

Physical Education teachers' assessment of invasion games and sports in Victorian secondary schools: A mixed methods study

By

David Gow

BEd, MEd

Thesis

Submitted to Flinders University

for the degree of

Doctor of Philosophy

College of Education, Psychology and Social Work, Flinders University

February 8, 2024

TABLE OF CONTENTS

TABLE OF CONTENTS	I
ABSTRACT	V
DECLARATION	VII
ACKNOWLEDGEMENTS	VIII
LIST OF FIGURES	IX
LIST OF TABLES	X
CHAPTER 1: INTRODUCTION	1
1.1 Introduction.....	1
1.1.1 Physical Education (PE), Games and Sports (GS)	2
1.2 The Three Message Systems	5
1.2.1 Curriculum.....	6
1.2.2 Pedagogy.....	9
1.2.3 Assessment	13
1.3 Research Question and Aim	15
1.4 Positionality.....	16
CHAPTER 2: ASSESSMENT IN PHYSICAL EDUCATION (PE) – A LITERATURE REVIEW	21
2.1 Introduction.....	21
2.2 An Overview of Assessment Concepts	22
2.2.1 Over-Conceptualised and Over-Complicated	22
2.2.2 Assessment, Feedback and Reporting	25
2.2.3 The Purpose of Assessment	27
2.2.4 Validity and Reliability.....	29
2.3 An Overview of Assessment Frameworks	31
2.3.1 Assessment Literacy.....	32
2.3.2 Quality Assessment.....	33
2.3.3 Authentic Assessment.....	35
2.4 Challenges of Assessment in PE	37
2.4.1 Challenges of Assessment in Games and Sports (GS).....	39
2.5 Conclusion and Further Research	43
CHAPTER 3: RESEARCH DESIGN	45
3.1 Introduction.....	45
3.1.1 Research Overview	45
3.2 Research Paradigm.....	48
3.2.1 Theoretical Lens (Perspective).....	48

3.2.2	Methodology.....	49
3.3	Conclusion	53
CHAPTER 4: SCOPING REVIEW (SCR) OF PEER-REVIEWED EVIDENCE-BASED ASSESSMENT TOOLS IN INVASION GAMES AND SPORTS (IGS) WITHIN A SCHOOL-AGE POPULATION		55
4.1	Introduction.....	55
4.1.1	Aim	55
4.2	Method.....	56
4.2.1	Eligibility Criteria and Guidelines	56
4.2.2	Search Strategy and Study Selection	57
4.2.3	Data Charting and Collating	61
4.3	Results	65
4.4	Discussion.....	77
4.4.1	Assessment Tool Characteristics.....	77
4.4.2	Assessment Tool Key Performance Criteria.....	81
4.4.3	Assessment Tool Applications.....	83
4.5	Recommendations	86
4.6	Strengths and Limitations	87
4.7	Conclusion and Further Research	89
CHAPTER 5: PHYSICAL EDUCATION TEACHERS’ ASSESSMENT PRACTICES IN INVASION GAMES AND SPORTS IN VICTORIAN SECONDARY SCHOOLS: A CROSS-SECTIONAL, QUANTITATIVE INQUIRY.....		92
5.1	Introduction.....	92
5.1.1	Aim	92
5.1.2	An Overview of Survey Research in PE	93
5.2	Method.....	94
5.2.1	Participants and Recruitment	94
5.2.2	Ethics	96
5.2.3	Data Collection	96
5.2.4	Data Analysis	101
5.3	Results	102
5.4	Discussion.....	111
5.4.1	Curriculum Alignment to Assessment.....	112
5.4.2	Assessment Tools, Users and Utility	114
5.4.3	Key Performance Criteria	119
5.4.4	Familiarity and use of the GPAI and TSAP	123
5.5	Recommendations	124
5.6	Strengths and Limitations	126

5.7	Conclusion and Further Research	127
CHAPTER 6: PHYSICAL EDUCATION TEACHERS' ASSESSMENT PRACTICES IN INVASION GAMES AND SPORTS IN VICTORIAN SECONDARY SCHOOLS: A QUALITATIVE INQUIRY		
128		
6.1	Introduction.....	128
6.1.1	Aim	128
6.2	Method.....	129
6.2.1	Participants and Sample	129
6.2.2	Ethics.....	131
6.2.3	Data Collection.....	131
6.2.4	Data Analysis.....	135
6.2.5	Rigor	141
6.3	Findings and Discussion.....	145
6.3.1	The Power of Accountability in Assessment.....	145
6.3.2	The Problems with Unilateral Assessment Practice	150
6.3.3	Students – the Missing Ingredient in Assessment	154
6.3.4	PE Teachers, Heal Thyself – Assessment Ideologies.....	158
6.4	Recommendations	164
6.5	Strengths and Limitations	166
6.6	Conclusions and Further Research.....	167
CHAPTER 7: A DOCUMENT ANALYSIS OF RUBRICS USED BY PHYSICAL EDUCATION TEACHERS IN INVASION GAMES AND SPORTS IN VICTORIAN SECONDARY SCHOOLS: A QUALITATIVE INQUIRY		
169		
7.1	Introduction.....	169
7.1.1	Aim	170
7.1.2	An Overview of Rubrics.....	170
7.2	Method.....	172
7.2.1	Participants and Sample	173
7.2.2	Ethics.....	175
7.2.3	Data Collection and Analysis.....	175
7.2.4	Rigor	181
7.3	Findings and Discussion.....	184
7.3.1	A Narrow Conceptualisation of Rubrics.....	184
7.3.2	Rubrics are for Reporting.....	193
7.3.3	Appropriate Criteria are Key	199
7.3.4	The Challenge of Describing Quality	207
7.4	Recommendations	215
7.5	Strengths and Limitations	217

7.6	Conclusion and Further Research	218
CHAPTER 8: CONCLUSION		220
8.1	Introduction.....	220
8.2	A Summary of Key Findings	220
8.2.1	Research Sub-Question 1	221
8.2.2	Research Sub-Question 2.....	223
8.2.3	Research Sub-Question 3.....	224
8.2.4	Research Sub-Question 4.....	226
8.3	Translation of Findings	228
8.3.1	Invasion Games and Sports Assessment Framework	228
8.3.2	The Rubric Design Framework	232
8.4	Recommendations	233
8.4.1	For Researchers.....	233
8.4.2	For Policy Makers.....	234
8.4.3	For PE Practitioners.....	236
8.5	Strengths and Limitations	237
8.6	Conclusion	238
REFERENCES		240
APPENDIX A: ETHICS APPROVAL AND MODIFICATION		264
APPENDIX B: CHARTING GUIDELINES FOR SCR (CHAPTER 4)		265
APPENDIX C: BIBLIOGRAPHIC DETAILS FOR INCLUDED STUDIES IN SCR (CHAPTER 4)		269
APPENDIX D: PRISMA – SCR EXTENSION CHECKLIST (CHAPTER 4)		273
APPENDIX E: AN OVERVIEW OF EVIDENCE-BASED ASSESSMENT TOOLS DEVELOPED IN A SCHOOL CONTEXT (CHAPTER 4)		275
APPENDIX F: QUESTIONNAIRE (CHAPTER 5)		276
APPENDIX G: SUPPORTING DOCUMENTATION (CHAPTER 5)		291
APPENDIX H: INTERVIEW SCHEDULE (CHAPTER 6)		297
APPENDIX I: STIMULUS MATERIAL (CHAPTER 6)		298
APPENDIX J: SAMPLE MEMOS (CHAPTER 6)		302
APPENDIX K: SUPPORTING DOCUMENTATION (CHAPTER 6)		303
APPENDIX L: SAMPLE MEMOS (CHAPTER 7)		307
APPENDIX M: SUPPORTING DOCUMENTATION (CHAPTER 7)		309
APPENDIX N: DESIGN ELEMENTS OF RUBRICS (CHAPTER 7)		314
APPENDIX O: CODING GUIDELINES (CHAPTER 7)		315
APPENDIX P: TRANSLATION OF FINDINGS (CHAPTER 8)		319

ABSTRACT

This thesis describes, explains and makes meaning of Physical Education teachers' understanding and practice of assessment. The uniqueness of the study is underpinned by its focus on a population of Victorian, secondary school teachers, and a context that holds a historically privileged position in most PE programs, invasion games and sports. The thesis employs an interpretive theoretical lens and a functional pragmatic worldview to create knowledge to further understanding with the intent to improve assessment practice. To respond to the overarching research question and aim, an explanatory, sequential, mixed methods approach is employed; this research design encompasses a Scoping Review, survey design and document analysis.

To provide a reference point for the assessment of invasion games and sports, the peer-reviewed literature is scoped to identify contextually relevant evidence-based assessment tools. This Scoping Review informs a two-phase survey design comprising a cross-sectional, quantitative inquiry of Physical Education teachers in Victorian secondary schools, and a qualitative inquiry to describe and explain participant understanding and practice of assessment in invasion games and sports. The thesis concludes with a document analysis of rubrics used by a nested sample of the population, with the aim of understanding how rubrics are constructed, and their degree of alignment to evidence-based criteria and the subject of Physical Education within the Victorian Curriculum.

The key findings of the thesis include the prevalence of evidence-based assessment tools that generate outcomes based on frequency-counts in the Scoping Review. This contrasts with the limited awareness and use of such tools by the sample of Physical Education teachers. There is congruence in the use of key performance criteria located in the Scoping Review and those indicated by the sample of teachers, however, the latter group understand rubrics to be the most useful tool in their assessment practice. Teachers generally use assessment to report student achievement to a parent audience rather than to identify the next steps in learning for students. Participants are inconsistent in their alignment of assessment to curriculum and describe the passive role of students throughout the assessment process; this includes the limited use of peer and self-assessment. The rubrics used by the sample are relatively uniform in their construction and thus considered narrow in their conceptualisation, they commonly include subjective criteria

like effort and wearing uniform, and all rubrics employ scale-type language that is reported as a key limitation of rubric utility.

In addition to the above findings, the significant and original contribution to knowledge resulting from this thesis includes the identification of 32 unique, evidence-based tools from the Scoping Review study, the subsequent charting of tool characteristics and their applications, and a series of recommendations aimed at improving assessment practices for the sample population. The research makes meaning of the understanding and practice of assessment by Physical Education teachers in Victorian secondary schools, giving voice to a largely silent community and context, and filling a gap in the literature.

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

(signature removed under advice from the Office of Graduate Research)

Date. February 8, 2024

ACKNOWLEDGEMENTS

The completion of this doctorate would not have been possible without the support, wisdom and generosity of a great number of people. I acknowledge that I am incredibly lucky and privileged to have been able to undertake this study.

To my supervisors; words seem inadequate to thank you for seven years of ongoing encouragement, intellectual provocation and mentorship. To Associate Professor Sam Elliott, I will forever be reminded that my pursuit of this PhD was as much about developing robust research skills as answering a compelling question. I cannot hear the word rigor without also seeing your face. To Dr. Kate Ridley, your ability to provide precise, concise, quality feedback was unparalleled. No matter how well I thought I had captured an idea, you offered insight to refine it. To Associate Professor Shane Pill, your unwavering support, complemented by constant reminders that this thesis was *my* thesis, sustained me throughout this journey. You gave me the freedom and scaffolding to wander down a myriad of rabbit holes, before expertly pulling me back up to a path of my choosing. I would also like to thank Dr. Lisa Young who verified my coding in the first study of the thesis and Dr. Helen Woodford for her careful and thorough editing of my thesis.

This thesis relied on the generosity of the managers of three recruitment organisations (Hilary at ACHPER Victoria branch; Rob at Peak Phys Ed; and Jarrod at ConnectedPE) to make the initial approach to participants; I acknowledge their support. To the participants, particularly those that contributed to multiple studies, I hope that I have honoured your voices throughout the studies. To my PE teacher colleagues that provided a sounding board for my developing ideas and piloted various studies; I am eternally thankful.

I am truly grateful for the love, support and understanding of my family over the duration of this journey. To my wife, Sarah, and my daughters, Matilda and Charlotte, thank you for sharing me with this PhD. I owe you all a great deal of quality time. To my mother, Rosemary, for always supporting my curious nature and love of learning. Your advice to get that 'piece of paper' through my undergraduate years at University has never left me. Lastly, I dedicate this thesis to every student that has ever picked up a ball in PE and wanted to get better at playing games.

LIST OF FIGURES

Figure 1.1 Teaching Games for Understanding (TGfU) model.....	11
Figure 2.1 Principles of Assessment in HPE	31
Figure 3.1 Levels of the Research Process in this Thesis	47
Figure 3.2 A Flow Chart of the Explanatory, Sequential, Mixed Methods Study Design	49
Figure 3.3 Positioning of this Research Design	52
Figure 4.1 Flow Chart for the PRISMA-ScR Extension.....	60
Figure 5.1 The Relative Usefulness of Indicators of Assessment Quality	119
Figure 7.1 Rubric 22 – A Generic Rubric for Invasion Games and Sports.....	189
Figure 7.2 Rubric 19 – A Generic Rubric for HPE	190
Figure 7.3 Rubric 10 – A Sport-Specific Rubric for Football (Australian)	192
Figure 7.4 Rubric 23 – An Instructional Rubric	198
Figure 7.5 Rubric 3 (Excerpt) - Criteria Linked to the VC: HPE.....	201
Figure 7.6 Rubric 4 (Excerpt) - Sample Qualitative and Evaluative Descriptors	210
Figure 7.7 Rubric 21 – A Novel Rubric.....	212
Figure 7.8 Rubric 24 – Inconsistent and Incongruent Descriptors	214

LIST OF TABLES

Table 1.1 Sample Invasion Games Rubric	19
Table 2.1 Ambiguous Terms Used in Communicating and Reporting Student Learning	23
Table 2.2 Sample References to Games and Sports in the VC: HPE	27
Table 3.1 Overview of the Research Sub-Questions and Emergent Design of the Thesis	46
Table 4.1 Working Definitions of Key Terms for the Purpose of this Study	57
Table 4.2 Search Strategy for this ScR	59
Table 4.3 Summary of Data Items and Definitions in the Scoping Review.....	62
Table 4.4 Journal Distribution of the Included Studies.....	66
Table 4.5 Bibliographic Characteristics of the Included Studies.....	67
Table 4.6 Assessment Tool Characteristics	68
Table 4.7 Assessment Tool Key Performance Criteria	70
Table 4.8 Assessment Tool Applications.....	72
Table 5.1 Overview of the Questions Comprising the Questionnaire	99
Table 5.2 Participant Demographics	103
Table 5.3 Curriculum Alignment to Assessment.....	104
Table 5.4 The Types of Assessment Tools Used and Users (Teacher, Self, Peer).....	105
Table 5.5 The Most Useful Assessment Tool	106
Table 5.6 The Utility of Select Aspects of Assessment	107
Table 5.7 Key Performance Criteria	108
Table 5.8 Select On-The-Ball Skills Assessed.....	109
Table 5.9 Familiarity and Use of the GPAI and TSAP	110
Table 5.10 On-The-Ball Skills Assessed in the Completed Studies of the Thesis.....	122
Table 6.1 Participant Characteristics	130
Table 6.2 Iterative Development of Themes	139
Table 7.1 Document and Participant School Characteristics.....	174
Table 7.2 Theme 1 - A Narrow Conceptualisation of Rubrics.....	185
Table 7.3 Single Point Rubric.....	188
Table 7.4 Theme 2 - Rubrics are for Reporting.....	194
Table 7.5 Theme 3 - An Overview of Rubric Criteria.....	199
Table 7.6 VC: HPE Invasion Game Rubric.....	202
Table 7.7 Theme 3 - References to Key Performance Criteria.....	204
Table 7.8 Key Performance Criteria Assessed in Relevant Studies Comprising the Thesis	206
Table 7.9 Types of Rubric Descriptors.....	208
Table 7.10 Theme 4 - The Challenge of Describing Quality	209

LIST OF KEY WORDS AND ACRONYMS

Acronym	Definition
ACARA	Australian Curriculum, Assessment and Reporting Authority
AC	Australian Curriculum
ACHPER	Australian Council for Health, Physical Education and Recreation
AfL	Assessment <i>for</i> Learning
AIESEP	Association Internationale des Écoles Supérieures d'Éducation Physique
AITSL	Australian Institute for Teaching and School Leadership
DET	Department of Education (Victorian)
FSG	Full-sided game
FUT-SAT	System of Tactical Assessment in Soccer
GBA	Game-Based Approach
GPAI	Game Performance Assessment Instrument
GPET	Game Performance Evaluation Tool
GS	Games and Sports
GTSC	Game Technical Scoring Chart
HPE	Health and Physical Education
IGS	Invasion Games and Sports
MAP	Movement Assessment in Practice
PE	Physical Education
PRISMA	Preferred Reporting Items for Systematic Reviews and Meta-Analyses
QPE	Quality Physical Education
ScR	Scoping Review
SEM	Sport Education Model
SSG	Small-sided game
TGfU	Teaching Games for Understanding
TSAP	Team Sport Assessment Procedure
UNESCO	The United Nations Educational, Scientific and Cultural Organization
VC	Victorian Curriculum
VCAA	Victorian Curriculum and Assessment Authority

CHAPTER 1: INTRODUCTION

1.1 Introduction

To help contextualise assessment in Physical Education (PE), this chapter begins by considering the historically contested nature, purpose and meaning of PE (Arnold, 1979; Capel & Blair, 2019; Gensemer, 1990; Kirk, 2010). It continues by describing games and sports in non-school settings and Games and Sports as a Focus Area within the Learning Area of Health and Physical Education (HPE) in the state and national curricula of Australia. This introductory chapter also provides an overview of the three message systems of schooling and dimensions of Quality Physical Education (QPE) (Penney et al., 2009) to provide an organising framework for the research. In considering the message system of curriculum, I address the role of Games and Sports in Australian HPE curricula with reference to both the Australian Curriculum (AC) as described by the Australian Curriculum Assessment and Reporting Authority (ACARA, n.d.-b) and the Victorian Curriculum (VC), as described by the Victorian Curriculum and Assessment Authority (VCAA, n.d.-f). Within the message system of pedagogy, relevant approaches that support the assessment of game performance are drawn from the literature on teaching styles (Mosston, 1966, 2002) and the Teaching Games for Understanding (TGFU) model (Bunker & Thorpe, 1986; Thorpe et al., 1984) are briefly outlined. Within the message system of assessment, perennial challenges faced by PE teachers are addressed within issues facing the wider field of education. Following the above background, I describe the aim of the research, my position in the research, and the contribution to knowledge the research provides.

PE for school-age children is a fundamental human right protected in the 1978 International Charter of Physical Education and Sport by the United Nations Educational, Scientific and Cultural Organization (UNESCO) (McLennan & Thompson, 2015). In their conceptualisation of QPE, the UNESCO consider PE as being vital to the development of youth and the foundation for life-long involvement in physical activity and sport (McLennan & Thompson, 2015). In Australia, PE is a school subject area that is taught within the broader Learning Area of HPE. Thus, HPE is one of eight Learning Areas in the national curriculum of Australia, and is considered a fundamental component of compulsory schooling (ACARA, n.d.-a).

Internationally, assessment is widely acknowledged as a key aspect of pedagogical practice and a major determinant of what knowledge is valued in schools (DinanThompson & Penney, 2015). Large class sizes, limited teaching time and catering to a wide variety of student abilities are key challenges in the assessment of practical aspects of PE (Braga & Liversedge, 2017; Gallo et al., 2006; Veal, 1988). Access to suitable spaces, availability of equipment, and the need to manage classroom behaviour provide further challenges for student assessment in PE (Veloo & Md Ali, 2016). The assessment in PE literature also recognises the prevalence of teachers assessing students based on a 'gut feeling' (for example, Annerstedt & Larsson, 2010; Hay & Macdonald, 2008; Svennberg, 2017), rather than clearly stated or well understood criteria. Specific to the context of Games and Sports within PE, assessment has historically lacked authenticity by emphasising the importance of technique through standardised skill or written tests (for example, Blomqvist et al., 2005; Siedentop et al., 2011), when assessment of game performance should instead be undertaken *during* game performance (Mitchell et al., 2013).

In Australia, some researchers have reported superficial assessment practices that indicate modest levels of assessment literacy among PE practitioners (DinanThompson & Penney, 2015). Others have reported the limited use of evidence-based assessments in Games and Sports in a school setting (Arias & Castejón, 2012; Georgakis et al., 2015; Williams et al., 2020). Collectively, the low level of assessment literacy and limited use of evidence-based assessments may diminish the valuable role of assessment in the teaching-learning cycle within PE (Collier, 2011; Kitts, 2003; Lund, 1992).

1.1.1 *Physical Education (PE), Games and Sports (GS)*

The lack of clarity in the meaning and/or purpose of PE among teachers is well established in the literature (Arnold, 1979; Capel & Blair, 2019; Gensemer, 1990; Kirk, 2010). More than thirty-years ago Gensemer (1990) noted the potential for confusion with the multitude of names for units or courses associated with tertiary PE (he named 60). In terms of the naming conventions more commonly associated with compulsory schooling, PE has been referred to as drill, training and movement (Arnold, 1979). In exploring the link between body and mind in the subject of PE, Gensemer (1990) argued that when the word *physical* is emphasised we can view PE as a type of education in which the body is educated, much like the 20th century focus of drill, calisthenics and gymnastics. Conversely, if we emphasise the word *education*, we view the body as the medium through which education takes place and thus the educative outcomes are far more expansive. It is a

premise of this thesis that although PE lessons necessarily involve physical movement of the body, physical activity in itself is not sufficient to justify PE as a subject worthy of inclusion in a curriculum. In Victoria and Australia, PE is fundamentally an educative enterprise (Penney et al., 2009) with an emphasis on its purpose to support “The progression and development of the disciplinary knowledge, understanding and skills” (VCAA, n.d.-e).

Consistent with other countries including England, Scotland and the USA (for example, Casey & Hastie, 2011; Gray et al., 2008; Mitchell et al., 2013; Ward & Griggs, 2011), Games and Sports have traditionally underpinned the educational rationale for PE in Australia (Kirk, 2006; Perlman & Forrest, 2015). The prevalence of GS in PE programs can be viewed in two contrary ways. One view is that as a Focus Area within the AC: HPE (ACARA, n.d.-c) and the VC: HPE (VCAA, n.d.-f) is that GS are a valid avenue for PE programs to support the developmental needs of students as well as any other school subject (Siedentop, 1987). As such, quality PE programs may be underpinned by a range of attributes including an emphasis on GS as long they “stand for something specific” (Siedentop, 1987, p. 25). In standing for something specific, PE programs could clearly define and identify GS as “the main theme that dominated the curriculum” (Siedentop, 1987, p. 25). PE teachers would promote their theme across the school and be proud of what they were doing (Siedentop, 1987).

Conversely, when GS in school PE is overly competitive and based on full-sided adult versions then less skilled students may be marginalised (for example, Bernstein et al., 2011; Bevans et al., 2010; Bryan et al., 2013; Garn et al., 2011). Further, the structuring of GS as a multi-sport experience may lead to disengagement for girls and low skilled boys (Ennis, 1999). The multi-sport approach, based on teaching a single sport over several lessons or weeks, is widely criticised for providing insufficient time for student learning and its focus on directive teaching (Pill, 2011; Pill et al., 2017; SHAPE, 2014). It has been argued that teachers adopting a multi-sport approach “cannot pretend to teach all sports or a multiplicity of sports and expect students to become knowledgeable and skilful in all of them” (Werner & Almond, 1990, p. 27). The prominence of GS in PE has led some researchers to question the performative sporting discourse and advocate for a multi-dimensional conceptualisation of the learning area that embraces functional, recreational and health-related pursuits as ‘informal sports’ (O'Connor et al., 2012; O'Connor & Penney, 2020).

Within the VC: HPE and the AC: HPE, there is no clear demarcation between what constitutes a game and what constitutes a sport. As an example, the following description from the AC: HPE states that the Focus Area of Games and Sports develops “movement skills, concepts and strategies through a variety of games and sports” (ACARA, n.d.-c) without providing any distinction between the two contexts. For clarity and consistency, the terms and meanings associated with Games and Sports are considered interchangeable in this thesis. As such, the following abbreviations will be used to refer to Games and Sports (GS) and Invasion Games and Sports (IGS) in community or school settings throughout.

The classification of Games and Sports within the state and national curricula (for example, ACARA, n.d.-c) is based on common tactical demands and includes four categories comprising: invasion games; net and wall games; striking and fielding games; and target games (Werner & Almond, 1990). As the IGS category generally comprise the largest number of games (Werner & Almond, 1990; Werner et al., 1996), and involve more dynamic and time-constrained environments than other categories (Inns et. al., 2023), IGS are the focus of this thesis. In my experience with secondary school PE teaching, these games are generally popular with teachers and students, and their variety and complexity make them a broad and rich topic for investigation.

QPE as described by the UNESCO suggests that competition and collaboration through the playing of GS provide students the opportunity to appreciate performance criteria, fair play and the benefits of teamwork (McLennan & Thompson, 2015). When QPE is assessed against various learning interventions on the psychomotor, cognitive and affective (social) domains, the use of Game-Based Approaches (GBAs), including the TGfU model (Bunker & Thorpe, 1986; Thorpe et al., 1984) and the Sport Education Model (SEM) (Siedentop, 1998; Siedentop et al., 2011) have been reported to provide some of the “greatest outcomes” on student learning (Dudley et al., 2022, p. 1). Thus, GS play a crucial role in the physical, social and emotional development of young people (Bailey, 2005; Pesce et al., 2013). In this thesis, GS are understood within a traditional lens; they involve individual, partner and team activities that may be recreational or competitive in nature and are likely to involve scoring, strategy and chance (Bailey, 2005). Of the four traditional game categories described by Werner and Almond (1990), historically invasion or territorial games have received the largest amount of teaching time in schools (Thorpe et al., 1984). Given that contemporary school settings have maintained the

positioning of IGS in PE programs (Gray et al., 2008), this context was considered relevant for PE teachers. In summary of key terms used in this thesis, GS are enjoyed in a range of settings, but within the state and national curricula of Australia, they are one of 12 Focus Areas that exist in the subject of PE taught within the Learning Area of HPE in Year Levels 3 -10.

1.2 The Three Message Systems

Education systems are built around three message systems of curriculum, pedagogy and assessment (Penney et al., 2009). Within this triad, a curriculum specifies the relevant knowledge, skills and understandings that are to be covered (Bailey, 2005); pedagogy or teaching relates to how the curriculum will be implemented; and assessment involves the gathering of evidence to make judgements about student learning (Hay & Penney, 2013). In recognising the interdependence of these three systems, each is briefly considered within an Australian context to provide an organising framework for the thesis. The focus of my research is assessment, which has been described as the “missing ingredient” within the three message systems (DinanThompson, 2013, p. 138).

Complementing this organising framework, the degree of congruence between the three message systems is referred to as instructional alignment (Cohen, 1987). Demonstrating the close links between instructional alignment and the organising framework of the thesis “Schooling is underpinned by a commitment to aligning curriculum, assessment and pedagogy and developing a common language and understanding of these three message systems” (Hayes, 2003, p. 225). The importance of instructional alignment is also shared by the AISEP (2020) in their position statement on assessment in PE. In reporting ‘poor’ levels of instructional alignment in Australia, the Netherlands, and the USA (AISEP, 2020), one exemplar case study from the USA revealed that “there was no alignment between the teachers’ espoused agenda, lesson tasks and assessments” (AISEP, 2020, p. 3). It is beyond the scope of the thesis to investigate the instructional alignment in participants’ assessment practices or examine models that evaluate the degree of instructional alignment (for example, Roach et al., 2008). However, where contextual data allows, references to instructional alignment, and the narrower conceptualisations of curriculum and pedagogical alignment, are made throughout the thesis.

1.2.1 Curriculum

To provide some context, the subjects of Health Education and Physical Education were amalgamated into Health and Physical Education (HPE) in the mid-1990s (Lynch, 2014). Currently, both Victorian and national curricula present HPE in two distinct Strands. For example, in the state of Victoria, Health is described and assessed within the strand of Personal, Social and Community Health, while PE is described and assessed within the strand of Movement and Physical Activity (VCAA, n.d.-f). Internationally, a variety of terms are used to describe HPE (Hay & Penney, 2013), with examples including Physical Education, Health and Physical Education, Physical Literacy and Health literacy (Lynch et al., 2016). My focus is the assessment of GS within the Movement and Physical Activity Strand of PE as a subject within the Learning Area of HPE in the state and national curricula of Australia. Given the widespread use of the term PE in the academic literature and curricula across educational systems (Hay & Penney, 2013), PE is the preferred term in this thesis. As such, all references to the state and national curricula Learning Area of the VC: HPE and AC: HPE refer specifically to the subject area of PE.

By way of definition, a curriculum includes the knowledge, skills and understandings that should be taught by teachers and learnt by students (Bailey, 2005). A curriculum is often presented in subjects and/or Learning Areas and is divided into Years or Levels that outline a developmental sequence of learning that progressively increase in sophistication. The following excerpt helps to define a curriculum and also indicate the close links between state and national curricula in Australia; “The Victorian Curriculum F–10 sets out what every student should learn during their first eleven years of schooling ... [it] incorporates the Australian Curriculum and reflects Victorian priorities and standards” (VCAA, n.d.-m). Emphasising the interconnectedness of *curriculum* and *assessment*, both the ACARA and VCAA acronyms include both these terms.

The three dimensions of movement as articulated by Arnold (1979) inform the AC: HPE and the VC: HPE. Education *about*, *through* and *in* Movement can be seen as an extension of the traditional view of school-based PE being purely *of* the physical (Gensemer, 1990). In broadening the view that PE has the sole aim of developing students in a physical sense to improve strength or fitness, Arnold (1979) argued that PE offers students more than enhanced biological outcomes. In Arnoldian terms, the first dimension of education *about* movement can be aligned to content knowledge, enquiry and

application of theory to practical situations. The second dimension of education *through* movement suggests that activities like games, dance, gymnastics, athletics, swimming and outdoor pursuits can be used as vehicles through which extrinsic benefits such as fitness, cognitive, social and moral development can be gained. The third dimension of education *in* movement posits that the actual activities, in and of themselves, are intrinsically worthwhile (Arnold, 1979). The latter dimension suggests that regardless of the theoretical content or extrinsic benefits, the activities in their purest form are inherently valuable (Arnold, 1979). While some question how well these dimensions are reflected in the state and national curricula of Australia (Brown, 2013), the conceptualisation by Arnold (1979) can be seen in the sub strands of the AC: HPE and VC: HPE: 'Learning through Movement'; 'Understanding Movement'; and 'Moving our Body' (for example, VCAA, n.d.-f).

In an Australian context, the ACARA was established in December 2009 to oversee the development of a national curriculum from Foundation to Year 10 (F - 10). Prior to 2009 each state and territory within Australia worked independently to develop their own curriculum across these year levels. The impetus for creating a national curriculum was to establish more uniformity in curriculum for students moving within Australia, to reduce duplication of time and resources for curriculum designers, and to address variation of student achievement and attrition across the states and territories (Reid & Price, 2018). Despite the establishment of the ACARA to develop an official national curriculum, the implementation of the curriculum remained the responsibility for each state and territory. This included each state being responsible for developing approaches to teaching, assessment, resources, and professional development.

The AC is built upon three intersecting elements that include eight Learning Areas (including HPE), seven general capabilities (such as personal and social capability), and three cross curriculum priorities (such as Aboriginal and Torres Strait Islander histories and cultures) (ACARA, 2023b). The subject of PE, within the AC, is comprised of two strands, six sub-strands and 12 Focus Areas. The Movement and Physical Activity strand was most relevant in this study as it comprises three sub-strands that align to the view of Arnold (1979). The sub-strand 'Moving our Body' aligns with education *in* movement, 'Understanding Movement' aligns with education *about* movement, and 'Learning through Movement' directly aligns with education *through* movement (Arnold, 1979).

In the AC, students are assessed against Achievement Standards that describe attainment levels in *two-year* bands, for example, Year 7 and 8. These are supported by content descriptors and elaborations that provide progressively more detailed examples of how the standards can be demonstrated. With reference to the AC: HPE, assessment is the most overlooked aspect of any discussion of the Learning Area (Reid & Price, 2018). This is consistent with the view of DinanThompson (2013), that assessment is the missing ingredient within the three message systems of schooling and QPE. At the time I was doing this thesis, the AC: HPE Version 8.4 was current until the end of 2022 before version 9.0 was released (ACARA, 2023a).

As the VC: HPE is the curriculum I am most experienced with and includes the Focus Area of GS from Year Level 3 – 10, the population sampled in this thesis was drawn from the Australian state of Victoria. In the state of Victoria, the VCAA operates within the Victorian Department of Education (DET). All government and Catholic schools are required to deliver the VC Foundation to 10 teaching and learning program (VCAA, n.d.-j). In contrast to the national curriculum, the Achievement Standards in the VC are presented in Levels, rather than Years (VCAA, n.d.-f). Thus, in the VC: HPE there is no expectation that a student completing Year 8 must demonstrate the stated Achievement Standards for Level 8. For consistency throughout this thesis, I will combine the naming conventions of both curricula and refer to ‘Year Levels’. Beyond this naming distinction, the structure, content and terminology of Achievement Standards, Strands, Sub-strands, Focus Areas, Content Descriptors and Elaborations are closely aligned in both PE curricula.

Specific to this thesis, the Victorian DET outline various conditions of QPE that demonstrate the inter-woven nature of curriculum and assessment. Relevant teaching and learning strategies directed to PE teachers include that the curriculum be implemented through the use of small-sided games (SSG), that assessment of student performance be ongoing, and that feedback is provided to individual students during class time on a regular basis (DET, 2021b). The DET also direct teachers to make overall, evidence-based and defensible judgements in line with the Achievement Standards described in the curriculum, by collecting data from a variety of assessment practices, interpreting the gathered data, and reflecting on findings that may include moderating judgements with colleagues (DET, 2021a).

1.2.2 Pedagogy

As the second message system (Penney et al., 2009), pedagogy is concerned with *how* knowledge, skills and understanding are to be taught. In recognising the inter-connectedness of the message systems, a brief overview of the Victorian Teaching and Learning Model (DET, 2023d), The Spectrum of Teaching Styles (Mosston, 1966, 2002) and the TGfU (Bunker and Thorpe, 1986) are presented to highlight the interdependence between pedagogy and assessment as message systems. In relation to instructional alignment, it has been reported that “Teaching and assessing have been institutionally dichotomized. Instead of being an integral part of instruction, assessment is separated institutionally as well as in practice” (Cohen, 1987, p. 19). While this dichotomy may be more accurately described as pedagogical alignment, as curriculum is not addressed, the following section seeks to address the institutional separation of teaching and assessment.

Pedagogical decision-making in Victoria is supported by the Victorian DET through the Victorian Teaching and Learning Model (DET, 2023d). This model is underpinned by the Framework for Improving Student Outcomes (FISO 2.0), that includes guidelines for teaching, learning and assessment (DET, 2023b). The Victorian Teaching and Learning Model was introduced in 2019 with the aim of providing various pedagogical principles and strategies for teachers to implement at the class-room level, regardless of the Learning Area taught. The Victorian DET encourages the use of the model to support school-wide improvement in all school sectors (DET, 2023d). Key pedagogical advice in this model includes the High Impact Teaching Strategies that are grounded in evidence-based practice. Given the focus of assessment in this thesis, the most significant High Impact Teaching Strategy is the provision of feedback (DET, 2023a). This highlights the inextricable links between feedback, as a product of assessment, and pedagogy.

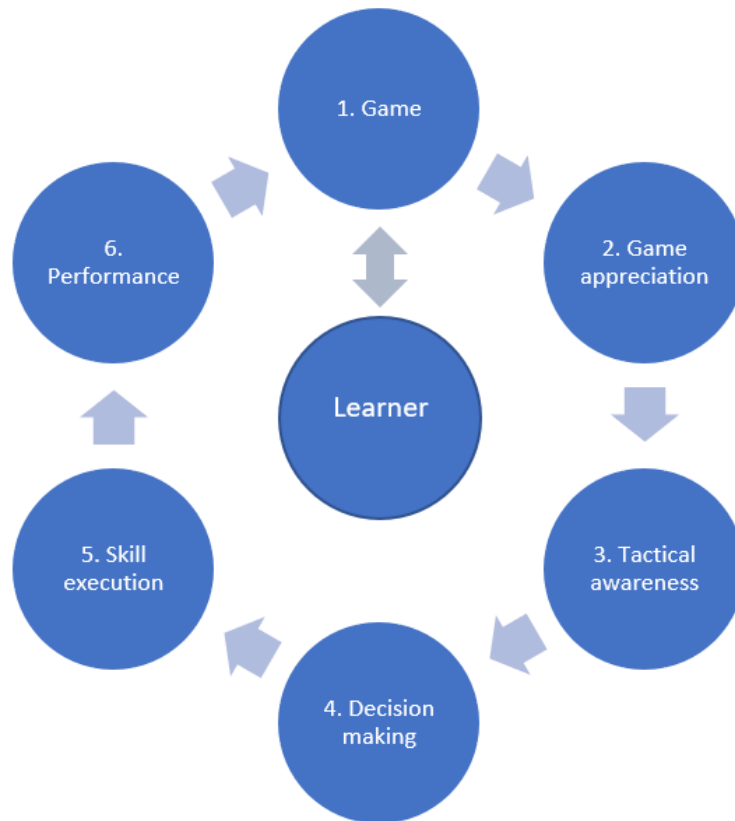
More specific to PE teaching, The Spectrum of Teaching Styles was originally developed in the 1960s, and was revised more recently, to describe teaching with reference to the degree of decision-making involved in the PE lesson (Mosston, 1966, 2002). A range of teaching styles exist along a continuum, with ‘teacher-centred’ approaches to the left and ‘student-centred’ approaches to the right. Teacher-centred styles have been criticised as lacking authenticity in the teaching of games as “practicing a skill in isolation or independently in a closed skill environment does not represent the real-world setting of team sports” (Jadeera et al., 2016, p. 167). In student-centred approaches,

students take more ownership of their learning and create solutions to movement problems that exist in games (Mosston, 2002). These student-centred approaches align closely to contemporary GBAs like the TGfU (Bunker & Thorpe, 1986; Thorpe et al., 1984). Regardless of the teaching style, each style includes a *post-impact phase* where information is gathered about student performance that includes assessing this information against criteria or standards and providing feedback to the learner (Mosston, 2002). The embedding of assessment into The Spectrum of Teaching Styles demonstrates the important relationship between teaching and assessment or pedagogical alignment (Mosston, 2002).

Widely considered the seminal GBA, the TGfU model (Bunker & Thorpe, 1986; Thorpe et al., 1984) is underpinned by continually modifying game forms to match students' learning needs. The teaching model is presented visually in Figure 1.1.

Figure 1.1

Teaching Games for Understanding (TGfU) model



Note. Adapted from The Curriculum Model, by D. Bunker and R. Thorpe, 1986, *Rethinking Games Teaching*, (p.2), Loughborough University of Technology. Copyright 1986 by Loughborough University of Technology. Adapted by permission.

In Figure 1.1 the six steps of the TGfU model begin with a developmentally appropriate game which is most likely to be a modified or SSG, rather than the adult game form. In step two, students acquire an appreciation of the particular rules or constraints the game is to be played within. In step three, students develop tactical awareness by grasping key principles of play, such as creating space for their team or denying space for the opposition. Decision making as step four includes distinguishing between choices “based upon ‘what to do?’ and ‘how to do it?’” (Bunker & Thorpe, 1986, p. 3). Based in the cognitive domain, this step recognises that an appropriate choice may not always lead to an appropriate skill execution or performance in the psychomotor domain. As step five, skill execution or technique is viewed separately from skill performance as step six. In applying the steps four, five and six to IGS, decision-making (step 4) may be judged as appropriate if a student

decides to shoot for a goal in basketball when unguarded, their skill execution (step 5) may be judged as appropriate if the shot is technically sound, and yet the performance (step 6) may be judged inappropriate, if the shot fails to get in the basket. The TGfU model allows for the delineation of cognitive and psychomotor elements so that in an assessment context, an assessor can consider the player's intent, technical skill and skill outcome as different components of game play.

In this thesis, GBAs refer to learner-centred teaching and coaching approaches through modified games that aim to develop thoughtful and skilful players (Teaching Games for Understanding Special Interest Group, [TGfU SIG], 2021). Along with the TGfU model, GBAs include a variety of approaches, such as Game Sense (den Duyn, 1996, 1997, 2000) Play with Purpose (Pill, 2007b, 2012, 2013), Play Practice (Lauder, 2001; Lauder & Piltz, 2013a, 2013b), and the Tactical Decision Learning Model (Gréhaigne, Wallian, et al., 2005). These approaches aim to develop game intelligence or tactical thinking by manipulating game forms to meet the challenge level of the students playing them (for example, den Duyn, 1996). Supporting the use of GBAs, including the TGfU model, a recent systematic review of what drives QPE reported that “If PE is to serve as a mechanism for the development of cognitive processes and cognitive learning outcomes, among other outcomes, a renewed focus on games-based pedagogies should be considered as part of QPE instruction” (Dudley et al., 2022, p. 18).

GBAs support a thematic approach to teaching GS by game categories (or classifications), rather than a multi-sport or multiple-activity model. The latter approach, also considered the technical or traditional model (Werner et al., 1996), involves students being taught a single sport, such as soccer, over a period of three to four weeks, and returning to the same sport year after year. The multi-sport model has been criticised for its inability to engage girls and low skilled boys (Ennis, 1999). Given that PE teachers cannot teach all games to all students in the time made available to them, a multi-sport approach may lead to practitioners choosing a small number of games that present “a narrow perspective of games education” (Werner & Almond, 1990, p.27). In a thematic approach, games are taught in categories with similar strategical and tactical demands over longer periods than the multi-sport model. The use of a thematic approach may also support the “transfer [of] movement concepts and strategies to new and challenging movement situations” identified in Year Levels 9 – 10 of the VC: HPE (VCAA, n.d.-d). In highlighting the nexus between pedagogy and assessment, the two most

frequently employed assessment instruments used in PE and youth sports, as reported by Arias and Castejón (2012), were both developed in pedagogical alignment to GBAs. These two tools, the Game Performance Assessment Instrument (GPAI) (Oslin et al., 1998) and the Team Sport Assessment Procedure (TSAP) (Gréhaigne et al., 1997) are addressed more fully in Chapter 2: Assessment in Physical Education (PE) – A Literature Review.

1.2.3 *Assessment*

While curriculum describes what is to be learnt by students (Bailey, 2005), assessment determines how much, or how well the curriculum has been learnt (Black & Wiliam, 2018). Like curriculum and pedagogy, assessment is contextual, and as such the ‘social dynamics’ of assessment should not be overlooked (Hay & Penney, 2013). Further, as performance and understanding are situational, assessment of performance and understanding is highly contextual (Wiggins, 1993b). The following introduction to assessment completes the overview of the three message systems of schooling and QPE (Penney et al., 2009) and presents various issues for the wider field of educational assessment. This section also introduces challenges specific to assessment in PE that are addressed more fully in the next chapter.

