient considerations) and educational identity (with acknowledgement of examinations, standards, and accreditation) *merged,* the approach was far more adaptive. Understanding the unique identity of the educator-clinician, whose two distinct identities are fused in this marriage of two very different worlds, will offer great insight into skill teaching approaches which surpass a single strategy.

9.5 Closing remarks

Beginning with a very simple question, "is 4SA more cost-effective than 2SA?", this program of study grew deeper and wider in the pursuit of a sound, reliable and valid answer. 4SA was not found to be more effective than 2SA in skills acquisition and retention, and the additional educator and facility booking time resulted in greater anticipated expense for 4SA. From an educator perspective, while 4SA is of some benefit for the novice educator, it may be restrictive for the master educator's practice, demanding greater cognitive capacity and potentially distracting from the content being delivered.

These studies have demonstrated that the 4SA method is neither more cost-effective than a more traditional, two-stage approach, nor symbiotic with the lived reality of a clinical educator. The retention of a requirement to use it by certain training courses and instruction manuals can no longer be seen to be evidence based, and may actually be counter-productive. This study has contributed to the body of knowledge in the craft of clinical education, and I trust its application will lead to benefits for all areas of the symbiotic framework – to students, teachers, educational institutions, health services, communities, governments, the profession, and the personhood of the educator. We are, after all, not human 'doings', but rather human 'beings'. Even when teaching and learning the most critical of skills, we are not merely 'do-ing', but rather 'who-ing'. We are moulding the identity of the student as much as their skill level, and ideally allowing in our pedagogical approach to be influenced by who we are as experienced or novice clinicians as we undertake this life saving teaching on behalf of the society we serve.

In this thesis, not only have I addressed crucial questions around cost-effective resuscitation education, but this has been achieved with reference to an established theoretical framework embedded within the practical tensions of clinical education, as a process which occurs within a network of needs which are sometimes competing, and other times complementary. Symbiotic clinical education allowed the question of cost effectiveness to evolve beyond a quantified measurement of scores and time measurements. Rather, an understanding of the subjective educator's experience, value statements surrounding assessment strategies, and challenging current visual and language-confined approaches to teaching part of *being* a health professional rather than the *doing* of a health professional, is also presented.

10 APPENDICES

10.1 Literature Chapter appendices

10.1.1 Comparative trial search strategy

Table 45: Database Search Strategy and Terms

	Search Terms ^	OVID*	Cochrane	Informit **	CINAHL	Pubmed	Science Direct	SCOPUS
1	"four stage" OR "four-stage" OR "four step" OR "four-step" OR "4 stage" OR "4-stage" OR "4 step" OR "4-step"	16628	475	0	460	3648	87,464	45 599
2	"Clinical education" OR "medical education" OR "clinical skill" OR "medical skill" OR manual skill"	108836	2757	0	7772	141022	138,187	316 650
3	1 AND 2	313	12	0	3	43	959	367
4	Manually identified*** as meeting inclusion/exclusion criteria according to title and/or abstract (including duplicates from other databases)	0	3	0	0	2	3	3

* Ovid search included the following databases: OVID Medline, Biological Abstracts and Journals@OVID,
** Informit search included the following databases: Australasian Medical Index (AMI), A+ Education Plus Text (AEIPT), Australian Public Affairs full text (APAFT)
^All search terms applied to all fields (eg *term*.mp)

10.2 Defibrillation trial (Chapter 3) appendices

10.2.1 Initial data collection upon registration

First Year Paramedic Skill Study

Participant Information Sheet

Student number:	
FAN (Flinders Authentication Name):	
First Name:	
Surname:	
My Primary Language is:	English or Other (Specify) :
Gender:	Male/Female
Date of Birth:	
I have been taught manual defib before:	Yes/ No
I am enrolled in (Circle all that apply) :	PARA1000 PARA2002

10.2.2 Defibrillation teaching guide

This guide is not intended to be recited word-for-word, but is intended to guide the content of the teaching session as this practice may differ to familiar on-road practice of the skill.

10.2.2.1 The skill

PPE

Approach the patient with the iSim

The CPR assistant is doing CPR

Tell them to continue with 30:2 CPR while you place the pads on the patient's chest and turn on the monitor.

Pad placement

Charge the machine (the CPR person is still doing CPR) to 200 Joules.

State when the defib is charging

Before **shocking**:

Make sure the oxygen is at least 1m away from the pads

Confirm a shockable rhythm

Identify the rhythm (fine VF, Coarse VF or VT)

Ask the CPR person to stand back for defib

Visualise and verbalise clear "bottom" (legs), "middle" (torso and abdomen) and "top" (neck and head) of patient are clear.

Make sure you can see your CPR person's hands (eye contact).

"Are you safe?" ("I am safe")

Tell them when you're delivering the shock

Instruct your person to start **CPR** again

10.2.2.2 Student briefing

Don't worry about DRABCDE. We'll assume that has been done. This session is purely on defibrillating safely in manual mode.

10.2.2.3 CPR assistant briefing

(For information only- they will be briefed outside the classroom)

You will do safe BLS level CPR

Keep doing CPR until you're told to stop (don't automatically stop for rhythm analysis, defib)

Stop CPR until you're told to start again (eg after defib, the person needs to tell you to start compressions again)

Don't offer "I'm clear", or "I'm safe" etc unless the person asks this of you.

Don't hold your hands up unless this is asked of you

Same with moving oxygen away

Don't help with pad placement, or anything which requires initiative. Nothing at all. If you see something done wrong, don't pull them up on it or the video will be void.

10.2.2.4 Intervention group (4SA) teaching

10.2.2.4.1 Step 1: Demonstrate the skill <u>without explanation</u>:

Hello, I have a defib

You keep doing 30:2 CPR while I put the pads on

I'm going to charge the defib to 200J, continue with CPR

(mutter) Move oxygen away

(mutter) The machine is charged

Confirming that it's (fine VF, Coarse CF, VT) which is a shockable rhythm

Bottom of the patient is clear

Middle is clear

Top is clear

(to CPR person) stop compressions- are you safe?

(give eye contact, hands in air, get "I am safe" response from CPR person)

Shock is delivered

Continue with CPR

10.2.2.4.2 Step 2: Demonstrate the skill with explanation at each step:

"that's a lot to take in, so I'm now going to walk you through the skill slowly, and explain each step as we go. Stop me at any point if you have any questions, but we'll also ask questions at the end."

So I'm going to introduce myself to the CPR person ("Hello, I have a defib")

We want them to continue with CPR because if the person is in cardiac arrest and their heart is not pumping properly, their brain suffers more damage if we increase the time without CPR, so I'm going to tell the CPR person to continue at a rate of 30:2 as per the ARC guidelines.

While they continue with compressions, I'm going to place the pads on. They have pictures to show where they go, if you forget.

I'll tell them when I'm charging, and to continue with compressions while I charge the machine. 200J is the level we defibrillate at with a biphasic monitor, according to ARC guidelines, so just check it's on the right number of joules. ("I'm going to charge the defib to 200J, continue with CPR).

We move the oxygen away, just for safety. It's really unlikely, but if there is any sort of spark with the defibrillator, we want this fuel source to be well clear of us and the patient. About 1 metre is all we need.

A beeping sound indicates that the machine is charged.

At that point, we need compressions to stop so we can analyse the rhythm properly

If it's flatline (asystole), it's not shockable, despite what House, ER, All Saints, home and away (etc) tell you! The shockable rhythms are VF (ventricular fibrillation) and VT (Ventricular tachycardia).

(Pictures of Fine VF, Coarse VF and VT- don't worry about going into the pathology of these, it's just a skills session. Just help them to recognise them.)

So we'll confirm that it's (fine VF, Coarse VF, VT) which is a shockable rhythm.

We need to make sure everyone is clear of the patient, so we sweep the whole patient visually and verbally:

Bottom of the patient is clear

Middle is clear

Top is clear once the CPR person stands back.

"stand back for defib" (or something to that effect)

I want to make eye contact with them

I want to see their hands so I know they're not touching the patient, and

I want to hear the words "I'm safe"

Deliver the shock, and then start CPR again straight away.

Any questions?

10.2.2.4.3 Step 3: Deconstruction

Now, I'm going to select one of you to tell me how to safely defibrillate in manual mode.

get one student to talk you through the entire procedure, and if they skip a step, get other students to contribute.

You can ask clarifying questions to make sure they understand.

Correct where necessary.

Any questions?

10.2.2.4.4 Step 4: Students perform the skill

Now one at a time each person will practice the skill.

Any questions?

10.2.2.5 Control group

10.2.2.5.1 Step 1: Demonstrate the skill <u>with explanation at each step:</u>

(exactly as per Stage 2 for intervention group)

10.2.2.5.2 Step 2: Students perform the skill

(exactly as per Stage 4 for intervention group)

10.2.3 Facilitator instructions

Facilitators who managed the skill performance sessions were asked to:

Ask the student "What is your name and student number?" (student to recite this. It will be cut from the videos- it just helps us line the students up correctly as we code the videos)

Instruct the student: "You can see the following adult patient is under CPR. Please demonstrate manual defibrillation of the patient." (CPR person starts 1-person CPR with OPA insitu and using a BVM)

You will state "end of skill" once the student has delivered 2 shocks, or at 5 minutes (whichever is soonest). (This was later updated to: You will state "end of skill" once the student has delivered a shock, after you have given the student an opportunity to initiate recommencement of CPR , or at 5 minutes (whichever is soonest).

Keep the manikin in coarse VF.

10.2.4 Skill assistant instructions

You are asked to perform adequate CPR at a ratio of 30:2, with an approximate compression rate of 100/minute.

Questions you can answer:

- Confirming VF? Confirming VT? (you may answer with the correct answer if the student gets it incorrect)
- Clear?

Questions you should answer “I don’t know” to:

- What rhythm is that?
- What do I do now?

So essentially, you’re fabulous at doing CPR but you won’t provide any direction. You won’t reinforce errors (for example don’t confirm an incorrect rhythm), or bring any “tricks” to the scenario, but you won’t pull them up on any errors other than rhythm checks. They don’t need to get a history or anything like that, so if they ask questions like that, just make up something really simple, like “I’m not sure, I just found them like this”.

10.2.5 Schedule

Monday 14 April 2014:

12:00	Access teaching room Collect 3 manikins, 3 I-sims, 3 OPAs and 3 BVMs Gloves, safety glasses, blankets to kneel on Collect ipads from nursing
12:30	Teacher arrives Run through 4-step Run through the skill
12:45	Practice teaching session with educator (Amy as ipad facilitator)
1:00	Facilitators arrive at registration room Sign confidentiality forms Run through roles, handouts/cheats I-sim training Setup rooms Door signs
1:30	Setup rego for participants Setup all ipads
1:40	Group 1 students arrive to register

	Assessment rooms	Instruction room	Registration room
1:40	Pre-teaching performance 1	Break	
2:00	Break	Instruction session 1	Group 2 arrives
2:20	Pre-teaching performance 2	Instruction session 1	
2:40	Post-teaching performance 1	Instruction session 2	
3:00	Break	Instruction session 2	
3:20	Post-teaching performance 2	Break	
	Pack all gear into Amy's office for tomorrow		

Tuesday 15 April 2014:

Time	Assessment rooms	Instruction room	Registration room
8:40			Group 3 arrives
9:00	Pre-teaching performance 3	Break	
9:20	Break	Instruction session 3	Group 4 arrives
9:40	Pre-teaching performance 4	Instruction session 3	
10:00	Post-teaching performance 3	Instruction session 4	Group 5 arrives
10:20	Pre-teaching performance 5	Instruction session 4	
10:40	Post-teaching performance 4	Instruction session 5	
11:00	Break	Instruction session 5	
11:20	Post-teaching performance 5	Break	Group 6 arrives
11:40	Pre-teaching performance 6	Break	
12:00	Break	Instruction session 6	Group 7 arrives
12:20	Pre-teaching performance 7	Instruction session 6	
12:40	Post-teaching performance 6	Instruction session 7	Group 8 arrives
1:00	Pre-teaching performance 8	Instruction session 7	
1:20	Post-teaching performance 7	Instruction session 8	
1:40	Break	Instruction session 8	
2:00	Post-teaching performance 8	Break	

10.3 Assessment tool development (Chapter 4) appendices

10.3.1 Global score mark sheets

This scoresheet was developed from GRITS presented by Doyle et al. (Doyle et al., 2007). The complete checklist is presented below, however use of assistants, communication skills, depth perception and Bimanual dexterity were omitted as there were no assistants available to the students, "the patient" in both skills was assumed to be under cardio-pulmonary resuscitation or otherwise unconscious, and neither procedures were laparoscopic in nature.

Table 46: Doyle et al's Global Rating Index for the Evaluation of Technical Skills (GRITS) Tool

Item	Score				
Respect for tissue	1 Frequent unnecessary force on tissues or caused damage by inappropriate use of instruments	2	3 Careful handling of tissue but occasionally caused inadvertent damage	4	5 Consistently handled tissue appropriately with minimal damage to tissues
Time and motion	1 Many unnecessary moves	2	3 Efficient time/motion but some unnecessary moves	4	5 Clear economy of movement. Maximum efficiency
Instrument handling/ knowledge	1 Tentative/awkward moves or inappropriate use	2	3 Competent use of instruments, occasionally awkward	4	5 Fluid moves with instruments. No awkwardness
Flow of operation	1 Frequently stopped, seemed unsure of next move	2	3 Some forward planning, reasonable progression	4	5 Obviously planned course, effortless flow
Knowledge of specific procedure	1 Deficient knowledge. Required specific instruction at most steps	2	3 Knew all important steps of operation	4	5 Demonstrated familiarity with all steps of operation
Use of Assistants (if applicable)	1 Consistently placed assistants poorly or failed to use	2	3 Appropriate use of assistants most of the time	4	5 Strategically used assistants to best advantage at all times
Communication skills	1 Frequent problems working with team or fails to communicate	2	3 Appropriate communication with team most of the time	4	5 Co-ordinates surgical team in a superior manner
Depth perception (laparoscopic procedures only)	1 Constantly overshoots, swings wide, slow correction	2	3 Some overshooting but quick to correct	4	5 Accurately directs instruments in correct plane
Bimanual dexterity (laparoscopic procedures only)	1 Uses only one hand, poor coordination between hands	2	3 Uses both hands but does not optimise their interaction	4	5 Expertly uses both hands to provide optimal exposure

Adapted from "A universal global rating scale for the evaluation of technical skills in the operating room" (2007) by Doyle, J. D., et al., *The American journal of surgery,* **193**(5) pp. 551-555

Table 47: Kneebone et al's Integrated Procedural Performance Instrument (IPPI)

	Below expectations		Borderline	**Meets expectations**	Above expectations		Unable to comment
Introduction/establish rapport	1	2	3	**4**	5	6	7
Explanation of intervention including patient's consent to proceed	1	2	3	**4**	5	6	7
Assessment of patient's needs before procedure	1	2	3	**4**	5	6	7
Technical performance of procedure	1	2	3	**4**	5	6	7
Maintenance of asepsis	1	2	3	**4**	5	6	7
Awareness of patient's needs during procedure	1	2	3	**4**	5	6	7
Closure of the procedure including explanation of follow-up care	1	2	3	**4**	5	6	7
Clinical safety	1	2	3	**4**	5	6	7
Professionalism	1	2	3	**4**	5	6	7
Overall ability to perform the procedure (including technical and professional skills)	1	2	3	**4**	5	6	7

Note: A score of 4 in each item is considered sufficient.

Adapted from "An Integrated Procedural Performance Instrument (IPPI) for learning and assessing procedural skill" (2008) by Kneebone, R., et al. *The clinical teacher* **5**(1) pp. 45-48.

10.4 IO and LMA Retention trial (Chapter 5) appendices

10.4.1 Facilitator instructions

Preparation: Check the mobile camera is on and at a good angle for viewing the table where the skills will be performed. If camera has just been turned on, clap loudly (this helps us synchronise the video at the end) Ensure the IO leg has glad wrap with anchor points marked Welcome in the next student (if none are waiting check in the rego room)
Introduction: Greet the student warmly, and acknowledge any nerves. Reassure them as necessary that this is not a test of them so much as a test on us. **What is your name and student number?** (student to state it aloud) **Have you completed the training session on Intraosseous insertion and LMA insertion today?** (Student answers yes or no)
1. LMA **The following patient is unconscious, and has been tolerating an OPA. Their airway is clear. Please insert an LMA into the patient. Talk us through your thoughts and process as you're doing it.** (End skill after 4 minutes unless the participant has already finished or obviously given up)
2. Intraosseous insertion Part A: placement **You are asked to gain IntraOsseous Access. On my leg, show me where you would insert an IO needle. As you do it, talk me through why you are selecting that particular placement, and what you are trying to avoid.** Part B: needle insertion **With the equipment provided, please insert the IO into the pig trotter provided. Again, talk us through your thoughts and rationale as you're doing it.** (End skill after 4 minutes unless the participant has already finished or obviously given up)
End of skill session: If the student hasn't yet had their training session, send them to wait outside N251B (teaching room). If they have, they are welcome to come back to the rego room to hang out, or they may leave.
Resetting equipment: Remove the IO from the chicken leg, and verbally comment whether it is stable in the bone or not. Stop the video recording Other notes: Try not to mention the cameras during the session. The participants know they're being recorded, but being reminded of it can cause anxiety.

If you are asked questions about the skill, just warmly tell the student that you're can't assist with that, and just do what they think it best. It is normal for people in the participants' situation to be nervous, so do your best to be supportive without giving anything away. Please try to keep the script word for word. Amy: (mobile number)

10.4.2 Registration cheat card

On arrival: Greet them warmly and reassure them that they're in the right place for the Skill study Confirm their name Mark their name off the list Give them a consent and participant registration form to complete Confirm their Teaching session time, and send them to wait outside the store room for a skill performance room to become available. If there's already a student waiting outside the performance rooms, they can hang around in the rego room to wait. Tell them that they will be asked their name and student number when they do their skill performance, but that this information will be edited out of the final videos (it's just to help us process the data).
The format of today (FYI) The students will have their initial skill performance session in either of the two performance rooms (whichever is available). After they have performed the skills, they will head to wait outside the training room for their teaching session. Then, they will return to a performance room and perform the skills again. After that, they may return to the rego room to hang out, chat, debrief Amy: (mobile number)

10.4.3 Allocation cards

Card 1: **First teach IO with 4SA** Demonstrate in real time Demonstrate again slowly, with explanation Get a student from the group to talk you through it as you do it Each student performs the skill under your supervision, with your feedback **Second, teach LMA with 2SA** Demonstrate LMA insertion slowly, with explanation Each student performs the skill under your supervision, with your feedback	Card 2: **First teach LMA with 4SA** Demonstrate in real time Demonstrate again slowly, with explanation Get a student from the group to talk you through it as you do it Each student performs the skill under your supervision, with your feedback **Second, teach IO with 2SA** Demonstrate IO insertion slowly, with explanation Each student performs the skill under your supervision, with your feedback
Card 3: **First teach LMA with 2SA** Demonstrate LMA insertion slowly, with explanation Each student performs the skill under your supervision, with your feedback **Second, teach IO with 4SA** Demonstrate in real time Demonstrate again slowly, with explanation Get a student from the group to talk you through it as you do it Each student performs the skill under your supervision, with your feedback	Card 4: **First teach IO with 2SA** Demonstrate IO insertion slowly, with explanation Each student performs the skill under your supervision, with your feedback **Second, teach LMA with 4SA** Demonstrate in real time Demonstrate again slowly, with explanation Get a student from the group to talk you through it as you do it Each student performs the skill under your supervision, with your feedback

10.4.4 Exposure survey

This survey was designed to help the researchers understand and quantify the skill exposure study participants have had between the initial Intraosseous (IO) and Laryngeal Mask Airway (LMA) skill training and re-performance.

It was not a condition of participation that involved students should forfeit other learning opportunities related to the skills taught, so the information provided did not exclude you from the study.

The exposure survey asked participants the following questions when they returned for the final performance:

How many times have you participated in training sessions since the study? (Please note whether these were a formal or informal part of your studies at Flinders, FUSPA training events, ALS courses, on SAAS placement, or specify other sources of training.)

LMA:	None	IO:	None
	One		One
	Two		Two
	Three or more		Three or more

How many times have you witnessed the skills performed in training sessions since the study? (Please note whether these were a formal or informal part of your studies at Flinders, FUSPA training events, ALS courses, on SAAS placement, on YouTube, or specify other sources of training.)

LMA:	None	IO:	None
	One		One
	Two		Two
	Three or more		Three or more

How many times have you witnessed the skills performed on a real patient since the study? (Please note whether these were on your SAAS placements or not.)

LMA:	None	IO:	None
	One		One
	Two		Two
	Three or more		Three or more

Please indicate, on a scale of 0 to 10 how confident you feel about performing the skills competently today (0 = not at all confident, and 10 = completely confident).

LMA: IO:

10.5 Educator perspective study (Chapter 7) appendices

10.5.1 Survey questions

1. **Have you learnt this teaching method (4SA) before?**
   □ Y/N
2. **How often do you teach with 4SA?**
   □ Never
   □ Sometimes (less than half the time)
   □ About half the time
   □ Often (more than half the time)
   □ Always

3. **That you can recall, how often have skills been taught to you with 4SA?**
   □ Never
   □ Sometimes (less than half the time)
   □ About half the time
   □ Often (more than half the time)
   □ Always

*note: If you answered *yes* to question 1, please reflect on the following questions as best you can, basing responses on your initial learning of 4SA*

4. **During your training session, how easy/difficult did you think 4SA was to learn?**
   □ Easy
   □ Moderate
   □ Difficult

5. **How easy/difficult did you expect 4SA would be to use in your teaching (before you had a chance to practice it)?**
   □ Easy
   □ Moderate
   □ Difficult

6. **How comfortable did you feel using 4SA?**
   □ Not comfortable at all (but I knew I had to)
   □ Slightly uncomfortable
   □ Moderately comfortable
   □ Quite comfortable
   □ It felt like second nature straight away

7. **After using 4SA for the first time, what did feedback indicate on your use of 4SA?**
   □ I used it perfectly
   □ I used it well
   □ It was ok, but there is plenty of room for improvement
   □ I need a lot more development to use 4SA properly

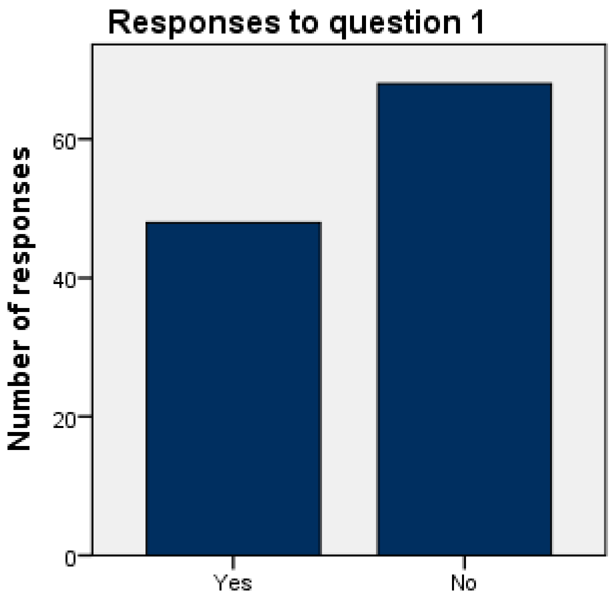
8. **After using 4SA and receiving feedback, do you feel that it was:**
   □ Easier than you expected?
   □ About as easy/difficult as you expected it would be?
   □ More difficult than you expected?

10.5.2 Survey results

117 surveys were collected to understand the initial hypotheses of this study. Further data collected was used to inform the early development of themes, which was further pursued in phone interviews, the focus group, and the educator debrief. This survey was in some ways the "kindling" which was later overcome by the flame of the greater data obtained.

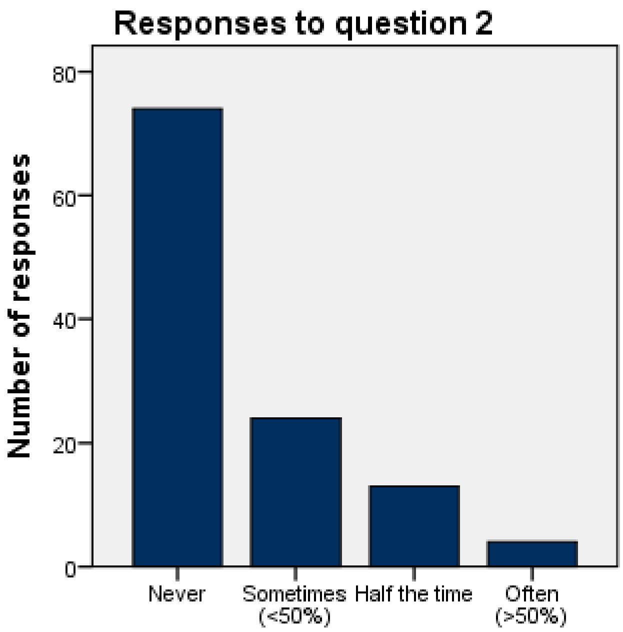
Surveys were numbered 1 to 106, and then 115 to 125. Upon collection of the last batch of surveys, I did not have convenient access to the earlier forms, so re-initiated numbering at 115, expecting that this may appear like surveys 107 to 114 were missing though they never existed. Re-numbering the surveys following analysis was seen as increasing the risk of erroneous numbering, so this was not done. Percentages noted are the percentage of valid responses (excluding missing values).

10.5.2.1 Q1 Have you learnt this teaching method (4SA) before?



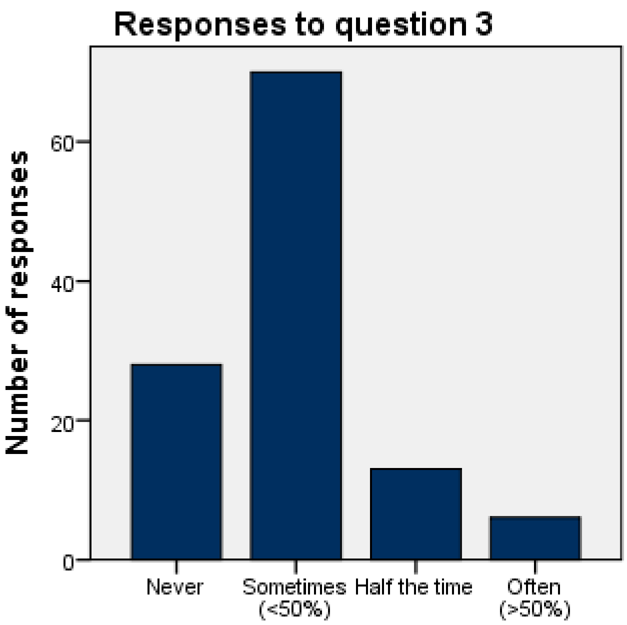
Of the 117 surveyed, 48 (41.4%) had learnt 4SA before and 68 (58.6%) had not. One response was missing.