Educational assessment has been defined variously as “Identifying appropriate standards and criteria and making judgements about quality” (Boud, 2000, p. 151) and “The process of gathering evidence of student achievement to inform education decisions” (Stiggins, 2018, p. 18). In a PE context, assessment has been defined as “Any action of information collection within education settings that is initiated for the purpose of making some interpretive judgements about students” (Hay & Penney, 2013, p. 6). Within the state of Victoria, the DET describe assessment as “the ongoing process of: gathering, analysing and interpreting evidence; reflecting on findings; [and] making informed and consistent judgments to improve student learning” (DET, 2021a). In summary of these definitions, assessment involves the collection of information about student achievement which allows for judgements to be made.

In a review of the assessment literature over the last 50 years, it has been observed that a lack of conformity and clarity in language associated with assessment has led to “definitional anarchy” (Cookson, 2018, p. 930). In the Australian context, the language used in relation to reporting has divergent meanings which contributes to confusion among stakeholders (Hollingsworth et al., 2019).

While the focus of the latter study was reporting, assessment provides the information presented in end of term or semester reports (Hollingsworth et al., 2019), and by implication, include language that is divergent in meaning and confusing. Through informal meetings, focus groups, workshops and surveys, the views of Australian students, parents, teachers and principals in the latter study revealed that common assessment terms like formative assessment, summative assessment, achievement, performance and standard “were used in different ways across the education community ... some terms appear to have multiple meanings, and some terms appear to be used interchangeably” (Hollingsworth et al., 2019). More alarmingly, the field of educational assessment has been described as “divided and in disarray” (Masters, 2013, p. 63) with low levels of assessment literacy reported across the wider education community. Although pressure to reform assessment in Australia has increased over recent years, different approaches and paradigms have led to a fractured field, devoid of a single unifying theory of assessment (Masters, 2013).

Unlike most academic subjects that assess student learning primarily in the cognitive domain (Lund & Veal, 2013), PE teachers are encouraged to assess students in three domains: cognitive; psychomotor; and affective (for example, Hay & Penney, 2013; Mitchell et al., 2013). While a fuller consideration of these domains is presented in Chapter 2, the cognitive domain refers to a student’s understanding and knowledge, the psychomotor domain refers to a student’s movement and motor skills, and the affective domain refers to a student’s personal and social skills, that might include teamwork and collaboration (Mitchell et al., 2013). The focus on domains of assessment in this thesis was guided by the literature on evidence-based game performance assessment tools (Chapter 4). It was hypothesised that these assessment tools would primarily employ criteria classified as psychomotor, and to a lesser degree, criteria that might be classified as cognitive.

The PE assessment literature has reported modest levels of assessment literacy among PE practitioners, typified by superficial assessment practices (DinanThompson & Penney, 2015; Williams et al., 2020). PE assessment practice has been described as subjective (Birky, 2012; Svennberg et al., 2014; Williams et al., 2020), lacking authenticity (Georgakis et al., 2015; Williams et al., 2020), internalised (Annerstedt & Larsson, 2010; Hay & Macdonald, 2008; Svennberg et al., 2014), poorly aligned to instruction (Lund & Veal, 2008), and criticised for taking up valuable instruction time (Braga & Liversedge, 2017; Gallo et al., 2006; Veloo & Md Ali, 2016). In the context of assessment *for*

learning (AfL) in PE, a recent literature review by Moura et al. (2021) found that teachers used assessment exclusively to grade students and that teachers did not have the required skills to use AfL effectively. As the use of AfL informs and improves PE instruction (McLennan & Thompson, 2015), the findings in the literature review of Moura et al. (2021) are counter to guidance for the ongoing application of student-centred assessment within QPE (McLennan & Thompson, 2015). The importance of assessment within the three message systems of schooling and QPE is underpinned by the following statement from assessment luminaries, Black and Wiliam (2018) in a broader educational context:

If students learned what they were taught, then assessment would not be unnecessary; we could simply document the educational experiences of each student secure in the knowledge that this would describe their capabilities accurately. But of course, students do not always learn what they are taught, so we need to develop processes of eliciting and interpreting evidence so that we can draw conclusions about what students have in fact learned. (Black & Wiliam, 2018, p. 570)

This statement indicates a fundamental purpose of assessment within the three message systems is to determine the degree of student learning. In turn, this allows teachers to make judgements as to the effectiveness of their pedagogical choices and implementation of the curriculum to improve instructional alignment. While a fuller discussion of assessment in PE is presented in the following chapter, the next section of this introduction describes the research questions and aims before concluding with a consideration of my positionality in the research.

1.3 Research Question and Aim

The overarching research question that drives the thesis is: *How do Physical Education teachers understand and practice assessment in the curriculum focus area of games and sports in Victorian secondary schools?* The choice of a mixed methods research design supported the broad research aim: *To describe, explain and make meaning of Physical Education teachers' understanding and practice of assessment in invasion games and sports in Victorian secondary schools (Year Levels 7 – 10).* Within this aim, the researcher sought to examine assessment practice and generate evidence-based recommendations for stakeholders. This examination can be viewed as a response to the many

invitations to undertake research to improve assessment practice in the extant PE literature (for example, Capel & Blair, 2019; Georgakis et al., 2015; Inns et al., 2023). Of note, invasion games are characterised by teams sharing a playing field and attempting to out-score their opponent by invading opposition field or court space (Werner & Almond, 1990). To allow for international comparisons and potentially support transferability in findings, students in Year Levels 7 -10 in Victoria are typically aged between 13 to 16.

To respond to the overarching research question and aim, the following four sub-questions were constructed:

1. What does the extant literature say are the defining characteristics of assessment tools developed for invasion games and sports?
2. How do Physical Education teachers view the assessment of invasion games and sports in Victorian secondary schools?
3. How do Physical Education teachers practice and/or want to practice assessment of invasion games and sports in Victorian secondary schools?
4. How are Physical Education teachers' assessment tools constructed for invasion games and sports in Victorian secondary schools?

1.4 Positionality

It is important to acknowledge my position in the study as this affects the nature of my observations and interpretations (Bourke, 2014; Bukamal, 2022; Mohler & Rudman, 2022). The lack of exposure to assessment in my secondary school experience as a student, the demonstration of the decontextualized assessment of skills in my undergraduate teacher education, and a focus on reporting as assessment in my teaching career, has led me to concur that assessment is “one of the most fraught and troublesome issues physical educators have had to deal with” (López-Pastor et al., 2013, p. 57). Contemporary strategies to promote instructional alignment in PE teacher education in the literature, and thus link assessment to teaching and curriculum, were absent in my experience in the 1990s. First, there was no modelling of good practice in instructional alignment; second, there was no encouragement of student teachers to embed this alignment in our practice; and third, there

was no overall buy-in from faculty stakeholders to enact instructional alignment (MacPhail et al., 2023). Further, there was no messaging to view “assessment as a means to engage students in the learning process ... [and instead it was seen] as an add-on to the learning experience” (Scanlon et al., 2023, p. 3).

I come to this thesis with a set of beliefs and values that underpin my “insider-outsider researcher positionality” (Bukamal, 2022, p. 327). For a researcher to be deemed an insider, they must share several similarities with their participants, and to be considered an outsider, the researcher should not belong to the same group of their study cohort (Braun & Clarke, 2013). I consider myself as an insider in this study due to my extensive secondary school PE teaching background, but I am also an outsider as I left secondary teaching mid-way through this doctorate. I bring my experience as a secondary school PE teacher and then PE Department Head to this thesis, so as an insider at the commencement of my thesis I bring to the study assumptions and perspectives born of this experience. As assumptions and positioning are integral aspects of research (Braun & Clarke, 2019), my view on the role of GS is that they should contribute to a well-balanced and inclusive PE curriculum; but that school PE should be more than just GS. I also believe that where GS are taught within a school PE, improving student game performance should be a core aim of those programs (Gray et al., 2008), and that assessment can contribute to that aim.

My views on assessment were established through an undergraduate PE teaching degree that made use of decontextualized assessment of discrete sport skills to assess our ability to *do* what was expected in the field. In the context of GS, on-the-ball skills were performed in a relatively closed environment that was generally devoid of opposition and complex decision-making. This approach supported a teacher-centred, command teaching style underpinned by a drill-based approach (Werner et al., 1996). Self-assessment was used sparingly and included counting things, like the number of successful basketball free throws, outside of a game situation. Peer assessment was rarely used and included checking off elements in the kinetic chain of a prescribed technical skill, for example, an overarm throw movement cue checklist. In both these examples, I recall my undergraduate colleagues questioning the value of self and peer assessment for students that might inflate their level of achievement, given there was no direct supervision by the teacher.

During the 26 years I have taught secondary school PE within Australia and internationally, I have continued to wrestle with issues surrounding effective practice in assessment of practical components of this subject. In my experience, performance-based assessment in PE was temporally bound, complex and dynamic. Given the inherent need for students to be moving in practical classes of PE, assessment is particularly problematic (Georgakis et al., 2015). In the latter half of my teaching career, rubrics became popular forms of assessment, and were well supported by colleagues due to their explicit criteria and expedient application. However, these rubrics generally used subjective language that did not adequately describe performance, and in some cases teaching colleagues admitted having a score in their mind before selecting descriptors in the rubric to add up to that score.

Like most of my secondary teaching colleagues, I made little use of assessment tools that generated frequency-based outcomes like the GPAI (Oslin et al., 1998) and the TSAP (Gréhaigine et al., 1997). It was the low level of familiarity and use of these tools among my colleagues that led me to this thesis. To help orient the reader, Appendix I provides an outline of both the GPAI and the TSAP frequency-count tools. Further, as rubrics were used widely in my secondary teaching experience, Table 1.1 is provided as a sample IGS rubric from the literature (Breed & Spittle, 2011). This rubric includes four *criteria* in the top row, five *levels* of achievement in the left-hand column, and *descriptors* in each cell to help define the levels of performance (Breed & Spittle, 2011).

Table 1.1*Sample Invasion Games Rubric*

Assessment	Technical	Tactical	Values	Behaviours
1	Excellent technical skills in passing, catching, shooting and dribbling.	Excellent decision-making with the ball (where, when and how to pass, dribble or shoot) and without the ball (when and where to move, and identifying space).	Excellent team work and was always able to work effectively with peers.	Was always enthusiastic, tried hard to improve and was always attentive.
2	Very good technical skills in passing, catching, shooting and dribbling.	Very good decision-making with the ball (where, when and how to pass, dribble or shoot) and without the ball (when and where to move, and identifying space).	Very good team work and was always able to work effectively with peers.	Was usually enthusiastic, tried to improve and was mostly attentive.
3	Competent skills in passing, catching, shooting and dribbling.	Competent decision-making with the ball (where, when and how to pass, dribble or shoot) and without the ball (when and where to move, and identifying space).	Competent team work and was often able to work effectively with peers.	Was often enthusiastic, and generally attentive.
4	Needs some improvement in technical skills in passing, catching, shooting and dribbling.	Needs some improvement decision-making with the ball (where, when and how to pass, dribble or shoot) and without the ball (when and where to move, and identifying space).	Needs some improvement in team work and was sometimes able to work effectively with peers.	Needs some improvement in enthusiasm and attentiveness.
5	Needs significant improvement in technical skills in passing, catching, shooting and dribbling.	Needs significant improvement decision-making with the ball (where, when and how to pass, dribble or shoot) and without the ball (when and where to move, and identifying space).	Needs significant improvement in team work; was rarely able to work effectively with peers.	Needs significant improvement in enthusiasm and attentiveness.

Note. From *Developing Game Sense through Tactical Learning: A Resource for Teachers and Coaches*, by R. Breed and M. Spittle, 2011, (p.183), Cambridge University Press. Copyright 2011 by Cambridge University Press. Reprinted by permission.

In acknowledging the three message systems of schooling and QPE as curriculum, pedagogy and assessment (Penney et al., 2009), it has been observed that assessment is the missing ingredient (DinanThompson, 2013). Despite the long-held view that IGS receive most curriculum time (Thorpe et al., 1984), there have been few studies investigating the assessment practices of secondary school PE teachers exclusively in this context. As such, this thesis can be viewed as a response to suggestions for

further research to scrutinise assessment practices in PE and invasion games in the literature (for example, Capel & Blair, 2019; Georgakis et al., 2015; Inns et al., 2023). The thesis is presented in eight chapters including the current Introduction. Chapter 2 provides an overview of assessment in the PE literature; Chapter 3 describes the research design, and Chapters 4 to 7 present each of the four studies. The thesis concludes in Chapter 8, with a translation of overall findings and final recommendations for stakeholders.

CHAPTER 2: ASSESSMENT IN PHYSICAL EDUCATION (PE) – A LITERATURE REVIEW

2.1 Introduction

The introductory chapter positioned the thesis within the three message systems of schooling and dimensions of QPE (Penney et al., 2009). This included a brief but necessary consideration of curriculum and pedagogy as the first two message systems. As the focus of the thesis is the assessment of IGS in the subject of PE, this chapter presents an overview of relevant literature regarding assessment concepts, frameworks and challenges in the wider field of education, before addressing them specifically in the context of PE. The review will show that educational assessment is an over-conceptualised and complicated field (Masters, 2014) that may benefit from greater consistency and clarity in commonly used assessment terms (Cookson, 2018; Hollingsworth et al., 2019). The review describes the conflated relationship between assessment, feedback and reporting before addressing the *purpose* of assessment as a key feature in assessment design. After acknowledging the contextual nature of assessment principles, the chapter describes validity and reliability as fundamental assessment concepts that underpin the broader thesis. Due to their specific development or application in a PE setting, frameworks for assessment literacy, quality assessment, and authentic assessment are then presented (Chappuis et al., 2012; Gulikers et al., 2004; Hay & Penney, 2013). In acknowledging Victorian PE teachers as the population under investigation in the thesis, references to the VC: HPE (VCAA, n.d.-h) are made throughout this chapter. It is noteworthy, however, that there is a dearth of literature relating to relevant assessment practices in this population. Consistent with the aim of literature reviews to identify problems, locate gaps in the literature and provide context for scholarly inquiry (Ary et al., 2014), this chapter seeks to identify problems specific to assessment in PE, especially GS, and to identify a program of research to address these problems.

To help situate this review, the Association Internationale des Écoles Supérieures d'Éducation Physique (AIESEP), an international professional association that aims to promote quality research in PE, recently provided a position statement on PE assessment through their website (AIESEP, 2020). Of relevance to this thesis, their position statement advocated for the development of assessment literacy for all stakeholders in schools. In doing so, the authors drew on the assessment literacy model

of Hay and Penney (2009) that is described later in this chapter. This position statement on assessment in PE also supported instructional alignment with reference to the three message systems of schooling and QPE: curriculum; pedagogy; and assessment (Penney et al., 2009). The position statement also endorsed the use of AfL in PE programs that was reported as an overlooked function of assessment by PE practitioners in the literature review of Moura et al. (2021). As such, the assessment in PE literature promotes the development of PE teacher assessment literacy, endorses pedagogical and curriculum alignment with assessment, and advocates the use of AfL to support improved assessment practice for PE teachers.

2.2 An Overview of Assessment Concepts

In Chapter 1, assessment was introduced as a process to help determine how much, or how well, a curriculum has been learnt (Black & Wiliam, 2018). Originating from the Latin verb *assidere*, to sit with or beside (Wiggins, 1993a), assessment may be considered a social activity, as assessment is required, developed, implemented and performed by people (Hay & Penney, 2013). Framed within a sociocultural perspective of assessment in PE, Hay and Penney (2013) define assessment as “Any action of information collection within education settings that is initiated for the purpose of making some interpretive judgements about students” (p. 6). It is this understanding and definition of assessment that informs the following review.

2.2.1 *Over-Conceptualised and Over-Complicated*

The need for greater clarity in the language and concepts used in educational assessment is widely reported in the educational assessment literature (for example, Cookson, 2018; Hollingsworth et al., 2019; Masters, 2014; Wyatt-Smith et al., 2014). These reports include descriptions of the field being in ‘disarray’, that may be partly attributed to an outdated vocabulary of “Assessment concepts and terminology introduced over the past half century [that] sometimes now function as impediments to clear thinking and good practice” (Masters, 2014, p. 1). Demonstrating this over-conceptualisation and confusion in assessment naming conventions, Table 2.1 is reproduced from an Australia-wide study by Hollingsworth et al. (2019). Their study involved multiple stakeholders in education and investigated how student progress in learning was communicated through reporting (Hollingsworth et al., 2019).

Table 2.1

Ambiguous Terms Used in Communicating and Reporting Student Learning

Terms		
achievement	growth	progress
attainment	improvement	progressive reporting
continuous reporting	indicator	progress task
formative assessment	level	standard
gain	normative	student report
grades	outcome	report card
grading	performance	summative assessment

Note. Terminology from *Communicating Student Learning Progress: A Review of Student Reporting in Australia*, by H. Hollingsworth, J. Heard and P. R. Weldon, 2019 (p.14), ACER.

Their national study indicated a lack of consistency and clarity in key assessment and reporting terms that hindered the development of coherent assessment and reporting practices (Hollingsworth et al., 2019). The table shows the 21 terms as a non-exhaustive list of common language found by Hollingsworth and colleagues in the context of assessment and reporting that had multiple meanings or were used interchangeably by stakeholders. No suggested definitions accompanied the table, instead the authors explained any contested terms within text.

An example of one of the many “fault lines” (Masters, 2014, p. 1) in commonly used assessment terms is the distinction between formative and summative assessment (for example, Boud & Soler, 2016; Cookson, 2018; Hollingsworth et al., 2019; Masters, 2014). One level of distinction in these terms is based on *when* the assessment occurs in the teaching-learning cycle, with formative assessment occurring during the learning cycle and summative assessment occurring after the learning period (Masters, 2014). A second distinction is based on the *manner* of assessment, with formative assessment stemming from regular teacher observations and summative assessment as more systematic and often externally administered (Masters, 2014). A third distinction is based on the *purpose* of the assessment, in this case formative assessment informs instruction and learning while summative assessment determines the degree of learning (Black & Wiliam, 1998b). Lastly, Cookson (2018) offers a classification based on *duration*, with formative evaluation (assessment) being brief,

and summative evaluation (assessment) being lengthier. Based on these four classification systems for formative and summative assessment there is support for the idea that some aspects of assessment may be “over-conceptualised” (Masters, 2014, p. 1) and lead to confusion among stakeholders.

In relation to other conflated terms found in Table 2.1, by way of another example, there may be misunderstanding between the concepts of *progressive* reporting and reporting on student *progress* (Hollingsworth et al., 2019). Hollingsworth and colleagues suggested that some stakeholders misconstrued progressive reporting as synonymous to reporting student progress; the key misunderstanding being that the ongoing tracking of performance was not the same as measuring any progress in learning (Hollingsworth et al., 2019). Further, their study found that many teachers were ill-prepared to measure and monitor learning gains or progress in their assessments (Hollingsworth et al., 2019). This suggests that these schools were not adhering to national requirements to make use of assessment data to report on achievement *and* progress as stipulated by the ACARA (n.d.-e). As the report of Hollingsworth et al. (2019) did not provide separate findings for each state, territory or Learning Area, the lack of clarity and consistency in communicating student achievement and progress may present a challenge for assessment in PE in Victoria. Central to this challenge is that much of the language used in educational assessment and reporting may fail to convey a shared mental representation of what is being discussed.

As an example of the confusion in assessment terms within PE *beyond* those listed in Table 2.1, different classifications of formal and informal assessment exist in the literature (Piotrowski & Capel, 2000; van der Mars et al., 2018b). *Informal* assessment has been described as assessment that takes place during routine PE class activities based on observation and verbal interaction, while *formal* assessment has been described as distinct from routine class activities and often involving standardised tests (Piotrowski & Capel, 2000). Others have described informal assessment in the same PE context as including non-verbal feedback, while formal assessment has been described as any recording of assessment information on paper or electronically (van der Mars et al., 2018b). These examples support the need for greater clarity and consistency in the language associated with assessment and reporting (Cookson, 2018; Hollingsworth et al., 2019; Masters, 2014; Wyatt-Smith et al., 2014). In summarising the key recommendations of the Hollingsworth et al. (2019) report that

apply to all Learning Areas (including the subject of PE), consistent terminology may support shared mental representations of assessment and reporting concepts used to communicate student learning, reporting should be aligned to an on-going cycle of assessment, learning progress should be linked to achievement standards, and students should be provided with directions for future learning.

Confusion also exists in terms used to describe common assessment *instruments* used in PE that include checklists, rating scales and rubrics (Lund & Veal, 2013). According to a review of rubrics in higher education (Brookhart, 2018), the key difference in these assessment instruments is based on a nuanced understanding of scales. Checklists work on a binary scale (for example, present and not present), rating scales use scale language that does not describe performance (for example, language is based on numerical, frequency or evaluative scales), and rubrics describe quality across levels of performance to help students identify their next steps in learning (Brookhart, 2018). Rubrics that describe performance without the use of scale type language have been termed “True rubrics” (Brookhart, 2018, p. 1). As an example of the conflation between rating scales and rubrics, the sample rubric presented in Table 1.1 of Chapter 1 employed evaluative scale language (for example, excellent, very good, competent) and frequency scale language (for example, always, often, rarely) (Breed & Spittle, 2011). According to Brookhart (2018), this sample rubric is more accurately described as a rating scale and may be more useful for grading than for learning, as it does not describe performance at various levels, or provide students the opportunity to envisage their next steps in learning.

2.2.2 *Assessment, Feedback and Reporting*

In untangling the relationship between assessment, feedback and reporting, feedback is defined by the Australian Institute for Teaching and School Leadership (AITSL) as “information about student’s progress towards a learning goal” (2017b, p. 5). Thus, feedback may be an interpretation of information collected from an assessment that indicates a student’s level of knowledge, skills or understanding. In a more expansive definition, Hattie and Timperley (2007) describe feedback as any “information provided by an agent (e.g., teacher, peer, book, parent, self, experience) regarding aspects of one’s performance or understanding” (p. 81). The latter definition is accompanied by a nuanced feedback model and useful conceptualisations of feed up and feedforward (Hattie & Timperley, 2007). Their feedback model is built upon three questions that help the learner determine

their level of learning against the intended learning goal and identify steps to close any gap (Hattie & Timperley, 2007):

Where am I going? (What are the goals?), How am I going? (What progress is being made toward the goal?), and Where to next? (What activities need to be undertaken to make better progress?) These questions correspond to notions of feed up, feedback, and feed forward. (Hattie & Timperley, 2007, p. 86)

In unpacking the above conceptualisations of feedback in relation to assessment, the first aspect refers to students understanding the learning intentions or success criteria for an assessment (feed up); the second aspect refers to students receiving information as to the quality of their learning as demonstrated through the assessment (feedback); and the third aspect refers to information from an assessment that aims to support students in future assessments or learning tasks (feed forward) (Hattie & Timperley, 2007). In summary of the relationship between assessment and feedback, assessment involves the collection of information, while the interpretation and use of this information may constitute feedback. It is feedback from an assessment task that may in turn generate an end of term, or end of semester, report. The interdependent relationship between assessment and reporting can also be seen in the naming of the Australian Curriculum, Assessment and Reporting Authority (ACARA).

In further considering the link between assessment and reporting in the local context, teachers in the AC and VC make use of the Achievement Standards at the end of a period of teaching to provide an on-balance judgement about the degree of learning demonstrated by students. In making on-balance judgements, the AC states that teachers should base their judgements on assessment data collected over the teaching period (ACARA, n.d.-e). Achievement Standards describe what students know, do, or can understand, while the content descriptors and elaborations provide more detail to help determine student progress and achievement against the standard(s) (VCAA, n.d.-i). The Victorian DET requires all schools to assess student achievement and progress across F to 10 and use this information to provide a minimum of two reports a year (DET, 2023c). Further, the DET states that reporting systems should be ongoing and include the regular sharing of assessment information with parents and guardians throughout each term or semester (DET, 2023c).

While reporting is a requirement in most schooling systems, historical and contemporary views of assessment in PE have indicated that assessment is used exclusively for reporting, as opposed to learning (Moura et al., 2021; Veal, 1988). Table 2.2 provides an example of some relevant aspects of the VC: HPE Year Levels 7 - 8 that relate to performance in GS (VCAA, n.d.-d). These excerpts provide direction for the learning and assessment practices of secondary school PE teachers in the state of Victoria and may reduce subjective assessment and/or assessment of non-performative criteria, which is described later in this chapter.

Table 2.2

Sample References to Games and Sports in the VC: HPE

Achievement Standards	Content Descriptors	Elaborations
Students demonstrate control and accuracy when performing specialised movement skills.	Use feedback to improve body control and coordination when performing specialised movement skills (VCHPEM133)	Analysing their own and others' performance using ICT and implementing feedback to enhance performance Participating in activities where vision is compromised to demonstrate the importance of auditory feedback Using visual and kinaesthetic feedback when coordinating eye–hand and eye–foot movements to control different pieces of equipment
They apply the elements of movement to compose and perform movement sequences.	Compose and perform movement sequences for specific purposes in a variety of contexts (VCHPEM134)	Designing and performing movement sequences to create, use and defend space Travelling, marking and intercepting to achieve and retain possession
They apply and refine movement concepts and strategies to suit different movement situations.	Practise, apply and transfer movement concepts and strategies (VCHPEM135)	Examining and demonstrating the similarities of strategies used in different physical activities and how they can be transferred to new movement situations Selecting previously successful strategies and applying the most appropriate when solving new movement challenges

Note. The above text is based on the VC: HPE Year Levels 7 – 8 as described by the VCAA (n.d.-d)

2.2.3 *The Purpose of Assessment*

As a fundamental concept in assessment design, the purpose of assessment is positioned as the first step in planning quality assessment in an educational context (Bearman et al., 2016; Chappuis et

al., 2012). In considering the purpose of assessment, it is suggested that teachers should determine what information will be collected, how this information will be used, and who will use this information (Chappuis et al., 2012). In the field of education there are a myriad of views on the purpose(s) of assessment (Earl, 2014). A common typology that italicises the relevant prepositions includes assessment *as* learning (Dann, 2014; Earl, 2014; Gibbons & Kankkonen, 2011), assessment *of* learning (Chiles, 2020) and assessment *for* learning (for example, Chng & Lund, 2018; Leirhaug & Annerstedt, 2015; Wiliam, 2011). The respective purposes of these assessments relate to the role of students in assessing themselves or their peers, using assessment to determine the degree of student learning, and using assessment to enhance learning.

Other purposes of assessment include assessment for teaching (Griffin, 2014), assessment as pedagogy (Hay et al., 2013), assessment from instruction (Wyatt-Smith et al., 2014), and assessment on instruction (Wyatt-Smith et al., 2014). More expansively, Newton (2007) has proposed a non-exhaustive list of 18 different purposes of assessment. This breadth of assessment purposes has led some to warn practitioners that they risk a metaphorical “death from a thousand prepositions” (Wyatt-Smith et al., 2014, p. 127), suggesting that the proliferation of assessment purposes has further complicated the field. In advocating for a universal and simplified understanding of the purpose of assessment Masters (2014) posited that:

The fundamental purpose of assessment in education is to establish and understand where learners are in an aspect of their learning at the time of assessment. There is no other purpose. Establishing where learners are in their learning usually means establishing what they know, understand and can do. (Masters, 2014, p. 1)

Masters argues that when this single purpose is accepted by stakeholders a range of other distinctions in educational assessment can be viewed in more constructive ways (2014). This understanding of assessment having a “single purpose” (Masters, 2014, p. 1) then positions other purposes as subsequent *uses* of assessment (Masters, 2014). Support for the single purpose view of Masters (2014) is offered in an Australian PE context, with advocates suggesting that “establishing assessment efficacy necessitates the transcending of divisions in purpose” (Hay & Penney, 2009, p. 391). In applying the single purpose view of assessment by Masters (2014) to the AIESEP position statement on assessment in PE (AIESEP, 2020), assessment may be *used* to support student learning,

inform teachers of their impact, certify student achievement, and determine stakeholder accountability.

2.2.4 *Validity and Reliability*

Just as a myriad of purposes or uses of assessment are suggested in the literature, there are no universally agreed assessment principles or guidelines in educational assessment. Indicative of the contextual nature of assessment (Wiggins, 1993b), various assessment principles are reported across different sectors of education in the state of Victoria. For example, the Victorian Teaching and Learning Model describes authenticity, planning for learning, linking to standards and analysis of data as assessment principles for schooling up to Year Level 10 (DET, 2023d). In the Victorian Certificate of Education (VCE) (VCAA, n.d.-I), the VCAA advocates for validity, equity, balance and efficiency as assessment principles for schooling in Year Levels 11-12. Representative of higher education, Victoria University identifies eight guiding principles of assessment on their website that include engaging students, considering student diversity, and preparing students for future employment (Victoria University, 2023). Given the over-conceptualised field of assessment (Masters, 2014), the scope of assessment *concepts* considered in the next section of the review is limited to validity and reliability as the two most important indicators of assessment quality (Brookhart, 2003, 2005).

In educational assessment that relates to the reporting of learning outcomes, validity may be described as the degree to which an assessment measures what it sets out to measure (Kervin et al., 2016). Rather than seeing this as a fixed property of any assessment, validity may help assessors judge the quality of their interpretations or justify any actions derived from inferences based on assessment results (Darr, 2005). Reliability of assessment has been described as the level of consistency in results from assessments in similar conditions (Kervin et al., 2016). Reliability may also describe the degree of confidence that assessors have that results earned by students accurately reflect their level of achievement, knowledge, or skills (Brookhart, 2005).

Synonymous with assessment principles or concepts, “conditions for assessment efficacy in physical education” (Hay & Penney, 2009, p. 389) include an assurance of validity. In PE, validity and/or reliability measures are commonly reported to indicate the robustness of standardised physical fitness tests and motor competence tests (Hulteen et al., 2015; Robertson et al., 2014; Wen et al., 2018). For example, the literature review of motor competence assessments in children and

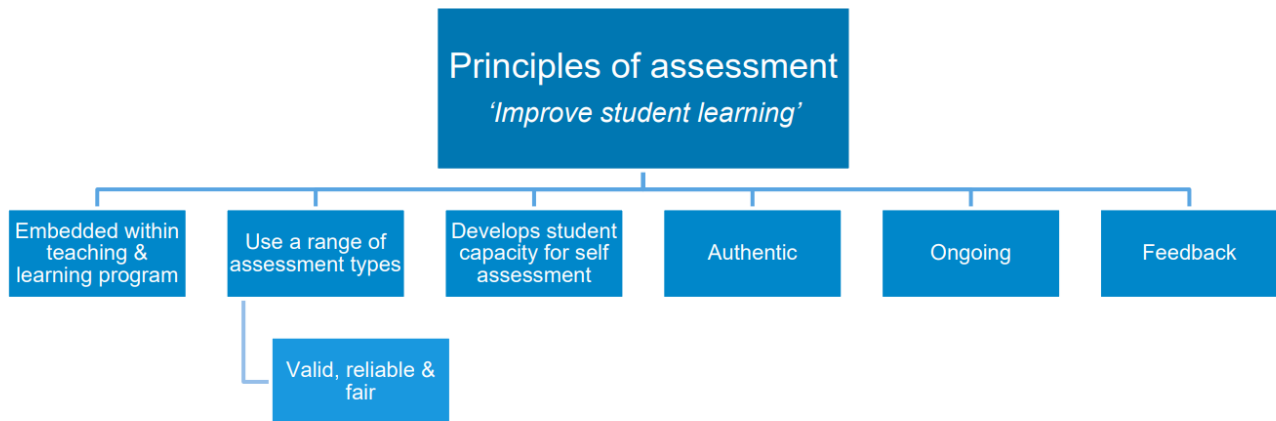
adolescents by Hulteen et al. (2020) located 57 different skill assessment tools from 107 studies and reported measures that included construct validity, content validity, criterion validity, intra-rater reliability and inter-rater reliability. Equally, assessment tools developed by researchers for use in GS performed in a school-age population often report specific measures of validity and/or reliability (for example, da Costa et al., 2011; Gréhaigne et al., 1997; Oslin et al., 1998). In GS assessment, content validity has been described as the use of experts to confirm that assessment variables represent performance (da Costa et al., 2011), ecological validity has been described as the authenticity or real-world nature of the assessment (Gréhaigne et al., 1997), and face validity has been described as the degree that assessment was perceived as being fair or reasonable by those assessed (Oslin et al., 1998). Beyond the contexts of fitness tests, motor competence tests, and GS assessment instruments, discussion of validity in PE assessment has been described as notably absent in the literature (Hay & Penney, 2013).

To promote reliability in assessment, strategies include ensuring the task comprises sufficient information to determine the level of performance, sharing and explaining criteria with students, making use of exemplars to provide 'anchors' for assessing performance at various levels, and having a second rater/assessor to compare judgements (Brookhart, 2005). In the case of the last strategy, comparing results between two assessors is referred to as inter-rater reliability, while a single assessor evaluating the same work at two different time points is referred to as intra-rater reliability (Ary et al., 2014). Reliability is often reported in assessment instruments used in the GS literature (for example, Amatria et al., 2016; Arias-Estero, 2013; Bredt et al., 2016). A fourth version of assessment principles located in the Victorian landscape within the VC: HPE indicates that validity and reliability support credible reporting on student achievement (see Figure 2.1).

Figure 2.1

Principles of Assessment in HPE

Principles of assessment



Outcome = Credible reporting on student achievement

Note. From *Victorian Curriculum's Achievement Standards and Assessment - Health and Physical Education*, by C. Clark, 2022, VCAA. Copyright 2022 by Victorian Curriculum and Assessment Authority. Reprinted by permission.

Figure 2.1 featured in a professional learning webinar for Victorian primary and secondary school PE teachers titled “Introducing the Victorian Curriculum F - 10 and assessment for Health and Physical Education” (Clark, 2022). The principles of assessment were described as a synthesis of the work of the Australian academic and researcher, Geoff Masters (C. Clark, personal communication, April 1, 2022). Of note, other aspects of the assessment principles espoused by Masters (2013) that are not represented in Figure 2.1 include the use of rubrics and describing student progress made over time.

2.3 An Overview of Assessment Frameworks

A detailed consideration of assessment frameworks in the wider educational literature is beyond the scope of this review. For example, assessment frameworks designed in higher education do not address a secondary schooling context (Boud, 2000; Boud & Soler, 2016; Lindberg-Sand &

Olsson, 2008) and AfL frameworks in PE provide a narrow conceptualisation of assessment (for example, Leirhaug & Annerstedt, 2015; Tolgfors, 2018a, 2018b). Instead, the three assessment frameworks presented shortly have been designed and/or applied in a PE setting and provide a suitable background for the thesis. Positioned first in this review, the assessment literacy framework by Hay and Penney (2013) describes effective assessment practice for PE teachers. Second, the quality assessment framework of Chappuis et al. (2012) is included in this review due to its application in a study of the assessment practices of secondary school PE teachers (Borghouts et al., 2016). Third, the authentic assessment framework of Gulikers et al. (2004) is included in this review due to its application in a study investigating the assessment practices of secondary school PE teachers in GS (Georgakis et al., 2015). These three frameworks provide a rich, inter-connected and contextually relevant conceptualisation of assessment in PE to support this literature review and the wider thesis.

2.3.1 *Assessment Literacy*

Research into assessment literacy dates back to the 1990s and assessment literacy is considered fundamental to a teacher's ability to facilitate assessment as a process to support learning (DinanThompson & Penney, 2015). Teacher awareness of what constitutes good assessment and the capacity to apply that knowledge are central to teacher assessment literacy (Stiggins, 1991). Teachers as *assessment literates* should understand the importance of establishing clear achievement targets, be aware of factors that interfere with accurate assessment results and be able to act appropriately on any results gathered (Stiggins, 1991).

The assessment literacy framework advocated by Hay and Penney (2013) in the context of PE comprises four interdependent elements of assessment comprehension, assessment application, assessment interpretation, and critical engagement with assessment. Cognisant that assessment literacy needs to be viewed from both a teacher and student perspective (Hay & Penney, 2013; Stiggins, 2014), within the scope of this review and broader thesis only a teacher perspective is described. The first element of assessment literacy, assessment comprehension, is concerned with a teacher's knowledge and understanding of assessment practice and how efficacy can be achieved. Second, assessment application concerns how teachers or students conduct assessment. Third, assessment interpretation involves understanding and acting on assessment information mindful of the social nature of assessment. Fourth, critical engagement with assessment involves being aware of

the power imbalance of assessment and the impact of any consequences of assessment (Hay & Penney, 2013). This assessment literacy framework has also been adapted for use with primary school PE teachers (DinanThompson & Penney, 2015) and in PE teacher education (Starck et al., 2018). This assessment literacy framework may be viewed as describing essential aspects in the *process* of assessment.

Contextually relevant studies of GS assessment within Australian secondary schools by Georgakis et al. (2015) and Williams et al. (2020) indicate low levels of assessment literacy in PE teachers. For example, the latter study employed focus group interviews of 19 PE teachers to report that participants often made subjective judgements and generally used assessment approaches that were not fit for purpose (Williams et al., 2020). These findings suggest a lack of assessment comprehension and assessment application as described in the assessment literacy framework of Hay and Penney (2013). In the Australian state of Queensland, a qualitative case study of 18 primary school PE teachers reported superficial assessment practices, inadequate teacher moderation of assessment tasks, and limited student involvement in assessment design to suggest that there was engagement with “elements of the assessment literacy framework, *albeit* with limitation” (DinanThompson & Penney, 2015, p. 485) (emphasis added).

More positively, it is widely agreed that PE teachers can develop their assessment literacy through ongoing professional development and the establishment of a school-wide assessment culture (for example, Collier, 2011; DinanThompson & Penney, 2015; Slingerland et al., 2014). In addition to the above, on-site coaching and prompting by experts has been found to increase the use of formal-formative assessment of student performance by teachers in secondary school PE (van der Mars et al., 2018a). The AIESEP position statement on PE assessment also endorse greater investment in the development of assessment literacy for practitioners of PE (AIESEP, 2020).

2.3.2 *Quality Assessment*

Where assessment literacy refers to the broad knowledge and skills that support the enactment of assessment as a process, the following quality assessment framework addresses assessment design, provision of feedback, and the promotion of student involvement (Chappuis et al., 2012). The quality assessment framework describes five sequentially linked “keys” (Chappuis et al., 2012, p. 5) to quality classroom assessment. The first key in assessment design, *clear purpose*, involves teachers

informing all users why they are completing the task and how the task fits into the learning goals. In key two, the authors propose that *clear targets* allow teachers to provide transparent expectations and a clear path forward for student success (Chappuis et al., 2012). Providing students explicit assessment criteria, as an example of this transparency, has been reported as an essential aspect of helping students understand what they are expected to learn in PE teacher education (MacPhail et al., 2023).

A cross-sectional, quantitative inquiry found that Belgian, secondary school PE students that were aware of upcoming assessment criteria felt greater autonomy, competence and connection to teaching staff than students that were not aware of assessment criteria (Haerens et al., 2019). The affective, or social/emotional outcomes for the same students also indicated they felt energised and enjoyed their PE lessons more (Haerens et al., 2019). This inquiry did not manipulate or measure how assessment criteria were presented in the sample, suggesting that students' *perceptions* of their awareness of assessment criteria led to the positive impact on their motivation and enjoyment (Haerens et al., 2019).

Key three within the quality assessment framework, *sound design*, involves establishing valid and reliable assessments with explicit connection to learning goals (Chappuis et al., 2012). As this thesis centres on performance-based assessment, this key is explored in greater detail within the third framework of authentic assessment that follows shortly. *Effective communication*, as key four, ensures that expectations and results can be accessed and understood by all participants (Chappuis et al., 2012). This key relates directly to the inter-woven relationship between assessment, feedback and reporting. As the fifth and final key, *student involvement* includes students unpacking learning intentions, assessment protocols, and being involved in self or peer assessment. This quality assessment framework fits within aspects of the assessment literacy framework (Hay & Penney, 2013), especially comprehension, application and interpretation. This quality assessment framework also supports the High Impact Teaching Strategies of setting goals, explicit teaching and feedback within the Victorian Teaching and Learning Model (DET, 2023d) outlined in Chapter 1.

The quality assessment framework also informed the design of The Physical Education Assessment Questionnaire (PEAQ) that examined the assessment practices of secondary school PE teachers in a cross-sectional, quantitative inquiry in the Netherlands (Borghouts et al., 2016). A total

of 260 PE teachers, each representing a single school, completed the PEAQ with results indicating a lack of alignment between intended learning outcomes, the limited use of AfL despite strong philosophical support, and an overall view that assessment in PE was of a moderate quality (Borghouts et al., 2016). Following their study, Borghouts and colleagues recommended that stakeholders in PE should examine their current assessment practices and work together to optimise PE assessment in relation to its instructional alignment and the promotion of student learning (Borghouts et al., 2016).

2.3.3 *Authentic Assessment*

In this review, authentic assessment is positioned as an elaboration of sound design as key three within the above quality assessment framework (Chappuis et al., 2012). Authentic assessment requires students to use the same knowledge, skills, and attitudes that they would use in similar situations beyond an educational setting (Gulikers et al., 2004). Terms often used interchangeably with authentic assessment in the literature include rich tasks, performance assessment, and alternative assessment (for example, Lund, 1997; Smith, 1997; Zessoules & Gardner, 1991). In the school subject of PE, authentic assessment of game performance should centre on movement and integrate movement-associated concepts that are valued in a real-world context (Hay & Penney, 2013). When viewed in Arnoldian terms, the focus of assessment in this thesis is learning *through* movement (Arnold, 1979). Authentic assessment in GS *through* movement would likely include activities in “naturalistic game settings ... instead of performance on a skill test” (Melograno, 1997, p. 34).

One way to consider authentic assessment in GS is to distinguish movement tasks that may be classified as “playing form” or “training form” (Ford et al., 2010, p. 484). Playing form is based on a developmentally appropriate representation of the adult sport, which typically involves some form of opposition, an aim (possibly to score), and decision-making. For example, in basketball, two stationary players passing the ball to each other without any opposition and limited choices to make would be considered training form, while a 2 versus 2 half-court game activity that involved dribbling, passing and shooting against an opposition would be considered playing form. In addition, assessment through written work (tests, examinations, assignments) lacks authenticity in the performance-based context of PE (Georgakis et al., 2015). Journals and portfolios may provide a catalogue of authentic

psychomotor assessments (Boyer & Sweeting, 1996; Lee & Hare, 2007; Lund & Veal, 2013; Smith, 1997), however, in Arnoldian terms these types of assessment represent learning *about* movement rather than *through* movement and are not the focus of this thesis. Due to the lack of authentic assessment reported in PE (for example, Backman & Pearson, 2016; Barrientos Hernán et al., 2022; López-Pastor et al., 2013), the authentic assessment framework proposed by Gulikers et al. (2004) is presented below to help contextualise the challenges to assessment in PE, including GS.