10.5.2.2 Q2 How often do you teach with 4SA?



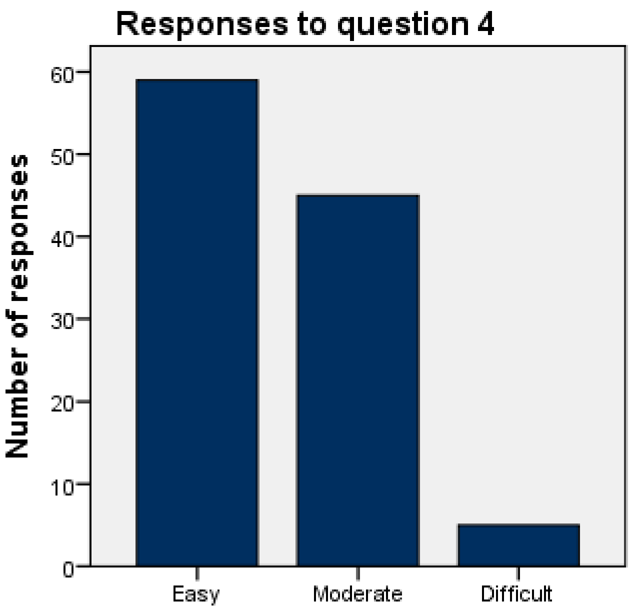
Of the 113 valid responses, 74 (64.3%) never use it to teach with, 24 (20.9%) sometimes, 13 (11.3%) about half the time, four (3.5%) used it often, and no one used 4SA every time. Two responses were missing.

10.5.2.3 Q 3 That you can recall, how often have skills been taught to you with 4SA?



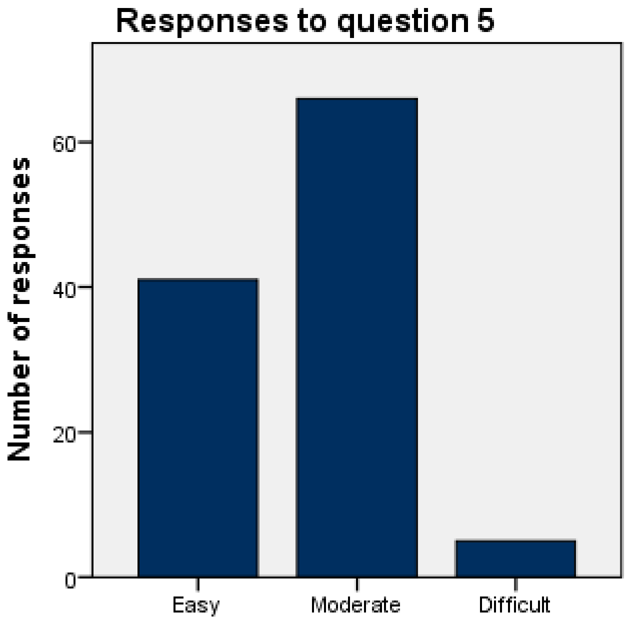
28 (23.9%) have never been taught skills with 4SA that they can recall, 70 (59.8%) have been taught with 4SA sometimes, 13 (11.1%) recall about half the time they have been taught skills with 4SA, and 6 (5.1%) state that it has been used often.

10.5.2.4 Q4 During your training session, how easy/difficult did you think 4SA was to learn?



When reflecting on their first training in 4SA, 59 (54.1%) thought it was easy to learn, 45 (41.3%) thought it was moderate, and 5 (4.6%) thought it was difficult to learn. 7 responses were missing, and one was considered invalid as multiple responses were marked.

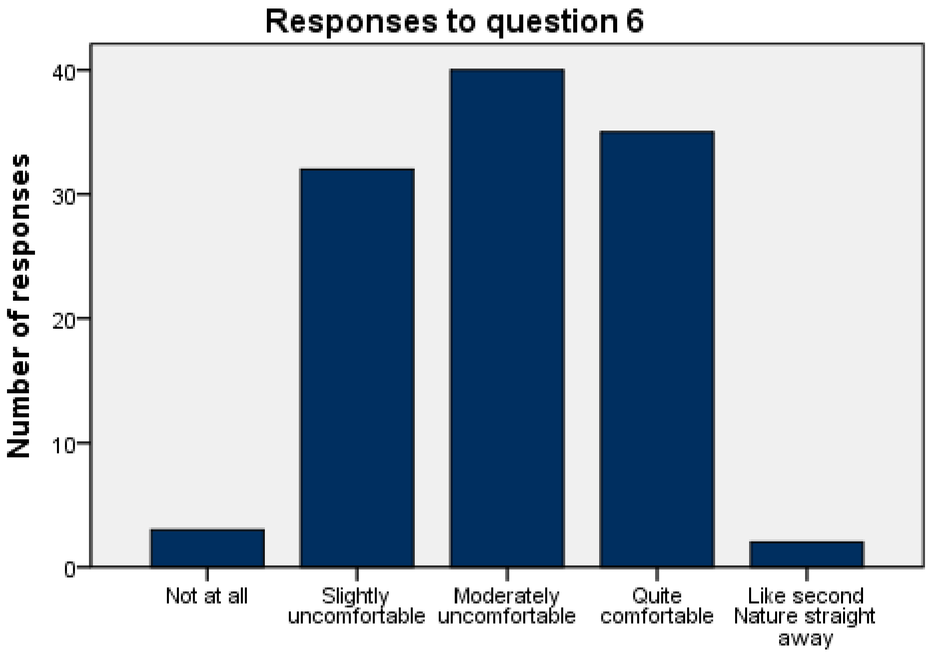
10.5.2.5 Q5 How easy/difficult did you expect 4SA would be to use in your teaching (before you had a chance to practice it)?



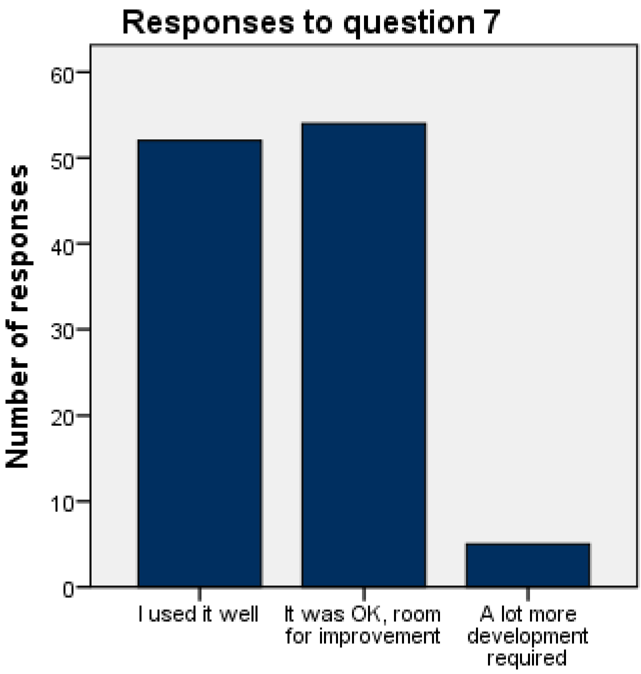
When reflecting on their first training in 4SA, 41 (35.0%) thought it was easy to learn, 66 (56.4%) thought it was moderate, and 5 (4.3%) anticipated that it would be difficult to put into practice. 5 responses were missing.

10.5.2.6 Q6 How comfortable did you feel using 4SA?

When they came to put it into practice, comfort levels using 4SA varied: three (2.6%) were not comfortable at all, 32 (27.4%) were slightly uncomfortable, 40 (34.2%) were moderately comfortable, 35 (29.9%) quite comfortable and 2 (1.7%) indicated that it felt like second nature straight away. 5 responses were missing.

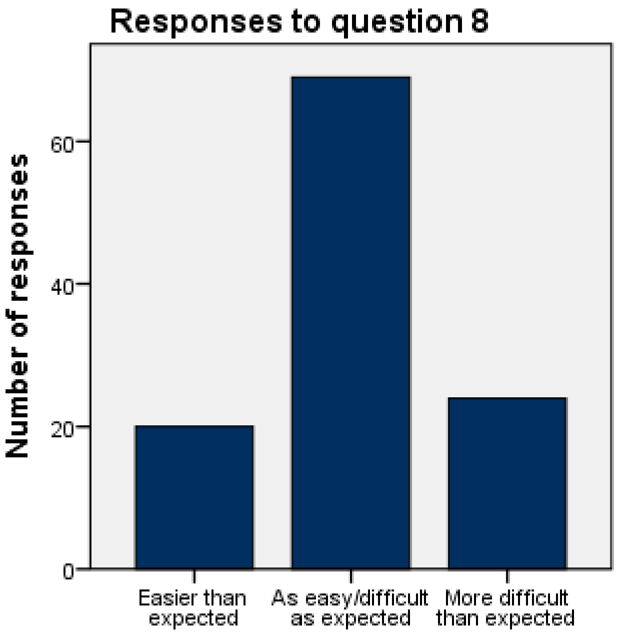


10.5.2.7 Q7 After using 4SA for the first time, what did feedback indicate on your use of 4SA?



No respondents indicated that their feedback reflected they had used 4SA perfectly. 52 (44.4%) note that they used it well, 54 (46.2%) said their performance was ok but with plenty of room for improvement, and 5 (4.3) noted that there was a lot more development needed to use 4SA properly. 6 responses were missing.

10.5.2.8 Q8 After using 4SA and receiving feedback, how easy or difficult was is on reflection?



On reflection of the learning, implementation and feedback of 4SA use, 20 (17.7%) indicated that it was easier than expected, 69 (61.1%) that it was about as easy or difficult as expected, and 24 (21.2%) that it was more difficult than expected. Four responses were missing.

10.5.2.9 Questions 4 and 7: do educators' expectations of how easy or difficult 4SA is to learn to use (during their training session) relate to their performance of it, based on feedback received?

The table indicates that of the participants who answered both questions (n=105), most respondents thought it was easy or moderate in difficulty to learn, and either used it well or felt that there was some room for improvement (n=31+17+24+25 = 97). No one received feedback that they performed the four steps perfectly, and only relatively few educators who thought it was easy or moderate to learn received the unexpected feedback that there was a lot more development required. These data suggest that even when 4SA seems easy to learn, to perform it properly in the early stages of practice should not be expected. Nearly half of the educators (44.6%) who found 4SA easy to learn did *not* apply it well, according to their perception of feedback received. Practice is indeed required, even if it seems straightforward to learn.

		Q7			
		Perfect use of 4SA	I used it well	It was OK, room for improvement	A lot more development required
Q4	Easy	0	31	24	1
	Moderate	0	17	25	2
	Difficult	0	3	1	1

10.5.2.10 Questions 5 and 7: do educators' expectations of how easy or difficult 4SA will be to implement relate to their performance of it, based on feedback received?

The relationship between these two sets of answers follows similar trends to those discussed in the relationship between questions 4 and 7. Of the surveys where valid responses existed for both questions (n=108), 100 (92.6%) thought it was easy or moderate in difficulty to learn, and received feedback that they used it well or felt that there was some opportunity for improvement. 24 (58.5%) educators who found 4SA easy to learn did *not* apply it well, according to their perception of feedback received. Only one educator who thought it would be easy received feedback that a lot of development was required, indicating that it was not very common to have a significantly inflated view of one's ability.

		Q7			
		Perfect use of 4SA	I used it well	It was OK, room for improvement	A lot more development required
Q5	Easy	0	17	23	1
	Moderate	0	32	28	2
	Difficult	0	2	2	1

10.5.2.11 Questions 4 and 8: do educators' expectations of how easy or difficult 4SA is to learn to use (during their training session) relate to their reflection of how easy or difficult it was?

66 educators (61.7% of the 107 surveys with valid responses for both questions) found that performing 4SA matched their perception of how easy it was to learn, whether they expected it would be easy, moderate or difficult to perform. Remaining responses were fairly evenly split between those who found it easier or harder than expected (20 and 21 responses respectively).

		Q8		
		Easier than expected	As easy/difficult as expected	More difficult than expected
Q4	Easy	13	36	9
	Moderate	6	27	11
	Difficult	1	3	1

10.5.2.12 Questions 5 and 8: do educators' expectations of how easy or difficult 4SA will be to implement relate to their reflection of how easy or difficult it was?

67 educators (60.9% of the 110 surveys with valid responses for both questions) found that performing 4SA matched their expectation of how easy it would be to perform, whether they expected it would be easy, moderate or difficult to perform. Again, the remaining responses were

fairly evenly split between those who found it easier or harder than expected (20 and 23 responses respectively). Unlike the previous comparison, of those who found it more difficult than expected to perform 4SA, most initially anticipated it would be easy to perform. While numbers are too low to perform sophisticated statistical analysis, this could indicate that educators who think 4SA will be easy to perform may be surprised to find it more difficult than expected.

		Q8		
		Easier than expected	As easy/difficult as expected	More difficult than expected
Q5	Easy	7	21	13
	Moderate	13	44	7
	Difficult	0	2	3

10.5.2.13 Questions 6 and 7: does the level of educator comfort using 4SA relate to feedback received?

Whether the educator felt comfortable using 4SA was generally a fair indicator of the feedback they state receiving. The central (shaded) boxes indicate responses where the comfort level using 4SA was generally consistent with the feedback received, amounting to 99 of the 108 surveys with a valid response to both questions (91.7%).

		Q7			
		Perfect use of 4SA	I used it well	It was OK, room for improvement	A lot more development required
Q6	Like second Nature straight away	0	2	0	0
	Quite comfortable	0	22	10	1
	Moderately uncomfortable	0	14	23	1
	Slightly uncomfortable	0	12	18	2
	Not at all	0	1	2	0

The survey gave us a glimpse into how easy or difficult ALS instructors (in training) find 4SA to learn and use. The study set out to explore the initial hypothesis that maybe 4SA is difficult to comply to because it's harder to implement than an educator first thinks during their educator development training. The survey was never intended to limit the study to just that one aspect, however. In the open comment sections (following each question), glimpses began to emerge of other explanations for lower 4SA compliance when compared to 2SA. Concepts of teaching habit and individual style, and practical considerations came to light very early in the data collection and analysis process. By the third level of analysis, the deeper notions of cognitive load and teaching craft had emerged, as outlined in Chapter 7.

10.5.3 Interview questions and analysis

Interview structure:

- Typical introduction: purpose and background of the study; interviewee's survey results
- Typical close: any other points you think are relevant, questions for me, and thanks.
- Most interviews around 10 minutes

10.5.3.1 Stage 1

Intention of first round of interviews:

- Our suspicion was that the 4SA seemed easier to learn than it actually is, leading educators to feel like they're doing it better than they actually are.
- Keep questions open
- Monitor rapport, question styles, review working (for bias)

Interview 1

1. As an educator, what do you think about the four-step method?
2. You said there were some parts of it that didn't come naturally and that might have made it a bit more difficult. Was there any part of it in particular that makes it unnatural? Or...?
3. As an educator would you expect the 4 step method to result in similar skill acquisition than the more traditional see one do one two stage approach?
4. Was there anything in particular about 4SA that made it easy to teach?
5. Did you feel that it flowed well or were there difficulties with that?
6. Do you feel that that's something that will be remedied the more you practised four stage or do you feel that it's not really worth pushing through to that point?
7. In terms of learning the method … in the first place and get it in your head, did it feel quite straightforward to learn, or was it a bit awkward to get into your head in the first place?
8. Do you know whether you think there would be any benefits or downfalls of using this particular method from the learner's perspective?
9. If there was a mandated requirement that this method be used in all clinical education setting … what successes or difficulties would you expect in your setting?

10. The final question was just if you had anything else to add, any other thoughts on the method or if you had any questions of me?

Interviewer reflection:

- Clinical educators are busy (clinical hours, shift work, family life, education role, additional activities like ALS instructing). Important to keep interviews brief.
- Easy to muddle steps up
- No predicted benefit with 4SA
- 4SA longer
- She seemed to become very nervous at the possibility that 4SA could be mandated
- Extra effort wasted when considering the simpler skills; might come into its own with more complex skills.
- Used a lot of qualifiers (a little bit, in some ways, I can't really comment on, potentially...)
- She is not a convert to 4SA but seems reluctant to overtly reject it
- I did some leading, eg using terms like awkward etc, and in the way I introduced some questions.

Interview 2

1. As an educator when you come to teach new skills to clinical students, what's important to you?
2. Would you expect that the 4 stage method would result in similar skill acquisition or retention than like a 2 stage see one do one?
3. Right so you think it has more of a place in the clinical skills lab rather than at the bedside teaching?
4. Do you feel that there were parts of the 4 stage approach that make it difficult to teach or easy to teach?
5. So you mentioned at the start there that you initially reverted to a 2 step method of teaching... Was that subconscious or was it something that...
6. Sure. So that your perception then is that the 4 stage approach takes more time from a preaching perspective?
7. When you think back to your session where you were taught how to use 4 stage, did you feel that it was a straightforward method to learn but just difficult to implement or did you think it was difficult to learn from the get go?

8. If, hypothetically, the 4 stage approach was mandated to be used in all clinical education settings you know across the country, do you predict that there would be particular successes or particular difficulties in the way it ... take up with students or clinicians?
9. So coming back to one of your earlier comments, you think it would be better taken up in the clinical skills lab but it would be difficult to employ in the clinical setting
10. Sure. And in that setting to you tend to teach novice students or do you teach people who have had some exposure to the skills in the past?
11. Do you feel that the 4 stage comes into its own with different skill complexities or with novice learners versus expert learners, or…?
12. So from what I'm hearing generally, just to sort of summarise, the four stage approach seemed quite easy to learn, but then when it came to practice it, it took quite a bit of reflection to practice it properly… Do you feel then that the take up would be any easier with sort of people who are earlier in their clinical education career? Or...

Interviewer reflection:

- Felt freer this time to follow tangents within the interview rather than stick religiously to my set questions, and will try to apply this flexibility more in interview 3.

Interview 3

1. I'm interested to know what's important to you when it comes to teaching clinical skills.
2. How do you think [the 4SA] might be well suited or poorly suited to that aim?
3. Right, so the repetition you see in the four stage approach you expect to be linked to better retention or acquisition of the skill?
4. As someone who's involved in clinical education, when you first learnt the 4 stage approach, was there any part of it that was particularly difficult or that made a lot of sense that you only considered for the first time?
5. So even though with the see one do one sort of method of teaching skills we're still required to break it down and explain it step by step, I'm hearing that the four stage method seems to make that easier for you?
6. When you came to teach with the 4SA you noted that you found it more difficult than you expected it to feel. Um can you remember what you were thinking or how that was processing at the time?
7. If across the whole nation we decided "right, in every medical education setting um educators are to teach with the 4SA every single time they are to teach a skill", what might

be some successes of that or some difficulties you might predict in your various work settings?

8. How well do you think it would be received by your teaching peers who might not be ALS instructors? Do you think they'd be receptive to it?
9. Do you encounter the four-step method in settings other than the ALS instructor course personally or have you just heard that it's widely adopted?
10. You mention that you used it to teach airway skills, is that mainly in the skills lab setting?
    a. So for those more complex skills then, did you have any specific suggestions as to what that teaching session would look like?

Interviewer reflection:

- During: Didn't feel like good rapport.
- Difficult to identify whether the interviewee was thinking or had finished responding;
- long pauses;
- many interruptions as a result.
- Difficult to transcribe as many sections of poor audio quality.

Debrief questions

(Questions of the educators)

I am keen to hear from you guys on how you felt that it went from an educator’s perspective.

1. Did anyone else find the same thing with the first step being quite difficult?
2. What were some benefits to having step one in there?
3. I'm interested to hear also from the people that chose not to use Peyton’s four step and some rational that you might have used to decide no I'm not going to use that um or anything that you noticed when you saw it be taught that because you didn't try it yourself you maybe had a different perspective on what you saw of others in your group?

(Questions for the students)

1. Have you ever been taught with that sort of method before or did you get a mix in your groups as to what method you were taught the skills with?
2. Was there any part of if you found most useful or did you find any parts awkward?

Themes from round 1

- Benefits to 4SA ("converts" to the method)
- Difficulties/barriers of 4SA (? Increased cognitive load)
- Practical considerations
- Educator's beliefs around teaching

10.5.3.2 Stage 2

Interview 4

1. As an educator I'm just really interested to hear what you think about that 4 stage skill teaching method that you were encouraged to use
2. Using that method for the first time, do you feel that it sort of took a lot of your concentration as an educator that it might have distracted you from the content of the course itself or did you feel that that wasn't an issue because you were so familiar with the course content? (investigating cognitive load)
3. Do you, as an educator yourself do you feel that it's kind of worth it? Obviously it's a bit difficult to grapple with at the start; for the students do you think it's worth being taught in this way? (investigating benefits/educator's beliefs)
    a. Right so rather than the skill teaching method itself just the ongoing exposure to the skills and implementing them you feel is what makes the difference?
4. I'm interested just to ask more about a comment you made earlier and I can't remember exactly what you said but it was talking about how some educators are naturally more gifted and engaging that others. Do you feel that um the 4 stage approach might have more relevance to those educators who maybe don't have that or do you think it's appropriate for all? Does that sort of make sense? (to investigate a comment re: educator style)

Interview 5

1. (review of survey responses) are you able to elaborate at all on that?
    a. so it comes down in your opinion to time pressures and just remembering to do it in the clinical setting as opposed to the lab setting?
2. As a teacher do you feel like there are any particular benefits to using the method either for your student or your patient or you yourself? (to investigate perceived benefits)
    a. Do you feel there for that the four stage method given that it's quite repetitive might be a bit more thorough therefore might have a bit more sticking capacity to be more traditional see one do one?

3. When you first learnt the method which I'm not sure how long ago that was if you've been doing EMST stuff um how did it feel to you the first few times you used it? (investigating learn-ability of 4SA from educator perspective)
    a. Is that a similar sort of impression you have with other colleagues learning this method? (moves away from purist phenomenological approach as it's not asking about the participant's personal experience)
4. In terms of your student level does that sort of change whether you're in the skills lab settings or um you know at the bedside do you tend to have more novice students in the lab or does that affect how you teach? (investigate some practical considerations)
5. It sounds like you subscribe to it um when you can because of the repetition involved and the memory building that that creates. Is there a particular part of it that you dislike? (participant seems to be a "convert", but are there difficulties despite that?)

Main themes from round 2

Expected compliance may be effected by:

- Indoctrination with 4SA
- Teaching habits/experience/art
- 4SA difficulties
- Practicalities

(for example how educators make the decision of how to teach)

10.5.3.3 Stage 3

Interview 6

1. You obviously found it easy but did you see did you feel that your peers found it as easy as you?
2. What do you think are some of the strengths that the method itself brings to the teaching?
    a. Given that it sort of dictates some of that structure to you, um do you think it makes it easier therefore because you can slot into a predetermined structure, or do you think it makes it harder for some people because they have to conform to something that they're not comfortable with you think it's an individual basis?
3. Do you think with that teaching methods itself but there's any learning principles or learning strategies that it neglects, or anything that could do better?
    a. have you since you've done your ALS instructor course have you instructed yet?

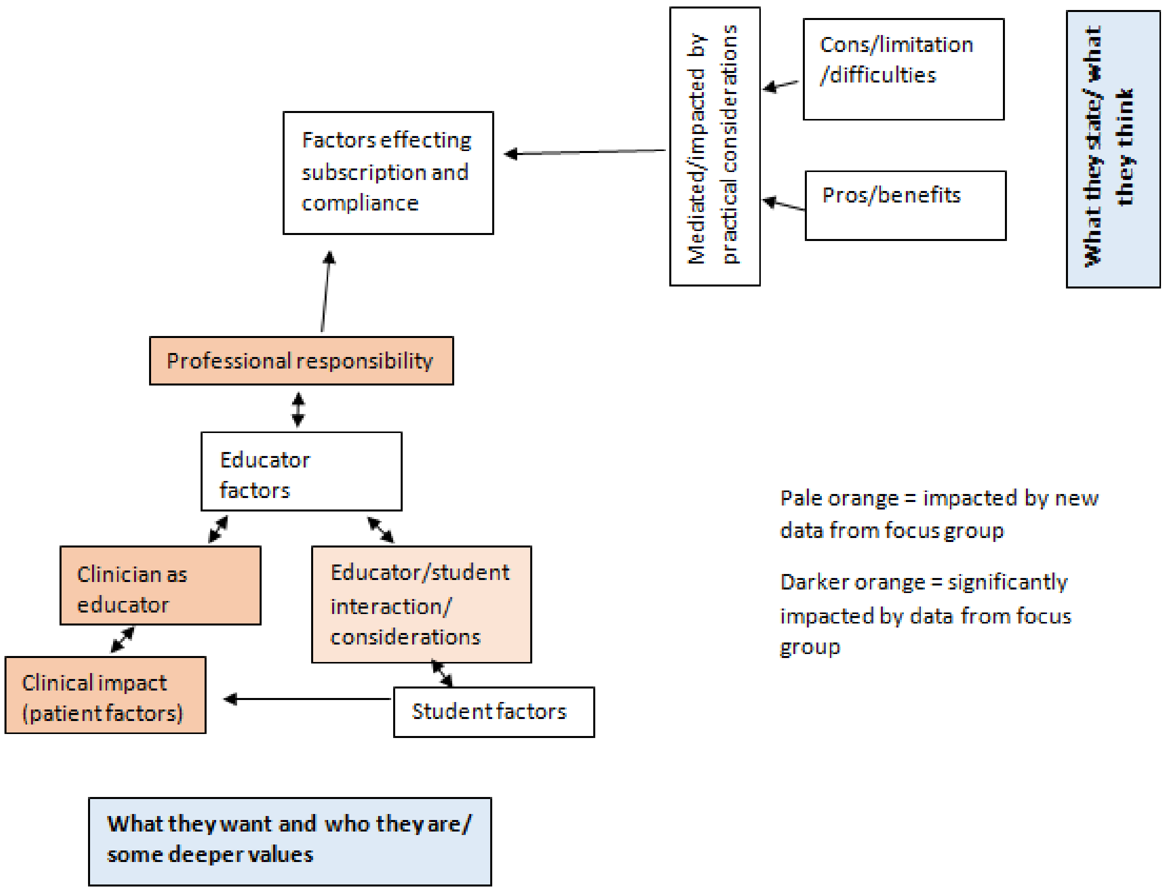
Themes from round 3

- A tool's benefits and difficulties can both impact cognitive load
- The individual teaching craft of educators
- Practicalities and aim of skills teaching

➔ These three things all impact the <u>internal conflict</u> between the craft/philosophy and the practicalities or skills teaching, for example:

  - Heart versus head
  - What should I do, versus what's comfortable, versus what's practical

Themes from round 3.2



10.5.3.4 Stage 4

Interview 7

1. You specifically mentioned not for on the floor teaching[36], [I'm interested to know] what your reasons for that were or what you're thinking behind that was? (simple recall from survey)
    a. So given that you sort of respond to the teaching context that you're in um in the skills lab setting do you think it's worth it to do the four stage approach or…?
    b. So a lot of the use for you as an educator comes from the structure of the teaching session kind of letting you order your thoughts a little bit?
    c. I'm assuming that given that you haven't used it too much in the sort of bedside theatre setting but you had a chance to use it in your ALS teaching?
2. If I get you to think about the most powerful educators that you've had [during your training], if they were told to use a 4 stage approach do you think that would have any effect on their teaching, either positively or negatively? (reflection on another's teaching art, and how that sits with 4SA)
3. So if we then come back to the course based settings when you feel that it is it does come into its own … what do you think the reason that would be? Do you think it would be more effective in aiding retention or um gaining sort of students attention or anything like that or predominately it's from the teachers' perspectives that helps you order your thoughts? (where 4SA is practical to do, why is that the case?)
4. You indicated humbly that there is still some room for improvement as you practised the four stage method [background noise] … how easily do you think that that could be picked up by an educator at your… point in their career? (investigating the alignment of 4SA with pre-existing art/habit)

Interview 8

1. Can you tell me what your main thoughts were when it came to the four stage approach?
    a. Was it anything in particular that you found difficult about that?
2. Is it more than it felt unnatural for you or did you also perceive that it felt unnatural for your student or a bit of both? (investigating the alignment of 4SA with pre-existing art/habit)
3. Where do you think it has most of its value: in a particular context or with a particular kind of student or should it be up to the discretion of the educator?
    a. Have you been able to use this on a course in or on your personal teaching or have you not have the opportunity yet?
4. If you think back to the teaching that you had throughout your career so far and the most inspiring teachers that you had or the most passionate or contagious teachers, have any of them

[36] "On the floor teaching" refers to teaching in the authentic, non-simulated clinical context, for example during ward rounds. Teaching in these settings is often unplanned and opportunistic, in contrast to skills laboratory sessions during formal teaching time.

used 4 step approach or something similar? (reflection on another's teaching art, and how that sits with 4SA)

   a. So to me your kind of describing um almost the craft of being an educator in that and the different things that an educator will be looking out for as they teach
   b. If we look at the four stage approach and particularly Stage 3 which a lot of people find a bit difficult by getting the student to talk you through it first, do you think that could give the Student a little bit of confidence by kind of demonstrating...