Informing a relevant Australian case study of secondary school PE teachers' assessment practices in GS (Georgakis et al., 2015), is the five-dimensional authentic assessment framework of Gulikers et al. (2004). Originally developed to determine practical competence of nursing students in higher education, the framework's basis in practical performance may be *transferable* across other contexts (Tracy, 2010). The framework may also be conceptually *generalisable* beyond the original nursing population (Smith, 2017). Collectively, this means that the framework may be of use in a different demographic where the conceptual underpinnings resonate with the discerning reader. Gulikers and colleagues identify *authentic assessment task(s)* as dimension one of their framework. This dimension involves a problem-based task that is perceived as representative of professional practice (the world beyond education); thus, the task should be relevant and meaningful to users.

As dimensions two and three, the authors contend that the assessment task should make use of a *physical* and *social context* that represents the conditions of the real-world (Gulikers et al., 2004). In a PE, this means that GS should be played in courts and fields with team members, just as GS are played outside of school. In dimensions four and five, the authors propose that the task should produce an assessment result based on a recognisable form of performance that is based on criterion-referenced judgement that students actually value, and are made aware of (Gulikers et al., 2004). In the PE context of GS, this might include a range of game-based performances that allow students to demonstrate the requisite knowledge, skills and dispositions underpinning quality game performance in the community. Authentic assessment criteria should be explicit, understood and valued by learners, and preferably based on criterion-referenced judgement that may be provided from sources other than the PE teacher (Gulikers et al., 2004).

Gulikers and colleagues' framework of authentic assessment provides a nuanced elaboration of performance-based assessment within the aspect of sound design as key three in the quality

assessment framework (Chappuis et al., 2012). Another link between these two frameworks is the notion that students should be made aware of the task and understand the success criteria. The importance of this in the wider assessment literature is described through the metaphor of assessment being a “secret garden” (Broadfoot et al., 2002, p. 155), that students are often denied entry to. In providing guidance for assessment in PE, the quality assessment and authentic assessment frameworks align with many elements described in the assessment literacy framework (Hay & Penney, 2013). The three frameworks presented in this chapter include many overlapping elements with other assessment models, guidelines and frameworks in the wider literature; for example, formative assessment (Black & Wiliam, 2009), conditions for assessment efficacy in PE (Hay & Penney, 2009), and the Assessment Design Decision Framework (Bearman et al., 2016). The three frameworks described in this chapter provide a holistic and coherent conceptualisation of assessment designed and/or applied in a PE context.

2.4 Challenges of Assessment in PE

Notwithstanding broader issues regarding assessment that apply equally to the wider field of education, assessment in PE presents some distinctive challenges given the inherent nature of students to move while learning (Georgakis et al., 2015). Key challenges to providing quality PE programs in schools, such as a lack of time, inadequate facilities and equipment, and large class sizes (Jenkinson & Benson, 2010; Morgan & Hansen, 2008), are similarly reported as challenges for assessment in PE (Braga & Liversedge, 2017; Gallo et al., 2006; Veloo & Md Ali, 2016). Other issues include the need to balance assessment within and across three learning domains (cognitive, affective and psychomotor) (Hay & Penney, 2013; Mitchell et al., 2013); the historically low level of engagement in assessment theory by PE teachers (Veal, 1988); and confusion in determining what students should be assessed on (Mohnsen, 2006). Further, when assessing students through observation of performance or verbal responses to questions, judgement may be impacted by a number of observational biases (Simon et al., 2017) and recollecting students’ actual performances or individual responses cannot be assured (Earl, 2014). Collectively, these challenges may lead to inconsistent assessment approaches in PE (MacPhail & Murphy, 2017).

Other challenges reported for assessment in secondary school PE programs in the late 1980s remain as relevant today and may stem from the same lack of theoretical understanding (Veal, 1988).

Historical issues for assessment in PE have included: PE teachers' narrow views of assessment; disparate understandings of the goals of PE programs; issues with record keeping; and limited professional preparation in teacher education courses (Veal, 1988). In a recent literature review examining the use of AfL in PE, teachers were found to lack the pre-requisite skills to apply AfL in their assessment practice, and instead they used assessment exclusively for grading and reporting purposes (Moura et al., 2021). The conflation of assessment and grading in the field of PE does not appear to have changed substantially in almost 40 years (Veal, 1988).

Further challenges for assessment in PE include a consideration of what students should be assessed and graded on (Mohsen, 2006). For example, traditionally prevalent standardised assessments like Physical Fitness tests have been associated with negative outcomes for students (Alfrey & Gard, 2017), while objective motor skill tests may lack authenticity (López-Pastor et al., 2013). Criticisms of the intuitive or subjective nature of PE assessment are widely reported in the literature (for example, Birky, 2012; Hay & Macdonald, 2008; Svennberg, 2017) and include reports of practitioners making judgements based on a 'gut-feeling' (for example, Annerstedt & Larsson, 2010; Hay & Macdonald, 2008). There is also evidence that in eschewing assessment based on criteria or standards in a curriculum, PE teachers assess non-performative criteria like attitude, effort and participation (Baghurst, 2014; Veal, 1988; Williams et al., 2020) and have difficulty locating appropriate assessment tools (Killian & Mays, 2021; Williams et al., 2020).

The relatively high degree of autonomy that PE teachers have in administering assessment in low stakes (non-examinable) PE (MacPhail & Murphy, 2017), may contribute to inconsistent assessment approaches within and across schools. Given this autonomy, some PE teachers may assess students through skill-driven measures of competency, while others may assess students on criteria outside of the curriculum (Baghurst, 2014). There is also evidence in a sports coaching context, that may apply equally to PE, that observation of players in games can be negatively affected by up to 20 different *heuristics*, or mental shortcuts used in problem-solving, that might lead to bias and reduce reliability in judgement (Simon et al., 2017). Two of these heuristics include the *primacy and recency bias* that states that events that happen early or late over a period of observation are more easily recalled than events that occur in the middle of the same time period. Second, the *confirmation bias* states that evidence supporting any a priori views may be overvalued in comparison to dis-confirming

evidence (Simon et al., 2017). As an example of the confirmation bias in PE, a teacher may consider a student to be a highly capable game player and attribute an error in game performance to the action of another student, rather than challenge their pre-conception of the student as highly capable.

Given the dynamic environment of the gymnasium, the general requirement for students to be moving, and the often weather-affected outdoors, a practical challenge for assessment in PE is how teachers gather data (Veal, 1988). The use of students in peer and self-assessment is one strategy that may support the gathering of data and address the view that teacher led assessment in PE is laborious (Braga & Liversedge, 2017; Gallo et al., 2006; Veloo & Md Ali, 2016). In support of engaging students as assessors, students in Grade 5-8 who completed peer assessment using the TSAP were found to have a “moderate to good level of precision and inter-observer reliability” (Richard et al., 2000, p. 90). Others have claimed that subjective assessment criteria reduce the reliability of peer assessment in GS (Georgakis et al., 2015; Gibbons & Kankkonen, 2011) and that training students to apply assessments and provide peers with constructive feedback takes up valuable instruction time (Gibbons & Kankkonen, 2011).

At a national level, teacher and student perceptions of assessment in PE reported in the Australian state of Queensland indicate that there was much less structure and clarity surrounding assessment in the middle years (Year Levels 4 – 9) than in the senior years (Year Levels 10 - 12) (Chan et al., 2011). The authors’ pilot study, comprising a series of one-on-one interviews with teachers and students in two state secondary schools, reported that while Year Level 9 students were aware that their PE grades were based on fitness and skill development, they were not sure when they were assessed, or what the exact criteria were (Chan et al., 2011). The authors found that these practices marginalised the academic merit of PE in the middle years and concluded their study by recommending that stakeholders develop more efficacious assessment practices that better represent valued aspects of learning in PE (Chan et al., 2011).

2.4.1 Challenges of Assessment in Games and Sports (GS)

Consistent with assessment in the broader PE landscape, assessment in GS has been reported to lack validity, reliability and authenticity (Georgakis et al., 2015; López-Pastor et al., 2013; Williams et al., 2020). Assessment in GS has generally neglected the importance of tactical awareness and instead focussed on the assessment of de-contextualised technical skills and non-performative criteria like

student effort, attitude or behaviour (for example, Blomqvist et al., 2005; Borghouts et al., 2016). In addressing the importance of tactics in the assessment of GS, relevant literature reviews that informed this chapter included investigations of tactical assessment, tactical performance, tactical learning, tactical evaluation and decision-making (for example, Ávila-Moreno et al., 2018; González-Víllora et al., 2015; Inns et al., 2023). Consistent with the above reviews (for example, Ávila-Moreno et al., 2018) the terms assessment *tool* and assessment *instrument* are used interchangeably to represent the means by which assessment data are captured throughout this review and the thesis.

Authentic assessment in the Focus Area of GS is predicated on the assessment learning and/or performance through the playing of games (for example, Georgakis et al., 2015; Gréhaigne et al., 1997). Appreciating the importance of game play supports the pedagogical alignment of GBAs to performance-based assessment, and the belief that authentic assessment can develop game appreciation, awareness of game play, individual performance and team performance (Slade, 2010). Authentic assessment in GS should involve aspects of decision-making, skill technique, movement, calling to teammates, and using body language (Slade, 2010). Specific to decision-making in IGS, the need to continuously perceive and respond to information in a dynamic and time-constrained environment means that assessment is “vastly complex” (Inns et al., 2023, p. 1). As an example of this complexity, a player may choose to dribble rather than pass a ball based on their physical abilities (psychomotor domain), even though they understand the advantage of making a long pass to an open team-mate closer to goal (cognitive domain) (Inns et al., 2023).

To address issues with the reported lack of validity, reliability and authenticity in GS assessment, a number of evidence-based assessment instruments have been developed for use within a school-age population. Despite the increased proliferation of such assessments, there is limited use of such tools in a GS context (for example, López-Pastor et al., 2013; Williams et al., 2020; Young, 2011). As prominent examples of these tools in the literature, the GPAI (Oslin et al., 1998) and TSAP (Gréhaigne et al., 1997) count or tally observable game behaviours to generate performance outcomes based on quantitative data. Authors of both tools report validity and reliability measures indicating tactical-technical performance of IGS in authentic assessment conditions (Gréhaigne et al., 1997; Oslin et al., 1998). These findings are supported in an international review of alternative assessment in PE that determined both tools were authentic, valid and reliable in measuring individual student learning and

performance in games (López-Pastor et al., 2013). In comparison to other evidence-based frequency-count tools developed for use within a school-age population, like the Basketball Learning and Performance Assessment Instrument (BALPAI) (Ibanez et al., 2019) and the Coding Instrument designed for the assessment of soccer (Blomqvist et al., 2005), the TSAP and GPAI are generic in nature, meaning they can be used in multiple IGS. A detailed description of both assessment tools, including their criteria, formulae for deriving performance outcomes, and sample coding sheets is provided in Appendix I.

Some criticisms directed at frequency-count tools include that they neglect situational details (Hastie et al., 2013), calculation of some performance outcomes such as efficiency indexes may distort performance (Barquero-Ruiz et al., 2019; Memmert & Harvey, 2008), off-the-ball actions are often overlooked (Arias & Castejón, 2012), they can be time consuming to administer (Hastie et al., 2013), and the fast pace and lack of natural breaks in IGS can mean that some behaviours are missed entirely (Mitchell et al., 2013). As an example of the laborious nature of frequency-count tools, a secondary school PE teacher in the Georgakis et al. (2015) study argued that replaying video tapes to tally individual student performance in applying the GPAI was untenable as it would take “hours upon hours” (p. 78). The respondent did not address the possibility of using peer assessment in real time to complete the assessment. While views on the utility of frequency-count tools are divided in the literature, with reference to the TSAP and GPAI, several studies concur that these tools are better suited to research given that PE teachers rarely use them (López-Pastor et al., 2013; Williams et al., 2020; Young, 2011). While the designers of the GPAI do not acknowledge the potential of video capture to support assessment in IGS in school settings, they do recognise that using the tool in real-time is untenable (Mitchell et al., 2013).

The tally system can be used with striking and fielding games and some net and wall games because they are played at a slower pace and have natural breaks (between pitches, bowls, or points), which gives the observer an opportunity to score or tally every event ... In invasion games and some net and wall games, this is impossible to do effectively because of the tempo, flow and unpredictability of such games ... when a lot of action might be missed in the recording or writing process. (Mitchell et al., 2013, p. 50)

In contrast to frequency-count tools, rubrics have also been reported as authentic, reliable and/or valid in the assessment of GS in secondary school PE (Penney et al., 2012; Williams & Rink, 2003). Rubrics are generally presented in a table format and comprise criteria, descriptors and standards of attainment to define what performance looks like at various levels of quality (Jonsson & Svingby, 2007). Table 1.1 in Chapter 1 presented a sample IGS rubric with four criteria and five levels of achievement (Breed & Spittle, 2011). The origins of the term rubric date back to the 15th century with reference to headings or important sections of mostly religious texts that were marked in red, based on the Latin term for red, “ruber” (Popham, 1997, p. 72). Thus, rubrics should identify what is important in the performance. More recently, rubrics designed for the qualitative rating of student work have been described as scoring guides (Siedentop et al., 2011) and instructional guides (Andrade, 2005). These terms indicate the capacity of the rubric to score (grade) and instruct (teach) students, respectively.

The review in this chapter located two relevant studies in an Australian secondary school PE context that examined the assessment of GS. The first study in the Australian Capital Territory (ACT) reported that some of the 19 PE teachers interviewed felt the AC: HPE lacked sufficient detail to assess specific sports and sport-related skills (Williams et al., 2020). The second study in the state of New South Wales (NSW), undertaken prior to the release of the Australian Curriculum, reported that many of the 17 PE teachers interviewed were critical of the relevance of assessment criteria to the state curriculum and the real world (Georgakis et al., 2015). These authors observed that some respondents partially assessed students on curriculum requirements, but there was a deficit in their assessment practice as tactics and team play were not addressed in their assessment (Georgakis et al., 2015).

Challenges specific to the assessment of IGS faced by PE teachers in Victorian secondary schools remain largely unknown, given the dearth of literature on this topic. For example, while studies in secondary school assessment of GS in PE were located in other Australian states and territories (Georgakis et al., 2015; Williams et al., 2020), no studies investigating the assessment of IGS in the state of Victoria were located in this review. With reference to particular assessment instruments for use in GS, the VC: HPE endorse the development and use of rubrics for formative assessment (AfL). As an example of this endorsement, the online resource “Using Formative Assessment Rubrics in Health

and Physical Education - Invasion, Net and Wall, Striking and Fielding Games Levels 7 - 10" (VCAA, 2019b) may be useful for rubric design. In this resource, four steps are suggested to implement formative rubrics that include: (a) describing a learning continuum; (b) developing a formative assessment rubric; (c) collecting, interpreting and using evidence to plan for teaching and learning; and (d) formative assessment rubrics in practice (VCAA, 2019b). It is unclear how well known, or widely used, this advice for rubric design is within PE teachers in Victorian secondary schools (Year Levels 7 - 10), and if any similar resources exist that support the use of frequency-count approaches for the same population.

2.5 Conclusion and Further Research

This chapter has provided a synthesis of key assessment concepts, frameworks and challenges in the assessment of PE literature. This review included a consideration of the Focus Area of GS within the VC: HPE to support the wider thesis. This chapter identified challenges faced by PE teachers in their assessment practice that included large class sizes (Jenkinson & Benson, 2010; Morgan & Hansen, 2008) and high levels of autonomy that may contribute to inconsistent assessment practice (MacPhail & Murphy, 2017). Professional capacity challenges in assessment were identified that included skill deficits in implementing AfL strategies (Moura et al., 2021), and PE teachers' proclivities for internalised or subjective assessment (Birky, 2012; Svennberg, 2017; Williams et al., 2020). Assessment quality challenges included a lack of authenticity (for example, Backman & Pearson, 2016; Barrientos Hernán et al., 2022) and the assessment of students through criteria like attitude and effort (Baghurst, 2014; Williams et al., 2020).

Several literature reviews investigating the assessment of GS in PE and/or youth sports informed this chapter (for example, Arias & Castejón, 2012; Barquero-Ruiz et al., 2019; González-Villora et al., 2015), however, these reviews also indicated gaps in the literature. While each review added to the body of knowledge regarding assessment in GS, one of those gaps is that no literature review was conducted to locate and describe evidence-based assessment tools applied exclusively in IGS within a school-age population. Other potential topics for further research include examining PE teachers' assessment literacy (Hay & Penney, 2013) and their use of quality assessment (Borghouts et al., 2016) and/or authentic assessment (Georgakis et al., 2015) frameworks. As the GPAI and TSAP are considered authentic game-based assessments (Gréhaigne et al., 1997; Oslin et al., 1998), identifying

reasons for their limited uptake in PE is worthy of further investigation. Given the lack of authentic GS assessment reported in secondary school PE in other Australian states and territories (Georgakis et al., 2015; Williams et al., 2020), and the paucity of research in this context in Victoria, investigations of PE teachers' assessment practice in this context are recommended.

This chapter has established problems and located gaps in the assessment in PE literature. It has described challenges specific to the assessment of GS and outlined potential areas and populations for further research. Further, it was shown that a number of authors recommend that researchers investigate and improve assessment practices in PE (Borghouts et al., 2016; Chan et al., 2011; Hughes et al., 2012). Thus, in response to these recommendations, the research design in the chapter that follows seeks to “further scrutinize assessment in PE and introduce reforming evidence-based practice” (Georgakis et al., 2015, p. 84).

CHAPTER 3: RESEARCH DESIGN

3.1 Introduction

This chapter presents the research design for the thesis that is considered an explanatory, sequential, mixed methods approach (Creswell & Plano Clark, 2018; Schoonenboom & Johnson, 2017). In outlining the research design, the methodological approach is described and justified as a guiding strategy or broad plan of action (Crotty, 1998; McChesney & Aldridge, 2019). This chapter begins by re-stating the research aim and question before providing an overview of the studies and the related research sub-questions. This includes justifying the use of a pragmatic paradigm and applying common typologies in mixed methods to clarify the approach of the thesis (Creswell & Plano Clark, 2018; Schoonenboom & Johnson, 2017). Consistent with the aim of furthering understanding in the field of assessment, the rationale for employing an interpretive theoretical framework is presented (Goldkuhl, 2012; McChesney & Aldridge, 2019; O'Donoghue & Farrelly, 2022). The specific tools, techniques or procedures used in each study constitute the methods (Crotty, 1998; McChesney & Aldridge, 2019) which are described and justified in the four studies that comprise Chapters 4 to 7 within the thesis.

3.1.1 Research Overview

The broad aim of this thesis is to: *Describe, explain and make meaning of Physical Education teachers' understanding and practice of assessment in invasion games and sports in Victorian secondary schools (Year Levels 7 – 10)*. In applying a pragmatic world-view, the thesis emphasises the production of useful knowledge based on real-world problems identified by the population under investigation (Kelly & Cordeiro, 2020). Thus, the following overarching research question asks: *How do Physical Education teachers understand and practice assessment in the curriculum focus area of games and sports in Victorian secondary schools?* In Table 3.1, an overview of the research sub-questions, study design, chapter number, emergent design, and a brief outline of the specific methods are presented. Of note, an emergent mixed method research design evolves over the duration of the research as earlier studies inform later studies (Schoonenboom & Johnson, 2017).

Table 3.1*Overview of the Research Sub-Questions and Emergent Design of the Thesis*

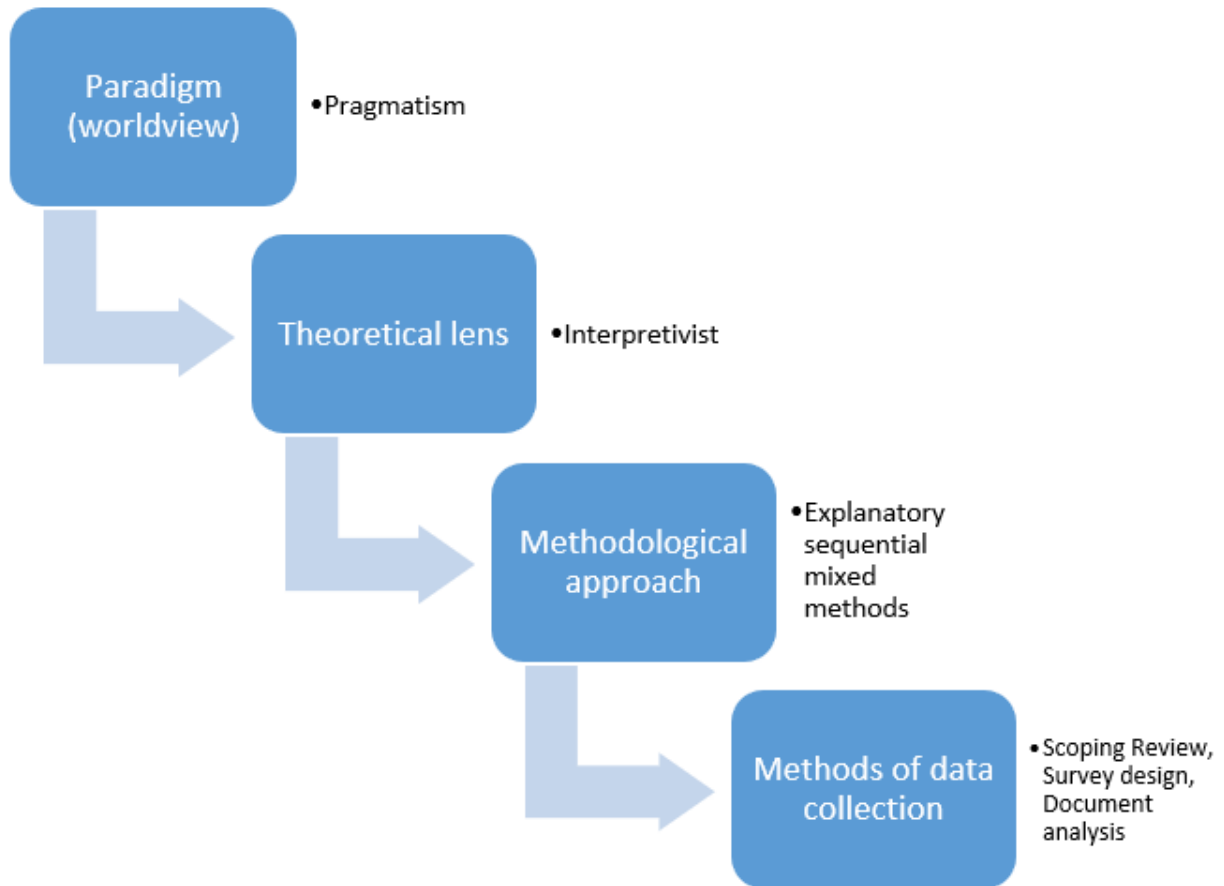
Research sub-questions	Study Design	Chapter	Emergent design	Brief outline of method
1. What does the extant literature say are the defining characteristics of assessment tools developed for invasion games and sports?	Quantitative – Scoping Review (ScR) following PRISMA guidelines	4	Informed by Chapter 2: Assessment in Physical Education (PE) – A Literature Review	A ScR of peer-reviewed literature to identify evidence-based assessment tools. Data charting to describe the assessment tools' key performance criteria, characteristics, and applications using descriptive statistics and analysis.
2. How do Physical Education teachers view the assessment of invasion games and sports in Victorian secondary schools?	Survey – Cross-sectional quantitative inquiry	5	Informed by the ScR that identified prevalent, evidence-based assessment tools (GPAI and TSAP) and key performance criteria.	Online questionnaire administered through Qualtrics to describe the views of assessment reported by PE teachers in Victorian secondary schools. Data presented using descriptive statistics and analysis.
3. How do Physical Education teachers practice and/or want to practice assessment of invasion games and sports in Victorian secondary schools?	Survey – Qualitative inquiry	6	Informed by the cross-sectional, quantitative inquiry that identified key topics for further explanation including curriculum alignment to assessment, types of tools used, use of assessment for learning and assessment utility.	Semi-structured interviews of a nested, purposive sample of participants from the cross-sectional, quantitative inquiry. Data interpreted through reflexive thematic analysis.
4. How are Physical Education teachers' assessment tools constructed for invasion games and sports in Victorian secondary schools?	Document analysis – Qualitative inquiry	7	Informed by both survey design studies that identified the prevalence and criticism of rubrics from participants.	Qualitative document analysis of assessment artefacts (rubrics) from a nested sample of participants from the cross-sectional, quantitative inquiry. Data interpreted through content and thematic analysis.

Note. The studies are considered emergent as earlier studies (chapters) informed the design of later studies (chapters) (Schoonenboom & Johnson, 2017).

In employing functional pragmatism (Goldkuhl, 2011, 2012) to answer the above research sub-questions, each of the four study chapters and concluding chapter provide several recommendations that address the real-world problems identified by participants. The research process of this thesis presented in Figure 3.1, is based on the four research *elements* described by Crotty (1998) and the subsequent reframing of these elements as *levels* in mixed methods research by Creswell and Plano Clark (2018). As the diagram in Figure 3.1 reflects the latter process more closely, only those authors are attributed in the note below the figure. The conventions of the source material are preserved by using single arrows to illustrate how the levels inform one another (Creswell & Plano Clark, 2018).

Figure 3.1

Levels of the Research Process in this Thesis



Note. Adapted from *Designing and Conducting Mixed Methods Research* (3rd edn.), by J. W. Creswell and V. L. Plano Clark, 2018, (p.35) SAGE. Copyright 2018 by SAGE publications. Adapted by permission.

3.2 Research Paradigm

In this thesis, pragmatism is considered a *paradigm* that embraces ontology as well as epistemology (Creswell & Plano Clark, 2018). Within this view of a paradigm, ontology refers to the nature of reality, while epistemology addresses how we gain knowledge or how we actually know (Creswell & Plano Clark, 2018). Pragmatism rejects a rigid position on using different research methods, instead preferring to use a combination of methods, in a pluralistic approach that best serves the research purpose (Goldkuhl, 2011, 2012). This understanding and application of pragmatism draws on descriptions in mixed methods research (for example, Creswell & Plano Clark, 2018; Johnson & Onwuegbuzie, 2004; Teddlie & Tashakkori, 2009) and pragmatic criteria and principles reported in the wider literature (for example, Henry & Feuerstein, 2003; Kelly & Cordeiro, 2020; Shusterman, 2010). Pragmatism was adopted as the research paradigm as it endorses an eclectic approach to research that values the inner world experience of participants and prefers action to philosophising (Johnson & Onwuegbuzie, 2004).

Key tenets of pragmatism in this thesis include the focus of the research in a real-world problem, the importance of the research question(s), the use of the most appropriate methods to collect and analyse data to respond the research question(s), valuing both quantitative and qualitative data, creating actionable knowledge, and encouraging stakeholders to provide solutions to their problems (for example, Creswell & Creswell, 2018; Goldkuhl, 2011, 2012; Kelly & Cordeiro, 2020). Functional pragmatism can be classified as *local* and *general* according to the population that the knowledge for action is directed towards (Goldkuhl, 2011, 2012). Given that three of the four studies were based on a sample of Victorian secondary school PE teachers, the thesis primarily employs local functional pragmatism (Goldkuhl, 2011, 2012). However, some of the findings and recommendations may constitute knowledge for action beyond the local population, consistent with general functional pragmatism (Goldkuhl, 2011, 2012).

3.2.1 Theoretical Lens (Perspective)

A theoretical lens (Creswell & Plano Clark, 2018) or perspective (Crotty, 1998) is a “way of looking at the world and making sense of it” (Crotty, 1998, p. 8). An interpretive theoretical lens was employed to make sense of the data as the research focussed on furthering understanding within a pragmatic paradigm (Goldkuhl, 2012; Potrac et al., 2014). In adopting an interpretivist theoretical

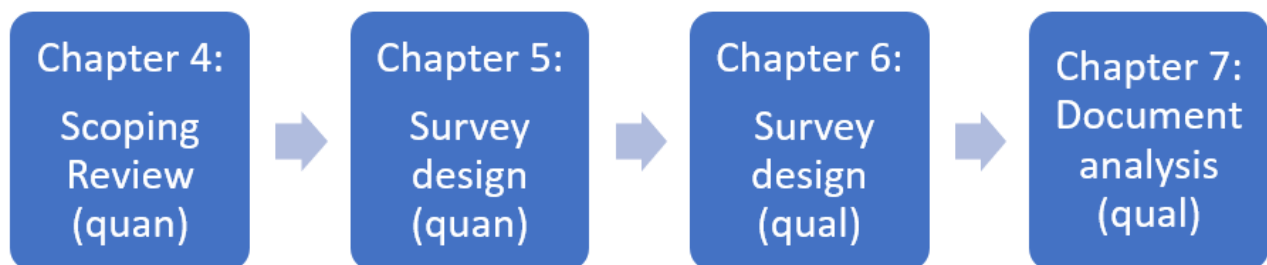
perspective, knowledge was viewed as socially constructed, subjective and interactional, rather than a “fixed and stable phenomenon” (Potrac et al., 2014, p. 33). In recognising the contested nature of terms that comprise the research process, interpretivism is often described as a paradigm in the wider literature (for example, Allemang et al., 2022; Christensen, 2020; Ryan, 2018; Shan, 2021). This thesis accommodates this alternate view by supporting ‘paradigm pluralism’ (Teddlie & Tashakkori, 2012), such that, pragmatism would be considered the ‘base’ paradigm of this thesis (Goldkuhl, 2012), thus eschewing any requirement to embrace the entire set of beliefs of any other complementary paradigm (Willis, 2012). In supporting the use of this interpretive lens, researcher positionality was addressed in Chapter 1 and qualitative rigor and researcher reflexivity are addressed in both qualitative studies (Chapter 6 and 7).

3.2.2 Methodology

The choice of a mixed methods approach supported the research aim to describe (quantitative studies) and explain (qualitative studies) teachers’ understanding and practice of assessment (Creswell & Plano Clark, 2018). In employing a widely recognised notation system to represent mixed methods research diagrammatically (Creswell & Plano Clark, 2018; Hesse-Biber et al., 2015; Teddlie & Tashakkori, 2009), Figure 3.2 presents a flow chart of the respective studies. Of key interest is the use of arrows and abbreviated text in parenthesis to represent the type of mixed methods approach and the broad nature of each study, respectively.

Figure 3.2

A Flow Chart of the Explanatory, Sequential, Mixed Methods Study Design



Specifically, the arrows between the studies in Figure 3.2 indicate the *sequential* design of the thesis, as opposed to the use of an addition sign (+) that would indicate that studies were completed concurrently. The use of the abbreviations ‘quan’ for quantitative and ‘qual’ for qualitative in the flow

chart indicates the predominant nature of each study/chapter (Creswell & Plano Clark, 2018; Hesse-Biber et al., 2015; Teddlie & Tashakkori, 2009). Thus, the flow chart in Figure 3.2 indicates that the studies in Chapters 4 and 5 were quantitative in nature and the studies in Chapters 6 and 7 were qualitative in nature. A possible limitation of the notation system of Creswell and Plano Clark (2018), is that a single study cannot be identified as mixed methods.

A key advantage of mixed methods research is that quantitative and qualitative research provides “complementary strengths” (Teddlie & Tashakkori, 2009, p. 98). This view posits that the strengths of one research method or data set can enhance another research method or data set (Maarouf, 2019; Teddlie & Tashakkori, 2009), and expand or consolidate research findings (Schoonenboom & Johnson, 2017). The two quantitative studies in this thesis provide context for the investigation of assessment practice in IGS, while the two qualitative studies help to explain and make meaning of this information within a nested sample of the population (Onwuegbuzie & Collins, 2007). As the intent of the thesis was to use qualitative data to explain quantitative findings, and the studies were completed in a sequence over a period of time, rather than completed concurrently, this approach is considered an explanatory, sequential, mixed methods design (Creswell & Plano Clark, 2018). In recognising that other nuanced aspects contribute to the typology of mixed methods research, the work of Schoonenboom and Johnson (2017) is applied to this thesis.

The seven major design dimensions of mixed methods research, described by Schoonenboom and Johnson (2017) provide greater clarity in the methodology. As dimension one, this mixed methods approach had two key *purpose(s)* that included making use of findings to shape future studies (development), and to elaborate on earlier findings (complementarity). In dimension two, the *theoretical drive* was equal-status, as quantitative and qualitative research contributed equally to answering the overarching research question. Equal-status, mixed methods research supported the broader pragmatic worldview to show that “paradigms can be mixed or combined, and that the incompatibility thesis does not always apply to research practice” (Schoonenboom & Johnson, 2017, p. 113).

As dimension three, the *timing* was considered sequential and dependant, as the studies occurred over time and the design of later studies was dependant on earlier studies. The sequential structure of the studies allowed time to “carefully listen to, consider, and continually dialogue with

qualitative and quantitative perspectives/epistemologies/values/methods and learn from the natural tensions between these while developing a workable solution for each mixed research study” (Johnson, 2017, p. 161). This time allowed for consultation with the supervisory team to design studies that were responsive to the preceding data sets.

In dimension four, the *point of integration* refers to how and when quantitative and qualitative components came together. Data were integrated in the research design phase, and in the findings and discussion sections of the qualitative studies. Further, the concluding chapter included a translation section that integrated the key findings from the four studies comprising the thesis. As an example of integration in the design stage, findings from the cross-sectional, quantitative inquiry (Chapter 5) informed the interview schedule in the following qualitative inquiry (Chapter 6). As dimension five, *typological design*, the naming conventions of Creswell and Plano Clark (2018) are supported in classifying this mixed methods approach as being explanatory and sequential in nature. This dimension also indicates that the research approach is multiphase, as it includes more than two studies or phases, and has the aim of explaining a phenomenon (Schoonenboom & Johnson, 2017).

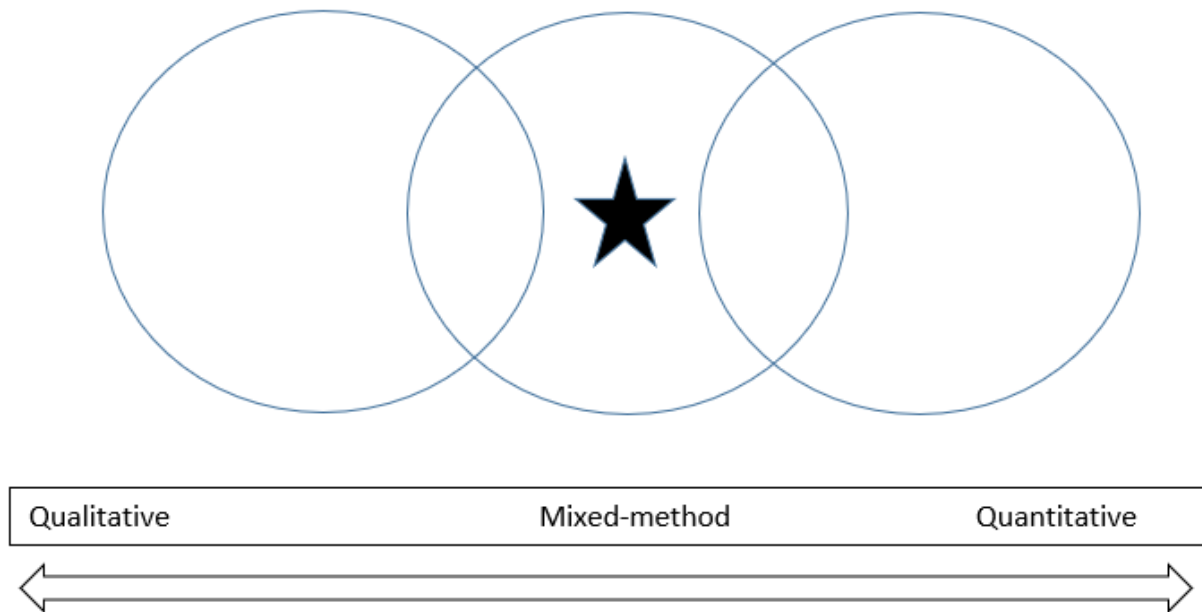
In dimension six, this mixed methods approach is considered *emergent* rather than planned, as latter studies could not be fully constructed until previous studies were completed. As an example of this emergent design, the final study, a qualitative inquiry that employed a document analysis of assessment rubrics (Chapter 7), was not conceived until the participants in the preceding study (Chapter 6) indicated that their rubrics lacked utility. As dimension seven, the *complexity* of this mixed methods approach is considered high due to its multi-phase nature and the multiple points of interface, where data informed both study design and enhanced understanding of the collected data (Schoonenboom & Johnson, 2017). For simplicity and consistency throughout the thesis, the methodological approach is termed an explanatory, sequential, mixed methods design (Creswell & Plano Clark, 2018; Schoonenboom & Johnson, 2017).

It has been suggested that mixed methods research designs should be assessed on a different set of criteria in comparison to monomethod approaches (Fàbregues & Molina-Azorín, 2017). While research design guidelines for mixed methods studies in the field of PE can be found in the literature (Anguera et al., 2017; Camerino et al., 2012), research criteria are historically and contextually bound, making their universal application problematic (Halcomb, 2019). With reference to the relative role of

quantitative and qualitative approaches in research design, the continuum in Figure 3.3 indicates that the range and scope of studies in this thesis are best described as mixed methods (Teddlie & Tashakkori, 2009).

Figure 3.3

Positioning of this Research Design



Note. Adapted from *Foundations of Mixed Methods Research: Integrating Quantitative and Qualitative Approaches in the Social and Behavioral Sciences*, by C. Teddlie and A. Tashakkori, 2009, (p. 28), SAGE. Copyright 2009 by SAGE publications. Adapted by permission.

In briefly summarising the role of quantitative and qualitative research and the use of an interpretive lens, quantitative data (numerical data) were collected in three of the four studies. The treatment of these data involved a basic approach to statistics, rather than a more complex, inferential approach. All four studies comprising the thesis were supported by an interpretive theoretical lens in the discussion (Chapter 4 and 5), or findings and discussion sections (Chapter 6 and 7). Presenting the findings and discussion together in the qualitative studies avoided repetition in the treatment of data that can occur when presenting these sections separately, and supported interpretive analysis (Braun & Clarke, 2013). The two qualitative studies employed thematic analysis to support interpretation of the understanding and practice of assessment by participants. As an example of the thematic analysis employed in these studies, the qualitative inquiry using a survey

design (Chapter 6), employed a reflexive thematic approach based on the influential work of Braun and Clarke (2021a) to make meaning of participant responses.

3.3 Conclusion

This chapter described the research design for this thesis as an explanatory, sequential, mixed methods design (Creswell & Plano Clark, 2018; Schoonenboom & Johnson, 2017). This design is informed by functional pragmatism that responds to real-world problems identified by participants and contributes knowledge to improve action (in assessment practice) (Goldkuhl, 2011, 2012). The use of an interpretivist lens was congruent with the pragmatic paradigm (Goldkuhl, 2011, 2012) and the aim of understanding a complex phenomenon (O'Donoghue & Farrelly, 2022). The research design was justified and described with reference to common mixed methods notational systems and typologies (Creswell & Plano Clark, 2018), including the seven major dimensions of mixed methods studies identified by Schoonenboom and Johnson (2017): purpose; theoretical drive; timing; point of integration; typological design; emergent nature; and the degree of complexity. Consistent with a recent mixed methods approach based on a population of teachers (McChesney & Aldridge, 2019), the studies involving PE teachers in Victorian secondary schools in this thesis “privileged teacher voice and accepted teachers’ accounts as reflecting their constructions of reality” (p. 231).

The thesis follows the conventions associated with reporting quantitative and qualitative research. The introduction and the conclusion (Chapter 1 and 8) and both qualitative studies (Chapters 6 and 7) are written using a first-person voice to situate the researcher in the study. The remaining chapters, including those that present the two quantitative studies (Chapters 2, 3, 4 and 5), are written using the third-person voice. The use of a first-person voice in qualitative work is well established and is justified by qualitatively driven researchers to be reflexive (Webb, 1992). For example, in employing reflexive thematic analysis (Braun & Clarke, 2021a) in Chapter 6, the qualitative research reported in that study employs a first-person active tense.

Each of the four chapters that follow respond to the four sub-questions presented in Table 3.1 and describe the specific methods, or how the studies were undertaken. These chapters describe, explain and make meaning of Victorian secondary school PE teachers understanding and practice of assessment. This knowledge includes findings and recommendations that constitute a significant and

original contribution to knowledge that may support the assessment practices of the local population. Consistent with *general* functional pragmatism (Goldkuhl, 2011, 2012), some of this knowledge may also support PE teachers, educators and assessment designers beyond the local context. The knowledge generated as a consequence of this explanatory, sequential, mixed methods design is underpinned by the belief that better assessment practices can improve teaching and enhance educative outcomes for students (Black & Wiliam, 1998a, 1998b; Earl, 2014). To provide a point of reference for making meaning of the understanding and practice of assessment by PE teachers in Victorian secondary schools, Chapter 4 describes a ScR of peer-reviewed evidence-based assessment tools used in the context of IGS within a school-age population that was conducted to answer research sub-question 1.

CHAPTER 4: SCOPING REVIEW (SCR) OF PEER-REVIEWED EVIDENCE-BASED ASSESSMENT TOOLS IN INVASION GAMES AND SPORTS (IGS) WITHIN A SCHOOL-AGE POPULATION

4.1 Introduction

The literature review of key concepts, frameworks and challenges in PE assessment in Chapter 2 acknowledged that the authentic assessment of GS in the subject of PE was limited (Georgakis et al., 2015; López-Pastor et al., 2013; Williams et al., 2020). The review acknowledged that validity and reliability were key considerations in determining assessment quality (Brookhart, 2005; Chappuis et al., 2012; Hay & Penney, 2013) and that some assessment rubrics and frequency-count tools have been reported as valid and/or reliable assessment instruments in GS. Chapter 2 did not aim to provide a full and thorough list of IGS assessment tools used in the published peer-reviewed literature.

Chapter 2 referred to other useful literature reviews that investigated the assessment of game performance and contributed to the field of study (Arias-Estero & Castejón, 2014; Ávila-Moreno et al., 2018; Barquero-Ruiz et al., 2019). However, none of the above reviews were conducted to locate and describe the characteristics of evidence-based assessment tools used exclusively in the context of IGS within a school-age population (age 5 to 19 years), which suggests this study constitutes an original contribution to knowledge. The identification of evidence-based tools was necessary to determine the performance criteria, characteristics, and applications of assessment instruments suggested in academic literature for use in IGS. These data provide a reference point to support a fuller consideration of the understanding and practice of assessment by PE teachers in Victorian secondary schools in the thesis. The findings of the study described in this chapter are followed by a series of recommendations that may be of use to researchers and PE teachers in the context of assessment of student performance in IGS.