5. What do you think would be an ideal way of teaching, or do you think there is an ideal way of teaching Clinical Skills?

11 REFERENCES

Ackoff, R. L. (1989). From data to wisdom. *Journal of applied systems analysis, 16*(1), 3-9.

Adams, L. (2016). Learning a New Skill is Easier Said Than Done. Retrieved from http://www.gordontraining.com/free-workplace-articles/learning-a-new-skill-is-easier-said-than-done/#

Ali, J., Cohen, R., Adam, R., Gana, T., Pierre, I., Ali, E., . . . Winn, J. (1996). Attrition of cognitive and trauma management skills after the Advanced Trauma Life Support (ATLS) course. *Journal of Trauma and Acute Care Surgery, 40*(6), 860-866.

Ali, J., Howard, M., & Williams, J. (2002). Is attrition of advanced trauma life support acquired skills affected by trauma patient volume? *The American journal of surgery, 183*(2), 142-145. doi:http://dx.doi.org/10.1016/S0002-9610(01)00862-5

Amaral, F., & Troncon, L. E. d. A. (2013). Retention of Knowledge and Clinical Skills by Medical Students: A Pro-spective, Longitudinal, One-Year Study Using Basic Pediatric Cardiology as a Model. *The Open Medical Education Journal, 6*(1).

American Heart Association. (2016, 2015). Part 12: Pediatric Advanced Life Support. *Emergency Cardiovascular Care Guidelines.* Retrieved from www.https://eccguidelines.heart.org

American Psychological Association. (2010). *Publication manual of the American psychological association* (6th ed.): American Psychological Association Washington.

Anderson, G. S., Gaetz, M., & Masse, J. (2011). First aid skill retention of first responders within the workplace. *Scand J Trauma Resusc Emerg Med, 19*(11).

Anderson, T. E., Arthur, K., Kleinman, M., Drawbaugh, R., Eitel, D. R., Ogden, C. S., & Baker, D. (1994). Intraosseous infusion: Success of a standardized regional training program for prehospital advanced life support providers. *Annals of Emergency Medicine, 23*(1), 52-55. doi:http://dx.doi.org/10.1016/S0196-0644(94)70008-7

Andrew, S. (2002). *The relationships among first year Bachelor of Nursing students' entry characteristics, self-regulated learning and academic performance for their science and nursing practice courses.* (Doctor of Philosophy), University of Wollongong, Wollongong, NSW. (1803)

ANZCOR. (2015). Summary of ALS changes.

ANZCOR. (2016a, Jan 2016). Guideline 11.2 Protocols for adult Advanced Life Support. Retrieved from www.http://resus.org.au/guidelines

ANZCOR. (2016b, Jan 2016). Guideline 11.6 Equipment and techniques in adult Advanced Life Support. Retrieved from www.http://resus.org.au/guidelines

ANZCOR. (2016c, Jan 2016). Guideline 12.4 Medications and fluids in Paediatric Advanced Life Support. Retrieved from www.http://resus.org.au/guidelines

Archer, E., Van Hoving, D. J., & de Villiers, A. (2014). In search of an effective teaching approach for skill acquisition and retention: Teaching manual defibrillation to junior medical students. *African Journal of Emergency Medicine*.

Ash, J. (2010). *A case of meaning: Change in clinical education.* (PhD), Flinders University, Adelaide, Australia.

Australian Commission on Safety and Quality in Healthcare. (2010). *National consensus statement: essential elements for recognising and responding to clinical deterioration*. Retrieved from https://safetyandquality.gov.au/wp-content/uploads/2012/01/national_consensus_statement.pdf.

Australian Resuscitation Council. (2015). Australian Resuscitation Council Advanced Life Support Instructor Course Regulations (Vol. 2017). Australia: Australian Resuscitation Council.

Baker, T. W., King, W., Soto, W., Asher, C., Stolfi, A., & Rowin, M. E. (2009). The efficacy of pediatric advanced life support training in emergency medical service providers. *Pediatric Emergency Care, 25*(8), 508-512.

Balmer, M. C., & Longman, L. P. (2008). A practical skill one day medical emergencies course for dentists and DCPs. *British Dental Journal, 204*(8), 453-456. doi:http://dx.doi.org/10.1038/sj.bdj.2008.300

Barelli, A., & Scapigliati, A. (2010). The four-stage approach to teaching skills: The end of a dogma? *Resuscitation, 81*(12), 1607-1608.

Benjamin, L. (2005). Selection, teaching and training in ophthalmology. *Clinical & experimental ophthalmology, 33*(5), 524-530.

Berg, D., Berg, K., Riesenberg, L. A., Weber, D., King, D., Mealey, K., . . . Tinkoff, G. (2013). The Development of a Validated Checklist for Thoracentesis: Preliminary Results. *American Journal of Medical Quality, 28*(3), 220-226. doi:10.1177/1062860612459881

Berg, K., Riesenberg, L. A., Berg, D., Mealey, K., Weber, D., King, D., . . . Tinkoff, G. (2013). The Development of a Validated Checklist for Adult Lumbar Puncture Preliminary Results. *American Journal of Medical Quality, 28*(4), 330-334.

Berg, K., Riesenberg, L. A., Berg, D., Schaeffer, A., Davis, J., Justice, E. M., . . . Jasper, E. (2014). The Development of a Validated Checklist for Radial Arterial Line Placement: Preliminary Results. *American Journal of Medical Quality, 29*(3), 242-246. doi:10.1177/1062860613492189

Berger, P. L., Luckmann, T., & Zifonun, D. (2002). *The social construction of reality*: na.

Billett, S. (1996). Situated learning: Bridging sociocultural and cognitive theorising. *Learning and instruction, 6*(3), 263-280.

Billett, S. (2001). Knowing in practice: Re-conceptualising vocational expertise. *Learning and instruction, 11*(6), 431-452.

Billett, S. (2008). Learning throughout working life: a relational interdependence between personal and social agency. *British Journal of Educational Studies, 56*(1), 39-58.

Billett, S. (2009). Realising the educational worth of integrating work experiences in higher education. *Studies in Higher Education, 34*(7), 827-843.

Birrell, R., & Khoo, S.-E. (1995). *The second generation in Australia: Educational and occupational characteristics*: Australian Government Pub. Service.

Bitsika, E., Karlis, G., Iacovidou, N., Georgiou, M., Kontodima, P., & Xanthos, T. (2013). Comparative analysis of two venipuncture learning methods on nursing students. *Nurse education today*.

Blandin, Y., Lhuisset, L., & Proteau, L. (1999). Cognitive processes underlying observational learning of motor skills. *The Quarterly Journal of Experimental Psychology: Section A, 52*(4), 957-979.

Bleakley, A., & Bligh, J. (2008). Students learning from patients: Let's get real in medical education. *Advances in Health Sciences Education, 13*(1), 89-107.

Boon, J., Gorry, D., & Meiring, J. (2003). Finding an ideal site for intraosseous infusion of the tibia: an anatomical study. *Clinical Anatomy, 16*(1), 15-18.

Boulkedid, R., Abdoul, H., Loustau, M., Sibony, O., & Alberti, C. (2011). Using and reporting the Delphi method for selecting healthcare quality indicators: a systematic review. *PloS one, 6*(6), e20476.

Bower, M. (2009). Lessons from cognitive science for computer science education. *Quick*(109), 17-27.

Brady, S., Bogossian, F., & Gibbons, K. (2015). The effectiveness of varied levels of simulation fidelity on integrated performance of technical skills in midwifery students — A randomised intervention trial. *Nurse education today, 35*(3), 524-529. doi:http://dx.doi.org/10.1016/j.nedt.2014.11.005

Brady, S., Bogossian, F., Gibbons, K., Wells, A., Lyon, P., Bonney, D., . . . Jackson, A. (2013). A protocol for evaluating progressive levels of simulation fidelity in the development of technical skills, integrated performance and woman centred clinical assessment skills in undergraduate midwifery students. *BMC Medical Education, 13*(1), 72.

Brink, H., Van der Walt, C., & Van Rensburg, G. (2012). *Fundamentals of Research Methodology for Health Care Professionals*. Cape Town: Juta and Company.

Bullock, I. (2000). Skill acquisition in resuscitation. *Resuscitation, 45*(2), 139-143.

Butler, C., & Harley, D. (2010). Primary, secondary and tertiary effects of eco-climatic change: the medical response. *Postgraduate medical journal, 86*(1014), 230-234.
Calkins, M. D., Fitzgerald, G., Bentley, T. B., & Burris, D. (2000). Intraosseous infusion devices: a comparison for potential use in special operations. *Journal of Trauma and Acute Care Surgery, 48*(6), 1068-1074.
Cangelosi, P. R., Crocker, S., & Sorrell, J. M. (2009). EXPERTtoNOVICE: Clinicians Learning New Roles as Clinical Nurse Educators. *Nursing education perspectives, 30*(6), 367-371.
Carefusion Corporation. (2014). Intraosseous infusion needle (4 ed.).
CareFusion™. (2014). Intraosseous Infusion Needle.
Carley, S., & Boyd, R. (2004). Screw tipped needles for intraosseous access. *Emergency Medicine Journal*(21). doi:10.1136/emj.2004.014902
Carse, J. (2012). *Finite and infinite games*: Free Press.
Cartwright, N. (2007). Are RCTs the gold standard? *BioSocieties, 2*(1), 11-20.
Cattaneo, L., & Rizzolatti, G. (2009). The mirror neuron system. *Archives of neurology, 66*(5), 557-560.
Cave, D. M., Aufderheide, T. P., Beeson, J., Ellison, A., Gregory, A., Hazinski, M. F., . . . Mosesso, V. N. (2011). Importance and implementation of training in cardiopulmonary resuscitation and automated external defibrillation in schools. *Circulation, 123*(6), 691-706.
Chan, P. S., Krumholz, H. M., Nichol, G., & Nallamothu, B. K. (2008). Delayed time to defibrillation after in-hospital cardiac arrest. *New England Journal of Medicine, 358*(1), 9-17.
Chang, A., Gardner, G., Duffield, C., & Ramis, M. (2010). A Delphi study to validate an advanced practice nursing tool. *Journal of advanced nursing, 66*(10), 2320-2330.
Classen, C. (2005). The witch's senses *Empire of the senses: The sensual culture reader* (pp. 71-84).
Clayton, M. J. (1997). Delphi: a technique to harness expert opinion for critical decision-making tasks in education. *Educational Psychology, 17*(4), 373-386.
Cook, D. A., Zendejas, B., Hamstra, S. J., Hatala, R., & Brydges, R. (2014). What counts as validity evidence? Examples and prevalence in a systematic review of simulation-based assessment. *Advances in Health Sciences Education, 19*(2), 233-250.
Cortina, J. M. (1993). What is coefficient alpha? An examination of theory and applications. *Journal of applied psychology, 78*(1), 98.
Creswell, J. W. (2013). *Research design: Qualitative, quantitative, and mixed methods approaches*: Sage publications.
Critical Appraisal Skills Programme. (2013). CASP RCT Checklist. Retrieved from http://www.casp-uk.net/
Cronbach, L. J., & Meehl, P. E. (1955). Construct validity in psychological tests. *Psychological bulletin, 52*(4), 281.
Cronbach, L. J., & Shavelson, R. J. (2004). My current thoughts on coefficient alpha and successor procedures. *Educational and psychological measurement, 64*(3), 391-418.
Crotty, M. (1998). *The foundations of social research: Meaning and perspective in the research process*: Sage.
Cruess, R. L., Cruess, S. R., & Steinert, Y. (2016). Amending Miller's pyramid to include professional identity formation. *Academic Medicine, 91*(2), 180-185.
Dalkey, N. C., Brown, B. B., & Cochran, S. (1969). *The Delphi method: An experimental study of group opinion* (Vol. 3): Rand Corporation Santa Monica, CA.
de Ruijter, P. A., Biersteker, H. A., Biert, J., van Goor, H., & Tan, E. C. (2014). Retention of first aid and basic life support skills in undergraduate medical students. *Medical Education Online*.
Dembo, M. H., & Howard, K. (2007). Advice about the use of learning styles: A major myth in education. *Journal of College Reading and Learning, 37*(2), 101-109.
Dimitrow, M., Mykkanen, S., Leikola, S., Kivela, S., Lyles, A., & Airaksinen, M. (2014). Content validation of a tool for assessing risks for drug-related problems to be used by practical

nurses caring for home-dwelling clients aged >65 years: a Delphi survey. *European Journal of Clinical Pharmacology, 70*(8), 991-1002.
Dooley, K. J. (1997). A complex adaptive systems model of organization change. *Nonlinear dynamics, psychology, and life sciences, 1*(1), 69-97.
Doyle, J. D., Webber, E. M., & Sidhu, R. S. (2007). A universal global rating scale for the evaluation of technical skills in the operating room. *The American journal of surgery, 193*(5), 551-555.
Dreyfus, S. E. (2004). The five-stage model of adult skill acquisition. *Bulletin of science, technology & society, 24*(3), 177-181.
Driscoll, P., Gwinnutt, C., & McNeill, I. (1999). Controversies in advanced trauma life support. *Trauma, 1*(2), 171-176.
Dunning, D., Johnson, K., Ehrlinger, J., & Kruger, J. (2003). Why people fail to recognize their own incompetence. *Current Directions in Psychological Science, 12*(3), 83-87.
Dyson, K., Bray, J., Smith, K., Bernard, S., Straney, L., & Finn, J. (2015). Paramedic exposure to out-of-hospital cardiac arrest is rare and declining in Victoria, Australia. *Resuscitation, 89*, 93-98.
Egan, T., & Jaye, C. (2009). Communities of clinical practice: the social organization of clinical learning. *Health:, 13*(1), 107-125.
Ehrlinger, J., & Dunning, D. (2003). How chronic self-views influence (and potentially mislead) estimates of performance. *Journal of personality and social psychology, 84*(1), 5.
Eoyang, G., & Holladay, R. (2013). *Adaptive action: Leveraging uncertainty in your organization*: Stanford University Press.
Eoyang, G. H. (2006). Human systems dynamics: Complexity-based approach to a complex evaluation. *Systems concepts in evaluation: An expert anthology*, 123-139.
Eva, K. W., Bordage, G., Campbell, C., Galbraith, R., Ginsburg, S., Holmboe, E., & Regehr, G. (2016). Towards a program of assessment for health professionals: From training into practice. *Advances in Health Sciences Education, 21*(4), 897-913.
Eva, K. W., & Regehr, G. (2005). Self-assessment in the health professions: a reformulation and research agenda. *Academic Medicine, 80*(10), S46-S54.
Field, A. (2009a). *Discovering statistics using SPSS*: Sage publications.
Field, A. (2009b). Exploring Assumptions *Discovering statistics using SPSS* (3rd ed., pp. 131-165): Sage publications.
Field, A. (2013). Comparing several means: ANOVA (GLM 1) *Discovering statistics using IBM SPSS statistics* (pp. 429-477): Sage.
Fitch, K., Bernstein, S., Aguilar, M., Burnand, B., LaCalle, J., Lazaro, P., . . . Kahan, J. (2014). The RAND/UCLA appropriateness method user’s manual. Santa Monica, CA: RAND Corporation; 2001. *THE JOURNAL OF BONE & JOINT SURGERY d JBJS. ORG, 98*.
Fitch, K., Bernstein, S. J., Aguilar, M. D., Burnand, B., & LaCalle, J. R. (2001). *The RAND/UCLA appropriateness method user's manual*. Retrieved from
Fitts, P. (1964). Perceptual-motor skill learning. *Categories of human learning, 47*, 381-391.
Fitts, P., & Posner, M. (1967a). Learning and skill performance *Human performance*.
Fitts, P., & Posner, M. (1967b). Skills of civilised man *Human performance*.
Forbes, H., Oprescu, F. I., Downer, T., Phillips, N. M., McTier, L., Lord, B., . . . Dayton, J. (2016). Use of videos to support teaching and learning of clinical skills in nursing education: A review. *Nurse education today, 42*, 53-56.
Fullan, M. (1993). Why teachers must become change agents. *Educational Leadership, 50*, 12-12.
Fullan, M. (2002). The change. *Educational Leadership, 59*(8), 16-20.
Fullan, M. (2006). The future of educational change: System thinkers in action. *Journal of Educational Change, 7*(3), 113-122.
Fullan, M., Galluzzo, G., Morris, P., & Watson, N. (1998). *The Rise & Stall of Teacher Education Reform*: ERIC.
George, J. H., & Doto, F. X. (2001). A simple five-step method for teaching clinical skills. *FAMILY MEDICINE-KANSAS CITY-, 33*(8), 577-578.

Gerard, J. M., Thomas, S. M., Germino, K. W., Street, M. H., Burch, W., & Scalzo, A. J. (2011). The effect of simulation training on PALS skills among family medicine residents. *Fam Med, 43*(6), 392-399.
Ghezzi, P., Chumber, S., & Brabazon, T. (2014). Educating medical students to evaluate the quality of health information on the web *The Philosophy of Information Quality* (pp. 183-199): Springer.
Giavarina, D. (2015). Understanding bland altman analysis. *Biochemia medica, 25*(2), 141-151.
Gingerich, A., Kogan, J., Yeates, P., Govaerts, M., & Holmboe, E. (2014). Seeing the 'black box'differently: assessor cognition from three research perspectives. *Medical education, 48*(11), 1055-1068.
Google. (2013). Google Scholar. Retrieved 8 October 2013, from Google Inc http://scholar.google.com.au
Govaerts, M., Van der Vleuten, C., Schuwirth, L., & Muijtjens, A. (2007). Broadening perspectives on clinical performance assessment: Rethinking the nature of in-training assessment. *Advances in Health Sciences Education, 12*(2), 239-260.
Green, B., Jones, M., Hughes, D., & Williams, A. (1999). Applying the Delphi technique in a study of GPs' information requirements. *Health & social care in the community, 7*(3), 198-205.
Greif, R., Egger, L., Basciani, R. M., Lockey, A., & Vogt, A. (2010). Emergency skill training—A randomized controlled study on the effectiveness of the 4-stage approach compared to traditional clinical teaching. *Resuscitation, 81*(12), 1692-1697. doi:http://dx.doi.org/10.1016/j.resuscitation.2010.09.478
Greif, R., Lockey, A. S., Conaghan, P., Lippert, A., De Vries, W., Monsieurs, K. G., . . . Castrén, M. (2015). European Resuscitation Council Guidelines for Resuscitation 2015: Section 10. Education and implementation of resuscitation.
Guadagnoli, M. A., & Lee, T. D. (2004). Challenge point: a framework for conceptualizing the effects of various practice conditions in motor learning. *Journal of motor behavior, 36*(2), 212-224.
Guest, G., Namey, E. E., & Mitchell, M. L. (2012). *Collecting qualitative data: A field manual for applied research*: Sage.
Hagen, N. A., Stiles, C., Nekolaichuk, C., Biondo, P., Carlson, L. E., Fisher, K., & Fainsinger, R. (2008). The Alberta Breakthrough Pain Assessment Tool for cancer patients: a validation study using a delphi process and patient think-aloud interviews. *Journal of Pain & Symptom Management, 35*(2), 136-152.
Hagler, D., Kastenbaum, B., Brooks, R., Morris, B., & Saewert, K. (2013). Leveraging the technology du jour for overt and covert faculty development. *The Journal of Faculty Development, 27*(3), 22-29.
Hallas, P. B., Mikkel; Folkestad, Lars;. (2012). Reasons for not using intraosseous access in critical illness. *Emergency Medicine Journal*(29). doi:10.1136/emj.2010.094011
Hallgren, K. A. (2012). Computing inter-rater reliability for observational data: an overview and tutorial. *Tutorials in quantitative methods for psychology, 8*(1), 23.
Hamdorf, J., & Hall, J. (2000). Acquiring surgical skills. *British Journal of Surgery, 87*(1), 28-37.
Harcke, T., Crawley, G., Mabry, R., & Mazuchowski, E. (2011). Placement of Tibial Intraosseous Infusion Devices. *Military Medicine, 176*(7), 824-827.
Harden, R. M. (1999). What is a spiral curriculum? *Medical teacher, 21*(2), 141-143.
Hargreaves, A., & Goodson, I. (2006). Educational change over time? The sustainability and nonsustainability of three decades of secondary school change and continuity. *Educational administration quarterly, 42*(1), 3-41.
Hartholt, K. A., van Lieshout, E. M., Thies, W. C., Patka, P., & Schipper, I. B. (2009). Intraosseous devices: a randomized controlled trial comparing three intraosseous devices. *Prehospital emergency care, 14*(1), 6-13.

Hartman, N., Wittler, M., Askew, K., & Manthey, D. (2014). Delphi method validation of a procedural performance checklist for insertion of an ultrasound-guided internal jugular central line. *American Journal of Medical Quality, 31*(1), 81-85.

Hartman, N., Wittler, M., Askew, K., & Manthey, D. (2016). Delphi Method Validation of a Procedural Performance Checklist for Insertion of an Ultrasound-Guided Internal Jugular Central Line. *American Journal of Medical Quality, 31*(1), 81-85.

Hasson, F., Keeney, S., & McKenna, H. (2000). Research guidelines for the Delphi survey technique. *Journal of advanced nursing, 32*(4), 1008-1015.

Hein, C. (2009). *Extraglottic Airway Devices and the Paramedic User: Teaching, Learning, Patient Safety and Device Suitability*: Flinders University, School of Medicine, Department of Paramedic and Social Health Sciences.

Hein, C., Owen, H., & Plummer, J. (2010). A Training Program for Novice Paramedics Provides Initial Laryngeal Mask Airway Insertion Skill and Improves Skill Retention at 6 Months. *Simulation in Healthcare, 5*(1), 33-39 10.1097/SIH.1090b1013e3181b1095c1093fb.

Herrmann-Werner, A., Nikendei, C., Keifenheim, K., Bosse, H. M., Lund, F., Wagner, R., . . . Weyrich, P. (2013). "Best practice" skills lab training vs. a "see one, do one" approach in undergraduate medical education: an RCT on students' long-term ability to perform procedural clinical skills. *PloS one, 8*(9), e76354. doi:10.1371/journal.pone.0076354

Herrmann, K. H., Kirchberger, I., Stucki, G., & Cieza, A. (2011a). The comprehensive ICF core sets for spinal cord injury from the perspective of occupational therapists: a worldwide validation study using the Delphi technique.[Erratum appears in Spinal Cord. 2011 May;49(5):676]. *Spinal Cord, 49*(5), 600-613.

Herrmann, K. H., Kirchberger, I., Stucki, G., & Cieza, A. (2011b). The comprehensive ICF core sets for spinal cord injury from the perspective of physical therapists: a worldwide validation study using the Delphi technique.[Erratum appears in Spinal Cord. 2011 May;49(5):676]. *Spinal Cord, 49*(4), 502-514.

Howes, D. (2005a). Culture tunes our neurons *Empire of the Senses* (pp. 21-24): Berg Publishers Oxford, UK.

Howes, D. (2005b). *Empire of the Senses*: Berg Publishers Oxford, UK.

Huang, G. C., Newman, L. R., Schwartzstein, R. M., Clardy, P. F., Feller-Kopman, D., Irish, J. T., & Smith, C. C. (2009). Procedural Competence in Internal Medicine Residents: Validity of a Central Venous Catheter Insertion Assessment Instrument. *Academic Medicine, 84*(8), 1127-1134. doi:10.1097/ACM.0b013e3181acf491

Iramaneerat, C., & Yudkowsky, R. (2007). Rater errors in a clinical skills assessment of medical students. *Evaluation & the Health Professions, 30*(3), 266-283.

James Cheung, W., Rosenberg, H., & Vaillancourt, C. (2014). Barriers and Facilitators to Intraosseous Access in Adult Resuscitations When Peripheral Intravenous Access Is Not Achievable. *Academic Emergency Medicine, 21*(3), 250-256.

Jenko, M., Frangež, M., & Manohin, A. (2012). Four-stage teaching technique and chest compression performance of medical students compared to conventional technique. *Croatian medical journal, 53*(5), 486-495.

Jeon, Y. H., Conway, J., Chenoweth, L., Weise, J., Thomas, T. H., & Williams, A. (2015). Validation of a clinical leadership qualities framework for managers in aged care: a Delphi study. *Journal of clinical nursing, 24*(7-8), 999-1010.

Johnston, M. (2016). What more can we learn from early learning theory? The contemporary relevance for behaviour change interventions. *British journal of health psychology, 21*(1), 1-10.

Kalyuga, S., Ayres, P., Chandler, P., & Sweller, J. (2003). The expertise reversal effect. *Educational psychologist, 38*(1), 23-31.

Kane, M. T. (2006). Validation. In R. L. Brennan (Ed.), *Educational measurement*: Praeger Pub Text.