4.1.1 Aim

The aim of this review is to describe the range and nature of IGS assessment tools used within a school-aged population and published in the peer-reviewed academic literature. This ScR answers research sub-question 1: *What does the extant literature say are the defining characteristics of assessment tools developed for invasion games and sports?* Within the family of purpose-specific

literature reviews and review typologies (Sutton et al., 2019), a ScR was the most appropriate form of review to locate and describe evidence-based assessment tools by making a preliminary assessment of the potential size and scope of available research. ScRs have been described as an increasingly common approach to examining the literature (for example, Colquhoun et al., 2014; Levac et al., 2010; Pham et al., 2014), specifically when the research need is to map a body of literature on a topic area (Arksey and O'Malley, 2005). ScRs are best suited for broad research questions, such as research sub-question 1, and help to identify research gaps (Arksey & O'Malley, 2005).

4.2 Method

The ScR followed the five-step framework described by Arksey and O'Malley (2005) and included identifying the research question; identifying relevant studies; selecting studies; charting the data; and collating, summarizing and reporting the results. Indicating the utility of this framework, two recent reviews indicated the framework has been used in more than half of all ScR studies (Pham et al., 2014; Tricco et al., 2016). The framework of Arksey and O'Malley (2005) was complemented by the PRISMA-ScR extension, a 22-point checklist that supports transparency, replicability and methodological rigor in ScR (Tricco et al., 2018). Both the PRISMA-ScR extension and the ScR framework of Arksey and O'Malley (2005) indicate that in contrast to a systematic literature review, there is no requirement for quality appraisal of the included studies. The omission of quality appraisal is predicated on the research question of a ScR being much broader than those informing systematic literature reviews (Peters et al., 2021), and the likelihood of a more heterogeneous data sample. As the (sub) question was considered broad, and the data sample was considered heterogeneous, quality appraisal of the included studies was omitted; this is consistent with a recent ScR of assessment practices in K to 12 PE in the United States (Killian & Mays, 2021).

4.2.1 *Eligibility Criteria and Guidelines*

The following terms comprised the inclusion and exclusion criteria for the ScR and were constructed in consultation with the researcher's supervisors:

1. Language – English language
2. Format – Journal article (full text)
3. Context – Applies a tool that assesses physical performance in invasion games and sports (IGS)
4. Population – Includes a school-age cohort (between an age range of 5 – 19 years)
5. Measurement Property – At least one measure of reliability or validity is described from the current study

In addition to the above inclusion criteria, studies were excluded if participants had additional learning or physical needs

To clarify important terms used in the above eligibility criteria, see Table 4.1 below.

Table 4.1

Working Definitions of Key Terms for the Purpose of this Study

Term	Working definition
Tool	An instrument or procedure used to gather assessment data. For example, frequency–count tools that tally behaviours or rating tools that describe performance across a numbered or descriptive scale
Physical performance	The assessment required the participant to physically participate as opposed to providing a written or verbal response after any participation
Invasion Games and Sports (IGS)	IGS are also referred to as territory games (Werner & Almond, 1990) and Team Invasion Games (TIG) (Gray et al., 2009). IGS are categorised by teams playing in a shared space with the aim of invading opposition territory to score (Werner & Almond, 1990). In this thesis, players needed to participate in “playing form” activities rather than standardised “training form” activities (Ford et al., 2010, p. 484). Playing form has been alternatively described as being ecologically valid (Gréhaigne et al., 1997), authentic (Georgakis et al., 2015), or demonstrating principles of representative task design (Inns et al., 2023). As an example of the two forms in soccer, two stationary players kicking a ball to each other 10 m apart would be considered a training form, while a 3 vs 3 SSG that involved a variety of skills, pressure from opposition and the aim of goal scoring would be considered playing form.

4.2.2 Search Strategy and Study Selection

The literature review of assessment in PE in Chapter 2 identified a number of evidence-based assessment tools from which the abstracts and keyword fields were scanned to create a series of pilot search strategies for this ScR. These tools included the two most frequently used in PE and youth

sports as reported by Arias and Castejón (2012), the GPAI and the TSAP. Pilot searches used various combinations of key words with the aim of locating a comprehensive range of primary studies (Arksey & O'Malley, 2005). The search strategy underwent an iterative process of piloting to allow for refinement (Arksey & O'Malley, 2005) before a final version was created and presented in this chapter.

Relevant studies were located through electronic database searches of SPORTDiscus, MEDLINE and ProQuest. SPORTDiscus was selected due to its sport specific discipline focus of sports and use in relevant literature reviews (for example, Ávila-Moreno et al., 2018; González-Víllora et al., 2015; López-Pastor et al., 2013), MEDLINE was selected due to its multidisciplinary nature and use in related reviews (Ávila-Moreno et al., 2018; González-Víllora et al., 2015; Killian & Mays, 2021), and ProQuest was included based on its inclusion of education within a wider multidisciplinary base and its success in the pilot searches. Given the total number of hits, included studies and unique assessment tools, these three databases were used to answer the research sub-question.

The search strategy was applied to each of the above databases in separate searches through the fields of title, keyword and abstract. Each of the key words in each theme was separated by the Boolean operator OR and each of the themes was combined with AND. To remove a large volume of extraneous studies a sixth thematic line was employed, this time with the Boolean operator of NOT (for example, including terms like nursing and clinical). This sixth thematic line excluded a number of studies investigating health related disciplines like medicine, nursing, rehabilitation and physiology. In addition, a limiter of peer-reviewed literature was selected for the SPORTDiscus and ProQuest databases, while in the case of MEDLINE, only peer-reviewed journals are indexed. No publication date restrictions were applied to any search, and all searches were completed on the 24th of April 2020. Applying the search strategy on this date met several design purposes regarding timing. This date was approximately half-way through the researcher's seven-year timeline to complete the thesis (part-time), meaning an equal amount of time was spent on pilot literature searches and completing the remaining studies. Second, this time point was approximately one month after the World Health Organisation (WHO) declared COVID-19 a pandemic. This unprecedented event provided an opportunity to view the included studies of the ScR as belonging to the pre-COVID-19 era, meaning that any update of this ScR in the future would allow researchers the opportunity to compare any

new studies with the data set. Thus, while the literature was scanned after April 2020, no studies that validated new tools or applied previously validated tools were included in this ScR. The choice not to update the search after this date was supported by the volume of studies and the expansive range of tools the search strategy identified in comparison to similar reviews (for example, Arias & Castejón, 2012; Barquero-Ruiz et al., 2019). The abundance of data the search strategy collected allowed the researcher to establish trends and patterns in tool characteristics, assessment criteria and applications to respond to the research sub-question. The final six themes informing the search strategy are presented in Table 4.2.

Table 4.2

Search Strategy for this ScR

Number	Theme	Search terms
1	Game	(game OR sport OR football OR soccer OR basketball)
2	Performance	(perform* OR ability OR skill OR tactic* OR techni*)
3	Assessment	(assess* OR analy* OR evaluat* OR measur* OR examin*)
4	Tool	(tool OR instrument OR procedure OR system OR rubric)
5	Measurement property	(valid* OR reliab*)
6	Exclusions	(clinical OR medical)

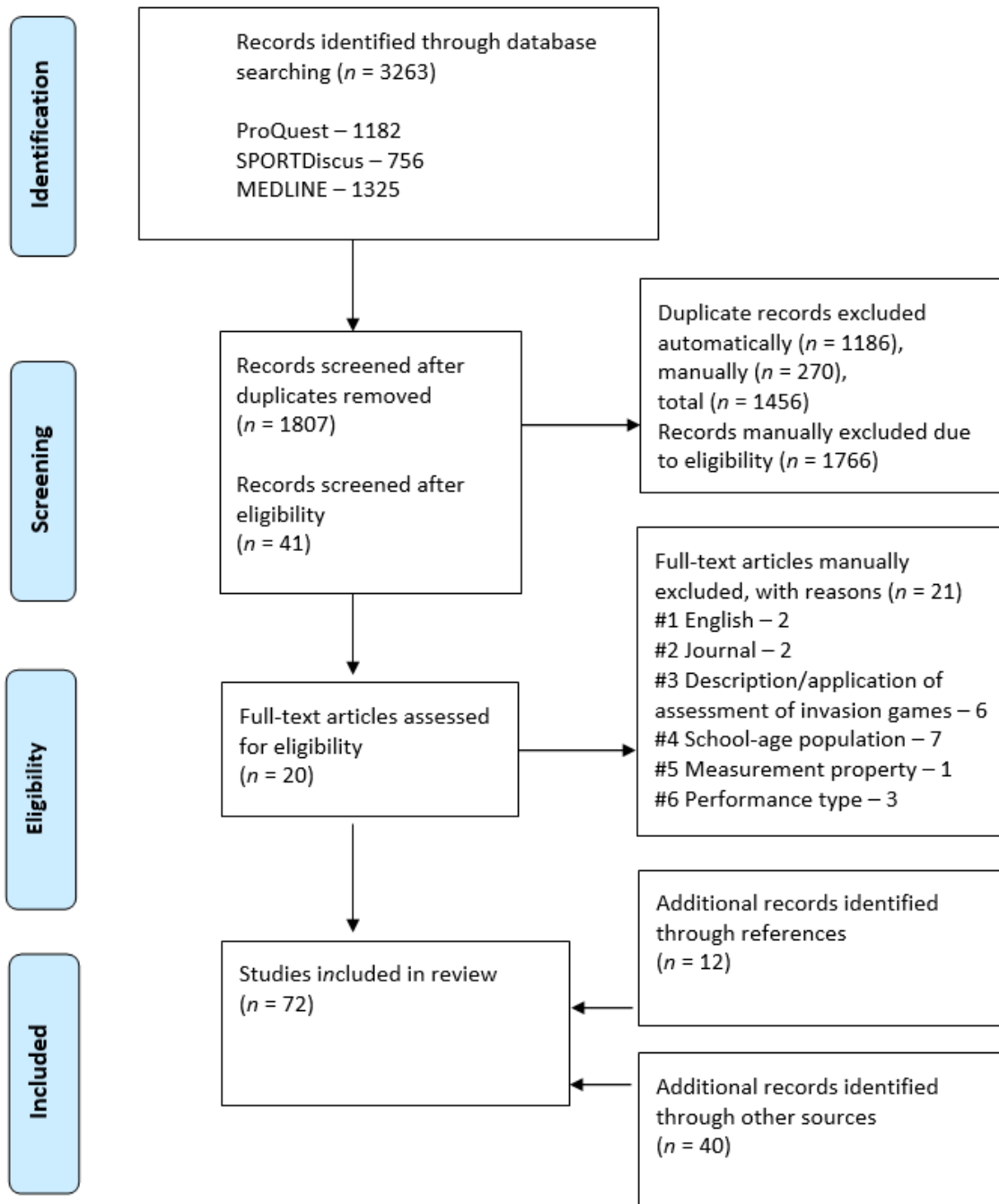
Note. Based on pilot searches and inconsistent reporting of search terms in studies, the three most prolific IGS located in pilot studies (basketball, soccer, football) were added to theme one to ensure the search captured the most comprehensive range of primary studies. Wild cards or truncations were used and are indicated with an asterisk (*).

The researcher independently imported all references to *EndNote* for sorting and management purposes ($n = 3263$). After this import, duplicates were removed automatically and then manually. Given the researcher's immersion in the literature and experience with the pilot searches, he independently screened all titles and abstracts ($n = 1807$) against the inclusion/exclusion criteria. Where the relevance of the study could not be determined from the title/abstract it was included to be screened at full text (Arksey & O'Malley, 2005). The resulting 41 studies were imported into Covidence™, software that supports the management of studies in literature reviews, before being screened for eligibility at full-text by the researcher and a second independent reviewer from the

three-person supervisory team (SP, SE, KR). A summary of the search process is presented in the PRISMA-ScR flowchart in Figure 4.1.

Figure 4.1

Flow Chart for the PRISMA-ScR Extension



Note. The PRISMA-ScR extension protocol flow chart illustrates the identification, screening, eligibility and inclusion stages of this review and represents stage three of the ScR framework (Arksey & O'Malley, 2005).

Provisions were made for a third member of the supervisory team to be involved in reaching consensus where disagreements could not be resolved by two reviewers (Fink et al., 1984), however, this step was not required. This screening for eligibility at full-text resulted in $n = 20$ studies for inclusion. The researcher then screened the reference lists of the included articles to locate additional studies that met the eligibility criteria (Arksey & O'Malley, 2005). This process yielded a further 12 studies, with an additional 40 studies located outside of this process from the researcher's extensive piloting and prolonged engagement with the literature. The large number of studies located beyond the search strategy and implications for replication of this study are addressed in the limitations section of this chapter.

4.2.3 *Data Charting and Collating*

The process of data *charting* in scoping reviews (Arksey & O'Malley, 2005) is synonymous with the process of data *extraction* in systematic literature reviews (Dalglish et al., 2021). Within stage four of the framework by Arksey and O'Malley (2005), the number of reviewers is not stated and this process is under-reported in the field (Pham et al., 2014; Tricco et al., 2016). More recent methodological advice for conducting ScRs advocates for at least two reviewers to chart data to reduce errors and minimise bias, however, there is no directive to report any measure of inter-rater reliability (Peters et al., 2021). Identifying what information would be required from each study was an iterative process that involved the researcher working closely with the supervision team through online meetings and email communication to consider potential data categories against the research sub-question and charted studies. As general bibliographic details were not subject to multiple interpretations, the researcher independently charted items including the authors' names, journal title, location of data collection, study title and the year of publication. Consistent with the above iterative process, data charting of information that responded to the research sub-question involved reviewing earlier studies to ensure consistent and accurate data were collected. An expansive total of 27 items comprised the final data charting form (Arksey & O'Malley, 2005); this form was created using the database program Excel and is presented in summary form in Table 4.3. These 27 items were then *collated* (Arksey & O'Malley, 2005) into three categories that comprised: Assessment tool characteristics; Assessment tool key performance criteria; and Assessment tool applications.

Table 4.3*Summary of Data Items and Definitions in the Scoping Review*

Data item	Definition
<i>Assessment tool characteristics</i>	
1. Tool name	The name of the tool
2. Original context	The original context for the tool's application - community or school - (earliest reference)
3. Developed for use in: (IGS)	The intended IGS the tool was developed for use in (earliest reference)
4. Tool outcome: Rating scale	Any reference to the tool rating performance
5. Tool outcome: Frequency count	Any reference to the tool counting actions/choices
6. Tool outcome: Indices	Any reference to a frequency-count tool generating a quantifiable performance index (ratio) as a summary of performance
7. Reported validity	Any reference to the type(s) of validity data of the tool analysed using the current study sample
8. Reported reliability	Any reference to intra-rater or inter-rater reliability data of the tool analysed using the current study sample
<i>Assessment tool key performance criteria</i>	
9. On-the-ball skills	Any reference to skill executions involving actions to receive, carry, pass, score or defend
10. Receive	Any reference to a reception, catch, trap or control the ball (excluding an intercept/steal that comprises defend below)
11. Dribble	Any reference to dribbling, carrying or running with the ball
12. Pass	Any reference to a pass, kick or throw
13. Score	Any reference to a score/shot
14. Defend	Any reference to defensive actions, for example, tackle, deflect, intercept, steal or save goal
15. Off-the-ball movement	Any reference to off-the-ball movement or positioning
16. Decision-making	Any reference to decision-making or the application of tactics/strategy
17. Other criteria	Any reference to criteria that cannot be classified as on-the-ball, off-the-ball or decision-making
<i>Assessment tool applications</i>	
18. Study aim* (abbreviated)	A brief summary of the study's most relevant aim
19. Sample size	The total number of participants used to generate the assessment data
20. Gender	Gender of participants (where indicated)
21. Age	Age of participants in years (within the age range of 5 – 19y)
22. Applied context	The context the tool was applied in (community or school)
23. Applied IGS	The IGS the tool was applied in
24. Game form	Small-sided game (SSG) or full-sided game (FSG)
25. Observation period	Total assessment time in minutes, plays or matches for player/team
26. Video capture	Any reference to the use of video capture during assessment
27. Peer assessor	Any reference to players completing the assessment

Note. *The abbreviated study aim was independently charted by the researcher.

The first category of Assessment tool characteristics included general features of the tool, such as the tool's name, the original context the tool was developed in, and the types of outcomes that were generated. To avoid having a large number of unnamed tools, tools that were not named within the study were given the title of the study (see Appendix B). This led to some lengthy tool names, for example the Australian Football Small-Sided Game Kicking Proficiency Assessment (Bonney et al., 2020), the Observation Instrument for Technical and Tactical Actions of the Offense Phase in Soccer (Ortega-Toro et al., 2019) and the Opportunities for and Success in Dribbling, Passing, Receiving, and Shooting in Youth Basketball (Arias-Estero, 2013). The two contexts were community, which included any sample drawn from leagues or competitions, and school, which included any sample of students identified by grades or Year Levels. Three assessment tool outcomes were charted based on their use of tallying discrete behaviours (frequency-counts), generating performance efficiency ratios (indices), and making a holistic judgement of performance (rating performance).

The second category of Assessment tool key performance criteria included four data items: (a) on-the ball skills; (b) off-the-ball movement; (c) decision-making; and (d) 'other' criteria. This classification of criteria has several parallels to the TGFU model (see Chapter 1, Figure 1.1). Specifically, step four in the model aligns directly to decision-making in the charting guidelines, and on-the-ball skills and off-the-ball movement may be viewed as sub-components of step six, performance. The use of 'other' criteria captured all other assessable measures in the tools, for example a player calling for the ball (Darnis & Lafont, 2013) and a player's attitude (Rowat et al., 2017). To support a more detailed and independent analysis of on the-ball skills, as the most prevalent key performance criterion, a subset of five data items were charted when they were assessed independent of decision-making. These five on-the-ball skills comprised receive, dribble, pass, score and defend. This approach of separating decision-making from specific on-the-ball skills, or performance, was consistent with the TGFU model (Bunker & Thorpe, 1986). As examples of data charting within this category "Make on-the-spot decisions to apply movement patterns in solving tactical problems" (Penney et al., 2012, p. 405) was charted as decision-making *and* off-the ball movement, while "Decision made: Player chooses to pass to an open teammate" (Oslin et al., 1998, p. 243) was charted as decisions-made and on the ball skills. The latter study also included the charting of specific skill of passing, but that was with reference to the criterion "Skill execution: Passing - Ball reaches target" (Oslin et al., 1998, p. 243).

In the third category, Assessment tool applications involved data items related to each study regarding the applied assessment context, the nature of the sample, and the protocols for the assessment. This category classified studies based on their *applied* context, which in some cases was different from the context they were *originally* designed for. This category also included the games the assessment tools were used in and some basic game conditions, such as the length of observation, any use of video capture, and the team size (represented as small or full sided games). As an example of the nuanced charting guidelines, the final charting guideline for the length of observation was 'The total time (in minutes) that a single player/team is assessed. Where time in minutes cannot be calculated, the figure will be based on the number of plays or matches. The observation period is reported for a single game, game format or reporting period, rather than aggregating any of these data points. In some cases, where the observation period cannot be established for a single player/team, the figure represents the observation period for assessment of the entire population' (see Appendix B).

As one of the data charting approaches commonly employed in ScRs, the researcher independently charted the above data items for each study before seeking verification from a second reviewer (Pham et al., 2014; Tricco et al., 2016). Consistent with the researcher's preference for the verification process to be completed in-person rather than online, a local research assistant completed the verification rather than any of the supervisors that were located interstate. The research assistant held a PhD in the field of PE, had experience in data extraction in a conceptual analysis of Physical Literacy (Young et al., 2020), and was able to meet face-to-face with the researcher for training and verification purposes.

The training was led by the researcher and involved an overview of the study, a detailed description of the two most prevalent assessment tools (the GPAI and TSAP), and an explanation of the charting guidelines and data charting form in an Excel document. The detailed set of charting guidelines (see Appendix B) contribute to the study's audit trail and support transparency and replicability of this study. To support verification of data items, the researcher shared an online folder with the research assistant that included highlighted and annotated copies of the 72 studies. Inter-rater reliability was measured through a percentage agreement figure of 0.98, indicating an excellent level of agreement (Ary et al., 2014). Queries and disagreements were tracked within the data sheet

and a series of face-to-face follow-up meetings allowed for further clarification. During the verification process, charting guidelines for data item six (Tool outcome: Generates indices) and item 25 (Observation period) were refined to help support consensus. This included the researcher creating further charting advice and clarifying definitions through email and in-person meetings. For example, clarifying the data charting of generating indices included additional reference material relating to ratios, proficiency percentages, and percentages of totals to help explain the data item. This included the clarification that a ratio might indicate the number of effective skill executions in relation to the number of ineffective skill executions. The process of clarifying charting guidelines and applying them to data items meant that disagreements were resolved by consensus between the two reviewers, thus making the provision for a third reviewer redundant (Fink et al., 1984). The total time for the research assistant's verification of the data charting, including training and debriefing, was 32 hours; this was funded by a Flinders University College of Education, Psychology and Social Work (EPSW) Research Award to support research costs for Higher Degree Research students.

4.3 Results

The ScR yielded 72 studies from 33 different journals to identify 32 unique, evidence-based assessment tools. A total of eight assessment tools were used in more than one study, and included additional validity or reliability testing of the tool on a separate sample of participants, and were published in different journal articles. A list of all included studies, sorted alphabetically and assigned a reference number (1 - 72), can be found in Appendix C. Further, the checklist for the PRISMA-ScR extension (Tricco et al., 2018) that complemented the five-step framework adopted in this chapter (Arksey & O'Malley, 2005) is provided in Appendix D. An overview of the general bibliographic information regarding the journal distribution across the included studies, location of data collection, and the years of publication by decade is presented in Table 4.4 and Table 4.5.

Table 4.4*Journal Distribution of the Included Studies*

Journal title	<i>n</i>	Reference number (1 to 72)
Asia-Pacific Journal of Health	1	47
Avante-Ontario	1	60
Cuadernos de Psicología del Deporte	1	32
European Journal of Sport Science	1	44
European Physical Education Review	3	18, 38, 39
Frontiers in Psychology Human Movement	2	31, 49
Human Movement	2	15, 56
International Journal of Environmental Research and Public Health	1	58
International Journal of Performance Analysis in Sport	2	7, 71
International Journal of Sports Physiology and Performance	1	68
International Journal of Sport Psychology	1	67
International journal of Sports Science and Coaching	3	2, 55, 70
Journal of Human Kinetics	1	1
Journal of Human Movement Studies	1	11
Journal of Physical Education and Sport	6	9, 24, 33, 51, 66, 46
Journal of Sport Psychology	1	21
Journal of Sports Science and Medicine	2	19, 37
Journal of Sports Sciences	4	5, 20, 40, 41
Journal of Teaching in Physical Education	8	17, 26, 35, 36, 48, 50, 72, 64
Motricidade	1	13
The Journal of Life and Environmental Sciences (PeerJ)	1	23
Perceptual and Motor Skills	1	8
Physical Education and Sport Pedagogy	12	27, 4, 14, 16, 25, 28, 30, 42, 43, 45, 53, 34
Public Library of Science (PLOS ONE)	1	57
Research Quarterly for Exercise and Sport	3	29, 61, 69
Revista Portuguesa de Ciências do Desporto	1	12
Revista Brasileira de Cineantropometria e Desempenho Humano	1	6
Revista Internacional de Ciencias del Deporte	1	54
Science and Medicine in Football	2	3, 10
South African Journal for Research in Sport, Physical Education and Recreation	2	52, 65
SPORT TK-EuroAmerican Journal of Sport Sciences	1	22
The Journal of Sports Medicine and Physical Fitness	1	62
The Physical Educator	1	59
TRENDS in Sport Sciences	1	63
Total Journal <i>n</i>	33	

*Note**: See Appendix C for full bibliographic details of the 72 included studies in the ScR.

Table 4.5*Bibliographic Characteristics of the Included Studies*

Location of data collection and year of publication	<i>n</i>	Reference number
Location of data collection		
Asia	3	
Malaysia		46, 47
Singapore		62
Australia	7	3, 5, 39, 40, 41, 53, 68
Brazil	4	6, 12, 13, 56
Canada	4	44, 45, 60, 61
Europe	36	
Belgium	1	67
France	2	14, 26
Germany	1	37
Greece	1	8
Italy	1	66
Netherlands	1	70
Poland	1	63
Portugal	8	7, 9, 16, 17, 18, 19, 24, 38
Spain	20	1, 2, 15, 22, 23, 27, 31, 32, 35, 42, 43, 49, 51, 52, 55, 57, 58, 64, 65, 71
United States	11	21, 28, 29, 20, 33, 34, 48, 50, 59, 69, 72
United Kingdom	4	10, 20, 25, 36
Location not stated	3	4, 11, 54
Years of publication by decade		
1980 – 1990	1	21
1991 – 2000	6	26, 29, 50, 60, 61, 69
2001 – 2010	13	4, 8, 11, 12, 28, 34, 36, 37, 44, 45, 48, 67, 72
2011 – 2020	52	1, 2, 3, 5, 6, 7, 9, 10, 13, 14, 15, 16, 17, 18, 19, 20, 22, 23, 24, 25, 27, 30, 31, 32, 33, 35, 38, 39, 40, 41, 42, 43, 46, 47, 49, 51, 52, 53, 54, 55, 56, 57, 58, 59, 62, 63, 64, 65, 66, 68, 70, 71

Note: See Appendix C for full bibliographic details of the 72 included studies in the ScR.

In Table 4.6 aggregated data are presented for the first category of Assessment Tool Characteristics. In this table each of the 32 tools are organised alphabetically by their *original* (i.e., first published) assessment context of either community or school. The data are presented descriptively using a combination of counts and percentages.

Table 4.6

Assessment Tool Characteristics

Assessment Tool name	Total <i>n</i> studies	Original context	Reference No	Developed for use in:	Tool outcome: Rating scale	Tool outcome: Frequency-count	Tool outcome: Indices	Reported validity (Reference No)	Reported reliability (Reference No)
Australian Football Small-Sided Game Kicking Proficiency Assessment	1	Community	5	Australian football	X	X	X	5	5
Coach Rating Tool	1	Community	68	Australian football	X	X	X	NR	68
Football Observation System (SOF)	1	Community	1	Soccer	-	X	-	NR	1
Game Technical Scoring Chart (GTSC)	2	Community	20	Soccer	X	-	-	20, 62	20
Game Test Situation 1: Taking Advantage of Openings	1	Community	37	Soccer	X	-	-	37	37
Game Test Situation 2: Offering and Orienting	1	Community	37	Soccer	X	-	-	37	37
Game-related Soccer Skill Assessment	1	Community	63	Soccer	X	-	-	63	63
Observation Instrument for Technical and Tactical Actions of the Offense Phase in Soccer	1	Community	49	Soccer	-	X	-	49	49
Observational Instrument	2	Community	21	Basketball	-	X	X	NR	21, 23
Opportunities for and Success in Dribbling, Passing, Receiving, and Shooting in Youth Basketball	1	Community	2	Basketball	-	X	-	2	2
Polar Coordinate Analysis	1	Community	32	Handball	-	X	-	32	32
Procedural Tactical Knowledge Test for Basketball (PTKT: Bb)	1	Community	54	Basketball	-	X	X	54	54
Rugby Attack Assessment Instrument (RAAI)	1	Community	35	Rugby union	-	X	X	35	35
Soccer Specific Behaviour Measurement Tool (S-SBMT)	1	Community	10	Soccer	-	X	-	10	10
System for Notational Soccer Analysis	1	Community	70	Soccer	-	X	X	70	70

Assessment Tool name	Total <i>n</i> studies	Original context	Reference No	Developed for use in:	Tool outcome: Rating scale	Tool outcome: Frequency-count	Tool outcome: Indices	Reported validity (Reference No)	Reported reliability (Reference No)
System of Tactical Assessment in Soccer (FUT-SAT)	5	Community	12	Soccer	-	X	X	13	6, 7, 12, 13, 56
Talent Identification Tool	1	Community	3	Soccer	-	X	X	NR	3
Basketball Learning and Performance Assessment Instrument (BALPAI)	1	School	31	Basketball	-	X	X	31	31
Coding Instrument	5	School	4	Soccer	-	X	X	25	4, 25, 38, 39, 41
Collective Game Efficacy and Individual Skill Level Tool	1	School	14	Basketball	X	X	X	NR	14
Cutting or Off-the-ball-Actions Coding Instrument	1	School	48	Invasion games	-	X	X	NR	48
Game Performance Assessment Instrument (GPAI)	13	School	50	All game categories	-	X	X	50	8, 15, 16, 18, 19, 24, 28, 36, 40, 42, 50, 59, 71
Game performance Coding Instrument	1	School	67	Basketball	-	X	X	67	67
Game Performance Evaluation Tool (GPET)	8	School	43	Modified invasion game	-	X	X	22, 27, 57, 58	22, 27, 43, 55, 57, 58, 64, 65
Game Performance Observation Instrument	1	School	17	Soccer	-	X	X	17	17
Game Play Observational Instrument	4	School	69	Field hockey	X	X	X	NR	11, 46, 47, 69
Observational Scoring Rubric	1	School	72	Movement forms	X	-	-	NR	72
Passing Decision Coding Instrument	1	School	48	Invasion games	-	X	X	NR	48
Standards Based Rubric	1	School	53	All game categories	X	-	-	53	53
Supporting Movement Tool	1	School	34	Tag rugby	-	X	X	NR	34
Team Sport Assessment Procedure (TSAP)	11	School	26	Team sports	-	X	X	26, 44, 45, 60	9, 26, 29, 30, 33, 44, 45, 52, 61, 66
Tool for Assessment and Learning of an Invasion Situation (TALIS)	1	School	51	Invasion games	-	X	-	51	51
Tools (<i>n</i> = 32)					10	26	20	22	32
Total as percentage					31.3	81.3	62.5	68.8	100.0

Note. The use of a dash (-) indicates the absence of the relevant item and NR indicates that no measurement property was reported.

In Table 4.7 aggregated data are presented for the second category of Assessment Tool Key Performance Criteria. In this table each of the 32 tools are organised alphabetically by their *original* (i.e., first published) assessment context of either community or school. The data are presented descriptively using a combination of counts and percentages.

Table 4.7

Assessment Tool Key Performance Criteria

Assessment Tool name	Developed for use in:	On-the ball skills	Receive	Dribble	Pass	Score	Defend	Off-the-ball movement	Decision-making	Other criteria
Australian Football Small-Sided Game Kicking Proficiency Assessment	Australian football	X	-	-	X	-	-	-	-	-
Coach Rating Tool	Australian football	X	X	-	X	-	-	-	-	-
Football Observation System (SOF)	Soccer	X	X	X	X	X	X	-	-	X
Game Technical Scoring Chart (GTSC)	Soccer	X	X	X	X	X	-	X	X	X
Game Test Situation 1: Taking Advantage of Openings	Soccer	-	-	-	-	-	-	-	-	X
Game Test Situation 2: Offering and Orienting	Soccer	-	-	-	-	-	-	-	-	X
Game-related Soccer Skill Assessment	Soccer	X	X	-	-	X	X	X	-	X
Observation Instrument for Technical and Tactical Actions of the Offense Phase in Soccer	Soccer	X	X	X	X	X	X	X	-	X
Observational Instrument	Basketball	X	X	X	X	X	-	-	X	-
Opportunities for and Success in Dribbling, Passing, Receiving, and Shooting in Youth Basketball	Basketball	X	X	X	X	X	-	-	-	-
Polar Coordinate Analysis	Handball	X	X	X	X	X	X	X	-	X
Procedural Tactical Knowledge Test for Basketball (PTKT: Bb)	Basketball	X	X	X	X	X	X	X	-	-
Rugby Attack Assessment Instrument (RAAI)	Rugby union	X	X	X	X	X	-	-	-	X
Soccer Specific Behaviour Measurement Tool (S-SBMT)	Soccer	X	X	X	X	X	X	-	-	-
System for Notational Soccer Analysis	Soccer	X	X	X	X	X	X	X	-	-
System of Tactical Assessment in Soccer (FUT-SAT)	Soccer	X	X	X	X	X	X	X	-	X

Assessment Tool name	Developed for use in:	On-the ball skills	Receive	Dribble	Pass	Score	Defend	Off-the-ball movement	Decision-making	Other criteria
Talent Identification Tool	Soccer	X	X	X	X	X	-	-	-	-
Basketball Learning and Performance Assessment Instrument (BALPAI)	Basketball	X	X	X	X	X	X	X	X	-
Coding Instrument	Soccer	X	X	X	X	X	X	X	X	-
Collective Game Efficacy and Individual Skill Level Tool	Basketball	X	-	X	X	X	-	X	-	X
Cutting or Off-the-ball-Actions Coding Instrument	Invasion games	X	X	-	-	-	-	X	-	-
Game Performance Assessment Instrument (GPAI)	All game categories	X	X	X	X	X	X	X	X	-
Game performance Coding Instrument	Basketball	X	X	X	X	X	-	X	X	-
Game Performance Evaluation Tool (GPET)	Modified invasion game	X	X	X	X	X	-	X	X	-
Game Performance Observation Instrument	Soccer	X	X	X	X	X	X	X	X	-
Game Play Observational Instrument	Field hockey	X	X	X	X	X	X	-	X	-
Observational Scoring Rubric	Movement forms	X	X	X	X	X	X	X	X	-
Passing Decision Coding Instrument	Invasion games	X	-	-	X	-	-	-	X	-
Standards Based Rubric	All game categories	X	-	-	-	-	-	X	X	X
Supporting Movement Tool	Tag rugby	-	-	-	-	-	-	X	-	-
Team Sport Assessment Procedure (TSAP)	Team sports	X	X	-	X	X	X	-	X	-
Tool for Assessment and Learning of an Invasion Situation (TALIS)	Invasion games	X	X	-	-	-	-	X	-	-
Tools (<i>n</i> = 32)		29	25	21	25	23	15	19	13	11
Total as percentage		90.6	78.1	65.6	78.1	71.9	46.9	59.4	40.6	34.4

Note. The use of a dash (-) indicates the absence of the relevant item.

The data items in Table 4.8, Assessment Tool Applications, summarise how each tool was applied in each of the 72 studies with reference to the individual aim of each study. The tools are again listed by their *original* (i.e., first published) assessment context, although the applied context is reported for each study.

Table 4.8

Assessment Tool Applications

Assessment Tool name	Reference No	Study aim (abbreviated)	Sample Size	Gender	Age (years)	Applied context	Applied game/sport	Game form	Observation period	Video capture	Peer assessor
Australian Football Small-Sided Game Kicking Proficiency Assessment	5	Develop a valid and reliable Australian football kicking proficiency assessment	145	M	U/13 to U/18	Community	Australian football	SSG	12 min	X	-
Coach Rating Tool	68	Investigate the relationship between fitness and skill in Australian football	156	M	10 to 15	Community	Australian football	FSG	Match	X	-
Football Observation System (SOF)	1	Investigate game structure on learning needs in soccer	NS	NS	8 to 10	Community	Soccer	SSG	30 to 50 min	X	-
Game Technical Scoring Chart (GTSC)	20	Evaluate technical attributes of soccer	16	NS	11	Community	Soccer	SSG	30 min	-	-
Game Technical Scoring Chart (GTSC)	62	Evaluate if physical characteristics are a better predictor of performance than technical skills in soccer	25	M	17	Community	Soccer	SSG	20 min	-	-
Game Test Situation 1: Taking Advantage of Openings	37	Evaluate tactical-oriented game test situations in soccer	195	NS	12 to 13	Community	Soccer	SSG	6 min	X	-
Game Test Situation 2: Offering and Orienting	37	Evaluate tactical-oriented game test situations in soccer	195	NS	12 to 13	Community	Soccer	SSG	6 min	X	-
Game-related Soccer Skill Assessment	63	Examine the relationship between skill tests and game skills in soccer	60	M	12 to 15	Community	Soccer	SSG	15 min	X	-
Observation Instrument for Technical and Tactical Actions of the Offense Phase in Soccer	49	Design, validate and test the reliability of an instrument to analyse technical/ tactical actions in offensive play in soccer	44	NS	U/12	Community	Soccer	FSG	3 Matches	X	-
Observational Instrument	21	Examine the relationship of knowledge to the development of skill in basketball	56	M	8 to 12	Community	Basketball	FSG	Half a match	X	-
Observation(al) Instrument	23	Analyse the impact of a decision training program on the development of decision-making and skill execution in basketball	11	M	13	Community	Basketball	FSG	5 to 11 Matches	X	-
Opportunities for and Success in Dribbling, Passing, Receiving, and Shooting in Youth Basketball	2	Analyse the relationship between the number of opportunities for, and success in, various on-the-ball skills in basketball	59	NS	11	Community	Basketball	FSG	10 Games	X	-
Polar Coordinate Analysis	32	Analyse counter-attack actions in handball	NS	M	U/16	Community	Handball	FSG	10 Games	X	-

Assessment Tool name	Reference No	Study aim (abbreviated)	Sample Size	Gender	Age (years)	Applied context	Applied game/sport	Game form	Observation period	Video capture	Peer assessor
Procedural Tactical Knowledge Test for Basketball (PTKT: Bb)	54	Validate an instrument for assessing procedural tactical knowledge in basketball	161	M	12 to 19	Community	Basketball	SSG/ FSG	4 min	X	-
Rugby Attack Assessment Instrument (RAAI)	35	Introduce an assessment for a player that is on-the-ball in rugby	NS	NS	U/12 to U/19	Community	Rugby	SSG	10 min	X	-
Soccer Specific Behaviour Measurement Tool (S-SBMT)	10	Assess the validity, objectivity, and reliability of an assessment tool for soccer	16	NS	11	Community	Soccer	SSG	30 min	X	-
System for Notational Soccer Analysis	70	Compose an objective and detailed notational analysis system for soccer	19	F	16	Community	Soccer	SSG	8 x 15 attack/10 defend trials	X	-
System of Tactical Assessment in Soccer (FUT-SAT)	12	Analyse tactical behaviours in soccer according to ten core tactical principles	300	NS	U/11 to U/20	Community	Soccer	SSG	4 min	X	-
System of Tactical Assessment in Soccer (FUT-SAT)	13	Report the development and preliminary validation of a tactical assessment tool for soccer	440	NS	U/11 to U/17	Community	Soccer	SSG	4 min	X	-
System of Tactical Assessment in Soccer (FUT-SAT)	7	Compare tactical behaviour in different SSGs in soccer	10	NS	U/11	Community	Soccer	SSG	8 min	X	-
System of Tactical Assessment in Soccer (FUT-SAT)	6	Investigate the tactical demands of SSGs in soccer	18	M	16	Community	Soccer	SSG	48 min	X	-
System of Tactical Assessment in Soccer (FUT-SAT)	56	Compare the development of tactical skills in different age groups in soccer	30	NS	14 to 15	Community	Soccer	SSG	36 min	X	-
Talent Identification Tool	3	Investigate the use of SSGs as a Talent Identification Tool in soccer	73	M	13	Community	Soccer	SSG	15 min	X	-
Basketball Learning and Performance Assessment Instrument (BALPAI)	31	Design and validate as assessment instrument for decision-making, technical execution and efficacy in basketball	6	M	11	School	Basketball	SSG	15 possessions	X	-
Coding Instrument	4	Assess decision-making and gameplay ability in soccer	12	M	14 to 15	School	Soccer	SSG	10 min	X	-
Coding Instrument	25	Investigate the effects of a tactical teaching approach on game playing performance in basketball	16	M/ F	13	School	Basketball	SSG	10 min	X	-
Coding Instrument	38	Examine the effect of a hybrid teaching model on decision-making, skill execution and overall game performance in soccer	26	M/ F	10 to 12	School	Soccer	SSG	NS	X	-
Coding Instrument	39	Evaluate the efficacy of a game-centered approach on netball	32	M/ F	9 to 12	School	Netball	SSG	4 min	X	-
Coding Instrument	41	Evaluate how playing form activity can affect game play outcomes in netball	41	F	9	Community	Netball	SSG	6 min	X	-
Collective Game Efficacy and Individual Skill Level Tool	14	Explore the role of co-operative learning in basketball	30	M/ F	9	School	Basketball	SSG	14 to 16 min	X	-
Cutting or Off-the-ball-Actions Coding Instrument	48	Examine the effect of instruction on off-the-ball actions in an invasion game	24	M/ F	Grade 4	School	Invasion game	SSG	7 - 8 min	X	-
Game Performance Assessment Instrument (GPAI)	50	Report on the development and validation of an instrument to measure tactical understanding and problem solving across three game categories	99	NS	Grade 6	School	Basketball, soccer	SSG	5 min	X	-

Assessment Tool name	Reference No	Study aim (abbreviated)	Sample Size	Gender	Age (years)	Applied context	Applied game/sport	Game form	Observation period	Video capture	Peer assessor
Game Performance Assessment Instrument (GPAI)	8	Investigate the effects of a technique and games approach on performance on soccer	72	F	12 to 13	School	Soccer	NS	NS	X	-
Game Performance Assessment Instrument (GPAI)	36	Examine how the tactical games model affects learning in netball	6	M/ F	9 - 10	School	Netball	SSG	14 min	X	-
Game Performance Assessment Instrument (GPAI)	28	Test the defensive off-the-ball aspects of the GPAI on soccer	34	M	14 to 18	School	Soccer	SSG	20 plays	X	-
Game Performance Assessment Instrument (GPAI)	59	Investigate the effectiveness of the SE tactical model during game play and performance in basketball	90	M/ F	Grade 6 to 7	School	Basketball	SSG	15 min	X	-
Game Performance Assessment Instrument (GPAI)	71	Compare the use of different teaching units on tactical learning in basketball and soccer	104	M/ F	11	School	Basketball, soccer	SSG	3 min	X	-
Game Performance Assessment Instrument (GPAI)	19	Examine how SE effects performance and involvement in basketball, handball and soccer	10	M/ F	12	School	Basketball, handball, soccer	SSG	10 min	X	-
Game Performance Assessment Instrument (GPAI)	42	Examine any improvement in performance after a TGfU unit on floorball	41	M/ F	12	School	Floorball	FSG	10 min	X	-
Game Performance Assessment Instrument (GPAI)	15	Assess tactical/ technical knowledge in futsal	57	M/ F	9 to 12	School	Futsal	SSG	5 min	X	-
Game Performance Assessment Instrument (GPAI)	16	Examine the development of game-play performance and transfer across basketball, handball and soccer	10	M/ F	12	School	Basketball, handball, soccer	SSG	10 min	X	-
Game Performance Assessment Instrument (GPAI)	18	Examine game performance according to tactical structures in basketball, handball and soccer	10	M/ F	12	School	Basketball, handball, soccer	SSG	5 min	X	-
Game Performance Assessment Instrument (GPAI)	24	Assess the effectiveness of the tactical games' approach/ technique-oriented approach on game performance in basketball and soccer	62	M/ F	13	School	Basketball, soccer	SSG/ FSG	10 min	X	-
Game Performance Assessment Instrument (GPAI)	40	Investigate how object control relates to game play competence in netball	107	M/ F	11	School	Netball	SSG	5 min	X	-
Game performance Coding Instrument	67	Examine the effectiveness of the Invasion Games Competence Model and a Traditional Approach on game performance in basketball	26	NS	10 to 11	School	Basketball	SSG	26 min	X	-
Game Performance Evaluation Tool (GPET)	43	Gauge student performance in a modified invasion game	27	M/ F	9 - 10	School	Modified invasion game	SSG	NS	X	-
Game Performance Evaluation Tool (GPET)	27	Evaluate possible gender differences in offensive tactical behaviour in a modified invasion game	74	M/ F	7 to 14	School	Modified invasion game	SSG	8 min	X	-
Game Performance Evaluation Tool (GPET)	22	Present the results of the validation and reliability processes for a tool that measures decision-making and skill execution in invasion GS	129	M/ F	7 to 14	Community/ School	Soccer	SSG	8 min	X	-

Assessment Tool name	Reference No	Study aim (abbreviated)	Sample Size	Gender	Age (years)	Applied context	Applied game/sport	Game form	Observation period	Video capture	Peer assessor
Game Performance Evaluation Tool (GPET)	65	Examine how game performance is impacted by representation and exaggeration in soccer	21	NS	8 to 9	Community	Soccer	SSG	8 min	X	-
Game Performance Evaluation Tool (GPET)	64	Analyse the impact of a TGfU approach on various SSGs in soccer	21	NS	9	Community	Soccer	SSG	8 min	X	-
Game Performance Evaluation Tool (GPET)	57	Analyse the effect of two teaching programs on decision-making and skill execution in soccer	19	NS	11	Community	Soccer	FSG	144 min to 336 min	X	-
Game Performance Evaluation Tool (GPET)	55	Analyse the effect of nonlinear pedagogy on decision-making and skill execution in futsal	8	M	15	Community	Futsal	FSG	120 min to 240 min	X	-
Game Performance Evaluation Tool (GPET)	58	Reveal the observation criteria for talent identification of gifted players in soccer	18	M	12	School	Soccer	SSG/ FSG	60 min	X	-
Game Performance Observation Instrument	17	Consider the impact of a hybrid teaching model on game performance in soccer	24	M/ F	10	School	Soccer	SSG	10 min	X	-
Game Play Observational Instrument	69	Test the validity of the TGfU model in comparison to a technique approach in field hockey	71	M/ F	Grade 6 to 7	School	Field hockey	SSG	30 min	X	-
Game Play Observational Instrument	11	Examine the transfer from a generic teaching of invasion games to floorball	36	M/ F	10 to 11	School	Floorball	SSG	12 min	X	-
Game Play Observational Instrument	47	Report on the development and testing of a hybrid model of teaching games in field hockey	108	M	13	School	Field hockey	SSG	NS	X	-
Game Play Observational Instrument	46	Investigate the effects of the TGfU model on ball control, decision-making and skill execution in field hockey	30	M	14 to 15	School	Field hockey	SSG	NS	X	-
Observational Scoring Rubric	72	Describe factors related to teachers' ability to use scoring rubrics accurately in basketball, flag football, soccer and ultimate frisbee	NS	NS	High school	School	Basketball, flag football, soccer, ultimate frisbee	SSG	10 min	X	-
Passing Decision Coding Instrument	48	Examine the effect of instruction on tactical passing decisions in an invasion game	24	M/ F	Grade 4	School	Invasion game	SSG	7 - 8 min	X	-
Standards Based Rubric	53	Investigate authentic digital forms of assessment in rugby union, soccer	16	NS	Year 11	School	Rugby union, soccer	SSG	NS	X	-
Supporting Movement Tool	34	Examine the effects of technique-focused and tactic-focused instruction on tactical learning in tag rugby	12	M/ F	12 to 14	School	Tag rugby	SSG	8 min	X	-
Team Sport Assessment Procedure (TSAP)	26	Present a pre-assessment and formative assessment procedure to assess individuals in team sports	27	NS	High school	School	Soccer	SSG	28 min	-	x
Team Sport Assessment Procedure (TSAP)	29	Examine the development of skill competence and tactical play during a SE unit in ultimate frisbee	6	M/ F	Grade 6	School	Ultimate frisbee	SSG	6 Games	X	-
Team Sport Assessment Procedure (TSAP)	60	Establish performance norms for the TSAP in basketball	561	M/ F	10 to 14	School	Basketball	SSG	10 to 14 min	-	X
Team Sport Assessment Procedure (TSAP)	61	Verify inter-observer reliability between performers and experts in basketball	82	M/ F	10 to 14	School	Basketball	SSG	24 to 30 min	X	X

Assessment Tool name	Reference No	Study aim (abbreviated)	Sample Size	Gender	Age (years)	Applied context	Applied game/sport	Game form	Observation period	Video capture	Peer assessor
Team Sport Assessment Procedure (TSAP)	44	Adapt the TSAP to ice hockey	103	NS	11 to 12	Community	Ice hockey	FSG	60 min	X	-
Team Sport Assessment Procedure (TSAP)	45	Establish the validity and reliability of a procedure of the TSAP in ice hockey	19	NS	14 to 17	Community	Ice hockey	SSG	6 min	-	X
Team Sport Assessment Procedure (TSAP)	33	Determine if SE improves game play in swirl ball	12	M/ F	10	School	Swirl ball	SSG	Match	X	-
Team Sport Assessment Procedure (TSAP)	9	Verify the technical accuracy of variables in the TSAP in basketball	42	M	14 to 18	Community	Basketball	FSG	3 Matches	X	-
Team Sport Assessment Procedure (TSAP)	52	Apply the TSAP as formative assessment in basketball	52	M/ F	11	School	Basketball	SSG	18 to 24 min	X	X
Team Sport Assessment Procedure (TSAP)	30	Test the effect of graded competition on game involvement and success rates in handball	106	M/ F	10 to 11	School	Handball	SSG	76 Matches	X	-
Team Sport Assessment Procedure (TSAP)	66	Assess the impact of gender and sport practice on technical and tactical skills in basketball	32	M/ F	13	School	Basketball	SSG	3 Matches	X	-
Tool for Assessment and Learning of an Invasion Situation (TALIS)	51	Design and validate a tool for peer evaluation in a passing game	22	M/ F	8	School	Passing game	NS	NS	X	X

Note. The reference numbers 1 to 72 were allocated to each of the included studies as found in Appendix C. NS – not stated, M – male, F – female, SSG - small-sided games, FSG- full-sided games. In the case of the use of video capture and peer assessors, the use of a dash (-) indicates the absence of the relevant item. No study reported any other category for gender other than male (boys) and female (girls).

4.4 Discussion

The following sections 4.4.1 through to 4.4.3 comprise the three discussion categories: Assessment Tool Characteristics; Assessment Tool Key Performance Criteria; and Assessment Tool Applications. These categories correspond directly to Tables 4.6, 4.7 and 4.8 in the results section of this chapter. To support readers in navigating the data and analysis, each category has been broken down into three further sub-categories.

4.4.1 *Assessment Tool Characteristics*

The data items charted in Table 4.6 indicated that of the 32 tools identified in the ScR, 24 were applied in one study within the corpus. The other eight tools appeared in multiple studies, representing over two-thirds of the total studies. The eight tools in order of prevalence comprised the: GPAI; TSAP; Game Performance Evaluation Tool (GPET); Coding Instrument; FUT-SAT; Game Play Observational Instrument; Game Technical Scoring Chart (GTSC); and the Observational Instrument. This ScR supported findings from a previous review of empirical studies that identified the two most frequently used tactical assessment instruments used in PE and youth sports were the GPAI and the TSAP (Arias & Castejón, 2012). Further support for the prevalence of these two tools is provided in the review of tactical learning game assessment tools by Barquero-Ruiz et al. (2019). This ScR is positioned as a nuanced update of earlier reviews, as this study focuses exclusively on assessment in IGS.

4.4.1.1 *Validity and reliability*

To ensure that the assessment tools were evidence-based, the inclusion criteria for this ScR required that at least one measure of validity or reliability of the assessment tool was reported for the population that was assessed. Table 4.6 indicated that almost one-third of tools did not report any type of validity measure. This finding supports previous literature reviews that have been critical of the lack of thorough validation processes for many GS assessment tools (Arias & Castejón, 2012; Barquero-Ruiz et al., 2019). Further, due to the myriad of adaptations across studies using the same tools, many tools were used in assessment contexts they were not validated in. For example, many studies that used the same tool often manipulated several of the following conditions, including the game, sample size and characteristics, team size, field size, game rules, period of observation, and key performance criteria; however, measures of validity were not always reported.

As an example of the manipulation of assessment conditions without reporting any validation, other than the original GPAI study (Oslin et al., 1998) no subsequent GPAI studies reported any measure of validity. Of the tools that did report validity, a total of eight different measures were identified (concurrent, construct, content, ecological, external, face, internal and logical validity). Several tools reported more than one type of validity, with the GPAI development paper (Oslin et al., 1998) reporting the most measures of validity (content, construct, ecological and face validity). Of note, no studies reported criterion validity, which would involve comparing the assessment tool scores or results to a 'gold standard' assessment (Terwee et al., 2007), as no gold standard was located in the literature.

In addressing the related inclusion criteria of reliability, all 32 assessment tools reported at least one measure of assessor reliability. Within the actual assessment criteria used in the tools, terms like appropriate, effective or efficient were used widely (for example, Farias, Mesquita, et al., 2019; Gray & Sproule, 2011; Oslin et al., 1998), however, these terms were often presented without any 'anchor' to support consistent judgement which may negatively impact inter-rater reliability (Memmert & Harvey, 2008). In considering the reporting of validity and reliability together, some measures for reliability may be of limited utility, given that almost one-third of all tools did not report any validation process. For example, consistent assessment results within a single rater or across multiple raters may not support student learning if the assessment criteria do not accurately represent the concept of game performance or content validity (Terwee et al., 2007).

4.4.1.2 School and community contexts

As the tools were developed almost equally in community and school settings; this distribution is reflected in the disciplinary nature of the journals the studies were published in. The researcher had hypothesised that more tools would be designed for the community than school environments, given the body of performance analysis literature linked to competitive or elite teams (for example, Hughes et al., 2012; Hughes & Franks, 2004; Nevill et al., 2008). In considering assessment tool characteristics across the two contexts, there were several points of convergence or similarity. With reference to performance outcomes, six community-developed tools rated performance, which is a similar number to the four tools developed in the school setting. The researcher had hypothesised that more tools would rate performance in the school setting given the widespread use of rubrics in education (for example, Brookhart, 2013; Jonsson & Svingby, 2007; Panadero & Jonsson, 2013). In both assessment contexts most tools made use of frequency-

counts, and the generation of indices was common. Given that indices have been criticised for obscuring performance and missing opportunities to enhance learning (Barquero-Ruiz et al., 2019; Memmert & Harvey, 2008), it is noteworthy that indices were used in all but one of the school-developed tools.

The lack of curriculum alignment in school-developed tools presents a potential barrier for their use by PE teachers. Namely, teachers may be required to assess and report student achievement on standards articulated in a curriculum to meet regulatory guidelines (for example, DET, 2023c). To promote curriculum alignment, the located tools may need to be critiqued against relevant achievement standards, or similar reference points in a curriculum, to ensure that the tool assesses what is to be taught. Beyond the broader challenge of aligning curriculum and assessment, frequency-count tools like the GPAI and TSAP may be better suited to research than school-based assessment due to their relative complexity (Arias-Estero & Castejón, 2014; López-Pastor et al., 2013). This complexity includes tallying multiple discrete events for individual students and the subsequent calculation of performance scores or indices (MacPhail et al., 2008; Williams et al., 2020).

4.4.1.3 Outcomes generated

The reporting of *how* criteria were measured in the studies was not always clear, however the charting of tools was predicated on two distinctly different approaches. In one approach, tools that used notational analysis of multiple, *discrete events* were categorised as using frequency-counts. These tools adopted a quantitative approach to the assessment of key performance criteria by counting things. As the most prevalent frequency-count assessment tool in the corpus, the GPAI validation study required passing and shooting actions to be tallied with reference to the appropriateness of the decisions made and the efficiency of the skill execution (Oslin et al., 1998). Appendix I provides the behaviours and actions to be tallied and the coding sheet that aggregated data under the various categories.

In a second approach, tools that assessed overall performance *holistically* were categorised as rating performance. These tools may be viewed as adopting a more qualitative approach to the assessment of key performance criteria, by rating or ranking total performance across levels of quality. More than three-quarters of all tools used frequency-counts to tally the number of behaviours and/or actions. Identifying a large number of frequency-based tools that collected quantitative data was not unexpected, given that the search strategy required each study to

report a measure of validity or reliability as hallmarks of quality assessment (Brookhart, 2005). As validity and reliability are standard markers of quality in quantitative research that involves numerical data, qualitative research that involves text-based, pattern-based or descriptive data may employ other markers of quality (Smith & McGannon, 2018; Tracy, 2010).

Almost two-thirds of frequency-count tools generated performance indices that summarised performance data after tallying the frequency of behaviours/actions. As an example of the indices generated, the above study calculated three different indices based on dividing the number of appropriate/efficient actions by the number of inappropriate/inefficient actions (Oslin et al., 1998) (Appendix I). Also known as “quantifiable indexes” (Arias & Castejón, 2012, p. 377), performance indices manipulate raw numerical results into a percentage or number that summarises the assessment criterion. Concerns regarding the use of indices include the soundness of mathematical formulas used in the GPAI (Memmert & Harvey, 2008), and the risk of indices skewing data and missing opportunities to support learning (Barquero-Ruiz et al., 2019; Harvey et al., 2010; Memmert & Harvey, 2008).

In regard to the soundness of the mathematical formulas in the GPAI, some users have refined formulas to include students that do not complete a single appropriate or inappropriate action (for example, Farias et al., 2018; Gouveia et al., 2019). Other users of the GPAI in the ScR applied the original formulae (for example, Chatzopoulos et al., 2006; Mesquita et al., 2012), while others did not report the formulae applied (for example, Evangelio et al., 2019). In one case, researchers deliberately avoided the generation of indices in the GPAI by citing issues with their utility (Harvey et al., 2010). The original formulas for the GPAI indices can be found in Appendix I. In regard to indices potentially skewing data, in combining skill execution for reception, passing and shooting into a single skill execution efficiency index, any differences between the three types of skill execution may be missed (Oslin et al., 1998).

In contrast to the high use of frequency-count approaches, less than one-third of all tools rated performance. As an example of an assessment tool based on a rating scale, the Game Technical Scoring Chart (GTSC) listed 10 performance criteria for soccer including “First touch, Awareness and overall control, Control from the air, Short passing (under 10m), Long passing (over 10m), Dribbling, Turning, Shooting accuracy, Two footedness, Attitude” (Rowat et al., 2017, p. 374). Two studies applied rubrics to rate overall performance (Penney et al., 2012; Williams &

Rink, 2003), which in contrast to a rating scale like the GTSC, included definitions or descriptors of performance that may provide more guidance for assessors and learners.

4.4.2 *Assessment Tool Key Performance Criteria*

In the field of performance analysis, identifying all the possible actions that may be performed in a single GS assessment tool is considered impossible (Williams, 2012). However, a comprehensive and standardised set of criteria and definitions may support replication of studies and comparison of data sets to enhance performance analysis research (Williams, 2012). Highlighting the contextual nature of assessment (Wiggins, 1993b), there was a considerable range in the number of assessment criteria in the tools located in this review and no standardised definitions of any criteria.

4.4.2.1 *Number of criteria*

The number of criteria in each tool was not charted, nor was the use of any specific software to collect and analyse this data beyond the widespread use of video capture. However, tools developed in the school setting tended to have fewer criteria than those developed in the community. For example, tools developed in the school setting like the TSAP (Gréhaigne et al., 1997) had six criteria, the GPAI (Oslin et al., 1998) had seven criteria (to select from), and the Coding Instrument (Blomqvist et al., 2005) had 10 criteria. In contrast, tools developed in the community like the System of Tactical Assessment in Soccer (FUT-SAT) (da Costa et al., 2011) comprised 76 criteria, the Polar Coordinate Analysis instrument (Jiménez-Salas et al., 2020) comprised 148 criteria, and the Observation Instrument for Technical and Tactical Actions of the Offense Phase in Soccer (Ortega-Toro et al., 2019) comprised 261 criteria. Consistent with other tools developed for use in the community, none of these tools were applied in a school context and as such their utility in a PE classroom remains uncertain.

4.4.2.2 *Nature of criteria*

On-the-ball skills was the most prevalent criterion across all tools; this supported the view that assessment of GS is often focused on skill execution or performance (Bennett et al., 2017; Koopmann et al., 2020; Pill, 2007a; Williams et al., 2020). As an example of a relatively objective operational definition that may contribute to consistent or reliable assessment, passing was defined by Arias-Estero (2013) as “Throwing the ball to a teammate with: a) the direction and b) the necessary force so that the receiver could receive it, from the waist to the top of the head, with the arms extended” (p. 705). In contrast, more abstract descriptions for passing included

terms like a “rainbow pass” (Farias, Mesquita, et al., 2019, p. 692) and “useful passing” (Darnis & Lafont, 2013, p. 466). The lack of clarity in operational definitions of performance criteria pervaded the assessment tools. Unclear assessment criteria are potential barriers to the construction of shared mental representations of the quality being observed. Similar to the use of ambiguous terminology in communicating student learning through assessment and reporting presented in Chapter 2 (Cookson, 2018; Hollingsworth et al., 2019), greater clarity and consistency in assessment criteria is encouraged.

As the second most prevalent key performance criterion, off-the-ball movement was assessed in almost two-thirds of all tools. While specific aspects of off-the-ball movement were not charted, references included players modifying their location according to the situation of their direct opponent (Ibanez et al., 2019), and movement required by the flow of the game (Gray & Sproule, 2011). Two of the five off-the-ball behaviours proposed in the GPAI were commonly used to assess attacking and defensive positioning. Specifically, defensive positioning was assessed with reference to the term *cover* (for example, Blomqvist et al., 2005; Oslin et al., 1998), and attacking positioning was assessed with reference to the term *support* (for example, Lee & Ward, 2009; Oslin et al., 1998). As an example of an unclear definition of these terms, *support* was defined in the GPAI development paper as “The player appeared to attempt to support the ball carrier by being in/ moving to an appropriate position to receive a pass” (Oslin et al., 1998, p. 243). This description does not describe what constitutes an *attempt* or what is *appropriate*. In contrast, the Game Performance Assessment tool (Miller et al., 2019) described support in netball with reference to players engaging in play, moving into space, and using space behind a defender if the player was covered.

As the third most prevalent key performance criterion, more than one-third of all tools assessed decision-making which is considered integral to team invasion game performance (for example, Inns et al., 2023). In recognising that decision-making in IGS must be assessed with reference to an action, most tools assessed decision-making with respect to off-the-ball movement and on-the-ball skills; for example, the Coding Instrument (Blomqvist et al., 2005), the GPET (García-López et al., 2013) and the Game Performance Coding Instrument (Tallir et al., 2007). In a smaller number of tools, decision-making was exclusively linked to on-the-ball skills; for example, the GPAI (Oslin et al., 1998), the Game Play Observational Instrument (Turner & Martinek, 1999), and the Passing Decision Coding Instrument (Nevett et al., 2001).

4.4.2.3 Links to domains

Considering the key performance criteria within the three learning domains in PE, all named key performance criteria may be classified exclusively within the psychomotor domain, as they relate to movement, skill selection and skill execution (Mitchell et al., 2013). Alternatively, when viewing decision-making as a form of tactical knowledge in action, this criterion may be classified within the cognitive domain (García-Ceberino et al., 2020; Turner & Martinek, 1992). While specific indicators were not formally charted within the 'other' criterion, examples like 'communicating with teammates' in the Collective Game Efficacy and Individual Skill Level Tool (Darnis & Lafont, 2013), and 'player attitude' in the Game Technical Scoring Chart (GTSC) (Rowat et al., 2017) may be classified within the affective domain. This finding partially supported the view that PE teachers should assess students across all three learning domains (Hay & Penney, 2013; Lund & Veal, 2013; Mitchell et al., 2013), however, assessment in the psychomotor domain was most prevalent.

4.4.3 Assessment Tool Applications

The data items reported in Table 4.8 indicated the total number of participants across the studies ranged from an $n = 6$ to $n = 561$. As with many variables across the ScR, several studies did not report this data item. Studies reporting the gender of the population came in three configurations including both genders, exclusively male, and exclusively female. As further evidence of omitting key data, over one-quarter of studies did not indicate the gender of the sample. In re-configuring the age-based data into two groups approximating the age of primary and secondary school students in a Victorian context, approximately one-quarter of studies used a population that was exclusively 13 and over (Year Level 7 and above). This data supported findings in the review of Arias-Estero and Castejón (2014) that studies involving secondary school students in games assessment are under-represented. In summary of these demographic findings, future studies involving populations of secondary school-age and exclusively female cohorts are needed.

4.4.3.1 Contextual alignment

According to the systematic review of assessment for tactical learning in games by Barquero-Ruiz et al. (2019), there should be alignment between where an assessment tool is developed and where it is applied. The importance of applying tools in the contexts they were developed in is attributed to potential differences in player attributes between youth sport contexts (community) and PE (school) (Diaz del Campo et al., 2011). This means that any operational definitions for assessment criteria in one context may not be transferable to another context (Barquero-Ruiz et

al., 2019). In the community context, each of the tools were applied in congruent settings while three school-developed tools were also applied in the community setting. A key challenge in applying community-developed tools in school settings is the expansive range of criteria that comprise some tools. Further, challenges with transferability may be linked to *feasibility*, which has been described as a third indicator of assessment information quality, along with validity and reliability (Brookhart, 2005). In this review, feasibility relates to how easy or manageable an assessment tool can be to administer (Robertson et al., 2014). These challenges might include the cost and expertise required to use sophisticated software (for example, Bredt et al., 2016; Castelão et al., 2014; da Costa et al., 2010) and the use of multiple cameras for video capture (for example, Bonney et al., 2020; Llobet-Martí et al., 2016; Pérez-Morales et al., 2018). Additionally, access to adequate space to cater to large class numbers and the use of multiple balls to maximise playing time may be problematic in a school setting (for example, Bonney et al., 2020; Fenner et al., 2016). As most assessment tools were used in the contexts they were originally developed, this review indicated a high level of congruence in the assessment design and assessment application contexts.

4.4.3.2 Game conditions

In terms of game conditions, game forms were classified as small-sided games (SSG) or full-sided games (FSG) based on game forms employed in the pilot searches of the literature. More than three-quarters of all studies applied their tools exclusively through SSG, suggesting potential alignment with GBAs used in a school setting. These GBAs might include the TGfU model (Bunker & Thorpe, 1986), Game Sense (Breed & Spittle, 2011; den Duyn, 1996), Play with Purpose (Pill, 2007b, 2012, 2013), Play Practice (Lauder & Piltz, 2013a, 2013b), and the Tactical Games Approach (Griffin & Butler, 2005; Mitchell et al., 2013). In studies that reported their observation period in minutes, the length of observation ranged from three minutes, in applying the GPAI (Viciano et al., 2017), to 336 minutes, in applying the GPET (Praxedes et al., 2018). The time in minutes for the latter study was calculated from the seven post-match assessment matches of 48 minutes duration and highlighted the inconsistent reporting of this item. In studies that reported their observation period in games or matches, the length of observation ranged from one match applying the TSAP (Layne & Hastie, 2014), to 76 matches applying the same tool (Hastie et al., 2017). Three other studies reported a series of plays or trials using the GPAI, the BALPAI, and the System for Notational Soccer Analysis (Harvey et al., 2010; Ibanez et al., 2019; van Maarseveen et al., 2017), while almost 10% of studies did not provide any indication of the period of observation.

The challenge in making sense of this data is that some studies clearly indicated the period of observation for each player/team, while other studies did not, and instead an aggregate of all players/teams was provided.

Given the prevalence of frequency-count approaches predicated on the number of actions or behaviours demonstrated, the amount of time that the ball was in play was rarely considered. As an exception, the validation paper of the Coding Instrument (Blomqvist et al., 2005) addressed this issue when stating that “The average effective playing time (the time ball was in play) was 61% of the total (10 minutes) playing time” (p. 114). Beyond reporting the average time, the same authors also provided the range of time that the ball was in play, with the highest rate being 73.8% and the lowest rate being 53.0% (Blomqvist et al., 2005). The issue of effective playing time was more broadly acknowledged in other studies (Evangelio et al., 2019; Farias, Harvey, et al., 2019; Farias et al., 2015; Farias, Mesquita, et al., 2019) that indicated the duration of observation was 10 min that included a minimum of 5 min of continuous play.

Video capture was used to complete the assessment in most studies, with outliers including two studies that applied the GTSC reporting that live performance enhanced the ecological validity of their instrument (Fenner et al., 2016; Rowat et al., 2017). The high level of video capture supported findings in the review of frequently used assessment instruments in GS by Arias-Estero and Castejón (2014) that indicated that three-quarters of studies involving the GPAI and TSAP used video capture. While video capture is advocated in PE and the assessment of GS in school settings (Koekoek et al., 2018; Weir & Connor, 2009), its use has implications for student privacy; logistics of camera placement, handling of equipment and uploading of footage; and cost.

4.4.3.3 Assessment users

Regarding assessment users, most studies relied on teachers, coaches, or researchers to complete the assessment. Several studies involving the TSAP (for example, Gréhaigne et al., 1997; Richard et al., 1998) and one study involving the Tool for Assessment and Learning of an Invasion Situation (TALIS) (Otero-Saborido & González-Jurado, 2015) made use of school-age participants as peer assessors. Researchers adapting this tool to the sport of ice-hockey found that students aged between 14 to 17 years could use the instrument with a “reasonable level of reliability” after less than one hour of training (Nadeau, Richard, et al., 2008, p. 77). It is hypothesised that this reliability figure is ‘reasonable’, as there are no decision-making or off-the-ball behaviours in the TSAP, meaning that fewer observable actions are counted. Equally, as the TALIS only assessed off-

the-ball movement and ball reception, it may be better suited to peer assessment than more complicated tools that include decision-making. The activating of students as judges may support student learning and is advocated in the AfL literature (AISEP, 2020; Black & Wiliam, 2009; for example, Gibbons & Kankkonen, 2011); however, it was under-utilised in the corpus.

4.5 Recommendations

A pragmatic aim of this study was to construct knowledge for action (Goldkuhl, 2011, 2012) for relevant stakeholders. As such, the following recommendations are directed at researchers and PE teachers respectively. Key recommendations for researchers include reporting population and assessment protocol data, providing their assessment instrument with a unique name, and/or acknowledging their tool as an adaption of a published instrument. Researchers are also encouraged to provide the name, total number, and definitions for all performance criteria. This information may be conveyed through sample coding sheets, appendices or supplementary material. Examples of tools that provided supplementary material include the Game Performance Coding Instrument (Tallir et al., 2007), the Observation Instrument for Technical and Tactical Actions of the Offense Phase in Soccer (Ortega-Toro et al., 2019), and the Observational Scoring Rubric (Williams & Rink, 2003). Appendix I also provides adapted coding sheets from the GPAI (Oslin et al., 1998) and TSAP (Gréhaigine et al., 1997) validation studies. Collectively, these recommendations may enhance the degree of rigor in research and provide readers greater clarity in determining the suitability or transferability of any tools.

While researchers have advocated raising awareness of evidence-based tools like the GPAI and the TSAP for PE teachers (Arias-Estero & Castejón, 2014; Williams et al., 2020), these tools have been found to be more likely used by researchers (Arias-Estero & Castejón, 2014; López-Pastor et al., 2013; MacPhail et al., 2008; Williams et al., 2020). Specific to the assessment of IGS, the developers of the GPAI have acknowledged that the tally system is “impossible” (Mitchell et al., 2013, p. 1) to use effectively because of the dynamic and complex nature of these games. Thus, the use of frequency-count tools in a PE setting may benefit from simplifying these instruments (Arias-Estero & Castejón, 2014). As an example, Harvey (2007) presented four rubrics adapted from the GPAI (Oslin et al., 1998) for the assessment of IGS in a school-age population. While these rubrics did not report any validity or reliability measures, adaption of other frequency-count tools into a rubric format may be a suitable way to simplify their use in a school setting. The application of the key performance criteria located in the ScR is encouraged, as is consideration of

the *minimum* number of criteria required to determine the degree of student learning in an effort to improve assessment utility (Arias-Estero & Castejón, 2014).

Further recommendations for PE teachers in the assessment of IGS include familiarising themselves with the published, peer-reviewed, evidence-based assessment tools. With greater awareness of the 15 assessment tools validated for use in a school context (see Appendix E), teachers may not have to ‘reinvent the wheel’ in designing assessment instruments. To support dissemination of this information to PE teachers, a series of publications in relevant journals and posts to social media forums are planned following the submission of this thesis. Supporting the implementation of these tools in school settings, organisations like the VCAA, ACHPER Victorian Branch and Peak Phys Ed are encouraged to create and share resource materials and professional learning opportunities. As a precedent, provision of professional learning opportunities for secondary school PE teachers in their assessment of GS using the GPAI was a key outcome in another Australian study examining the same assessment context (Williams et al., 2020).

Integral to the provision of the above support material and professional learning opportunities, PE teachers may require support in adapting and simplifying these tools to suit their student cohort. Modification or adaptation of existing tools to support a specific context is endorsed in the PE and performance analysis literature (for example, Brewer & Jones, 2002; Fernandes et al., 2019; Nadeau, Godbout, et al., 2008). An essential component in adapting any existing GS assessment tool is to establish objective and observable operational definitions (Williams, 2012). Any adaptation of tools and/or assessment protocols, for example, the criteria used, criteria definitions, outcomes generated, team size, game rules and the length of observation, may also benefit from the involvement of those being assessed. Beyond establishing face validity, by determining the degree of acceptability by those that are being assessed (Oslin et al., 1998), providing students a degree of voice and choice in their assessment can develop their assessment literacy (DinanThompson & Penney, 2015; Smith et al., 2013), promote student agency (Vaughn, 2020), and help meet their psychological needs (Weeldenburg et al., 2021).

4.6 Strengths and Limitations

Other reviews have used different search strategies to identify GS assessment tools (for example, Arias & Castejón, 2012; Barquero-Ruiz et al., 2019), however, a key strength of this ScR was its identification of the most expansive collection of IGS assessment tools in the literature ($N = 32$). For example, the review of Arias and Castejón (2012) identified just six assessment

instruments used in the context of IGS. These instruments were also located in this ScR and comprised the GPAI; TSAP; Coding Instrument; Game Play Observational instrument; Cutting or Off-the-Ball-Actions Coding Instrument; and the Passing Decision Coding Instrument. In a more recent review, Barquero-Ruiz et al. (2019) identified four assessment instruments (GPAI, GPET, FUTSAT, TSAP) and an instrument category (spatial location). All four instruments were located in this ScR, while the instrument category (spatial location) established by Barquero-Ruiz et al. (2019) did not identify any tools by name. It is noteworthy that tools in this category of spatial location were considered to lack utility in a school PE environment, as they focussed their assessment on *where* an action was performed on a playing space. Overall, the volume of tools identified, the detailed description of their characteristics, the reporting of key performance criteria and study-specific applications constitute a key strength of the study.

A potential limitation of this ScR was that studies were only included if written in the English language, whereas a multi-lingual approach, like that used in the systematic review of the evaluation of tactical performance in IGS by Ávila-Moreno et al. (2018), may have yielded other studies and assessment tools. Alternative key word search terms, other databases (for example, ERIC), and the use of grey literature, may have also located other useful studies. Limited references in abstracts to tools/instruments and measurement properties in the search strategy led to many studies being located outside of the formal search. The omission of these key terms in the title, abstract and key word search may be attributed to the primary aim of many papers to investigate a teaching intervention, rather than validate an assessment tool. In hindsight, these themes should have been refined by the researcher, such that the reporting of validity and reliability measures would remain in the inclusion criteria, not the search strategy. Further, just as qualitative research may be assessed against a wide range of criteria that are different from validity and reliability commonly reported in quantitative research (for example, Nelson, 2016; Sparkes & Smith, 2009; Tracy, 2010), it is possible that a range of qualitative assessment tools were missed by the ScR due to the search strategy. Given the large volume of studies located outside of the search strategy, replication of this ScR is likely problematic.

Regarding the data charting, omission of reported data was widespread. While most data items were low inference, some items were challenging to chart due to inconsistent and unclear reporting in the studies; a feature observed in other ScRs (Arksey & O'Malley, 2005). As an example, the TSAP was not formally identified by name until its fourth application in the ScR (Richard et al., 2000). A detailed set of charting guidelines to support the verification of the

charting by the research assistant can be found in Appendix B. While the absence of quality appraisal in a ScR may be viewed as a potential limitation (Arksey and O'Malley, 2005), the aim of this ScR was "to map the available evidence rather than provide a synthesized and clinically meaningful answer to a question" (Peters et al., 2021, p. 7).

4.7 Conclusion and Further Research

The aim of this study was to determine the characteristics, key performance criteria, and applications of evidence-based assessment tools used in the context of IGS within a school-age population, as located in the published peer-reviewed literature. To achieve this purpose, the researcher completed a ScR based on the five-step framework of Arksey and O'Malley (2005) that was aligned to the PRISMA-ScR (Tricco et al., 2018). A total of 27 data items from 72 included studies were charted in order to meet the aim of the study. All items were independently charted by the researcher and subsequently verified by a research assistant with experience in data extraction and a PhD in the field of PE.

This ScR located a total of 32 unique assessment tools from 72 studies across 33 peer-reviewed journals. The vast number of tools identified in comparison to similar reviews (for example, Arias & Castejón, 2012; Barquero-Ruiz et al., 2019), poses a potential problem for researchers and practitioners; namely, locating and selecting the most appropriate assessment tool for a specific purpose. A possible solution to address the challenge of locating the most appropriate tool from such a broad array may be to construct an assessment *framework*. This framework could be based on assessment tool characteristics, criteria and applications across the corpus, to establish a single, yet flexible, evidence-based *approach* to IGS assessment. Such a framework may support the sharing of mental representations of key assessment criteria that has been criticised in the wider field of educational assessment and reporting (Cookson, 2018; Hollingsworth et al., 2019; Masters, 2014; Wyatt-Smith et al., 2014).

In summary of the key findings, single sport assessment tools were more prevalent than multiple-purpose tools, soccer was the most prevalent game that tools were designed for, and approximately half of all tools were developed for use in a school-setting. Reliability was reported more widely than measures of validity, and frequency-count tools that generated indices were more prevalent than tools that rated the quality of performance. On-the-ball skills was the most prevalent criterion across the corpus, and receiving and passing were the most frequently assessed game skills within this criterion. Off-the-ball movement was assessed more widely than

decision-making, and 'other' criteria were assessed more widely in community-developed than school-developed tools. Few studies drew on wholly female cohorts, and a relatively modest number of studies targeted a population exclusively over the age of 13. A total of 16 different games were assessed, primarily in SSGs that employed video capture to facilitate assessment. The findings from this ScR provide a reference point for the following three studies in this thesis by allowing comparisons between the assessment tools, key performance criteria, and protocols used by PE teachers in Victorian secondary schools and those located in this study.

As this ScR privileged validity and reliability as the "two most important indicators of assessment information quality" (Brookhart, 2005, p. 11), other ScRs that include feasibility as a third indicator of assessment quality are encouraged (Brookhart, 2005). Such a review may be useful to practitioners in the field given the limited use of frequency-count assessment tools by PE teachers (for example, Arias-Estero & Castejón, 2014; López-Pastor et al., 2013; MacPhail et al., 2008). Other ScRs are advocated to describe the characteristics, key performance criteria and applications of assessment tools used in other game categories. Systematic literature reviews that include appraising the quality of studies of the more prevalent assessment tools including the GPAI, TSAP, GPET and FUT-SAT are supported. These systematic reviews may provide a fuller examination of the validity and reliability of these tools to help inform future use in research and teaching practice.

Given the timing of the current ScR, with the included studies and tools classified as belonging to the pre COVID-19 era, an update of the current ScR may allow for comparisons in the design, characteristics and application of assessment tools post COVID-19. Some gaps identified in this ScR may be addressed by studies investigating the degree of curriculum alignment with assessment practices and applying tools in under-reported games, for example lacrosse, water polo, tag rugby and speedball. Studies drawing on exclusively female cohorts and research set exclusively in secondary schools are also invited due to their under-representation in this ScR. Examination of assessable criteria in the affective domains aligned to GS is encouraged; this may include investigations of teamwork, fair play, inclusivity, leadership, communication, and collaboration, that are supported in the state and national PE curricula of Australia (for example, VCAA, n.d.-d). As a potential avenue for this data, re-scanning the located assessment tools/studies in this ScR that were coded as using 'other' criteria may reveal some evidence-base for assessment in the affective domain.

Given the focus of the thesis within a PE teacher population and school context, the high number of studies located in Physical Education and Sport Pedagogy indicate that this journal may be receptive to any potential publications that arise as a consequence of this study. Further, as studies from Australia comprised less than 10% of the total studies, further investigations in IGS assessment in a local and national setting may be warranted. As Spain was identified as the greatest contributor of studies from a single country, multi-lingual collaborations in the design and development of assessment tools and/or literature reviews of assessment tools may be worthwhile. For example, the literature search in the review of tactical performance in invasion team sports within an elite adult population by Ávila-Moreno et al. (2018) was conducted in English, Spanish and Portuguese.

To bridge the divide between quantitative and qualitative assessment tools located in his study, investigating the potential of assessment formats that combine frequency-counts and rating performance is also endorsed. This could involve closer examination of the tools that used a combination of frequency-counts and rating performance in this study (see Bonney et al., 2020; Darnis & Lafont 201; Tangalos et al., 2015). It is possible that further study may be able to construct and validate an assessment framework predicated on the characteristics, key performance criteria and applications across the corpus to establish a single, yet flexible, evidence-based approach to IGS assessment. The findings presented in this ScR provide a substantive addition to the literature in the context of IGS assessment that constitute a significant and original contribution to knowledge for researchers, the discipline of PE, and practitioners in the field. The following chapter draws on the findings of this ScR in describing the construction of a bespoke questionnaire for practicing PE teachers in the context of IGS assessment.

CHAPTER 5: PHYSICAL EDUCATION TEACHERS' ASSESSMENT PRACTICES IN INVASION GAMES AND SPORTS IN VICTORIAN SECONDARY SCHOOLS: A CROSS-SECTIONAL, QUANTITATIVE INQUIRY

5.1 Introduction

The overview of assessment in PE literature in Chapter 2 revealed the limited use of evidence-based assessments in school settings and the prevalence of subjective assessment based on non-performative criteria. The ScR in Chapter 4 located 32 evidence-based assessment tools in the peer-reviewed published literature and indicated that the alignment of assessment to curriculum was limited; that assessment based on generating frequency-based outcomes was prevalent; and that the evaluation of on-the-ball skills, off-the-ball movement and decision-making was widespread. Both chapters supported findings from an earlier literature review that identified the GPAI and the TSAP as the two most frequently employed assessment tools for the tactical assessment of students in PE and youth sports (Arias & Castejón, 2012).

It remains unclear how teachers currently understand and use assessment tools in their daily practice, and how this understanding and use compares to the tools, practices and issues arising in the academic literature. Consistent with the interpretive theoretical framework of the thesis, to make meaning of participants' understanding and practice of assessment, the purpose of this study was to answer research sub question 2: *How do Physical Education teachers view the assessment of invasion games and sports in Victorian secondary schools?*

5.1.1 Aim

The aim of this inquiry was to describe and understand assessment practices used by PE teachers in Victorian secondary schools (Year Levels 7 - 10) in the Focus Area of GS. Specifically, this involved considering aspects of assessment that were understood as being useful in the assessment of IGS. Questions used in this study sought to address gaps in the literature and further understanding in areas of interest identified in the thesis. To answer the research sub-question, the key foci that were determined a priori comprised: (a) curriculum alignment to assessment; (b) assessment tools, assessment users and assessment utility; (c) key performance criteria; and (d) familiarity and use of seminal assessment tools (GPAI and TSAP). As the utility of assessment was a key component of the research sub-question, items were informed by evidence-based options identified in the literature review of assessment in PE (Chapter 2).

5.1.2 *An Overview of Survey Research in PE*

Survey design, comprising questionnaires and interviews, is a common approach in social research (Vogt et al., 2012). Questionnaires are widely used in the field of education (Ary et al., 2014), and often precede interviews in explanatory, sequential, mixed methods studies (Creswell & Plano Clark, 2018; Ivankova et al., 2016). In an international PE context, questionnaires have been used within a survey design to determine assessment quality in secondary schools in the Netherlands (Borghouts et al., 2016), to report student views on grading in Sweden (Redelius & Hay, 2012), and to describe student views on the value and enjoyment of IGS in Scotland (Gray et al., 2008). Respectively, these studies have reported modest assessment quality and limited use of AfL (Borghouts et al., 2016), that secondary school students were unaware of the criteria informing their PE grades (Redelius & Hay, 2012), and that primary school students valued IGS more highly than their secondary school peers (Gray et al., 2008).

No survey designs comprising questionnaires related to the assessment of IGS could be found in the extant literature. However, studies that utilised questionnaires with a sample of Australian PE teachers included those that explored teacher perceptions of physical literacy (Essiet et al., 2022), the assessment of Fundamental Movement Skills (FMS) (Lander et al., 2015; Lander et al., 2016), and teacher efficacy in senior PE (Whittle et al., 2017b). While two relevant Australian studies have explored the assessment of GS in PE using individual, semi-structured interviews and focus group interviews respectively (Georgakis et al., 2015; Williams et al., 2020), there was no use of an online questionnaire from which to select a purposive sample. In addition to these differences in study design, one study involved the state-based curriculum of NSW, and the other the AC: HPE, while this study primarily involved the VC: HPE curriculum. Additionally, while both studies considered secondary school PE teachers' use of the GPAI and the TSAP, neither study invited participants to provide a detailed critique of these tools, which occurred in this study. In summary of their relevant findings, modest use of the GPAI and TSAP was reported by Georgakis et al. (2015), while none of the 19 secondary PE teachers in the study by Williams et al. (2020) used either assessment tool. As such, this chapter provides an original contribution to knowledge by investigating teachers' understanding and practice of assessment of IGS within a curriculum not previously investigated in the PE literature.