Kaptchuk, T. J. (2001). The double-blind, randomized, placebo-controlled trial: gold standard or golden calf? *Journal of clinical epidemiology, 54*(6), 541-549.
Karpicke, J. D. (2012). Retrieval-based learning active retrieval promotes meaningful learning. *Current Directions in Psychological Science, 21*(3), 157-163.
Karpicke, J. D., & Roediger, H. L. (2008). The critical importance of retrieval for learning. *science, 319*(5865), 966-968.
Keeney, S., Hasson, F., & McKenna, H. (2006). Consulting the oracle: ten lessons from using the Delphi technique in nursing research. *Journal of advanced nursing, 53*(2), 205-212.
Keeney, S., Hasson, F., & McKenna, H. (2011). The Delphi technique. *The Delphi Technique in Nursing and Health Research*, 1-17.
Kennedy, T. J., & Lingard, L. A. (2006). Making sense of grounded theory in medical education. *Medical education, 40*(2), 101-108.
Kim, M., Oh, J., Min, J., Lee, K., & Lee, J. (2014). A randomised comparison of the i-gel™ and the Laryngeal Mask Airway Classic™ in infants. *Anaesthesia, 69*(4), 362-367.
Kleinman, M. E., Chameides, L., Schexnayder, S. M., Samson, R. A., Hazinski, M. F., Atkins, D. L., . . . Freid, E. B. (2010). Part 14: pediatric advanced life support. *Circulation, 122*(18 suppl 3), S876-S908.
Kneebone, R., Nestel, D., Yadollahi, F., Brown, R., Nolan, C., Durack, J., . . . Darzi, A. (2006). Assessing procedural skills in context: exploring the feasibility of an Integrated Procedural Performance Instrument (IPPI). *Medical education, 40*(11), 1105-1114. doi:10.1111/j.1365-2929.2006.02612.x
Kolb, D. A. (1984). *Experiential learning: Experience as the source of learning and development*. Englewood Cliffs, NJ: Prentice Hall.
Kors, J. A., Sittig, A. C., & van Bemmel, J. H. (1990). The Delphi method to validate diagnostic knowledge in computerized ECG interpretation. *Methods of Information in Medicine, 29*(1), 44-50.
Krautter, M., Weyrich, P., Schultz, J.-H., Buss, S. J., Maatouk, I., Jünger, J., & Nikendei, C. (2011). Effects of Peyton's Four-Step Approach on Objective Performance Measures in Technical Skills Training: A Controlled Trial. *Teaching and learning in medicine, 23*(3), 244-250.
Krischer, J. P., Fine, E. G., Davis, J. H., & Nagel, E. L. (1987). Complications of cardiac resuscitation. *Chest, 92*(2), 287-291.
Lafer-Sousa, R., Hermann, K. L., & Conway, B. R. (2015). Striking individual differences in color perception uncovered by 'the dress' photograph. *Current Biology, 25*(13), R545-R546.
Lake, F. R., & Hamdorf, J. M. (2004). Teaching on the run tips 5: teaching a skill. *Medical journal of Australia, 181*, 327-328.
Lamb, G. S., & Huttlinger, K. (1989). Reflexivity in nursing research. *Western Journal of Nursing Research, 11*(6), 765-772.
Lammers, R. L., Byrwa, M. J., Fales, W. D., & Hale, R. A. (2009). Simulation-based assessment of paramedic pediatric resuscitation skills. *Prehospital emergency care, 13*(3), 345-356.
Larsen, D., Butler, A., & Roediger III, H. (2009). Repeated testing improves long-term retention relative to repeated study: a randomised controlled trial. *Medical education, 43*(12), 1174-1181.
Larsen, M., Eisenberg, M., Cummins, R., & Hallstrom, A. (1993). Predicting survival from out-of-hospital cardiac arrest: a graphic model. *Annals of Emergency Medicine, 22*(11), 1652-1658.
Latman, N. S., & Wooley, K. (1980). Knowledge and skill retention of emergency care attendants, EMT-As, and EMT-Ps. *Annals of Emergency Medicine, 9*(4), 183-189.
Lavis, M. V., A; Tozer, C;. (2000). Adult intraosseous infusion in accident and emergency departments in the UK. *Journal of accident & emergency medicine*(17), 29-32.
Lederer, W., Mair, D., Rabl, W., & Baubin, M. (2004). Frequency of rib and sternum fractures associated with out-of-hospital cardiopulmonary resuscitation is underestimated by conventional chest X-ray. *Resuscitation, 60*(2), 157-162.

Lee, J. C., Boyd, R., & Stuart, P. (2007). Randomized controlled trial of an instructional DVD for clinical skills teaching. *Emergency Medicine Australasia, 19*(3), 241-245.
Lemberg, I., Kirchberger, I., Stucki, G., & Cieza, A. (2010). The ICF Core Set for stroke from the perspective of physicians: a worldwide validation study using the Delphi technique. *European journal of physical & rehabilitation medicine., 46*(3), 377-388.
Lewis, P., & Wright, C. (2015). Saving the critically injured trauma patient: a retrospective analysis of 1000 uses of intraosseous access. *Emergency Medicine Journal, 32*(6), 463-467.
Lincoln, Y. S., & Guba, E. G. (1985). *Naturalistic inquiry* (Vol. 75): Sage.
Lincoln, Y. S., Lynham, S. A., & Guba, E. G. (2011). Paradigmatic controversies, contradictions, and emerging confluences, revisited. *The Sage handbook of qualitative research, 4*, 97-128.
LMA Classic Excel™. In T. L. M. C. Ltd (Ed.).
LMA Classic™, LMA Flexible™, LMA Flexible™ Single Use & LMA Unique™. In T. L. M. C. Ltd (Ed.).
Logue, M. D., & Effken, J. A. (2013). Validating the personal health records adoption model using a modified e-Delphi. *Journal of advanced nursing, 69*(3), 685-696.
Lund, F., Schultz, J.-H., Maatouk, I., Krautter, M., Möltner, A., Werner, A., . . . Nikendei, C. (2012). Effectiveness of IV cannulation skills laboratory training and its transfer into clinical practice: a randomized, controlled trial. *PloS one, 7*(3), e32831.
Mackety, C. K. (1973). Orientation program for the new OR nurse. *AORN Journal, 17*(5), 88-90.
Maconochie, I. K., Bingham, R., Eich, C., López-Herce, J., Rodríguez-Núnez, A., Rajka, T., . . . Nolan, J. P. (2015). European Resuscitation Council Guidelines for Resuscitation 2015: Section 6. Paediatric life support. *Resuscitation, 95*, 223-248.
Mazzio, C. (2005). The sensess divided *Empire of the senses: The sensual culture reader* (pp. 85-105).
McKenna, H. P. (1994). The Delphi technique: a worthwhile research approach for nursing? *Journal of advanced nursing, 19*(6), 1221-1225.
Milan, F. B., Parish, S. J., & Reichgott, M. J. (2006). A Model for Educational Feedback Based on Clinical Communication Skills Strategies: Beyond the" Feedback Sandwich". *Teaching and learning in medicine, 18*(1), 42-47.
Miller, G. A. (1956). The magical number seven, plus or minus two: some limits on our capacity for processing information. *Psychological review, 63*(2), 81.
Miller, G. E. (1990). The assessment of clinical skills/competence/performance. *Academic Medicine, 65*(9), S63-67.
Mok, M., & Ker, J. (2015). Developing clinical skills bundles. *The clinical teacher, 12*(6), 403-407.
Moore, C. M. (1987). *Group techniques for idea building*: Sage Publications, Inc.
Murphy, M. A., Neequaye, S., Kreckler, S., & Hands, L. J. (2008). Should we train the trainers? Results of a randomized trial. *Journal of the American College of Surgeons, 207*(2), 185-190.
Naik, V. N., & Brien, S. E. (2013). Review article: simulation: a means to address and improve patient safety. *Canadian Journal of Anesthesia/Journal canadien d'anesthésie, 60*(2), 192-200.
Nelson, M. S. (1989). Medical student retention of intubation skills. *Annals of Emergency Medicine, 18*(10), 1059-1061. doi:http://dx.doi.org/10.1016/S0196-0644(89)80931-X
Newton, J. M., Billett, S., Jolly, B., & Ockerby, C. M. (2009). Lost in translation: barriers to learning in health professional clinical education. *Learning in Health and Social Care, 8*(4), 315-327.
Nicolis, J. S., & Tsuda, I. (1985). Chaotic dynamics of information processing: The "magic number seven plus-minus two" revisited. *Bulletin of Mathematical Biology, 47*(3), 343-365. doi:http://dx.doi.org/10.1016/S0092-8240(85)90031-X
Nikendei, C., Huber, J., Stiepak, J., Huhn, D., Lauter, J., Herzog, W., . . . Krautter, M. (2014). Modification of Peyton's four-step approach for small group teaching - a descriptive study. *BMC Medical Education, 14*, 68.
Noblat, L., Oliveira, M. G., Santos-Jesus, R., Noblat, A. C., & Badaro, R. (2006). Validation of criteria for nosocomial use of amikacin in Brazil with the Delphi technique. *Brazilian Journal of Infectious Diseases, 10*(3), 173-178.

Nolan, A., Barry, S., Burke, S., & Thomas, S. (2014). The impact of the financial crisis on the health system and health in Ireland. *WHO/European Observatory on Health Systems and Policies Case Study*.
Norcini, J. J. (2003). Setting standards on educational tests. *Medical education, 37*(5), 464-469.
Oermann, M. H., Kardong-Edgren, S., Odom-Maryon, T., Hallmark, B. F., Hurd, D., Rogers, N., . . . Dowdy, S. W. (2011). Deliberate practice of motor skills in nursing education: CPR as exemplar. *Nursing education perspectives, 32*(5), 311-315.
Oermann, M. H., Saewert, K. J., Charasika, M., & Yarbrough, S. S. (2009). Assessment and grading practices in schools of nursing: national survey findings part I. *Nursing education perspectives, 30*(5), 274-278.
Oermann, M. H., Yarbrough, S. S., Saewert, K. J., Ard, N., & Charasika, M. (2009). Clinical evaluation and grading practices in schools of nursing: National survey findings part II. *Nursing education perspectives, 30*(6), 352-357.
Oestergaard, J., Bjerrum, F., Maagaard, M., Winkel, P., Larsen, C. R., Ringsted, C., . . . Soerensen, J. L. (2012). Instructor feedback versus no instructor feedback on performance in a laparoscopic virtual reality simulator: a randomized educational trial. *BMC Medical Education, 12*(1), 7.
Oksan, D., & Ayfer, K. (2013). Powered intraosseous device (EZ-IO) for critically ill patients. *Indian pediatrics, 50*(7), 689-691.
Orde, S., Celenza, A., & Pinder, M. (2010). A randomised trial comparing a 4-stage to 2-stage teaching technique for laryngeal mask insertion. *Resuscitation, 81*(12), 1687-1691.
Oriot, D., Darrieux, E., Boureau-Voultoury, A., Ragot, S., & Scépi, M. (2012). Validation of a Performance Assessment Scale for Simulated Intraosseous Access. *Simulation in Healthcare, 7*(3), 171-175. doi:10.1097/SIH.0b013e31824a5c20
Ota, F. S., Yee, L. L., Garcia, F. J., Grisham, J. E., & Yamamoto, L. G. (2003). Which IO model best simulates the real thing? *Pediatr Emerg Care, 19*(6), 393-396.
Owen, H., & Plummer, J. L. (2002). Improving learning of a clinical skill: the first year's experience of teaching endotracheal intubation in a clinical simulation facility. *Medical education, 36*(7), 635-642. doi:10.1046/j.1365-2923.2002.01260.x
Oyebode, F. (2013). Clinical errors and medical negligence. *Medical Principles and Practice, 22*(4), 323-333.
Palter, V. N., Graafland, M., Schijven, M. P., & Grantcharov, T. P. (2012). Designing a proficiency-based, content validated virtual reality curriculum for laparoscopic colorectal surgery: a Delphi approach. *Surgery, 151*(3), 391-397.
Parratt, J. A., Fahy, K. M., Hutchinson, M., Lohmann, G., Hastie, C. R., Chaseling, M., & O'Brien, K. (2016). Expert validation of a teamwork assessment rubric: A modified Delphi study. *Nurse education today, 36*, 77-85.
Pasley, J., Miller, C. H., DuBose, J. J., Shackelford, S. A., Fang, R., Boswell, K., . . . Matsuura, M. (2015). Intraosseous infusion rates under high pressure: A cadaveric comparison of anatomic sites. *Journal of Trauma and Acute Care Surgery, 78*(2), 295-299.
Patton, M. Q. (2002a). Particularly appropriate qualitative applications. In M. Q. Patton (Ed.), *Qualitative research & evaluation methods.* (3a Ed. ed.). Thousand Oaks/London/New Delphi: : Sage Publications.
Patton, M. Q. (2002b). *Qualitative evaluation and research methods* (3rd ed.): SAGE Publications, inc.
Paxton, J. H. (2012). Intraosseous vascular access: A review. *Trauma, 14*(3), 195-232. doi:10.1177/1460408611430175
Peyton, R. (1998). *Teaching and learning in medical practice*: Manticore Europe.
Peyton, R. (2008). Best Practice in Surgical Education. In D. P. Saltzman (Ed.), (Vol. 2015): Ethicon Endo-surgery.
Phillips, L., Brown, L., Campbell, T., Miller, J., Proehl, J., & Youngberg, B. (2010a). Consortium on Intraosseous Vascular Access in Healthcare Practice: Recommendations for the use of