5.2 Method

The population for this study comprised PE teachers in Victorian secondary schools; this group were the most appropriate demographic given the location of the research. Consistent with the emergent research design, the questions that comprised the cross-sectional, quantitative inquiry described in this chapter were informed by the literature review of assessment in PE (Chapter 2) and the ScR (Chapter 4). An online questionnaire was developed that involved a sample of the population, at a single point of time, and was cross-sectional in nature (Ary et al., 2014; Creswell, 2012). As the second of four studies comprising the explanatory, sequential, mixed methods design (Creswell & Creswell, 2018; Schoonenboom & Johnson, 2017), this study is also positioned as the first phase of a two-phase survey design. The survey design in this chapter focusses on quantitative data, which is followed by a second-phase, qualitative inquiry using semi-structured interviews described in Chapter 6.

5.2.1 *Participants and Recruitment*

PE teachers in Victorian secondary schools were selected as the target population. The inclusion criteria required the participants to have taught IGS to students in Year Levels 7 - 10 in the calendar year of 2019. These Year Levels were selected as they typically represent the final four years of compulsory PE in secondary (high) school in Australia. These Year Levels also include an under-represented group in the application of evidence-based assessments as identified in the ScR (Chapter 4). The ScR indicated that less than a quarter of the included studies applied assessment tools with children aged 13 years or over, equivalent to Year Level 7 and above.

There was no way to determine the total number of practicing PE teachers in the state of Victoria, thus the potential population could not be accurately established. As the exact number of PE teachers in Victorian secondary schools taking classes at Year Levels 7 - 10 is unknown, there are 249 Primary/Secondary schools and 342 Secondary schools, that means there are a total of 583 relevant school sites in Victoria (State Government of Victoria, 2018). Thus, in collecting data from a portion of the above population a “sample survey” approach was adopted (Ary et al., 2014, p. 400). With advice from the supervisory team, based on their collective experience and knowledge of staffing in secondary schools, an estimate of three PE teachers taking at least one class of Year Level 7 - 10 in each school was made to allow for differences at each school site. This estimation identified a potential target population of 1749 secondary school PE teachers in the

state of Victoria for 2019. In applying a confidence level of 95% with a confidence interval of 10%, this proposed a target sample size of 91.

Participant recruitment was supported from the Australian Council for Health Physical Education and Recreation (ACHPER) Victorian Branch (<https://achper.vic.edu.au/>), Peak Phys Ed (www.peakphysed.com.au/), and ConnectedPE (<https://connectedpe.com/>). In a Victorian context, the first two of these organisations were also used by Lander et al. (2015) and Essiet et al. (2022) to recruit participants, thus providing precedence for their selection. Identifying the target population from membership details for each of the organisations was not possible due to the limited demographic information available. For example, ACHPER Victorian branch as the largest organisation, reported having up to 1000 financial members, that fluctuated at any point of time in the year. These members included teachers that taught exclusively in Primary school, and secondary teachers that taught Outdoor Education, Health Education and/or PE (K. Borrie, personal communication, August 26, 2018). Of the indeterminate number of secondary school PE teachers, there remained no way to discern how many teachers taught Year Levels 7 - 10, meaning that there was no way to establish the return rate from the members of the recruitment organisations. Each organisation gave permission for recruitment materials to be disseminated through their regular communication channels, that included online newsletters and social media. The materials for recruitment included a brief text summary and flyer which detailed the aim of the study and participant inclusion criteria (Appendix G). In Victoria, as with other Australian states and territories, the school year is based on the calendar year, and extends from late January to December. The first approach for the recruitment of participants was made on November 2019, which allowed participants to reflect on approximately 35 weeks of teaching in the school year.

In recognising the summer break that followed the first recruitment of participants, a second approach was made to participants early in the following school year. This approach was made through the same peak bodies' newsletters and social media platforms in February, 2020. Consistent with other approaches to maximise response rates (SueSee et al., 2018), a planned third approach was originally scheduled for March 2020. This was postponed to October 2020 due to a series of extended COVID-19 lockdowns and school closures in Victoria that ultimately led to Melbourne being classified as the most locked-down city in the world (Tuffield, 2021). Given these unprecedented events that extended the time the questionnaire was open, the first participant completed the questionnaire on November 20th 2019, and the final participant's response was submitted on December 15th 2020.

5.2.2 *Ethics*

Ethics approval for the study was provided (project number 8434) by the Social and Behavioural Research Ethics Committee (SBREC) of Flinders University and can be found in Appendix A. The original approval was provided on November 6th 2019. All participants in the study were provided with an information sheet to read at the start of the online questionnaire. Participants provided informed consent to ensure their right to privacy, anonymity and confidentiality by choosing this option prior to the content questions. Copies of the scripts for recruitment, letter of introduction, information sheet and consent can be found in Appendix G.

5.2.3 *Data Collection*

The online platform Qualtrics was used to construct and administer the self-report questionnaire. A self-report questionnaire provided participants with the opportunity to respond to a series of questions or statements (Ary et al., 2014). An online approach offered the potential to capture a state-wide sample and has been used in research involving PE teachers in Victorian secondary schools (Lander et al., 2015; Whittle et al., 2017b). To support validity and reliability, researchers advocate the use of established questionnaires (Kervin et al., 2016). The review of the assessment in PE literature presented in Chapter 2 located a contextually aligned survey design that used an online questionnaire (Borghouts et al., 2016). However, this investigation of secondary school PE teachers' assessment practices in the Netherlands did not focus on assessment in IGS, it made no reference to the seminal tools of the GPAI and TSAP, and was considered inappropriate to meet the aim of this study. As with other self-report questionnaires used in a Victorian, secondary school, PE teacher population (Lander et al., 2015), a bespoke questionnaire was designed to respond more appropriately to the research sub-question. In designing the questionnaire, content validity was established through the iterative development of items with the researcher's supervisors and reference to the literature, while face validity and reliability were established through protocols endorsed by Lander and colleagues (2015), that are explained presently.

The questionnaire included five sections and a total of 16 questions, some of which had multiple parts. In addition to the four foci described in the aims of this chapter, a fifth section collected data on participant demographics and the prevalence of IGS in participants' PE programs. Close-ended responses were used to reduce subject burden (Emmananouilidou et al., 2012) and increase efficacy of data treatment. To maximise participant completion rates of the

questionnaire, a modest number of investigative questions were classified within four categories that were most appropriate to respond to the research sub-question. These categories comprised: (a) curriculum alignment to assessment; (b) assessment tools, assessment users and assessment utility; (c) key performance criteria; and (d) familiarity and use of seminal assessment tools (GPAI and TSAP). Determining the nature of assessment tasks, the design of assessment tools, the assessment of non-performative criteria, the relative use of formative or summative assessment, and the frequency of assessment were not stated aims of the current study.

Specifically, the questionnaire was composed of multiple-choice and Likert scale questions. A five-point scale was preferred over a four-point scale; the latter typically omits a neutral option and subsequently forces participants to agree or disagree with statements (Allen & Seaman, 2007; Ary et al., 2014). Given that some participants may have been unsure, indifferent or ambivalent about question items, the neutral option was provided to cater for such participants and potentially reduce participant drop out. The scale of the questions used the terms strongly agree, agree, neutral, disagree and strongly disagree. The items were constructed in consultation with the researcher's supervisors and involved critically examining the questions in their capacity to answer the research-sub-question. This consultation included meetings, email communication seeking feedback on the clarity of questions and the suitability of response options.

In addition to the ongoing iterative process to refine the questionnaire and establish content validity with the supervisory team, the researcher invited three experienced secondary school PE colleagues, who were not currently teaching Year Levels 7 - 10 PE, to complete the questionnaire. A similar pilot testing approach to check face validity was employed in the self-report questionnaire in a study of FMS assessment within a Victorian, secondary school, PE teacher population (Lander et al., 2015). In the current study, the researcher prefaced the questionnaire by outlining the aim of the instrument was to understand the assessment practices used by PE teachers in Victorian secondary schools in IGS at Year Levels 7 - 10. While completing the questionnaire, the researcher's colleagues were asked to concurrently respond to a feedback form that invited their opinions of features of the questionnaire including question structure, question order, multiple choice options and the total time required to complete the questionnaire. One of the three participants agreed to use a mobile device, as opposed to a laptop or desktop computer, and provide feedback on the user interface given the truncated screen size of such devices.

Acting on their responses, definitions of the assessment types were removed on the basis that the terms were well understood, they occupied too much space on mobile device screens, and they increased reading time which increased the time required to complete the questionnaire. As an example of items that were adjusted based on participant feedback, observation as an assessment tool was not included as an option. All three participants explained that observation of itself, did not constitute an assessment tool, and was instead a *means* to collect data rather than a specific tool like rubrics, checklists, rating scales and frequency-count approaches. This rationale was also supported in the literature when acknowledging that “Physical Education teachers observe students every day, but observation, by itself is not assessment” (Lund & Veal, 2013). The brief descriptions of the GPAI and TSAP remained on the basis that only one of the participants was familiar with these tools; the limited awareness of these tools is supported in other Australian studies (Georgakis et al., 2015; Williams et al., 2020).

A final step in pilot testing the questionnaire was based on the repeatability testing protocol in a similar self-report questionnaire that also drew on a small sample size (Lander et al., 2015). In the current study, three secondary school PE colleagues completed the questionnaire at two times, between seven and ten days apart, and the researcher compared their responses for consistency. As with the findings from Lander et al. (2015), this test-retest protocol, representing intra-rater reliability, indicated consistency across items and led to the final version of the questionnaire being uploaded to Qualtrics (see Appendix F). In summary of this pilot testing, the researcher, the supervisory team and pilot participants agreed the questionnaire had an acceptable level of content and face validity. An overview of the questionnaire items in Table 5.1 identifies the 16 questions by number and letter, the broad topic investigated, the question format, and the rationale for each question.

Table 5.1*Overview of the Questions Comprising the Questionnaire*

Question	Broad topic	Question format	Rationale for question
1-6	School system, PE degree, Qualification, Teaching experience, Gender, Curriculum followed	Single multiple-choice	Demographics to describe the sample and help select nested purposive sample for future studies
7	Focus Areas	Single multiple-choice	Determine the prevalence of Games and Sports as a Focus Area within the curriculum
8	Game category	Single multiple-choice	Determine the prevalence of invasion games within the Focus Area of Game and Sports
9	Role of curriculum	3-part 5-point Likert rating scale	Determine degree of assessment alignment to curriculum
10	Types of assessment and users	10-part 5-point Likert rating scale	Determine prevalent assessment tools and users
11	Most useful type of assessment	Single multiple-choice	Determine what is understood to be the most useful assessment tool
12	Key performance criteria	8-part 5-point Likert rating scale	Determine the most used key performance criteria in invasion games and sports
13	On-the-ball skills	9-part 5-point Likert rating scale	Determine the most used on-the-ball skills
14	GPAI		
a	GPAI awareness/use	Single multiple-choice	Determine awareness of GPAI, use, and non-use Help to identify purposive sample of GPAI users/non-users
b	Limitations of GPAI	Various multiple-choice	Determine reasons for non-use
c	GPAI format	Single multiple-choice	Determine format of GPAI used (frequency-count, modified rubric)
d	GPAI criteria used	Various multiple-choice	Determine GPAI criteria/performance indicators used
e	Skills assessed	Various multiple-choice	Determine skills assessed

Question	Broad topic	Question format	Rationale for question
15	TSAP		
a	TSAP awareness/use	Single multiple-choice	Determine awareness of TSAP, use, and non-use Help to identify purposive sample of TSAP users/non-users
b	Limitations of TSAP	Various multiple-choice	Determine reasons for non-use
c	TSAP format	Single multiple-choice	Determine format of TSAP used (frequency-count, nomogram, modified rubric)
d	TSAP criteria used	Various multiple-choice	Determine TSAP criteria/performance indicators used
16	Useful aspects in assessment	7-part 5-point Likert rating scale	Determine what aspects of assessment are understood to be useful (independent of tool)

Note. GPAI – Game Performance Assessment Instrument (Oslin et al., 1998), TSAP – Team Sport Assessment Procedure (Gréhaigne et al., 1997).

The questionnaire items were informed by the review of PE assessment literature (Chapter 2) and the ScR (Chapter 4) that located 32 evidence-based assessment tools. Both these chapters provided detailed references to contextually related systematic literature reviews of assessment in GS (Arias & Castejón, 2012; Barquero-Ruiz et al., 2019; López-Pastor et al., 2013). These chapters also described common assessment forms, assessment criteria, and assessment protocols that were used in the assessment of IGS that informed the construction of the questionnaire items. Questions in the first category of the questionnaire, curriculum alignment to assessment, acknowledged that as assessment determines how much, or how well a curriculum has been learnt (Black & Wiliam, 2018), participants' views on the utility of curriculum were important. Questions in this category related to the capacity of the participants' curricula to describe learning goals, measure student progress and inform their assessment practice.

Questions in the second category of the questionnaire, assessment tools, assessment users and assessment utility, included identifying assessment tools that were used by participants from a range of tools that were located in the review of PE assessment literature (Chapter 2) (for example, Arias & Castejón, 2012; Desrosiers et al., 1997; Harvey, 2007). This category also asked participants to indicate which aspects of assessment they found useful. For example, the need for

assessment in PE to be time efficient (Braga & Liversedge, 2017; Gallo et al., 2006; Veloo & Md Ali, 2016) and suitable for peers to use (Chng & Lund, 2018; Gibbons & Kankkonen, 2011; Melograno, 1997) are widely reported in the PE assessment literature. The intention of questions in this category was to determine which assessment forms and aspects of utility were considered to be most used and useful.

Questions in the third category of the questionnaire, key performance criteria, included options that were established from the ScR (Chapter 4), the AC: HPE (ACARA, n.d.-b), and the VC: HPE (VCAA, n.d.-d). For example, the options of decision-making, positioning off-the-ball and skill execution (result of performance and technique), were identified in most of the school-developed tools located in the ScR. As a key performance criterion, the term 'decision-making' was deliberately chosen over references to strategy and/or tactics, which are often confused by PE teachers (Gréhaigine et al., 1999). This naming convention was also consistent with the criterion, decisions-made, found in the GPAI (Oslin et al., 1998). As an example of criteria described by the VC: HPE across Year Levels 7 - 10, working in teams (teamwork) and creating solutions (creativity (VCAA, n.d.-d), were included as item options. To reduce subject burden, only criteria that were supported in the literature were offered as questionnaire items, meaning that any assessment of non-performative criteria by the sample remains unknown.

Questions in the fourth category of the questionnaire, familiarity and use of seminal assessment tools (GPAI and TSAP), were based on the tools identified in the review of the instruments most frequently employed to assess tactics in physical education and youth sports (Arias & Castejón, 2012) and findings in the ScR (Chapter 4). As with all questions, participants were asked about their use of the GPAI and TSAP in the current school (calendar) year to reduce potentially inaccurate recollections of practice. The questionnaire concluded by offering participants the opportunity to express their interest in a follow-up interview to explain key findings more thoroughly. A copy of the questionnaire is provided in Appendix F.

5.2.4 Data Analysis

The participants' responses were exported from Qualtrics before being tabulated and organised into a Microsoft Excel file based on the four categories. These categories were the most appropriate to answer the research sub-question and comprised: (a) curriculum alignment to assessment; (b) assessment tools, assessment users and assessment utility; (c) key performance criteria; and (d) familiarity and use of seminal assessment tools (GPAI and TSAP). As the researcher

did not intend to explain relationships between any groups in the data set, data analyses were limited to the use of “basic descriptive statistics” (Kilborn et al., 2016, p. 27), rather than inferential statistics that can report variance in and across data sets. The data included frequency tabulations in the form of response totals and percentages.

5.3 Results

A total of 80 participants gave consent to participate in the questionnaire and response rates for individual items varied from 80 (100%) to 67 (84%). The maximum available sample for each item is reported, such that, where the number of participants is the same for all items in a table, the *n* is reported as a table note. Where the number of participants that responded to items within a table varies, the *n* for each item is reported in the final column of that table. Table 5.2 presents participant demographics including: level of education; years of teaching experience; PE teaching degree; participant gender; school type; and the curriculum followed. For the last question, participants were asked to select the curriculum that was *most* influential in their assessment of PE across Year Levels 7-10.

Table 5.2*Participant Demographics*

Demographics	<i>n</i> =	%
Level of education		
Bachelor's degree	53	66.3
Masters	13	16.6
Diploma of teaching	12	15.0
PhD	1	1.3
Other	1	1.3
Years of teaching experience		
Less than 5	8	10.0
Between 5 and less than 10	18	22.6
Between 10 and less than 15	20	25.0
Between 15 and less than 20	11	13.8
More than 20	23	28.8
PE teaching degree		
Yes	77	96.3
No	1	1.3
No response	2	2.5
Gender		
Male	50	62.5
Female	30	37.5
I do not identify with either of the above	0	0.0
School type		
Government	38	47.5
Independent	31	38.8
Catholic	11	13.8
Curriculum followed		
Victorian Curriculum (VC)	55	68.8
School-developed	13	16.6
Australian Curriculum (AC)	11	13.8
Middle Years Program (MYP)	1	1.3

Note. (*N* = 80 participants)

All respondents included the Focus Area of GS in their Year Level 7 - 10 PE programs. Further, 75 out of 80 respondents (93.8%) indicated that the GS Focus Area occupied the most curriculum time. When asked to indicate the game category that occupied the most curriculum time within the GS Focus Area, 76 of 80 respondents (95.0%) indicated the category of IGS. In contrast, net and wall games and striking/fielding games were each selected by two respondents representing (2.5%) of the population. In Table 5.3 data are presented for the first category, Curriculum Alignment to Assessment in IGS.

Table 5.3*Curriculum Alignment to Assessment*

Curriculum statement	Strongly agree	%	Agree	%	Neutral	%	Disagree	%	Strongly disagree	%	n
The curriculum clearly outlines student progress in invasion games and sports across Year Levels	7	8.6	34	42.5	18	22.5	17	21.3	4	5.0	80
The curriculum clearly sets out learning goals for invasion games and sports	4	5.0	43	53.8	14	17.5	15	18.8	4	5.0	80
The curriculum directly informs my assessment practice of invasion games and sports	4	5.0	33	42.3	20	25.6	17	21.3	4	5.0	78

Data for the second category, Assessment Tools, Users and Utility, is presented in three tables and comprises the most substantive category in the discussion section. In Table 5.4 participants were asked to indicate which assessment tools they used, and who completed the assessment.

Table 5.4

The Types of Assessment Tools Used and Users (Teacher, Self, Peer)

Assessment tool	Teacher	%	Self	%	Peer	%	Not used	%
Rubrics	68	89.5	26	34.2	17	22.4	5	6.6
Verbal responses	49	64.5	20	26.3	20	26.3	19	25.0
Checklists	47	61.8	17	22.4	23	30.3	24	31.6
Skill tests	32	42.1	16	21.1	17	22.4	35	46.1
Rating scales	26	34.2	10	13.2	7	9.2	45	59.2
Written test/ assignment	23	30.3	13	17.1	9	11.8	41	53.9
Frequency count tools	7	9.2	3	3.9	3	3.9	67	88.2
Journal	5	6.6	10	13.2	6	7.9	60	78.9
Other	5	6.6	5	6.6	5	6.6	66	86.8
Portfolio	2	2.6	5	6.6	5	6.6	66	86.8

Note. n = 76 participants. Multiple users could be selected.

In Table 5.5 participants were asked to select one assessment tool as the most useful in their assessment practice.

Table 5.5

The Most Useful Assessment Tool

Assessment tool	Most useful	%
Rubrics	37	50.7
Rating scales	11	15.1
Checklists	8	11.0
Verbal responses	7	9.6
Skill tests	4	5.5
Portfolio	2	2.7
Other	2	2.7
Frequency count tools	1	1.4
Journal	1	1.4
Written test/ assignment	0	0.0

Note. $n = 73$ participants. Only one assessment tool could be selected.

In Table 5.6 participants were asked to indicate the degree to which various aspects of assessment were considered useful.

Table 5.6

The Utility of Select Aspects of Assessment

Aspect of assessment	Strongly	%	Agree	%	Neutral	%	Disagree	%	Strongly	%	<i>n</i>
	agree								disagree		
Is time efficient to use	49	71.0	10	14.5	7	10.1	3	4.3	0	0.0	69
Can be completed electronically	26	37.7	19	27.5	19	27.5	4	5.8	1	1.4	69
Can be completed by student peers	25	36.2	33	47.8	10	14.5	0	0.0	1	1.4	69
Is linked to the curriculum	24	34.3	31	44.3	13	18.6	2	2.9	0	0.0	70
Suggests the next steps in learning	24	35.3	32	47.1	8	11.8	3	4.4	1	1.5	68
Accurately indicates performance level	21	30.0	39	55.7	5	7.1	4	5.7	1	1.4	70

Data for the third category, Key Performance Criteria, is presented in two tables. In Table 5.7 participants were asked to indicate the extent to which their assessment of IGS was based on the following items.

Table 5.7

Key Performance Criteria

Key performance criteria	Strongly agree	%	Agree	%	Neutral	%	Disagree	%	Strongly disagree	%	<i>n</i>
Decision-making	38	53.5	31	43.7	1	1.4	1	1.4	0	0.0	71
Teamwork	33	47.1	32	45.7	3	4.3	2	2.9	0	0.0	70
Attacking play	21	29.6	39	54.9	9	12.7	2	2.8	0	0.0	71
Skill execution (result)	18	25.0	42	58.3	6	8.3	6	8.3	0	0.0	72
Skill execution (technique)	18	24.7	42	57.6	9	12.3	3	4.1	1	1.4	73
Positioning off-the-ball	17	25.0	40	58.8	6	8.8	4	5.9	1	1.5	68
Defensive play	16	22.9	44	62.9	8	11.4	1	1.4	1	1.4	70
Creativity	7	10.4	25	37.3	26	38.8	7	10.4	2	3.0	67

In Table 5.8 participants were asked which of the following on-the-ball skills they assessed in IGS.

Table 5.8

Select On-The-Ball Skills Assessed

Ball skills	Used in assessment	%
Pass	62	88.6
Dribble/carry	57	81.4
Intercept/steal	54	77.1
Receive	50	71.4
Block an attacking play	43	61.4
Clear the ball	35	50.0
Score	34	48.6
Tackle	22	31.4
Other	8	11.4

Note. $n = 70$ participants. Multiple items could be selected.

Data for the fourth category, Familiarity and use of the GPAI and TSAP, is presented in the following table and text. In Table 5.9 participants were asked to indicate their level of awareness and use of the two most prevalent assessment tools located and described in the ScR (Chapter 4), the GPAI and the TSAP.

Table 5.9

Familiarity and Use of the GPAI and TSAP

Familiarity and use of the GPAI and TSAP	Number of participants	%	<i>n</i>
I am not familiar with the GPAI	47	65.3	72
I am familiar with the GPAI and have not used it	20	27.8	72
I have used the GPAI	5	6.9	72
I am not familiar with the TSAP	57	80.3	71
I am familiar with the TSAP and have not used it	12	16.9	71
I have used the TSAP	2	2.8	71

Note. GPAI – Game Performance Assessment Instrument (Oslin et al., 1998), TSAP – Team Sport Assessment Procedure (Gréhaigne et al., 1997).

Due to the participants' low level of familiarity and use of both tools the following results are presented in text rather than tabular form. For 20 of the 72 respondents (27.8%) that were aware of the GPAI but had not used it, key limitations included that the tool was too time consuming ($n = 11$), that student-led peer assessment lacked accuracy ($n = 5$), and that it was too complicated ($n = 4$). For 5 of the 72 respondents that used the GPAI, three found the frequency count approach most useful while two respondents indicated that a modified rubric was most useful. These five respondents indicated their use of the following GPAI performance indicators: decisions-made ($n = 5$); skill execution ($n = 4$); support ($n = 2$); and adjust ($n = 1$). In regard to specific on-the-ball skills that were assessed, the five respondents indicated their use of: pass ($n = 5$); clear the ball, dribble or carry the ball; intercept or steal ($n = 4$); block an attacking play; catch or trap the ball ($n = 3$); and score and tackle ($n = 2$). The respondents also indicated their use of the following indices or performance scores: game performance ($n = 4$); game involvement ($n = 3$); decisions-made index ($n = 1$); and skill execution index ($n = 1$).

For 12 of the 71 respondents (16.9%) that were aware of the TSAP but had not used it, the following limitations were identified: it was time consuming ($n = 8$); student-led peer assessment lacked accuracy ($n = 4$); and it was too complicated ($n = 2$). For the two respondents that used the TSAP, one responded to a series of follow-up questions while the other participant did not. One of the two users of the TSAP indicated that a modified rubric was the most useful format, and that they assessed the following performance criteria: attack ball; received ball; lost ball; played ball; and successful shot.

5.4 Discussion

Prior to responding to the research sub-question, this cross-sectional, quantitative inquiry of PE teachers in Victorian secondary schools found that GS occupied the most curriculum time across Year Levels 7 - 10 for most participants. This supports other findings in Australia (Kirk, 2006; Perlman & Forrest, 2015; Williams, 2016) and England (Casey & Hastie, 2011; Ward & Griggs, 2011) that GS are entrenched in school PE programs. Further, IGS occupied the most curriculum time within the Focus area of GS, which is similarly supported in the PE literature (Gray et al., 2008; Thorpe et al., 1984). As the study did not seek to determine the breadth of PE programs, it was not clear if participants held a traditionally myopic view that PE *was* sport (O'Connor et al., 2012; Pill, 2007a; Pill & Stolz, 2017). It is also not clear if limited facilities, equipment, or teacher interest and expertise led to less curriculum time for other Focus Areas (Challenge and Adventure

Activities, Lifelong Physical Activities and Rhythmic and Expressive Activities) (VCAA, n.d.-f). The prevalence of GS in participants' PE programs supported the view that "what has been taught 'in the name of PE' has changed little since the mid-twentieth century. In other words, PE during this period has emphasised skill development using team games and modified sports" (Williams, 2016, p. 230). These findings support the deep-rooted value of GS for many PE teachers (Green, 2002). The following sections 5.4.1 through to 5.4.4 discuss the four categories: Curriculum Alignment to Assessment; Assessment Tools; Users and Utility; Key Performance Criteria; and Familiarity and use of the GPAI and TSAP.

5.4.1 Curriculum Alignment to Assessment

Findings in category one included that some participants in government and Catholic schools did not adhere to the mandate to follow the VC (VCAA, n.d.-j). The actual number of participants in breach of this mandate may be greater than reported, as five participants from government and Catholic school sectors failed to indicate which curriculum they followed. In cases where participants from these sectors indicated their use of the AC: HPE rather than the VC: HPE, there may be minimal differences in what is being taught and assessed, as there is considerable alignment between the Achievement Standards, Strands, Sub-strands, Content Descriptors and Elaborations in both curricula (ACARA, n.d.-c; VCAA, n.d.-j). A further five respondents in government and Catholic sectors indicated their use of a school-developed curriculum so that their adherence to the VC: HPE cannot be determined. It is not clear what advice the Early Years–10 Curriculum and Assessment Committee provide to the VCAA on "the monitoring and reporting of student participation and performance in Early Years–Year 10 assessment programs" (VCAA, n.d.-a), to tackle this issue of curriculum non-compliance.

As it was not an aim of the study to compare group data based on curricula or other demographic attributes, the following discussion of curriculum refers to the entire cohort with an awareness that multiple curricula were used. Approximately half of all respondents reported some alignment between their curriculum and assessment. It is not clear how closely teachers followed their curricula, nor if assessment was based on criteria outside of the curriculum. The assessment of non-performative aspects of PE such as effort, participation, punctuality, attendance, uniform and bringing equipment, are widely reported nationally (Georgakis et al., 2015; Williams et al., 2020) and in the USA (for example, Baghurst, 2014; Young, 2011). While PE teachers have considerably more autonomy regarding assessment than teachers of numeracy and literacy in

Australia (DinanThompson & Penney, 2015), it is not clear what reference point PE teachers use for assessment if their practice is not informed by a common curriculum.

In terms of implementing a PE curriculum the “realisation of progressive intentions embedded in official curriculum texts is far from assured” (Lambert & Penney, 2020, p. 378). This view suggested that what is described in curricula may not be taught or assessed within and across school PE programs. This finding has been reported in PE curricula in countries including Australia, Canada, England, and Ireland (Georgakis et al., 2015; Herold, 2020; Kilborn et al., 2016; MacPhail & Murphy, 2017). As an example of the challenges in implementing a curriculum, a sample of secondary PE teachers in England were critical of guidance for teaching and assessment provided in a new national curriculum (Herold, 2020). This study revealed that there was no substantive change to what teachers delivered, and that a single two-line assessment statement in the curriculum was seen as unhelpful (Herold, 2020). The challenge of implementing a PE curriculum may contribute to the pervasive reporting of subjective assessment practices that are based on elements outside of any official discourse (Annerstedt & Larsson, 2010; Svennberg, 2017; Svennberg et al., 2014). Any failure to align assessment to a curriculum may also contribute to the widely reported practice of grading PE students based on a gut feeling (Annerstedt & Larsson, 2010; Hay & Macdonald, 2008; Svennberg et al., 2014). It is possible that the limited direction for assessment within national and state curricula texts, may have contributed to the lack of curriculum alignment to assessment reported by half of the respondents in the current study.

In an Australian example, researchers found that despite the AC being mandated in secondary schools within the ACT, some secondary PE teachers had trouble implementing this curriculum because it did not describe sport-specific skills (Williams et al., 2020). It is possible that this lack of specificity contributed to PE teachers assessing students subjectively based on criteria that were not grounded in the AC: HPE (Williams et al., 2020). There may be a similar lack of specificity within the VC: HPE at Year Levels 7 – 10, demonstrated by the absence of references to common sport skills like catch, throw and score. Instead references are made to “travelling, marking and intercepting to achieve and retain possession” and “experimenting with the manipulation of force and speed applied to an object to examine the difference created in movement paths” (VCAA, n.d.-d). It has also been observed in the context of instructional alignment that “if the chosen curriculum is not one that is challenging, exciting, and meaningful to learners then, no matter how well aligned it is, and despite learning taking place, how much application it will have for young people and their lifestyles is questionable” (MacPhail et al., 2023,

p. 154). In summary, curriculum alignment to assessment in IGS within the sample population was modest.

5.4.2 *Assessment Tools, Users and Utility*

Key findings in this category included that teachers were the greatest users of assessment, and that rubrics were used more widely and seen as more useful than frequency-count tools. Of these findings, the prevalence of rubrics contrasted with findings from the ScR in Chapter 4 that reported the prevalence of assessment tools based on frequency-counts. In locating 15 unique evidence-based assessment tools developed in a school context, only two tools were identified as rubrics (see Appendix E). The limited use of evidence-based frequency-count assessment tools by the sample population may pose issues for the validity and reliability of their assessment; a finding that has been reported in similar Australian studies (Georgakis et al., 2015; Williams et al., 2020).

5.4.2.1 *Assessment tools*

In a survey investigating assessment quality in secondary school PE in the Netherlands, assessment *modes* were reported rather than assessment *tools* as termed in this study. In the former study, teacher observation was identified as the most prevalent assessment mode, but it was not clear what tools were used when observing student performance (Borghouts et al., 2016). As the Netherlands survey investigated PE assessment broadly, it is not surprising that evidence-based tools like the GPAI and the TSAP were not mentioned, but considering the findings of the current study, it is noteworthy that the use of rubrics was not reported anywhere.

Endorsing the utility of rubrics, half of all respondents identified rubrics as being the *most* useful assessment tool in IGS. The use of rubrics in the extant PE assessment literature includes contexts such as personal and social learning (Gibbons & Robinson, 2004), FMS (Mohnsen, 1998), generic sport skills (Hensley, 1997), specific sport skills (Chen et al., 2016), badminton (Casebolt & Zhang, 2020), basketball (Shaw, 2014), flag football (Robinson & Melnychuk, 2009), a generic invasion game (Harvey, 2007), racket games (Harvey & van der Mars, 2010) and tag rugby (Harvey & Hughes, 2009). Support for the pervasive use of rubrics within Australian secondary school PE teachers in the context of GS assessment has also been reported in a study in the ACT (Williams et al., 2020). In that study, the authors identified rubrics as more prevalent than observation, peer assessment, and self-assessment (Williams et al., 2020). While a body of evidence into rubric efficacy and validity continues to grow, much of the corpus draws on populations in higher education and the skill of writing (Brookhart & Chen, 2015; Panadero & Jonsson, 2013). This

means that studies in the area of secondary school PE, including IGS, are under-represented or absent in many literature reviews of rubrics (Brookhart, 2018; Brookhart & Chen, 2015; Panadero & Jonsson, 2013, 2020).

As the second most useful assessment tool reported by the sample population, rating scales indicate the extent to which a criterion behaviour has been met (Lund & Veal, 2013). However, unlike rubrics, rating scales do not provide a description of performance quality (Brookhart, 2018). Rating scale language can be based on numerical scales (for example, 1 to 5), evaluative scales (for example, excellent to poor), and frequency scales (for example, consistently to rarely) (Brookhart, 2018). There is evidence in a review of rubrics in higher education that some rating scales are inaccurately identified as rubrics, as they fail to provide descriptors of quality (Brookhart, 2018). As examples of this erroneous labelling of rubrics, the Observational Scoring Rubric (Williams & Rink, 2003) and the Standards Based Rubric (Penney et al., 2012) located in the ScR in Chapter 4 both use rating scale language. The former rubric uses evaluative scale language (for example, proficiently, good, poor) and frequency scale language (for example, consistently, usually, rarely) (Williams & Rink, 2003). In contrast, the latter rubric uses a numerical scale, as it does not provide any descriptors of performance across the numbered levels 0 to 5 (Penney et al., 2012).

As the third most useful assessment tool indicated by the sample population, checklists may be viewed as a single-level rating scale that allow users to indicate the presence or absence of an aspect of performance. Checklists have commonly been used to identify critical elements or criterion behaviours within FMS (Sgrò et al., 2013), and sport-specific skills (Lund & Veal, 2013; Martin et al., 2015). Through the observation of the sequencing and timing of body movements, checklists may help to diagnose timing or fluency issues with critical elements (Mitchell et al., 2013), or sub-routines in motor skills (Pinheiro, 1994). The use of checklists and rating scales by respondents in this study is also found in a widely cited Canadian investigation that reported that 80% of the assessment tools used by secondary school PE teachers were checklists and rating scales (Desrosiers et al., 1997). Of note, Desrosiers and colleagues (1997) did not report any use of rubrics, suggesting that rubrics are a relatively modern assessment phenomenon. This observation is consistent with a literature review of rubrics in performance assessment that indicated that less than 10% of the total studies were published prior to 1997 (Jonsson & Svingby, 2007). In the context of IGS, checklists are unlikely to describe the quality of work, capture the totality of game play, or provide detailed information on game performance. Despite the relatively high use of checklists by teachers in this study, checklists may be better suited to the assessment of sport-

specific skills in their capacity to indicate discrete elements of technique, rather than assess complex, dynamic and authentic game play (Lund & Veal, 2013; Martin et al., 2015).

The potential confusion between checklists, rating scales and rubrics observed by Brookhart (2018) was not investigated in this study after the pilot testing suggested these terms were widely used and universally understood. As such, it is unclear if participants' rubrics used scale language or provided descriptors of performance. It is suggested that the use of scale language in rating scales and checklists in PE is "simply too crude in showing that an improvement was actually made" (Darst, 1989, p. 7). The use of scale language in rubrics has been similarly criticised for its inability to show student improvement and direct student learning (Brookhart, 2018). Use of scale language has implications for assessment quality and utility, if the rubrics used by the sample population more accurately represent rating scales. With less than 10% of respondents in the sample indicating any use of frequency-count tools, and only one respondent identifying these approaches as being the most useful, the findings reported in the current study are incongruent with the use and inferred utility of frequency-count approaches reported in the ScR (Chapter 4).

5.4.2.2 Assessment users

Rates of peer and self-assessment were generally much lower than teacher-led assessment across the listed assessment tools. To tackle the widely reported view that assessment in PE is difficult due to large class sizes (Chng & Lund, 2018; Gibbons & Kankkonen, 2011; Melograno, 1997) and is time consuming (Braga & Liversedge, 2017; Gallo et al., 2006; Veloo & Md Ali, 2016), training students to complete self and peer assessment offers a potential solution. For example, peer assessment in primary school PE has been reported as complementing teacher feedback, promoting learning, and engaging non-performing students (Chng & Lund, 2018). In secondary school PE there are benefits reported for performers and assessors, by providing opportunities for assessment *for* learning and assessment *as* learning, respectively (Gibbons & Kankkonen, 2011). By engaging students in self and peer assessment, formative assessment practices are supported that include activating learners as instructional resources for one another and owners of their learning (Black & Wiliam, 2009; Wiliam & Leahy, 2015).

Originally developed in the context of higher education, sustainable assessment principles refer to the development of self-assessment practices (Boud, 2000; Boud & Soler, 2016). Sustainable assessment practices are underpinned by the formative purpose of assessment and the aim of developing students' self-regulation skills (Boud, 2000; Boud & Soler, 2016). Sustainable

assessment may also support student learning beyond an educational setting, when success criteria may lack transparency and judgment might not be externally provided (Boud, 2000; Boud & Soler, 2016). Sustainable assessment challenges the unilateral application of assessment by teachers, instead advocating for a bilateral arrangement between teachers and students to empower learners to take greater responsibility for their assessment and/or learning (Boud & Soler, 2016). In applying sustainable assessment practices, assessment may be seen as being more time efficient, which is discussed more fully below.

5.4.2.3 Assessment utility

To determine assessment utility *independent* of the assessment tool, participants were provided with the following prompt and six options that were described as being desirable in the literature review of PE assessment (Chapter 2).

When assessing an individual student's performance in invasion games and sports it is useful if the tool:

1. is time efficient to use
2. can be completed electronically
3. can be completed by student peers
4. is linked to the curriculum
5. suggests the next steps in learning
6. accurately indicates performance level

When combining responses for strongly agree and agree for each item, there was minimal difference in most items. However, when viewing participant responses based on the strongly agreed option, almost three-quarters of participants strongly agreed that time efficiency was useful for assessment, and completing assessment electronically ranked second in the same response category. The latter finding regarding the value of electronic assessment supports the view that "keeping hundreds of pieces of paper on students is not feasible, [but] tracking progress electronically is" (SHAPE, 2014, p. 102). In combining the respondents' views on the importance of time efficiency and electronic data collection, efficacious assessment practice was inferred as paramount by participants.

While the nature of the electronic data capture and/or management was not stipulated in the question item, mobile smart devices that capture audio-visual data (Robinson & Randall, 2017)

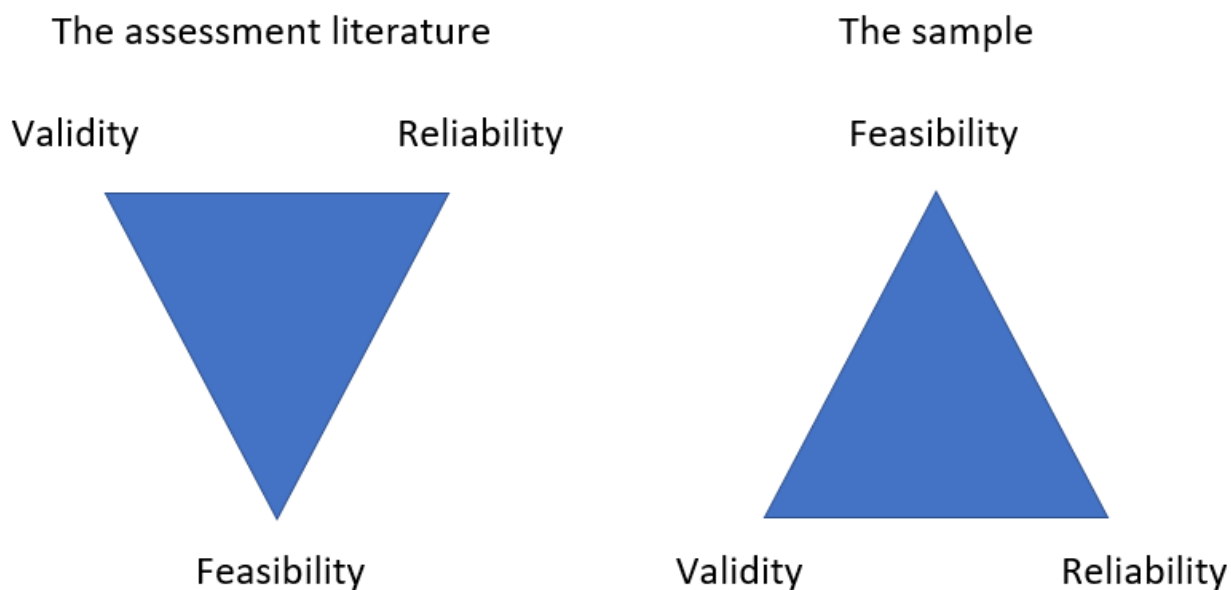
and applications like Dartfish™ that ‘tag’ (mark and track) student performance on iPads (Shaw, 2015) may be useful for teachers. In a Western Australian senior PE context, the use of digital performance-based assessment was viewed as an authentic approach by students and its piloting met technical, pedagogic, manageability and functional viability measures (Penney et al., 2012). However, embracing digital technology may challenge existing teacher beliefs, is likely to take time to develop expertise, and may present issues with equity/access in the assessment of PE (Robinson & Randall, 2017).

Less than one-third of participants strongly agreed that assessment is useful if it ‘accurately indicates performance level’, positioning this as the lowest of the six desirable options. The naming of this item was an attempt to represent features of validity and reliability using a layman term (accurately). While participants may not have shared this understanding, it is possible that feasibility of assessment was viewed as more useful than validity and reliability by respondents. This interpretation presents a tension in the sample and the wider assessment literature, where validity and reliability are positioned as the two most useful indicators of classroom assessment quality (Brookhart, 2003, 2005). In support of the significance of validity, reliability and feasibility in assessment within the VC: HPE, all three aspects are used to endorse a recently developed suite of resources linked to the VCAA website. The Movement Assessment in Practice (MAP) platform includes a bank of evidence-based FMS tests to provide “Health and Physical Education teachers with valid, reliable and feasible movement skill assessments that can be incorporated into their teaching and learning program” (VCAA, n.d.-h).

A diagrammatic representation of the potential tension in the literature and the sample is provided in Figure 5.1. The relative utility of these three indicators of assessment quality is based on their superior (more useful) or inferior (less useful) positioning on the triangles. The first triangle represents the superior positioning of validity and reliability in the wider assessment literature (for example, Brookhart, 2003, 2005), which is also supported in the assessment of FMS, sport-specific skills and games (Currell & Jeukendrup, 2008; Eddy et al., 2020; Hulteen et al., 2020). In contrast, the second triangle is inverted to suggest that feasibility *may* be more useful for the sample, based on their responses to the question on assessment utility. This interpretation has implications for the findings of the ScR (Chapter 4) given that validity and reliability informed the eligibility criteria for the studies. Accordingly, the 15 school-based tools located in the ScR may benefit from an examination of their feasibility to support their use by the sample.