intraosseous vascular access for emergent and nonemergent situations in various health care settings: a consensus paper. *Critical care nurse, 30*(6), e1-e7.
Phillips, L., Brown, L., Campbell, T., Miller, J., Proehl, J., & Youngberg, B. (2010b). Recommendations for the use of intraosseous vascular access for emergent and nonemergent situations in various healthcare settings: a consensus paper. *J Emerg Nurs, 36*(6), 551-556. doi:10.1016/j.jen.2010.09.001
Prideaux, D., Worley, P., & Bligh, J. (2007). Symbiosis: a new model for clinical education. *The clinical teacher, 4*(4), 209-212.
Rapp, R. S., & Queri, S. (2016). [Validation of Personal Factors in Breast Cancer Rehabilitation - A National Delphi Survey]. *Gesundheitswesen, 7*, 7.
Resuscitation Council UK. (2008). *Pocket guide to teaching for medical instructors* (I. Bullock, M. Davis, A. Lockey, & K. Mackway-Jones Eds.): Blackwell Publishing Oxford.
Resuscitation Council UK. (2011). Generic Instructor Course Regulations. In A. L. S. Group (Ed.), (pp. 6).
Reznick, R. K., & MacRae, H. (2006). Teaching Surgical Skills — Changes in the Wind. *New England Journal of Medicine, 355*(25), 2664-2669. doi:doi:10.1056/NEJMra054785
Riener, C., & Willingham, D. (2010). The myth of learning styles. *Change: The magazine of higher learning, 42*(5), 32-35.
Riesenberg, L. A., Berg, K., Berg, D., Davis, J., Schaeffer, A., Justice, E. M., & Tinkoff, G. (2013). The Development of a Validated Checklist for Femoral Venous Catheterization Preliminary Results. *American Journal of Medical Quality, 29*(5), 445-450.
Riesenberg, L. A., Berg, K., Berg, D., Mealey, K., Weber, D., King, D., . . . Tinkoff, G. (2013). The Development of a Validated Checklist for Paracentesis: Preliminary Results. *American Journal of Medical Quality, 28*(3), 227-231. doi:10.1177/1062860612460399
Riesenberg, L. A., Berg, K., Berg, D., Schaeffer, A., Mealey, K., Davis, J., . . . Tinkoff, G. (2013). The development of a validated checklist for nasogastric tube insertion: preliminary results. *Am J Med Qual, 28*(5), 429-433. doi:10.1177/1062860612474488
Roberts, L. (2005). The death of the sensuous chemist *Empire of the senses: The sensual culture reader* (pp. 106-127).
Rosenberg, H., & Cheung, W. J. (2013). Intraosseous access. *Canadian Medical Association Journal, 185*(5), E238-E238.
Rosetti, V. A., Thompson, B. M., Miller, J., Mateer, J. R., & Aprahamian, C. (1985). Intraosseous infusion: an alternative route of pediatric intravascular access. *Annals of Emergency Medicine, 14*(9), 885-888.
Ryle, G. (1945). *Knowing how and knowing that: The presidential address.* Paper presented at the Proceedings of the Aristotelian society.
Saaty, T. L., & Ozdemir, M. S. (2003). Why the magic number seven plus or minus two. *Mathematical and Computer Modelling, 38*(3–4), 233-244. doi:http://dx.doi.org/10.1016/S0895-7177(03)90083-5
Sacks, O. (2005). The mind's eye: What the blind see *Empire of the senses: The sensual culture reader* (pp. 25-42).
Saewert, K. (2013). Socialization to Professional Nursing. In E. E. Friberg & J. L. Creasia (Eds.), *Conceptual Foundations: The Bridge to Professional Nursing Practice* (pp. 42-70): Elsevier Health Sciences.
Sahlberg, P. (2011a). *Finnish lessons*: Teachers College Press.
Sahlberg, P. (2011b). The fourth way of Finland. *Journal of Educational Change, 12*(2), 173-185.
Sahlberg, P. (2012). How GERM is infecting schools around the world. *The Washington Post, 29*.
Salamonson, Y., & Andrew, S. (2006). Academic performance in nursing students: influence of part-time employment, age and ethnicity. *Journal of advanced nursing, 55*(3), 342-349. doi:10.1111/j.1365-2648.2006.03863_1.x

Santos, D., Carron, P.-N., Yersin, B., & Pasquier, M. (2013). EZ-IO® intraosseous device implementation in a pre-hospital emergency service: A prospective study and review of the literature. *Resuscitation, 84*(4), 440-445.
Sawyer, T., White, M., Zaveri, P., Chang, T., Ades, A., French, H., . . . Kessler, D. (2015). Learn, see, practice, prove, do, maintain: an evidence-based pedagogical framework for procedural skill training in medicine. *Academic Medicine, 90*(8), 1025-1033.
Schalk, R., Schweigkofler, U., Lotz, G., Zacharowski, K., Latasch, L., & Byhahn, C. (2011). Efficacy of the EZ-IO® needle driver for out-of-hospital intraosseous access-a preliminary, observational, multicenter study. *Scand J Trauma Resusc Emerg Med, 19*, 65.
Schuwirth, L., & Ash, J. (2013). Principles of Assessment. In K. Walsh (Ed.), *Oxford Textbook of Medical Education*: Oxford University Press.
Schuwirth, L., & Van der Vleuten, C. (2011). Programmatic assessment: from assessment of learning to assessment for learning. *Medical teacher, 33*(6), 478-485.
Schuwirth, L., & van der Vleuten, C. (2012). Programmatic assessment and Kane's validity perspective. *Medical education, 46*(1), 38-48.
Sermeus, W., Gillet, P., Gillain, D., Grietens, J., Laport, N., Michiels, D., . . . Van den Heede, K. (2009). Development and validation of nursing resource weights for the Belgian Nursing Minimum Dataset in general hospitals: a Delphi questionnaire survey approach. *International journal of nursing studies, 46*(2), 256-267.
Shadish, W. R. (1995). Philosophy of science and the quantitative-qualitative debates: Thirteen common errors. *Evaluation and Program Planning, 18*(1), 63-75.
Smith, F. C., & Greenwood, S. R. (2012). Modern ways to enhance surgical teaching skills. *Surgery (Oxford), 30*(9), 471-476.
Soar, J., Nolan, J. P., Böttiger, B. W., Perkins, G. D., Lott, C., Carli, P., . . . Smith, G. B. (2015). European Resuscitation Council Guidelines for Resuscitation 2015: Section 3. Adult advanced life support.
Soifer, N. E., Borzak, S., Edlin, B. R., & Weinstein, R. A. (1998). Prevention of peripheral venous catheter complications with an intravenous therapy team: a randomized controlled trial. *Archives of internal medicine, 158*(5), 473-477.
Sopka, S., Biermann, H., Druener, S., Skorning, M., Knops, A., Fitzner, C., . . . Beckers, S. (2012). Practical skills training influences knowledge and attitude of dental students towards emergency medical care. *European Journal of Dental Education, 16*(3), 179-186.
Søreide, E., Morrison, L., Hillman, K., Monsieurs, K., Sunde, K., Zideman, D., . . . Soar, J. (2013). The formula for survival in resuscitation. *Resuscitation, 84*(11), 1487-1493.
Sorensen, N. T. (2014). *Improvisation and teacher expertise: a comparative case study.* Bath Spa University.
Spearpoint, K. G., McLean, C. P., & Zideman, D. A. (2000). Early defibrillation and the chain of survival in 'in-hospital'adult cardiac arrest; minutes count. *Resuscitation, 44*(3), 165-169.
Squires, D. (2014). The global slowdown in health care spending growth. *Jama, 312*(5), 485-486.
Steen, H. L. (2008). Effective eLearning design. *Journal of Online Learning and Teaching, 4*(4), 526-532.
Stevens, R., & Davies, M. (2012). "Do one, teach one": the new paradigm in general surgery residency training. *Journal of surgical education*.
Stewart, S. (2005). Remembering the senses *Empire of the senses: The sensual culture reader* (pp. 59-69).
Streiner, D. L. (2003). Starting at the beginning: an introduction to coefficient alpha and internal consistency. *Journal of personality assessment, 80*(1), 99-103.
Sullivan, G. M. (2011). Getting off the "gold standard": randomized controlled trials and education research. *Journal of graduate medical education, 3*(3), 285-289.
Sweller, J., Van Merrienboer, J. J., & Paas, F. G. (1998). Cognitive architecture and instructional design. *Educational psychology review, 10*(3), 251-296.

Tettamanti, M., Buccino, G., Saccuman, M. C., Gallese, V., Danna, M., Scifo, P., . . . Perani, D. (2005). Listening to action-related sentences activates fronto-parietal motor circuits. *Journal of cognitive neuroscience, 17*(2), 273-281.
Thomas, I. (2012). Improving suturing workshops using modern educational theory. *The clinical teacher, 9*(3), 137-142.
Thomson, S., Figueras, J., Evetovits, T., Jowett, M., Mladovsky, P., Maresso, A., . . . Kluge, H. (2014). *Economic crisis, health systems and health in Europe: impact and implications for policy*: WHO Regional Office for Europe.
Toouli, J. (2006). What Is the role of "Live Surgical Demonstrations" at conferences? *HPB: the official journal of the International Hepato Pancreato Biliary Association, 8*(3), 163.
Tuovinen, J. E., & Sweller, J. (1999). A comparison of cognitive load associated with discovery learning and worked examples. *Journal of educational psychology, 91*(2), 334.
Turner, N. M., Lukkassen, I., Bakker, N., Draaisma, J., & ten Cate, O. T. J. (2009). The effect of the APLS-course on self-efficacy and its relationship to behavioural decisions in paediatric resuscitation. *Resuscitation, 80*(8), 913-918. doi:http://dx.doi.org/10.1016/j.resuscitation.2009.03.028
Van de Walle, D. (2004). A goal orientation model of feedback-seeking behavior. *Human Resource Management Review, 13*(4), 581-604.
Van der Vleuten, C. (2000). Validity of final examinations in undergraduate medical training. *BMJ: British Medical Journal, 321*(7270), 1217.
Van der Vleuten, C., & Schuwirth, L. (2005). Assessing professional competence: from methods to programmes. *Medical education, 39*(3), 309-317.
Van der Vleuten, C., Schuwirth, L., Driessen, E., Dijkstra, J., Tigelaar, D., Baartman, L., & Van Tartwijk, J. (2012). A model for programmatic assessment fit for purpose. *Medical teacher, 34*(3), 205-214.
Vnuk, A. K. (2013). *Going through the motions: Medical students' experiences of learning physical examinations.* (Doctor of Education), Flinders University, Adelaide.
Von Hoff, D. D. (1991). Intraosseous infusions: an important but forgotten method of vascular access. *Cancer investigation, 9*(5), 521-528.
Walker, M., & Peyton, J. (1998). Teaching in the Theatre. In J. Peyton (Ed.), *Teaching and learning in medical practice* (pp. 216): Heronsgate Rickmansworth: Manticore Europe.
Wall, D. (1999). Twelve tips on teaching the consultant teachers to teach. *Medical teacher, 21*(4), 387-392.
Walters, L., Greenhill, J., Richards, J., Ward, H., Campbell, N., Ash, J., & Schuwirth, L. (2012). Outcomes of longitudinal integrated clinical placements for students, clinicians and society. *Medical education, 46*(11), 1028-1041.
Walters, L., Prideaux, D., Worley, P., & Greenhill, J. (2011). Demonstrating the value of longitudinal integrated placements to general practice preceptors. *Medical education, 45*(5), 455-463.
Walters, L., Prideaux, D., Worley, P., Greenhill, J., & Rolfe, H. (2009). What do general practitioners do differently when consulting with a medical student? *Medical education, 43*(3), 268-273. doi:10.1111/j.1365-2923.2008.03276.x
Wang, T., Schwartz, J., Karimipour, D., Orringer, J., Hamilton, T., & Johnson, T. (2004). An Education Theory-Based Method to Teach a Procedural Skill. *Archives of Dermatology, 140*(11), 1357-1361.
Wayne, D. B., Butter, J., Siddall, V. J., Fudala, M. J., Wade, L. D., Feinglass, J., & Mcgaghie, W. C. (2006). Graduating internal medicine residents' self-assessment and performance of advanced cardiac life support skills. *Medical teacher, 28*(4), 365-369.
Wayne, D. B., Siddall, V. J., Butter, J., Fudala, M. J., Wade, L. D., Feinglass, J., & McGaghie, W. C. (2006). A longitudinal study of internal medicine residents' retention of advanced cardiac life support skills. *Academic Medicine, 81*(10), S9-S12.

Wearne, S. (2011). Teaching procedural skills in general practice. *Australian family physician, 40*(1-2), 63.
Whelan, R. R. (2007). Neuroimaging of cognitive load in instructional multimedia. *Educational Research Review, 2*(1), 1-12. doi:http://dx.doi.org/10.1016/j.edurev.2006.11.001
Wiles, M. (2015). ATLS: Archaic Trauma Life Support? *Anaesthesia, 70*(8), 893-897.
Williams, N. M. (2011). Advanced life support training and assessment: A literature review. *Australasian Emergency Nursing Journal, 14*(4), 240-245. doi:http://dx.doi.org/10.1016/j.aenj.2011.07.001
Windle, J. (2004). The ethnic (dis) advantage debate revisited: Turkish background students in Australia. *Journal of Intercultural Studies, 25*(3), 271-286.
Winkler, A., Spillmann, L., Werner, J., & Webster, M. (2015). Asymmetries in blue–yellow color perception and in the color of 'the dress'. *Current Biology, 25*(13), R547-R548.
Winocur, G., & Hasher, L. (2004). Age and time-of-day effects on learning and memory in a non-matching-to-sample test. *Neurobiology of aging, 25*(8), 1107-1115.
Wisher, R., Sabol, M., Ellis, J., & Ellis, K. (1999). *Staying sharp: Retention of military knowledge and skills*. Retrieved from
Wong, A., Marcus, N., Ayres, P., Smith, L., Cooper, G., Paas, F., & Sweller, J. (2009). Instructional animations can be superior to statics when learning human motor skills. *Computers in Human Behavior, 25*(2), 339-347. doi:http://dx.doi.org/10.1016/j.chb.2008.12.012
Worley, P. (2002a). Integrity: the key to quality in community-based medical education?(Part two). *EDUCATION FOR HEALTH-ABINGDON-CARFAX PUBLISHING LIMITED-, 15*(2), 129-138.
Worley, P. (2002b). Relationships: a new way to analyse community-based medical education?(Part one). *EDUCATION FOR HEALTH-ABINGDON-CARFAX PUBLISHING LIMITED-, 15*(2), 117-128.
Worley, P. (2003). *The immediate academic impact on medical students of basing an entire clinical year in rural general practice.* (Ph.D), Flinders University.
Worley, P., Prideaux, D., Strasser, R., Magarey, A., & March, R. (2006). Empirical evidence for symbiotic medical education: a comparative analysis of community and tertiary-based programmes. *Medical education, 40*(2), 109-116. doi:10.1111/j.1365-2929.2005.02366.x
Yardley, S., Teunissen, P., & Dornan, T. (2012). Experiential learning: AMEE guide No. 63. *Medical teacher, 34*(2), e102-e115.
Zhu, F., Poolton, J., Wilson, M., Hu, Y., Maxwell, J., & Masters, R. (2011). Implicit motor learning promotes neural efficiency during laparoscopy. *Surgical endoscopy, 25*(9), 2950-2955.
Zollo, J. (1998). Reflective practice in nurse education: a step towards equity in education and health care. *Collegian, 5*(3), 28-33. doi:http://dx.doi.org/10.1016/S1322-7696(08)60298-X
Zwaan, L., Monteiro, S., Sherbino, J., Ilgen, J., Howey, B., & Norman, G. (2016). Is bias in the eye of the beholder? A vignette study to assess recognition of cognitive biases in clinical case workups. *BMJ quality & safety*, bmjqs-2015-005014.