Figure 5.1

The Relative Usefulness of Indicators of Assessment Quality



Note. The terms validity, reliability and feasibility are identified as indicators of assessment information quality by Brookhart (2005). The first triangle illustrates that validity and reliability are more useful than feasibility in the assessment literature; in contrast, the second triangle indicates that feasibility may be more useful than validity and reliability, as inferred by responses from the sample.

5.4.3 Key Performance Criteria

Key findings in this category included that most participants assessed the key performance criteria located in the ScR (Chapter 4), and the criterion of teamwork that was derived from the state and national curricula documents (for example, VCAA, n.d.-d). In contrast, creativity as exemplified in the content descriptor to “create solutions to movement challenges” (VCAA, n.d.-d), was the only criterion used by less than half of the respondents. Regarding the key on-the-ball skills located in Chapter 4, most items were well supported, with scoring and tackling the only named skills that were assessed by less than half of all respondents. As questionnaire items describing criteria were supported in the literature or curriculum documents, the widely reported assessment of non-performative criteria in PE remain unknown for the current population (for example, Blomqvist et al., 2005; Borghouts et al., 2016). As an example of the historical approach to assessing PE, almost 50 years ago Morrow (1978) identified absence, attitude, effort and showering as grading criteria in high school PE programs in the United States. Thus, it is possible that a range of inappropriate criteria were assessed in addition to the empirically supported options reported in this study.

5.4.3.1 Decision-making

It is interesting that decision-making was the most used assessment criterion while creativity was the least used. It is possible that the modest use of creativity by participants is due to the myriad of definitions of creativity within sport (Fardilha & Allen, 2019), which may lead to confusion with the term. For participants with a clearer understanding of creativity, it may be that pen and paper tasks associated with the assessment of creativity in sports may be seen as lacking authenticity (Fardilha & Allen, 2019). When viewing decision-making as synonymous to tactical play in GS, there appeared to be differences in the assessment focus of secondary school PE teachers in Victoria and the ACT. In ranking themes based on frequency-counts of language used in focus group interviews, the ACT study by Williams and colleagues found that skills (with 72 frequencies) was ranked the number one theme, while tactics (with 2 frequencies) was ranked the last of seven themes (Williams et al., 2020). This contrasts with the findings of this study that indicated decision-making was the most assessed aspect of game performance. Given the similarities in the achievement standards, content descriptors and elaborations in the AC: HPE (ACARA, n.d.-b) and VC: HPE (VCAA, n.d.-d), this difference cannot be explained in the respective curricula.

5.4.3.2 Off-the-ball movement

A literature review of assessment for tactical learning in all game categories posited that player roles can be generally classified as attacking (team with the ball) and defending (team without the ball) (Barquero-Ruiz et al., 2019). Consistent with this understanding, defensive play and attacking play were assessed identically by respondents when affirmative statements were aggregated. This finding suggested that teachers valued both phases of play equally, which contrasts with other literature reviews that have reported higher rates of assessment for attacking play (Arias & Castejón, 2012; Barquero-Ruiz et al., 2019). To capture the totality of game play, PE teachers and researchers are encouraged to assess players in attacking and defending roles (Barquero-Ruiz et al., 2019). Consistent with the assessment of players in attacking and defensive roles, a high rate of assessing players off-the ball was also reported in this study. The assessment of this criterion by participants recognised the large amount of time players spend off-the-ball in GS (Oslin et al., 1998).

5.4.3.3 On-the-ball skills

Participants assessed skill execution based on technique, and skill execution based on the result or outcome, at similar rates. This suggested that participants supported the distinction

between skill-execution and performance described in the TGfU model (Bunker & Thorpe, 1986). Whether this support was linked directly to their use of TGfU pedagogy remains unclear. It is also difficult to reconcile how teachers assessed students that demonstrated contradictory performances in technical execution and skill outcomes. For example, it is unclear how teachers assessed a student that demonstrated technical proficiency in passing to teammates and shooting at goal in basketball, but rarely completed passes or scored goals. Equally, it is not clear how PE teachers assessed a student that completed passes to teammates or scored goals in basketball through immature or unsound technique.

Another key issue regarding skill execution based on an outcome is the pass-catch dynamic that has been described as complicated, interactive, and co-dependent (MacPhail et al., 2008). In the current study, passing was the most assessed on-the-ball skill, while receiving was assessed by almost three-quarters of respondents. The interdependence between passing and catching can be seen in the concept of the *catchable* pass which has been defined as a “pass between receiver's knees to just above the head, arm's length to side with the appropriate force/ touch on the pass” (Nevett et al., 2001, p. 356). This definition requires subjective judgement on the assessor's behalf and considerable understanding of each player's catching ability. This level of understanding may be difficult for teachers with a new class, and for students completing peer assessment. Given the complexity in assessing individual performance in the relational skills of catching and passing, MacPhail et al. (2008) has argued that “perhaps the most appropriate unit of analysis in games programs should be the game itself ... [which] may have radical implications for how physical educators make judgments and record the learning progressions of their students” (MacPhail et al., 2008, p. 113).

The implications of the above are significant in school contexts, however the notion of assessing the whole team or a playing unit (for example, the mid-field), has been advocated elsewhere in the assessment of game performance literature (Barquero-Ruiz et al., 2019). The authors of the latter study reported that game assessment instruments should generally assess a range of tactical levels comprising: the *match level* that represented the team; the *partial forefront level* that represented a sub-set of the team; and the *primary level* that represented the actions of a single player (Barquero-Ruiz et al., 2019). It is uncertain how this multiple-level approach to assessment might be received by the population of Victron PE teachers.

In order to allow for a comparison of the on-the-ball skills assessed by participants in this study, and those identified in assessment tools from the ScR (Chapter 4), Table 5.10 was constructed. The parallel presentation of data collected from different research methods is consistent with the overarching research design and a key benefit of mixed methods approaches (Creswell & Plano Clark, 2018).

Table 5.10

On-The-Ball Skills Assessed in the Completed Studies of the Thesis

On-the-ball skills	Scoping Review (Chapter 4)		Cross-sectional quantitative inquiry (Chapter 5)	
	Tools	%	Participants	%
Pass	11	73.3	62/70	88.6
Receive	11	73.3	50/70	71.4
Score	10	66.7	34/70	48.6
Dribble	9	60.0	57/70	81.4
Defend	7	46.7	various	55.0*

Note. Only the 15 tools designed for use in the school setting from the ScR are reported. *The term defend was used to represent several skills in the ScR, in the relevant item in this study the skill of defend comprised: intercept/steal; block an attacking play; clear the ball; and tackle. Responses were reported as an average of user rates.

Table 5.10 highlights the participants' use of evidence-based on-the-ball skills at mostly similar rates to those found in the school-developed tools located in the ScR. A discrepancy in the rate of scoring across both studies is perhaps the most unexpected finding. Given a key aim of IGS is to outscore an opponent, it is not clear why this skill was assessed by less than half of the sample. It is possible that respondents may have assessed students in games like netball, where the principles of play preclude all players from scoring, or assessed students in full-sided games using adult sized fields which may have reduced scoring opportunities. Given that scoring "and offensive success is a key motivational aspect of games playing" (Mitchell et al., 2013, p. 7), its modest level of use among the cohort is an interesting finding.

As points of convergence in the assessment of key performance criteria in the two completed studies in this thesis, decision-making was assessed in 11 of the 15 school-developed assessment tools in the ScR (73.3%), compared to 97.2% of respondents in this study. Further, off-

the-ball movement was assessed in 12 of the school-developed assessment tools in the ScR (80.0%), compared to 83.8% of the sample in this study. In addition to the similarities in the assessment of on-the-ball skills described in Table 5.10, these findings suggest that the sample of PE teachers in Victorian secondary schools assess students consistent with the empirical evidence provided in the ScR (Chapter 4). A key difference exists in the generation of performance outcomes across both studies that is described with specific reference to the GPAI and the TSAP below.

5.4.4 *Familiarity and use of the GPAI and TSAP*

As the final category in this discussion, participants' level of awareness and use of the two most frequently employed tools in the assessment of tactics in PE and youth sports in the literature was considered (Arias & Castejón, 2012). Given the lack of familiarity and low level of use of both tools by participants, this category is relatively modest in length and substance. To help orient the reader to the specifics of each tool an outline is presented in Appendix I.

The low rates of use and familiarity with the GPAI and TSAP supports similar findings within Australia (Georgakis et al., 2015; Williams et al., 2020) and international reviews of assessment in PE and games (for example, Barquero-Ruiz et al., 2019; López-Pastor et al., 2013). In comparing the use of these tools in the current sample with secondary school PE teachers in other Australian states and territories, Williams et al. (2020) reported that no teachers from a sample of 19 used either tool, while Georgakis et al. (2015) reported that eight of 15 teachers used at least one of these tools (50.3%). The user rates within this sample of Victorian teachers sit in between the other Australian studies, with five of 72 respondents indicating their use of the GPAI (6.9%) and two of 71 respondents indicating their use of the TSAP (2.8%). Within the three studies, there is no explanation provided for the varying rates of use across states and territories. In contrast, for the small number of participants that were familiar with the tools and did not use them, the time taken to administer the assessments and their complexity were offered as common reasons for non-use.

As the AC: HPE and VC: HPE are both standards-based curricula, underpinned by a learning continuum that describes levels of quality, frequency-count tools that collect quantitative data and generate indices may be considered less appropriate to reflect their curricula outcomes. As assessment describes *how well* a curriculum has been learnt (Black & Wiliam, 2018), it is possible that respondents considered rubrics to be more useful than frequency-count tools in serving that

purpose. Given the pervasive association between assessment and reporting in the wider field of education (Hollingsworth et al., 2019) and the PE assessment literature (Moura et al., 2021; Veal, 1988), frequency-count tools may lack transferability to reporting achievement and progress in school-based PE. A potential middle-ground may be to adapt the criteria of the GPAI and/or the TSAP into a rubric format for less complicated and more expedient assessment. While lacking validation, a generic invasion game rubric based on the GPAI has been suggested by Harvey (2007).

5.5 Recommendations

The following series of recommendations fulfil the pragmatic aim of the thesis to produce actionable knowledge (Kelly & Cordeiro, 2020). These recommendations are provided for policy makers and practitioners and are intended to improve assessment practice in IGS for the sample population. As most participants in this study indicated their use of the VC: HPE, and less than half the cohort agreed that curriculum informed their assessment, the following curriculum recommendations are directed towards the designers of the VC: HPE.

First, the Early Years–10 Curriculum and Assessment Committee that provide advice to the VCAA on monitoring and reporting of student involvement in the VC (including the Learning Area of HPE), should consider auditing schools for curriculum aligned assessment compliance. Such auditing occurs in other programs, including the Victorian Certificate of Education (VCE) that comprises senior school Year Levels 11 to 12, and is described below:

As part of the VCAA's ongoing monitoring and quality assurance program for the VCE, assessment tasks for school-based assessment in each VCE study and scored VCE Vocational Education and Training (VET) program can be requested for audit from schools. The VCAA's audit of school-based assessment is conducted in line with the powers set out in sections 2.5.3 and 2.5.5 of the Education and Training Reform Act and the requirements set out in section 4.1.2 of the VRQA [Victorian Registration and Qualifications Authority] Guidelines and Standards for the Registration of Awarding Bodies and the Accreditation of Senior Secondary Qualifications (VCAA, n.d.-a).

Second, the VCAA is encouraged to provide further assessment support for PE teachers that aligns with their curriculum. Specific to the subject of PE, the VCAA has provided helpful online resources to develop formative assessment rubrics in IGS for students Levels 7 - 10 (VCAA, n.d.-g)

and a range of online modules to support rubric design more broadly (VCAA, 2019a). It is not clear if participants were familiar with these resources, or to what extent any of these materials supported participant understanding and practice of assessment. It is noteworthy that the VCAA provides the following caveat for the online modules that indicates their tacit support for contemporary research like that being undertaken in this thesis:

These materials were prepared in 2019. Please note that this area of research is evolving fast, therefore these materials should be supported with additional evidence bases that more accurately reflect best practice after 2024. It is therefore recommended that these materials be used with consideration of updated research after this date. (VCAA, 2019a)

Third, as part of providing assessment support the VCAA is encouraged to make use of the language and key performance criteria reported in the ScR (Chapter 4). Specifically, reference to the terms 'on-the-ball skills', 'off-the-ball movement' and 'decision-making' is advocated. In addition, the five game skills identified in the ScR and widely supported in this study (receive, dribble/carry, pass, score and defend) could also be utilised. These terms may be most helpful in the 'elaborations' of the VC: HPE, or assessment support material. Embedding this language in curriculum and assessment material may support greater consistency in assessment and address criticism that curriculum documents lack reference to specific sports and skills (Williams et al., 2020).

Regarding assessment tools and users, participants are advised by the Victorian DET to use a range of assessments *for*, *as* and *of* student learning (DET, 2021a). As rubrics, verbal responses, checklists, skill tests, rating scales and written tests/assignments were widely used, participants are encouraged to continue to use a range of assessments in their practice. Given the limited use of self-assessment reported by teachers in this study, assessment *as* learning may be underutilised by the sample. Activating students as agents for their own learning, and their peers, is widely supported in the literature (Dann, 2014; Earl, 2014; Gibbons & Kankkonen, 2011; Wiliam & Leahy, 2015). As such, participants are encouraged to introduce and scaffold the use of assessment tools for peer and self-assessment in their PE classes.

As recommendations regarding the seminal assessment tools of the GPAL and TSAP, it is possible that professional learning opportunities may develop greater awareness and support for

their implementation. In the study of Williams et al. (2020), a professional learning session supporting the use of the GPAI was a consequence of their study; it was widely attended and well received by participants. It is possible that similar professional learning opportunities may be provided by the researcher post thesis to support Victorian teachers. In addition to the GPAI and TSAP, other school-developed tools located in the ScR (Chapter 4) (see Appendix E), may be worthy of inclusion in any professional learning sessions. Consistent with advice in the performance analysis and PE assessment literature, these sessions would provide teacher support in modifying and simplifying these existing tools (Brewer & Jones, 2002; Fernandes et al., 2019).

5.6 Strengths and Limitations

A key strength of this work was that the results and discussion sections have shed light on a largely overlooked group and context in the literature, the assessment of IGS by PE teachers in Victorian secondary schools. In providing insights where none previously existed, participant responses indicated modest use of curriculum to inform their assessment, that students played a limited role in their assessment, and that teachers made greater use of rubrics than frequency-count tools. While the approach to generating performance-based outcomes differed between the sample and most tools located in the ScR (Chapter 4), the key performance criteria were consistent across both studies.

A limitation of any self-administered questionnaire is social desirability bias (Ary et al., 2014; Whittle et al., 2017b). That is, the responses of participants may not constitute the reality in the field and instead, an idealistic approach to assessment may be reported. While the letter of introduction stated that assessment was a long-standing challenge for PE teachers (López-Pastor et al., 2013), it is not clear if this helped to normalise any self-identified deficits in the assessment practice of participants. Future research is encouraged in the form of extended field observations to support or challenge the findings generated by the questionnaire.

Regarding the representativeness of the sample, it is possible that teachers with a low level of self-efficacy in their assessment practice/assessment literacy may have opted out of the questionnaire, leading to a skewed data set. In the broader context of grading in an educational setting, it has been found that survey return rates by teachers with poor grading practices were lower than teachers with good grading practices (Thomas et al., 2015). Second, with a maximum sample of 80 teacher participants per question item, the response rate was under the estimated sample size of 91. This may be partly attributed to one of the recruitment organisations modifying

the original participant invitation and breaking the link to the online questionnaire upon the initial recruitment in November 2019. While this was rectified within days, it is not clear how many participants attempted to complete the questionnaire at this time and did not return to the revised invitation. Lastly, with the impact of COVID-19 in March 2020, and the substantial impact of lock-downs across Melbourne (Tuffield, 2021), there was a significant extension to the original timeline for the third recruitment that may have hindered participant uptake.

5.7 Conclusion and Further Research

This chapter outlined the method for the design of a bespoke questionnaire and presented the results of a cross-sectional, quantitative inquiry. The four categories in the discussion described and made meaning of the understanding and practice of assessment by PE teachers in Victorian secondary schools in the context of IGS in Year Levels 7 - 10. The categories informing the questionnaire design were determined a priori after reviewing the assessment in PE literature (Chapter 2) and completing a ScR of evidence-based assessment tools (Chapter 4). In doing so, this chapter addressed a gap in the literature by furthering understanding of assessment practice in the subject area of PE.

Potential areas of future inquiry include examining how PE teachers align their assessment of GS to a curriculum, identifying the characteristics of assessment tools used in this context, and exploring how students can be more involved in peer and self-assessment. Regarding the key performance criteria, examining how teachers understand the various terms is warranted. Notwithstanding the complexity and potentially time-consuming nature of the GPAI and TSAP, as reported by users in this study, inviting practicing PE teachers to adapt these tools is advised. As most participants used rubrics, and rubrics were rated as the most useful assessment tool, an examination of the construction and use of rubrics in IGS assessment is advocated. As with the authors of the cross-sectional, quantitative inquiry of assessment practice in PE in the Netherlands (Borghouts et al., 2016), this researcher also acknowledges the need for more in-depth analysis of qualitative data from secondary school PE teachers. To address this need, the following chapter draws on the findings of the current study to describe the design and implementation of a series of interviews employing a nested sample of participants from the same teacher population.

CHAPTER 6: PHYSICAL EDUCATION TEACHERS' ASSESSMENT PRACTICES IN INVASION GAMES AND SPORTS IN VICTORIAN SECONDARY SCHOOLS: A QUALITATIVE INQUIRY

6.1 Introduction

The cross-sectional, quantitative inquiry in Chapter 5 provided an overview of the understanding and practice of PE teachers in Victorian secondary schools in their assessment of IGS in Year Levels 7 - 10. Within the sample, it remains unclear what assessment frameworks are followed, the purpose(s) of assessment, the level of student involvement in assessment, and the specific design of assessment tools. Further, given their low level of familiarity, it is not known how practitioners might understand the strengths and limitations of seminal assessment tools (GPAI and TSAP), or implement them in their practice. Cognisant of the interpretive imperative to investigate one core concept, at the heart of this inquiry was a consideration of participant understandings of assessment (O'Donoghue & Farrelly, 2022).

This chapter describes the analytical methods that supported the construction of four substantive themes to make meaning of participant understandings. It provides recommendations for practitioners before acknowledging the study's strengths, potential limitations, and possible areas for future research. As no similar qualitative inquiries investigating the assessment practices of secondary school PE teachers in an IGS context within Victoria were located in the assessment in PE literature review (Chapter 2), this study constitutes an original contribution to knowledge. This qualitative inquiry also fills a gap in the literature by allowing for comparisons of assessment practices in GS across other Australian states and territories (Georgakis et al., 2015; Williams et al., 2020). These comparisons may provide useful information to researchers, curriculum designers and secondary school PE teachers, given that each state and Territory in Australia is responsible for implementing or adapting the AC (ACARA, n.d.-d).

6.1.1 Aim

Consistent with the purpose of an explanatory, sequential, mixed methods study design (Creswell & Plano Clark, 2018), the aim of this qualitative study employing a series of semi-structured interviews was to explain the quantitative findings of the ScR (Chapter 4) and the cross-sectional, quantitative inquiry (Chapter 5). Within a pragmatic paradigm and an interpretive theoretical framework, a series of individual, online semi-structured interviews were selected to

meet this aim. The chapter presents four constructed themes to answer research sub-question 3: *How do Physical Education teachers practice and/or want to practice assessment of invasion games and sports in Victorian secondary schools?* To answer the research sub-question, the key foci that were determined a priori comprised participant understanding and practice of: (a) assessment processes including links to curriculum; (b) assessment strategies including links to feedback and reporting; (c) AfL or formative assessment; (d) rubrics and frequency-count tools.

6.2 Method

The population for this study comprised a sub-set of the PE teachers that completed the cross-sectional, quantitative inquiry (Chapter 5). Given the emergent research design of the thesis, the questions that comprised the interview schedule were informed by the two previous studies and the literature review of assessment in PE (Chapter 2). As the third of four studies comprising the explanatory, sequential, mixed methods design (Creswell & Creswell, 2018; Schoonenboom & Johnson, 2017), this study is also positioned as the second phase of a two-phase survey design. The survey design in this chapter focusses on qualitative data collected through semi-structured interviews to explain and make meaning of the preceding quantitative data collected through a questionnaire described in Chapter 5.

6.2.1 Participants and Sample

This qualitative inquiry employed a series of individual online semi-structured interviews from a nested purposive sample ($N = 8$) (Onwuegbuzie & Collins, 2007). Typical of sequential mixed methods studies (Tanner, 2023), the sample was recruited from questionnaire respondents that volunteered for further involvement in the study ($n = 22/80$). An advantage of this nested sampling approach was that participants were familiar with the broader topic of assessment and interested in discussing their views further. To explore a wide range of participant views, a heterogeneous sample of participants with a range of teaching experience from different school sectors was preferred over a homogenous sample (Onwuegbuzie & Collins, 2007). Table 6.1 provides participant demographics collected from the cross-sectional, quantitative inquiry and lists participants in the current study alphabetically by gendered pseudonym.

Table 6.1*Participant Characteristics*

Pseudonym	Gender	PE degree	Years of teaching	School category	Year Levels taught (PE)	Curriculum taught	Position of responsibility
Amy	Female	Yes	Between 5/ less than 10	Government	7, 8	VC: HPE	Head of PE
Betty	Female	Yes	Between 10/ less than 15	Catholic	9, 10	VC: HPE	NA
Chris	Male	Yes	More than 20	Government	7, 8, 9, 10	School developed*	Head of PE
Ed	Male	Yes	Between 5/ less than 10	Independent	9	VC: HPE	Head of Outdoor education
Frank	Male	Yes	Between 10/ less than 15	Independent	7, 9	VC: HPE	Head of PE
Greg	Male	Yes	More than 20	Government	7, 8, 9	AC: HPE*	Senior leader
Harry	Male	Yes	Between 15/ less than 20	Government	7, 8, 10	School developed*	Head of PE
Ian	Male	Yes	Less than 5	Government	9, 10	VC: HPE	Assistant Head of PE

Note. The three participants marked with an asterisk (*) indicated they used the above curriculum in the cross-sectional, quantitative inquiry presented in Chapter 5, however, in the qualitative inquiry each participant indicated they used the VC: HPE. Their use of the VC: HPE is consistent with the mandate from the Victorian Department of Education and Training for teachers in government and Catholic schools (2021b). Originally these three participants were recruited to provide a breadth of views on curricula, however their responses in this study allowed for focussed consideration of the VC: HPE, as 100.0% of participants in this study referred to this curriculum. To avoid participants being able to select multiple years of experience options, the terms less than and more than were used.

The sample described in Table 6.1 was considered heterogenous, as it comprised one to two participants in each of the five categories reflecting years of teaching experience. Participant representation in school sectors was also commensurate with demographic information provided in the cross-sectional, quantitative inquiry (Chapter 5), with teachers from the government sector featuring prominently ($n = 5$), and teachers from the Catholic sector ($n = 1$) the least prominent. While participant use and awareness of the GPAI and TSAP was identified in the questionnaire, the number of users of both tools was too low to select the sample based on users and non-users of these tools as originally planned.

6.2.2 *Ethics*

Consistent with the standards of Social and Behavioural Research Ethics Committee (SBREC) at Flinders University, all eight participants from the cross-sectional, quantitative inquiry were sent an invitation via email having previously indicated their willingness to be involved in further study. This invitation was accompanied by an information sheet and consent form for their consideration (Appendix K). Each participant included in the interview study returned the signed form via email to provide their informed consent. In the interview transcripts anonymity was supported by using pseudonyms and de-identifying any information including school names, intranet systems that suggested a particular school sector, and the names of any colleagues. Participant privacy and confidentiality was addressed by storing audio-recorded data on a password protected Flinders university network in a de-identified format. Copies of the above documentation can be found in Appendix K.

6.2.3 *Data Collection*

The use of individual interviews in qualitative research in the context of sport and exercise is common, with a recent literature review in the above context reporting that 85% of studies used this method ($N = 1941$) (McGannon et al., 2021). Further, the choice of individual interviews for data collection in this study was consistent with qualitative research within the explanatory, sequential, mixed methods study design, in its aim to explain and make meaning of findings from the previous quantitative inquiry (Creswell & Plano Clark, 2018). The purpose of this study was to provide a rich and thick description (Kervin et al., 2016; Roy et al., 2015) of the complex phenomenon of assessment. To achieve this purpose, a nested, heterogeneous sample of participants from the first

phase of the survey (Chapter 5) were interviewed. Where the cross-sectional, quantitative inquiry focussed on *what* participants understood by assessment and the nature of the assessment they used, this qualitative inquiry focussed on *how* participants understood and practiced assessment.

International individual interview studies have examined Swedish secondary school PE teachers' grading practices to reveal the arbitrary awarding of grades and assessment practices that lack transparency, validity and reliability (Annerstedt & Larsson, 2010). In a Spanish primary and secondary school setting, PE teachers reported that successful implementation of AfL and authentic assessment required planning and structuring assessment as part of their instruction (Barrientos Hernán et al., 2022). Within Australia, individual interviews studies of secondary school PE teachers in the state of NSW indicated modest use of the TSAP and GPAI in the assessment of student performance in GS (Georgakis et al., 2015). Consistent with these studies, the current qualitative inquiry also adopted a semi-structured approach to the interviews. This was seen as a middle ground between a rigid, structured interview approach, that does not allow deviation from the interview schedule, and the unstructured interview approach that does not use predetermined questions (Ary et al., 2014; Kervin et al., 2016). The semi-structured interview approach was selected to provide a guide or flexible framework that was responsive enough to shift in line with participant responses (Braun & Clarke, 2013). In this approach, the use of probes, as complementary questions, were used to elicit additional information according to the responses that participants gave (Kervin et al., 2016). The use of probes was consistent with the interpretive framework of the thesis (O'Donoghue & Farrelly, 2022) as it allowed respondents the opportunity to provide a rationale for their assessment practice.

Due to the impact of the COVID-19 global pandemic, and specifically the enforcement of numerous lockdowns (where residents could not leave their homes) and social distancing restrictions (Tuffield, 2021), the interviews were held online rather than face-to-face, through the ZOOM™ platform. Despite a traditional view that virtual interviews are a poor substitute for face-to-face interviews (Braun & Clarke, 2013), there are those that advocate for the use of online conference platforms (Archibald et al., 2019; Gray et al., 2020). Benefits of an online approach to interviews, including use of the ZOOM™ platform, have included the ease of use by researchers and participants, and the sense of empowerment for the participant (Backman & Pearson, 2016; Gray et al., 2020). Regarding the latter aspect, participants in online interviews through ZOOM™ have reported high

levels of autonomy, based on their ability to choose the venue of the interview and leave the interview if they chose to do so (Gray et al., 2020). I used ZOOM™ as the software platform for the interviews in this study and agree with Gray et al. (2020) that the greatest practical advantage of an online approach was accessibility to participants. Conversely, there were also technical issues that included periodic drops in sound quality and occasional lags in the live feed that have been reported elsewhere (Archibald et al. , 2019).

Consistent with the emergent design of the mixed methods approach of the thesis (Schoonenboom & Johnson, 2017), a draft interview schedule was based on the literature review of assessment in PE (Chapter 2), findings from the ScR (Chapter 4), and findings from the cross-sectional, quantitative inquiry (Chapter 5). Examples of findings in the latter inquiry that informed the interviews included: less than half of the participants indicated they aligned their assessment to a curriculum; the majority of participants indicated that student assessment was useful, yet the use of self and peer assessment was largely under-utilised; and while rubrics were the most widely used assessment tool, approximately half the respondents indicated a different assessment tool as being the *most* useful. Questions investigating these findings are representative of the mixing of data that led to the final interview schedule (see Appendix H).

This drafting of the interview schedule was done in consultation with my supervisors via online meetings and email. As with the design of the questionnaire described in Chapter 5, interview questions were piloted to clarify questions and address any misunderstandings. In keeping with general interview protocols described by Ary et al. (2014), the participants were familiar with the topic and from the target population. Two experienced PE teachers piloted the interview schedule and reported that the questions were straight forward, logically ordered and conversational rather than confronting. I recorded the pilot interviews via ZOOM™ in September 2020 and then transcribed them verbatim within a week of the interview taking place. The recordings and the transcriptions were then shared with the supervisory team for feedback. From the feedback, I was encouraged to listen more, talk less, and to use probes more diligently. The drafting of the schedule following the piloting involved the removal of questions investigating how teachers defined and assessed specific components of game play, like decision-making, to instead focus on broader issues of assessment processes, strategies and utility.

The four areas of inquiry in the interview schedule were described as *topics* to help distinguish them from the *themes* that were constructed within the findings and discussion section of this chapter (Braun & Clarke, 2021a). These topics comprised open-ended questions investigating participant understanding and practice of assessment comprising: (a) assessment processes including links to curriculum; (b) assessment strategies including links to feedback and reporting; (c) AfL or formative assessment; and (d) rubrics and frequency-count tools. Specifically, topic 1 of the interview schedule asked participants to define assessment and describe any links between assessment and curriculum. Topic 2 addressed assessment strategies including the collection of assessment data, the generation of feedback, and any use of assessment moderation. Topic 3 explored participant AfL practices, including any involvement of students in the assessment process. Topic 4 investigated participant use of rubrics and participant understanding of the strengths and limitations of the GPAI and TSAP. Given the low level of familiarity and use of these tools identified in the previous study, a brief synopsis of both tools was sent to the participants as pre-reading to support meaningful engagement with relevant questions (see Appendix I).

The online interviews ($n = 8$) were held in my home office from early December 2020 to March 2021 according to participant availability either side of the summer break (late December to late January). In recognising the need to build rapport and establish a non-threatening environment (Braun & Clarke, 2013; Kervin et al., 2016), at the outset I introduced myself as a PE teacher and the nature of my interest in the study. In explaining my position as an insider-outsider (Bukamal, 2022), I was able to engage in authentic professional conversation based on our shared experience as PE teachers. To avoid “doing expert” (Braun & Clarke, 2013, p. 96), by presenting myself as an accomplished practitioner of assessment, I informed participants that my motive in this study was to address a challenge that I faced in my teaching career, assessment of IGS. Participants were reminded that their involvement would be confidential, and their anonymity would be protected by a pseudonym in the write-up. Participants were told that the focus of the interview was to understand their assessment practice in IGS across Year Levels 7 - 10. To establish context, I listed a range of games, for example, basketball, soccer, football, rugby, handball, hockey and lacrosse.

As ice-breakers that also served to gather demographic information, each of the interviews opened with close-ended questions about the Year Levels participants taught during 2019, the

curriculum they followed, and any leadership role they may have held (see Appendix H). For example, participants were asked, “Which Year Levels (between 7 - 10) did you teach practical PE in 2019?” Following these introductory questions, the ZOOM™ session was recorded, and I asked the participants open-ended questions that linked to the four topics. An example of a main question from topic 1 was, “How do you define assessment?” When specific forms of assessment were offered within the definition, for example summative and formative, I used the participant’s terms throughout the interview. The adoption of participant language was a means of minimising any power imbalance in the interview dynamic, in order to make participants feel more comfortable (Kervin et al., 2016).

Probes were used throughout the interviews when participants were asked to explicate a response, often including the question “would anything else make X more useful?” The final question in the interview allowed participants to present ideas that may have been overlooked, with the following invitation “Is there anything else that you think could be useful when assessing students in invasion games and sports?” I engaged in an ongoing dialogue via email and ZOOM™ with two supervisors (SP, SE) throughout the series of interviews which served as a de-briefing of preliminary findings and helped me recognise researcher bias. Following the interviews, the ZOOM™ recordings were uploaded to a password protected Flinders university network in a de-identified format (using the participant’s pseudonym) before being deleted from my laptop. In total, the interviews produced 62,000 words of textual data.

6.2.4 Data Analysis

The following method for data analysis was based on a six-phase process for reflexive thematic analysis (Braun et al., 2019). This type of thematic analysis has been described as both an accessible and flexible interpretive approach for treating qualitative data (Braun & Clarke, 2012). An earlier version of this type of analysis is considered seminal (Braun & Clarke, 2006), however, the more recent iteration emphasises that themes are actively *constructed* by the researcher, rather than seen as passively *emerging* from the data (Braun & Clarke, 2019, 2021a; Braun et al., 2019). Reflexive thematic analysis is an alternative to *codebook* or *coding reliability* approaches that focus on the early establishment of codes, or group consensus in coding, respectively (Braun & Clarke, 2021a). As such, codes were identified organically in response to the data, and all data were coded independently.

Supporting my application of reflexive thematic analysis within an interpretive theoretical framework, a worked example of this approach within a similar population of teachers proved very helpful (Byrne, 2022).

In keeping with reflexive thematic analysis (Braun & Clarke, 2021a), preliminary data analysis began during the interview process as I asked questions and made brief notes. In the first phase of this process, “Data familiarisation and writing familiarisation notes” (Braun & Clarke, 2021a, p. 331), I conducted each of the eight interviews and transcribed them verbatim. To promote accuracy in transcription, the process was completed manually, and as close as possible to the time of recording. This process took me between six to eight hours for each participant, which is consistent with other time guides for transcription (Braun & Clarke, 2013). On several occasions, the audio broke up due to a weak internet connection which was indicated as buffering in the transcripts and has been reported as an issue with online interviews (Archibald et al., 2019). This process of manual transcription allowed me to immerse myself in the data; this involved pausing and replaying the audio and engaging in multiple readings of the transcription (Maguire & Delahunt, 2017). During the process of transcription, memos were used to help identify my assumptions, biases and questions discovered in the process of transcription (Braun & Clarke, 2013; Taylor, 2014). In addition to promoting researcher reflexivity, these memos also helped to establish an audit trail (Tracy, 2010). A sample of these memos is provided in Appendix J, providing further evidence of the detailed data collection and analysis process that may indicate *rich rigor* (Tracy, 2010).

The second phase of analysis, “Systematic data coding” (Braun & Clarke, 2021a, p. 331), involved making a choice between manual or electronic coding of data predicated on the size of the data-set, the time available, and the proclivities of the researcher (Basit, 2010). In recognising that qualitative data can be “unstructured, messy, intimidating, confusing and, at times, contradictory” (Taylor, 2014, p. 181), and given my preference for electronic storing and management of data, I imported all interview transcripts into the electronic software platform of NVivo9. This form of Computer-Assisted Qualitative Data Analysis Software (CAQDAS) has been used to organise transcripts and support reflexive thematic analysis in an Australian sporting context (Elliott et al., 2023). Through prolonged online training in NVivo9 (10 hours), I found that the software afforded me efficient data management and retrieval capability. Research designers, Maher et al. (2018), have

argued that traditional coding approaches using coloured pens, sticky notes and display boards *along* with NVivo9 software, provide greater kinaesthetic and visual engagement with data. However, I determined that over 60,000 words of transcription constituted a large data set (Maguire & Delahunt, 2017) and was more manageable when treated exclusively in NVivo9.

The software allowed me to highlight text and identify this text as separate codes. In generating codes, data was reduced into small chunks of meaning (Maguire & Delahunt, 2017). These codes were given labels, and provided with definitions and exemplar quotes to help establish their uniqueness. As codes were the building blocks of theme development, code terms were semantically driven, in that, they remained very close to the language used in the raw data (Braun et al., 2019). The final number of codes amassed 205 in total, with many preliminary codes split or combined by dragging and dropping them into the appropriate code or code branch to allow for further classification. Within NVivo9 this involved classifying each of the codes into various families and establishing a hierarchy in their overarching patterns of shared meaning. This coding approach was considered qualitative rather than quantitative in nature, with the latter approach described as content analysis underpinned by determining the frequency of occurrence in data (Vaismoradi et al., 2013). In my application of abductive coding within reflexive thematic analysis, the number of times an idea was presented was not as important as the substance of the idea in helping to construct themes that supported my understanding of the participants' assessment practice (Braun & Clarke, 2006).

In recognising that the phase approach of reflective thematic analysis is not a rigid process, and that the phases can blend together, the process of analysis became more recursive (Braun & Clarke, 2021a). This included replaying the video and audio of interview transcripts, reading earlier memos, reviewing the current codes, and re-engaging with literature on the chosen method. I employed an abductive approach to code the transcripts by oscillating between inductive and deductive analytical processes (Mitchell, 2018). The combination of generating codes from the data (an inductive approach), and from ideas, concepts and language identified in the literature (a deductive approach) (Braun et al., 2019) was consistent with the pragmatic paradigm of the thesis. The abductive approach acknowledged that my coding did not occur in "a theoretical vacuum" (Braun & Clarke, 2021a, p. 331) and that I was unable to distance myself from prior knowledge that supported coding data to respond to the research sub-question (Byrne, 2022). Two aspects of the research process supported my use of

abduction in this qualitative inquiry (Earl Rinehart, 2020). First, I spent considerable time immersed in the data (Earl Rinehart, 2020) with the first interview occurring in December 2020 and the writing/drafting process extending into 2023. Throughout this period, I regularly returned to the transcripts and my memos to review and refine my coding in NVivo9. Second, having time away from the data throughout this period allowed off-task prompts and influences to bring fresh ideas to my understanding and interpretation (Earl Rinehart, 2020).

In the third phase of analysis, “Generating initial themes from coded and collated data” (Braun & Clarke, 2021a, p. 331), I aimed to identify similarities, differences, overlaps, and clusters of ideas. Early attempts to generate themes led to the tentative identification of candidate themes. To support my independent coding of transcripts, I regularly consulted with two supervisors with a combined 30 years of qualitative research and reflexive thematic analysis experience (SE and SP). This involved email and ZOOM™ meetings to discuss portions of my coding and the evolving construction of themes. It was not an exercise to find consensus, but to ‘sense-check’ ideas, consider alternatives and locate blind spots (Byrne, 2022).

In phase four, “Developing and reviewing themes” and phase five “Refining, defining and naming themes” (Braun & Clarke, 2021a, p. 331), I again worked iteratively. This involved shifting between the data set, memos, and the research sub-question to hone and clarify what constituted a theme. This development and refinement of themes was based on sorting many, one-dimensional codes into a series of more complex, multi-faceted and abstract themes with a central organising concept (Braun & Clarke, 2021a). The process was supported by understanding that there was no correct number of themes, and that there need only be as many themes as required to tell a coherent, plausible and detailed story (Byrne, 2022). Theme naming and refining involved sharing various iterations of the developing codes and preliminary themes with my supervisory team to ask provocative and clarifying questions regarding my interpretation. The naming of themes across the drafting process presented in Table 6.2 became less superficial and increasingly more latent, meaning that the names became increasingly more conceptual rather than literal.

Table 6.2*Iterative Development of Themes*

Draft 1	Draft 2	Draft 3	Draft 4	Final theme names
Teachers 'have to' follow school policy	What school wants	Big picture	Assessment and accountability	The power of accountability in assessment
Teachers assess for reporting	Teachers assess to report	Curriculum, everything and nothing		
An 'airy fairy' curriculum	Curriculum, yes, no, maybe			
	Nuts and bolts	Perennial challenges of assessment	Unsustainable assessment practices	The problems with unilateral assessment practice
Assessment is done to students	The passive role of students in assessment	Students as passengers	Students as passengers in assessment	Students – the missing ingredient in assessment
Assessment done with, for and by students	A better tool kit for teachers	In a perfect world	Assessment ideologies	PE teachers, heal thyself – assessment ideologies

The iterative construction of latent themes may be viewed as evidence of my developing engagement and interpretation of the data (Braun et al., 2019). The final theme names captured patterned meaning across the data set and avoided a common challenge in reflexive thematic analysis of presenting domain summaries as themes (Braun & Clarke, 2021b). As an example of the construction of theme 4, 'PE teachers – heal thyself', some of the 54 codes comprising this theme included: feedback in a perfect model; the potential of rubrics; assess fewer students; and teach umpiring. The central organising concept that served as a working definition for this theme as documented in NVivo9 was as follows:

This theme captures all the hopes/dreams/potential for participants' improved assessment practice. Some ideas are obvious to participants, others remain hidden in plain sight. e.g. increase student umpiring, make self-peer assessment safe, use student friendly language in rubrics, use video exemplars, create a hybrid frequency-count-tool or rubric/frequency-count-tool, increase student voice/choice, assess slivers of game play at a time, greater links to curriculum, clarity around the 'standard', use video capture, assess outside of class time, use ball skills linked to TSAP, separate assess for decision-making, consider team work/fair play and creativity. It recognises an ideal or perfect world of assessment that was not part of their current reality.

Phase six of reflexive thematic analysis, "Writing the report" (Braun & Clarke, 2021a, p. 331), constituted the writing of this chapter. In keeping with qualitative work that applies pattern-based analysis through thematic analysis, the findings and discussion were integrated into a single section in this chapter (Braun & Clarke, 2013). This meant that interview excerpts were contextualised within the literature as they were reported, rather than addressing the literature in a separate discussion section (Braun & Clarke, 2013; Byrne, 2022; Terry et al., 2017). This integrative approach allowed for strong and cogent links between participant views and the extant literature on assessment in PE, avoided repetition of data, and supported the interpretive theoretical framework of the wider thesis (Braun & Clarke, 2013). The four themes were presented in a deliberate sequence to support a logical narrative that responded to the research sub-question (Byrne, 2022). This narrative began with addressing assessment challenges at a systemic level, before considering issues with teacher-led assessment that marginalised the potential role of students, and concluded with participant views on the ways and means to improve assessment practice. While the following analysis employed a mix of *illustrative* and *analytic* approaches to understand participant views (Terry et al., 2017), as most participant quotes were analysed with close reference to the literature on assessment, the findings and discussion section of the chapter is more consistent with an analytic approach (Braun & Clarke, 2013). This format allowed me to "weave together data, analysis, and connections to scholarly (and other) literature" (Terry et al., 2017, p. 31) to address the research sub-question. All data analysis was underpinned by an interpretive view that teachers *reported* their constructed understanding and practice of assessment, rather than provided an *absolute* truth (McChesney & Aldridge, 2019).

6.2.5 Rigor

The appraisal of social research quality often includes the core criteria of reliability, validity and replicability (Bryman, 2016). While these terms are generally associated with quantitative work, qualitative research is more often associated with the debated concept of rigor (Smith & McGannon, 2018). Within an allied sport and exercise context, some authors argue that there are no absolute, universal or pre-ordained criteria for qualitative inquiry, problematising any judgement of rigor for this study (Smith & McGannon, 2018; Sparkes & Smith, 2009). The position adopted in this thesis is that criteria are historically and contextually bound, making their universal application problematic (Halcomb, 2019). Instead, qualitative researchers embrace the subjectivity of the research process by selecting a set of criteria that supports their underlying philosophical assumptions (Elliott et al., 2023). The following consideration of rigor begins by addressing the conceptual depth of the study (Nelson, 2016), as an alternative to the contested notion of saturation (Hennink et al., 2017; Rowlands et al., 2015). It continues by describing the employment of several quality criteria identified in the seminal work of Tracy (2010) that includes: worthy topic; rich rigor; credibility; and sincerity (including reflexivity). As with all qualitative research, the reader is tasked with determining the degree of methodological rigor in this study (Elliott et al., 2023).

Saturation is a nebulous aspect of qualitative research that includes conceptualisations of theoretical saturation (Rowlands et al., 2015), data/thematic saturation (Guest et al., 2006), code saturation and meaning saturation (Hennink et al., 2017). It is typically associated with a *stage* in qualitative research of “information redundancy” (Braun & Clarke, 2021b, p. 201), where no new information emerges or when no further codes, themes or theory are developed (Low, 2019). This notion appears to be built upon a “logical fallacy” (Low, 2019, p. 131), as theoretical insights may continue for as long as data are collected. In the case of researchers returning to reanalyse data, it is impossible *not* to be able to discover something new (Braun & Clarke, 2021b). Given the fluid and organic coding process of reflexive thematic analysis “codes are never fully fixed” (Braun & Clarke, 2021b, p. 207) as they develop, shrink, expand or are re-named, and hence the concept of data saturation is incompatible with reflexive thematic analysis. Ultimately, codes represent my interpretation of trends and patterns across the data set (Byrne, 2022), and these interpretations may continue to evolve for as long as data are analysed.

As there was “no magical number” (Low, 2019, p. 135) of interview participants, I established a provisional lower and upper range of participants that might potentially yield an adequate data set. Consistent with cautious guidelines in another qualitative inquiry (Guest et al., 2006), it was hypothesised that between six and 12 participants would provide a rich and complex story regarding the phenomenon of assessment. After eight interviews I made an “*in situ* decision” (Braun & Clarke, 2021b, p. 211) that the data quality, diversity and complexity was adequate to answer the research sub-question. This decision was predicated on most codes in transcripts interview seven and eight being classified within the existing 200 plus codes. Further, participant views in these interviews supported patterns that were well established in memos and candidate themes. This number of participants was consistent with the interpretive aim to understand the complex phenomenon of assessment, rather than to test hypotheses or make generalisable findings to a total population that would require a larger number of participants (O’Donoghue & Farrelly, 2022). In purposively selecting participants across the three school categories, with a range of experience including the Year Levels taught, the sample had a ‘richness’ that demonstrated qualitative integrity and rigor. Ultimately the level of acceptance of these qualitative criteria, and those that follow, rest with the reader (Braun & Clarke, 2013; Roy et al., 2015).

To avoid “perpetuating the myth of data saturation” (Braun & Clarke, 2021b, p. 212), that theoretical insights have an end point in qualitative inquiry (Low, 2019), conceptual depth as articulated by Nelson (2016) is used to describe the degree of completeness of this inquiry. Conceptual depth or density, as an alternative to saturation, provides several criteria that allow the adequacy of the data in qualitative research to be judged. As with other alternatives to saturation like item salience (Weller et al., 2018) and information power (Malterud et al., 2016), conceptual depth criteria can support researchers in determining if their conceptual categories, or themes in reflexive thematic analysis, are sufficiently robust (Nelson, 2016). Notwithstanding that there is no widely agreed set of criteria for qualitative inquiry (Smith & McGannon, 2018; Sparkes & Smith, 2009), the following reflexive thematic analysis demonstrates a high level of conceptual density in responding to the criteria of range, complexity, subtlety, resonance and validity (Nelson, 2016). The numerous participant excerpts covering four substantive themes indicates a *range* of concepts and the detailed coding network that moved progressively from concrete to abstract demonstrated its inherent *complexity*. In employing a latent approach to the construction of themes, the findings and discussion

section of this chapter demonstrates *subtlety* in making meaning of participants' nuanced views of assessment. In linking participant views to the extant literature, the analysis demonstrated a high level of *resonance* with historical and contemporary views on assessment in PE. Lastly, given the findings support much of the wider assessment in PE literature, the recommendations may have external *validity*, making them applicable beyond the immediate population (Nelson, 2016). As such, the findings and recommendations of this study may resonate with the lived experience of educators beyond the sample and potentially support their transferability and generalisability (Smith, 2017; Tracy, 2010). It is possible that the use of PE as “the medium ... to highlight the principles of instructional alignment” (Scanlon et al., 2023, p. 5), may support the generalisability of the findings to a wider audience.

In employing several quality criteria as articulated by Tracy (2010), the nature of the investigation and the research sub-question is considered a *worthy topic*. Given the prevalence of GS in PE curricula nationally (Kirk, 2006; Perlman & Forrest, 2015; Williams, 2016) and internationally (for example, Casey & Hastie, 2011; Gray et al., 2008; Ward & Griggs, 2011), the content area is likely to be important beyond the sample. In focussing on assessment utility, challenges faced by PE teachers that include: balancing assessment across multiple learning domains (cognitive, affective and psychomotor) (Hay & Penney, 2013; Mitchell et al., 2013); implementing a curriculum to meet learning objectives (Georgakis et al., 2015; Lambert & Penney, 2020; Williams et al., 2020); and limited use of evidence-based assessment tools in the assessment of GS (López-Pastor et al., 2013; Williams et al., 2020; Young, 2011), suggest that the topic is worthwhile.

In demonstrating *rich rigor*, the sample ($n = 8$) was considered heterogeneous as it included PE teachers with differing experiences based on years of teaching, school category, and Year Levels taught. The common use of the VC: HPE, while not part of the original selection strategy, allowed for a richer and fuller consideration of the VC: HPE, which may be particularly useful for the sample and their colleagues in the state of Victoria. The liberal use of participant views in the following findings and discussion section of the chapter, selected from over 60,000 words of verbatim transcripts, illustrates an abundance of data. These extracts are closely linked to the extant literature on assessment in PE including authentic assessment (Gulikers et al., 2004) and quality assessment

frameworks (Borghouts et al., 2016; Chappuis et al., 2012). The data collection and analysis process provides further evidence of *rich rigor* within this study (Tracy, 2010).

As the researcher is the data collection instrument in qualitative studies (Dodgson, 2019; Kervin et al., 2016; Maguire & Delahunt, 2017), and any series of interviews are highly contextual, my background, positionality and motivation were outlined in Chapter 1. My positioning, combined with detailed participant demographics and transparency in the methods employed in this study, help to address issues of credibility and sincerity. It is my subjectivity or “humanness” (Braun & Clarke, 2013, p. 36) that helped shape this qualitative research. To support the credibility of the study, thick description and multivocality were demonstrated in the richness of the quotes and variety of the participant views offered in the findings and discussion section of this chapter. This involved drawing on my supervisors, as more experienced qualitative researchers, to review transcriptions, coding, quote selection, and thematic construction. This was not an exercise to determine inter-rater reliability or triangulate findings (Smith, 2017), but to engage their support as critical friends (Carlson et al., 2018; Deuchar, 2008) and raise awareness of potential blind spots and support my self-reflexivity within the quality criterion of *sincerity* (Tracy, 2010).

Underpinning the role of a critical friend, two supervisors (SE and SP) posed provocative questions, reviewed data and generally critiqued my work (Carlson et al., 2018). This involved an ongoing dialogue via emails and ZOOM™ meetings throughout the development of the interview schedule, the data transcription and code/theme generation. This dialogue promoted consideration of alternative explanations and interpretations that might be plausible (Smith, 2017), but ultimately, the final reflexive thematic analysis that follows is based on my interpretation of the data set. The considered reflection of experienced researchers was considered a “prized source of knowledge and understanding” (Willis, 2012, p. 16) that supported my interpretive positioning in this study. As further support of researcher reflexivity, the prolific use of memo writing within NVivo9 provided an ongoing internal dialogue for developing my understanding of findings (see Appendix J). As reflexivity involved critical consideration of the knowledge constructed and the researcher’s role in its production (Braun & Clarke, 2013), this allowed me to continue to challenge my understanding of the complex phenomenon of assessment as described by participants.

6.3 Findings and Discussion

A total of 205 codes led to the construction of four substantive themes that served to answer the research sub-question. The four themes are considered “informative, and memorable” (Byrne, 2022, p. 1408) and described by punchy or evocative titles. The themes are coherent and distinct from one another (Maguire & Delahunt, 2017) and serve to capture key patterns in the data set (Braun & Clarke, 2006). The themes are deliberately ordered to create a narrative (Byrne, 2022) that show the potentially hierarchical relationship between systemic influences, individual practice, the marginalisation of students and potential solutions. The themes also increase in their complexity and substance, culminating in an expansive fourth theme that provides a wide range of suggestions to improve assessment practice. The following sections 6.3.1 through to 6.3.4 consider the four themes: The Power of Accountability in Assessment; The Problems with Unilateral Assessment Practice; Students – the Missing Ingredient in Assessment; and PE teachers, Heal Thyself – Assessment Ideologies.

6.3.1 *The Power of Accountability in Assessment*

To help establish a common understanding of assessment, interviewees were initially asked for their definition of assessment. The naming conventions used by each participant were then used by the interviewer to minimise any power imbalance (Kervin et al., 2016). Responses to this introductory question supported the view of Masters (2014), that the field of assessment is ‘over-complicated’ as terms like diagnostic, formative, summative, progress, growth, learning, performance and achievement were all used by participants. Despite the disparate conceptualisations of assessment by respondents, a common ground was found when assessment was described as being helpful in forming an understanding of where students were in their learning. As an outlier, Chris described the etymology of assessment meaning:

To sit beside, which I thought was quite an interesting concept. Uh, the idea of sort of looking in the same direction together, opposed to sitting opposite them and putting out a challenge for them to meet. So, sitting side by side, and working through a task or a challenge is probably how I look at it, assessment. However, that ideology and the reality don't always line up.

In recognising the incongruence of ideal and the real-world practices, Chris encapsulated a pervasive notion that participant assessment philosophies and ideologies were quite distinct. This division between real and imagined assessment practices led to the construction of final theme of this chapter.

After establishing a shared understanding of assessment, participant views on assessment within their school and the role of curriculum in informing assessment were considered. This theme supported the strong links between assessment and reporting in PE (Moura et al., 2021) and the impact of school mandates to report student achievement via rubrics. It was accompanied by the counterpoint, that there was not always the same sense of obligation or accountability for participants to align their assessment to the VC: HPE as mandated by the state. This theme is deliberately presented first as it captures the sociological perspective of assessment in PE (Hay & Penney, 2013) that recognises that assessment practices “Do not operate independently of, or unaffected by, the systemic environments in which they are situated” (2013, p. 2). This perspective recognises that assessment operates within the various social, cultural and political contexts of schools (Hay & Penney, 2013). Thus, any changes in assessment practice must address these contexts before any substantive change can be made in the field. The essence of the theme was the juxtaposition of accountability between two different organisations on respondents. From a sociocultural perspective, most participants felt bound to follow school mandates, but not necessarily the state-wide mandate to implement and assess the VC: HPE in government and Catholic schools (VCAA, n.d.-j).

6.3.1.1 School mandates

While the literature suggested that PE teachers have great autonomy in their assessment practice (MacPhail & Murphy, 2017), many participants in this study spoke of *reporting* rather than *assessment* and felt constrained by their school’s reporting policies. Typical of this view was Frank’s comment that reporting was merely an exercise in “making the school happy”. In Amy’s case, school-wide reporting mandates meant that “we have to give a grade” based on a standardised five-level rubric. The rationale for this rigid, school-wide approach was to make it easier for parents to understand student achievement, but this was an impediment to Amy’s desire to adapt her rubrics because:

That's not what our school wants ... It doesn't match up with our school, so you're trying to scale that back. For our school's purposes, to be common with whatever everyone else was doing, so it's easy for parents and what not, that's been our biggest challenge.

The pervasiveness of assessment being linked to reporting for a parent audience was encapsulated in the following statement from Greg:

Oh, assessment is twofold, um, obviously, you want to get an idea of where the student is, where they're at, where they're coming from, I guess, so you might do pre-test, post-test ... and it's also important we use it to report back to parents and, you know ... obviously we report back to parents regularly with reports and, you know, interim, um, progress report, reports.

Highlighting the resentment felt by some participants towards the accountability of school-wide reporting policies and the inability to assess or report on learning was captured by Frank:

We're just assessing them on whatever they've previous skills are so that's not really giving them, it's not giving us any idea of how well we're teaching, and it's not giving them any idea of how well they're learning, so we're sort of, we're just doing it to tick a box to make the school happy really.

Further, there was a strong suggestion that the emphasis on reporting subjugated efforts to develop ongoing assessment or AfL during the learning cycle as reported by Ed:

I think my plan is to try to use a range of different tools and assessment types to keep that variety ... So, I think probably GPAI ... movement patterns maybe ... and then maybe a TSAP... and I guess here's the trick, we still need to do rubrics so, you know ... you kind of have to do that. Therefore, if you have that and GPAI and videoing, do you assess on all of them? Do you write a comment on all of them? Is that then too much time you're using for assessment? So, I think that's one big challenge. We, like, the school wants us to do a rubric for every unit ... Therefore, you end up with a lot of assessment and a lot of time talking and showing, and yeah.

A rigid five-level rubric format that supported the awarding of letter grades was also seen as constraining by some; and may be contrary to the view that rubrics should only contain as many levels as can be accurately discerned by users (Brookhart, 2018). Teachers in higher education have reported the requirement to work within assessment parameters established by more senior colleagues as a barrier to modifying their assessment practice (Deneen & Boud, 2014); a finding that may apply equally to secondary school teachers. As such, in conforming to school reporting policy imposed by senior leaders it is possible that strategies for AfL may have been negatively affected (Deneen & Boud, 2014).

6.3.1.2 State mandates

In contrast to the accountability participants expressed regarding school reporting mandates, there were mixed views towards the requirement to align their assessment/reporting to the VC: HPE. Representative of views that indicated adherence to the VC: HPE, Amy stated that:

The Vic Curric informs everything. So, the 5 criteria that we assess against the invasion games is straight from the Vic curriculum ... when we rewrote our units we had the Vic curriculum on the board and was almost using it as a checklist, like what can go in this unit, what can go in this unit. To make sure that we ticked everything off.

Others expressed the need for greater clarity in curriculum documents to support their alignment of assessment. As an example, Harry reported that the language of the VC: HPE was:

Very airy fairy [colloquial for unclear] ... the content descriptors, so you know, being able to manipulate a skill, they won't say being able to pass in-step in soccer, and then the next level of the continuum is outside of your foot in soccer at a volley level. You know, they're not, it's not that descriptive.

Other participants indicated the VC: HPE played a limited role in their assessment practice, for example Chris felt his department's approach to assessment was "fairly casual" before suggesting a reason for not following the curriculum closely:

Well, I guess we don't really have a pressure to have to do that. I guess in my time like we've seen a number of different frameworks or educational models for, you know, what the best

approach is and what sort of content we should be covering. So, I think you know after a while people, they realize that we're playing the same games and probably fall into those traps of let's just go and do this activity and give them a score at the end. Most people seem to be fairly happy with that from PE.

When Chris was asked if he could provide examples of specific references to the VC: HPE in his assessment he responded, "Unfortunately, not (laughs)". Chris's lack of accountability in implementing a curriculum contrasted with the wider adherence of most participants to support their schools' reporting policies. The difficulty of implementing a curriculum and ergo, aligning assessment to curriculum, is well established in educational literature including the learning area of PE (Herold, 2020; Lambert & Penney, 2020; Williams et al., 2020). Historically, it has been reported that PE "Curriculum Guides ... are perhaps seen by teachers as non-practical for they do lack the necessary procedural referents. To this end, as an innovation they have had very limited use in curriculum practice" (Tinning, 1980, p. 247). For some of the participants, this view remains as relevant today as it did more than 40 years ago.

While most participants spoke at length about their use of valid criteria based on performance in games as described by the VC: HPE, some also indicated the assessment of other criteria including: effort; participation; punctuality; attendance; uniform; and bringing equipment. While effort and participation underpin involvement in PE, the other criteria support findings of the pervasive assessment of poorly aligned criteria reported nationally (Georgakis et al., 2015; Williams et al., 2020) and internationally (Baghurst, 2014; Morrow, 1978; Veal, 1988; Young, 2011). Unlike the auditing of the VCE program, there was no suggestion by any participant that the VCAA, or any other regulatory body, had ever audited their teaching and learning programs at Year Levels 7-10, nor had the VCAA ever requested any student assessments or reports.

Contrary to the requirements of a standards-based curriculum to assess and report student achievement described in the curriculum, several participants mistook norm-based assessment for the standards-based assessment required of the VC: HPE. As an example, Harry commented that he deliberately imposed a bell-shaped curve when reporting to counteract that fact that "PE teachers ... are pretty nice" and several staff would put all students at the standard, which Harry felt should never

happen. Ian also applied a normative approach to reporting that was in direct conflict with the state mandate to assesses students against the achievement standards based on merit:

I generally have a pretty good bell curve...I actually deliberately try to get that because if in a, in a class of 25 students. Um ... I would have, you know, maybe two to three that might be just beyond, um, I might have 10, you know, thereabouts at standard and then your, the remaining probably working towards below.

This misunderstanding of the standards-based framework of the VC: HPE, along with participants' strong alignment of assessment with reporting, present a challenge for AfL and the provision of meaningful feedback for students. Specifically, students may not get adequate assessment information to help them improve their performance and/or any information they do get may not accurately reflect the achievement standards they should be assessed against.

6.3.2 The Problems with Unilateral Assessment Practice

Most participants acknowledged that they did most of the assessment in their classes and that this was challenging. In recognising assessment practice was an individual pursuit led by teachers, this theme considered perennial problems for assessment in PE, the limited use of moderation, and the challenge of collecting and tracking assessment data. Overall, my inference was that in managing the assessment process as individuals, opportunities for participants to collaborate and innovate were missed.

6.3.2.1 Perennial problems

Commonly described barriers to quality assessment practice included traditional issues like large class sizes and time constraints that are well supported in the literature (Braga & Liversedge, 2017; Gallo et al., 2006; Veloo & Md Ali, 2016). As an example of the problem large class sizes presented for assessing student performance in IGS, Ed reflected:

I probably try a method that's not very sustainable, which is trying to get to as many students as possible, which at times, probably leads to less good feedback ... I know the whole idea of choosing five kids this lesson, five kids the next lesson. And honestly probably my organization could be better, I've got to, if I'm going to do that, I've got to write down which kids I've talked to.

The same practice of assessing an entire class in a lesson was also described by Greg with reference to the use of a rubric:

It sounds great, but in a game, if I'm watching, oh, thing is you've got to sit down with one of these in front of me, and watch every kid and quickly work out where they are. But, you know, sometimes in one lesson you don't pick it all up ... like sometimes in the game, some kids don't get involved in all these aspects that we're trying to assess, you know like, they might not get an opportunity to score, or you know, yeah to you know, do certain things.

As an anomaly, Ian spoke of assessing five students at a time which also included the admission that "There's probably occasions where I have, in terms of being time poor ... I will have made general observations in my mind". Ian's internalised assessment approach is well supported in the PE assessment literature where assessment based on a 'gut-feeling' is widely reported (Annerstedt & Larsson, 2010; Svennberg, 2017; Svennberg et al., 2014). It was inferred from this internalised assessment process that students were unaware that they were being assessed and did not receive any feedback.

Lack of time for assessment was another common barrier raised by participants, with many citing this as the reason for their limited use of formative assessment. While the limited use of formative assessment by secondary school PE teachers is reported elsewhere (van der Mars et al., 2018a; 2018b), the lack of ongoing formative assessment is contrary to the AIESEP position statement on assessment in PE (AIESEP, 2020). Specifically, their position statement contends that "purposeful learning in PE should always include (aspects of) AfL ... and aid in legitimising the subject-area within the educational system and society as a whole" (AIESEP, 2020, p. 8). Some participants spoke of formative assessment practices in other activities like gymnastics and dance, and in other game categories like net/court and striking/fielding games, suggesting that the use of AfL in PE was especially challenging in IGS. These challenges have been attributed to the complexity of tactics in IGS as opposed to other game categories (Werner et al., 1996) and the dynamic pace and lack of natural breaks in play that exist in other game categories (Mitchell et al., 2013).

It is possible that time management and embedding assessment into learning activities may have been a barrier to assessment, rather than having a lack of time. As an example, the following comment from Greg indicated the use of games in his teaching practice, but it was devoid of any consideration of assessment:

We just haven't got the time, you know, often, it's you know, get in, just get into a game, get into some skill development, game situation and you know, almost time to get changed, go and get changed. We don't have time to give that feedback back to the kids. The feedback generally comes, um, you know, report writing time, I guess.

6.3.2.2 A lack of moderation

Time constraints were also blamed for a lack of moderation of assessment in most participants' assessment practices. As such, through all stages of assessment, from design, to implementation, to evaluation, and moderation, teachers tended to act independently. In doing so, opportunities to question and improve assessment practice may have been overlooked. Amy summarised her department's lack of moderation by stating "We have their rubrics and we mark their rubrics, but there's no discussion with the teams about how other groups have gone, or what skills that we're lacking, or what we should be teaching better". The lack of moderation by participants supports findings in other Australian PE contexts (DinanThompson & Penney, 2015).

There were also indications that some participants lacked experience and skill, rather than time, to explain the absence of moderation in their practice. Initially, Harry admitted that he "wouldn't have a clue" how to do moderate assessment of IGS before offering a fuller response:

Is there a moderation process that English use for every assessment task, or Humanities or whatever it is? So that's actually a, a knowledge deficiency in my teaching area, is, other than, you know, yeah, I can cross mark someone else's test or assessment? Um, but is there a set procedure that we could do in that curriculum time after work? Because I think that's where it would be best to do it.

Frank commented that he had not "Thought about it [moderation] in a prac sense at all, to be honest", suggesting that Frank had the skill and the motivation to do this for his high stakes PE within the VCE, but not his low-stakes Year Levels 7 - 10 PE classes. As moderation of assessment informs

Standard 5.3 of the Australian Professional Standards for Teachers (AITSL, 2017a), these findings are of concern. At the lowest graduate standard, there is an expectation that teachers demonstrate an understanding of moderation and how it can be applied to promote consistent judgment of student learning (AITSL, 2017a).

6.3.2.3 Data management

A major problem in the unilateral approach to assessment through teacher-led assessment was the associated challenge of collecting and tracking data. When Amy was asked how she tracked student assessment she replied self-deprecatingly, “Not very well at all, badly if at all. I don't know if we do. We don't track student assessment”. As with many respondents, Amy's use of ‘we’ when asked about her practice indicated that participants often spoke on behalf of their team. Yet, there were also times when participants could not fully articulate how other colleagues approached aspects of assessment. Amy went on to acknowledge that “Every teacher is different and so this is where we don't really have a, I guess consistent approach”. She reflected that some colleagues would track data by making “a little note in their Chronicle” and others would use “post-its” and that these ad hoc approaches would not be done in every class. This type of data collection was mirrored by Frank in his admission that when he was assessing he would “just be sort of sitting back and, trying to take notes”.

Ed recognised the value of tracking assessment data and reflected that this was “Another area where I would like to improve and I would really like to do it in an app or on my iPad, but I'm still really old school with sheets”. Ultimately, Frank's response when asked about tracking assessment data was representative of other participants:

There's got to be a useful way. I don't, I don't know what it is though, because I don't have a way at the moment, so I've got nothing that I can really, use, but I mean there's always an app or something.

While participants were equally divided in their preference for hard copy and electronic collection/tracking of assessment data, it was clear that no participant had an approach that they were satisfied with. Part of this issue may have been the pervasive view that teachers had to be the sole managers of data collection and tracking, when students may have been asked to play a supporting role. Participants may have benefited from a more strategic approach to identify which

students, assessment tasks, assessment forms and assessors would be selected across their IGS unit. The general lack of data collection and management is counter to Standard 5.5 of the Australian Professional Standards for Teachers to make use of accurate and reliable records when reporting on student achievement (AITSL, 2017a). The lack of systemic formative assessment is also contrary to advice from the Victorian DET for schools to “ensure there is ongoing assessment of each student’s performance, and that this assessment is embedded in the school’s curriculum program” (DET, 2023c).

6.3.3 Students – the Missing Ingredient in Assessment

This theme was an adaptation of assessment being described as the “missing ingredient” within the three message systems of QPE and education (DinanThompson, 2013, p. 138). In this adaptation, students are seen as missing from their assessment, as teachers constructed the assessment tasks, selected the assessment tools, chose the assessment timing, and were the chief assessors. In contrast, students were often reported as being unaware of when they were being assessed, unclear about the assessment criteria, and often not provided with any feedback until their end-of-semester report. In capturing the essence of this theme, assessment was generally “performed on the learner” (Boud, 2000, p. 156).

6.3.3.1 Student agency

Developing student voice and assessment literacy are positioned as integral to twenty-first century lifelong learning (Charteris & Thomas, 2016; Tillema, 2014). Equally, providing students with some degree of autonomy in their learning and assessment may help meet students’ psychological needs (Weeldenburg et al., 2021). However, the assessment practices of the cohort indicated very limited levels of student involvement, voice, or autonomy regarding their assessment. The pervasive use of rubrics by participants was accompanied by acknowledging that students played no role in their design, and a limited role in determining the criteria they were assessed on. The omnipresence of rubrics across this study supported findings in a similar Australian investigation where rubrics were identified as the second most popular theme after skills (Williams et al., 2020). However, none of the participants indicated students contributed to rubric construction, which is supported in the wider literature on rubrics and student agency (Fraile et al., 2017; Zheng et al., 2019). As an anomalous example of student choice in assessment with reference to rubrics Amy indicated that:

Our students already choose a rubric and they can choose to learn a new skill, work as a team member, show leadership or like officiates like rules and tactics. Um, so they already get to choose the focus area, so ... giving students like ownership and agency over their assessment.

While giving students a choice demonstrated more student agency than any other participant, Amy's students at Year Levels 7 - 8 still had to be summatively assessed on a generic-analytical rubric, with five levels of quality and five equally weighted criteria. The students' actual choice was to select one of four different criteria that comprised 5 of the 25 marks within the school mandated rubric format. It is not clear to what extent these students perceived any sense of agency with this level of choice. Promoting student agency is a key driver in the creation of modern assessment models (Nieminen & Tuohilampi, 2020) and developing student voice in assessment and learning is supported in education (Tillema, 2014), including the learning area of PE (Lorente & Kirk, 2013).

6.3.3.2 The need for transparency

Epitomising the theme of students as the missing ingredient in assessment, some participants indicated that students were unaware they were being assessed. For example, Chris reflected that in his classes "we probably don't really articulate that we're assessing invasion games". As an outlier in this sample, Chris was the only participant that that did *not* use rubrics in his assessment of IGS. Some participants that did use rubrics indicated that the rubrics were only available on the school's Learning Management System (LMS) or intranet, and were not openly shown to students. As an example, Frank commented:

Um, so it would be at the end of the unit ... that's probably the first time the kids would see the rubric too. It's up on the portal, like it's up on the LMS, but none of them would look at it, and we don't explicitly show them.

Several others provided equivocal responses to student awareness of assessment results, demonstrated in the following statement from Betty:

Our students didn't really see the rubrics, um at all, we just put it on their [LMS] ... They see it. We show them what they're being assessed on, but they don't, they didn't see exactly what we circled. Or maybe they did? Sometimes they might have, sometimes they

might not have, because it might have been rushed. You know, things getting in the way of classes and stuff like that, so it might have been a bit rushed at the end to actually give them the rubrics back.

While Amy was confident her students were aware of the assessment criteria in their rubrics, she was not confident of their level of understanding of the criteria, noting “If they had to explicitly say what each stage would look like and what they should be able to do at each stage, I don't know if they'd be able to do that very well”. Low levels of student understanding of assessable criteria in PE is supported beyond this study (Chan et al., 2011; Redelius & Hay, 2012). Amy's view highlighted the difficulty in constructing descriptive (qualitative) criteria to help students understand what different levels of performance in IGS looked like.

6.3.3.3 Students as assessors

As a contributing factor to the above issue of students being unaware or unclear of assessment criteria, empowering students to act as assessors was not widely reported. Further, most participants did not make use of their rubrics for self-assessment which is endorsed in the broader field of education (for example, Andrade, 2007; Andrade & Valtcheva, 2009; Panadero & Romero, 2014). While clearly an underutilised practice, most participants saw value in self and peer assessment as indicated by Chris:

I think there is value in it. Again, that diagnostic capacity where kids are checking off, yes you can do this or yes you can do that, that sort of thing. I think is quite positive, especially for skill development and skill clarity, performance in particular, skills within a sport. I think that's very positive. But yeah, we're not there yet.

Amy responded positively that engaging students as assessors would provide them with “ownership” of their learning, which supported her willingness to develop student agency described earlier in this theme. Others felt that training students to self and peer-assess would take time away from their teaching. In acknowledging the need to train assessors, participants supported the PE assessment literature that indicates that students require scaffolding to provide specific and accurate feedback to peers (Butler & Hodge, 2001). Further, there is support in the literature that with training,

school-age team sport players can reliably assess their peers in games like soccer (Holt et al., 2012), and apply frequency-count tools like the TSAP (Richard et al., 2000).

In some cases, an ad hoc approach to peer assessment was described that made use of students that did not get changed into their PE uniform or presented with medical notes. As an example, Ian responded that:

If students are out of uniform for whether it be medical reasons, they just have forgotten to bring a uniform ... I'll have a, some sort of questionnaire or chart and they will select a student that they will be viewing for that lesson and they, there will be various questions on, rate this student from one to five on their communication. Um, you know skill execution ... and sort of give them some things to focus on for when they're viewing a peer.

In the above example, assessment is an alternative task for the non-active participant. The lack of rigor in this approach to assessment led Amy to conclude that the amount of peer assessment opportunities her students received was “not enough”. Amy continued to describe her use of peer assessment was:

Mainly for the students who are not changed. That would maybe be a job that they get given, and the guidelines are like whatever our success criteria is. These are, this is the two things or one thing that I want you to look for. Can you watch such and such for the next 5 minutes? Record them on the iPad and then have a discussion about whether they're, um, you know, showing that particular skill?

The findings in this study support similar research in Norway, where PE teachers of the equivalent of Year Level 8 and 10 students reported no systematic use of self or peer assessment strategies. Instead teachers indicated the occasional use of peer assessment while a student “sits on the sidelines” (Aarskog, 2020, p. 881). Peer assessment in GS in primary school PE can supplement teacher feedback; enhance learning; and engage observers as active participants (Chng & Lund, 2018). There is also support for peer assessment in secondary school PE, where students assessing other students provides “Assessment for learning for the performer and assessment as learning for the assessor” (Gibbons & Kankkonen, 2011, p. 10). Pragmatically, peer assessment can also help

overcome the logistical issues of large class sizes and only one PE teacher (Chng & Lund, 2018; Gibbons & Kankkonen, 2011; Melograno, 1997). The incidental assessment practices reported in the current study and supported in the literature suggest that students are likely to get inequitable access to peer assessment opportunities and whole-class scaffolding support to foster skilful peer assessment may also be overlooked.

In summary of the support for peer assessment in GS the developers of the GPAI suggest that:

Teachers can and should assess, but they should not do all the assessment, all the time ...

Students can also assess and should do so in various formative situations ... Peer

assessment adds value because students may accept feedback from one another that they might not take seriously from their teacher. (Mitchell et al., 2013, p. 250)

6.3.4 PE Teachers, Heal Thyself – Assessment Ideologies

In addressing the final theme, a key finding from the interviews was that participants felt a degree of frustration with aspects of their current assessment practice. In many cases teachers were able to articulate solutions, while in other cases, solutions remained elusive. This theme presents participant solutions to some of the challenges in assessing PE. These solutions included: building student capacity to manage games; prioritising individual feedback; embracing technology; narrowing the assessment task; and improving rubric design. The draft theme title, 'In a perfect world', represented that participants' idealised assessment practices were central to developing a response to the research sub-question.

6.3.4.1 Build student capacity to manage games

A common suggestion to alleviate time constraints, large class sizes, and attentional focus for teacher to assess, was to develop student capacity to manage themselves in games. Chris captured this sentiment when stating:

To look for a particular student and their role in an activity isn't always easy, especially when you're facilitating the game itself. So, in an ideal world you'd be able to sit back from the activity and actually go about the assessment. But sometimes you're just so busy running it and actually getting it up and happening and managing it as it's occurring that you don't really get effective time for that.

While Chris was unable to articulate a solution, Amy suggested “We really need to build up our students so that we didn't need to be running each game. We need those games to be running almost autonomously so that we can assess properly”. In developing student capacity to umpire themselves, much like many students do while playing games during recess and lunch periods at school, participants might be able to give greater attention to assessment. This approach would be consistent with curriculum statements in Year Levels 7 - 10 in the VC: HPE. For example, a relevant content descriptor at Year Levels 7 - 8 states that students should “Practise and apply personal and social skills when undertaking a range of roles in physical activities (VCHPEM139)” (VCAA, n.d.-b). Greg summed up the feelings of other participants when suggesting:

If I have to teach and assess at the same time it would be difficult, but if I, if the game or activity could be, if the kids can run it or, you know, be, be running the activity and I can stand up in the background just observing and assessing, it wouldn't be so time consuming, it wouldn't be such a drag.

6.3.4.2 Prioritise individual feedback

Given the limited use of formative assessment reported by participants, feedback frequency and timing were also described as being poor. The frequency of feedback ranged from “every lesson” (Amy) to “They get none, literally none” (Frank). Amy spoke for many when suggesting that most of her feedback was praise and “that probably doesn't mean enough now that I'm reflecting on it, a lot of well done, good job, nice pass”. Participants were also critical of their timing of student feedback that did not always allow students time to implement it. As an example, Frank felt that:

It's easy to see the class progress, I think, but it's harder to pinpoint 24 kids to say you are now doing this better than what you were, particularly in the model that we're running, where, in the model, a perfect model, every kid would know, I think, what they were sort of lacking early, and what little bits they've added throughout that year or semester or unit or whatever it is. Um, which, which our kids aren't getting certainly.

A common desire among participants was to give more regular personalised feedback to students linked to the use of AfL. This view was encapsulated by Ed when stating that he wanted to:

Find ways to embed it [assessment feedback] so that they do get, you know, preferably, you know, best case scenario, it's not just, I'm assessing, here's your end assessment or even here's me, like I'm assessing, here's one check in, here's your assessment. In an ideal world they will be having multiple check-ins and finding out ... it's just such a challenge, I think.

Despite the widespread support for greater provision of personalised feedback among participants, it has been reported that corrective feedback from PE teachers may limit student reflection and the associated opportunity to self-assess their learning (Aarskog, 2020). The implication by the authors of the latter study is that in addition to providing explicit corrective feedback, PE teachers also need to “provide space and encourage students to negotiate [understanding] between themselves” (p. 885). Within the sample in this chapter, there was little mention of providing students with this space or setting tasks that promoted reflection.

6.3.4.3 Embrace technology

Participants generally supported the use of video capture as another way to address the challenge of large class sizes and limited time to assess students in class. However, the actual use of video capture for assessment was limited due to conflicting school policies and the logistics of managing the equipment. As an example of these views, getting individual student footage during game play was considered both “ideal” and “tough” by Betty. The use of mobile devices and video capture are supported in the wider field of education (Franklin & Smith, 2015), including the learning area of PE (Weir & Connor, 2009), and more specifically within GS (Koekoek et al., 2018). While data security, student consent, and privacy must be considered when recording video footage, some school structures constrained participants’ efforts to develop their assessment practice, making the use of video capture problematic. As an example, Ed indicated that video capture was:

Something I really want to do. I really want to do, I looked at doing it at the start of the year. The main thing is that, like I have a school iPad that I could use and that would be great, but they're not supporting, there's a really good video program, can't remember what it was called ... it looked great, but they're not supporting it. They don't want us using it because the data is not secure. And I went over and try to get them to put it on my

tablet, they wouldn't do it, and then it just got to a point where I just, you know, alright, let's move on and look at other things.

A common suggestion to improve rubric utility, beyond the problematic nature of descriptors, was the use of video exemplars to help describe performance otherwise presented in text. Some participants discussed the value of using a portable television or projector to show video footage of performance linked to certain levels in their rubrics. Ian was an outlier that had already begun establishing a video data base of various performance levels in a Google Doc library. Further, Ian suggested a digital library of exemplars of game performance created by the VCAA could help Victorian PE teachers come to a common understanding of the Achievement Standard. Ian's request suggested he was not familiar with the Digital Assessment Library (DAL), a suite of online assessments to support PE teachers in their assessment of VC: HPE (VCAA, n.d.-c). It is not clear if other participants were familiar with this resource, or felt it lacked utility, as currently the DAL has no assessments aligned to GS as a Focus Area (VCAA, personal communication, August 15, 2022).

6.3.4.4 Narrow the assessment task

Some participants felt that their assessment tasks were too broad and did not always elicit evidence of their intended assessment criteria. As an example of broad assessment tasks, some participants recognised that assessing both attack and defence phases in game play was difficult, but not all of these participants recognised the possibility of assessing a single phase of play. Assessing repeat sets in a single phase of game play to reduce the observable criteria was successfully employed in studies involving the GPAI (Harvey et al., 2010) and the System for Notational Soccer Analysis (van Maarseveen et al., 2017), and may be useful for participants in this study. While simplifying the task was one solution to observing assessment criteria, Ed offered another solution by suggesting he could focus on fewer elements in his rubrics in a single lesson because:

Individually talking to each kid and saying this is one row, one skill that you can work on within the rubric and I think you can improve that by doing X, Y and Z. That's the ideal world, I think.

6.3.4.5 Improve rubric design

In contrast to narrowing the task or narrowing the criteria of the tool, Chris was critical of rubrics that limited the range of assessable criteria and presumably failed to capture the totality of game play:

The rubric needs to be broad enough to be able to consider different possible responses from a student, ... consider that bigger picture of, well, what ... if they show this type of activity, what assessment would they get? ... So, I think you just sort of need to work through the different possibilities.

As the only participant that did not use rubrics to assess IGS, it is possible that Chris's negative view explained his non-use of rubrics. Criticism of rubrics constraining performance by offering a finite range of criteria and performance levels exist in the wider literature beyond the field of PE (Cooper & Gargan, 2009; Martin et al., 2015; Panadero & Jonsson, 2020). While not articulated, Chris's suggestion of working through what students might submit, or perform in the assessment context, tacitly suggested the need to understand the likely range of game performance of students at each Year Level before constructing a rubric.

Despite the prolific use of rubrics, Frank expressed limited interest and expertise in rubric design, and was not sure of the origins of his assessment criteria. When asked specifically how he came to have rubrics with four levels of quality and their associated titles from "excellent" to "developing" Frank responded:

Yeah, good question, I think just (colleague) and I were, maybe we were speaking, there's a girl at school who's humanities based who's done at a rubric course, so we actually we've sat with her, she went to Uni and just did a 6 months or something on, which wouldn't float my boat, but she was the expert at our school and she gave us some of those ... But in terms of how the exact reason I don't, I don't know the answer, sorry.

When asked how useful Frank found rubrics he provided a score of six and a half; a view shared by Amy when suggesting a utility score of six. In both cases these scores indicated a commonly held view by participants that rubrics could be improved. Frank felt that a shortcoming of his rubrics was that the language used was too subjective and not "student-friendly". By this, Frank meant that

students found it difficult to make judgements based on the descriptors in his rubrics. The language in Frank's rubrics, like many others, included frequency-based language such as "consistently" and "always", and evaluative language such as "developing" and "beginning". This use of rating scale language in rubrics has been described as potentially more useful for grading rather than for learning, as this language may fail to describe the quality of performance (Brookhart, 2018). It is the absence of rating scale language and instead the exclusive use of qualitative descriptions that constitute a "true rubric" (Brookhart, 2018, p. 1). As an example of the limited utility of frequency-based descriptors, Frank posed the following question when scanning one of his rubrics that he was able to access due to the online interview format:

So, as we go down our list, it's just we've got consistently, usually does it, developing or is beginning to do it, um, you know, how, what, how often is consistently compared to usually, what are we looking at, 90%, are we're looking at 80%?

Given the pervasive use of rubrics by participants in this study and the earlier phase of the survey (Chapter 5), a generic rubric based on the GPAI and TSAP was proposed by participants (Frank, Amy and Ian). With specific reference to the GPAI, Amy commented that this tool:

Would be helpful for us as teachers but would then for our school would still need to translate for a rubric on some level so I guess you know you may be using that to inform at least a rubric, I could see that happening, yeah?

As the GPAI has been adapted into a rubric format for use in a generic invasion game (Harvey, 2007), Amy's suggestion has considerable merit. Frank also felt that the criteria of the GPAI "would fit into rubrics" as they were what his staff were "trying to embed, we want our kids to be able to make the correct decisions, hopefully perform that skill execution well and then be able to move to a position after that". Further, Ian suggested that a "hybrid" version of the GPAI and TSAP could be "a powerful tool" in assessment. Participant suggestions to modify tools for their students is endorsed in the wider PE and performance analysis literature (for example, Brewer & Jones, 2002; Nadeau, Godbout, et al., 2008; Nadeau, Richard, et al., 2008). While participants spoke of adapting these tools into rubrics to meet school reporting policies, there was no indication that any participant would act on their suggestions.

6.4 Recommendations

As the knowledge created in interpretively framed mixed methods research is inextricably linked to the participants and the research context, the practices examined through the four themes are not suggested as being universal (Willis, 2012), nor are the following recommendations necessarily generalisable beyond the immediate population (McChesney & Aldridge, 2019). While this study comprised a sample of PE teachers in Victorian secondary schools, and the “shorter the step to generalisation the better” (Willis, 2012, p. 31), some of the following recommendations may resonate with PE teachers beyond the local context. Thus, the following recommendations may be contextually and analytically transferable (Smith, 2017; Tracy, 2010). The application of any findings or recommendations beyond the target population requires thoughtful reflection of the possibilities for other educators in their own practice (Willis, 2012).

Teachers in this sample are encouraged to work closely with senior leaders in their school to help co-create assessment *and* reporting policy that supports the use of AfL. Given the limited use of ongoing formative assessment, the assessment practices among this cohort appears to contradict the AIESEP position statement on assessment in PE that advocates the ongoing use of AfL to support purposeful learning and legitimisation of the subject (AIESEP, 2020). In collaborating with senior school leaders, assessment for purposes and audiences beyond reporting to parents/ guardians may be promoted, thus addressing a widely reported issue in the PE assessment literature (Moura et al., 2021; Veal, 1988). This collaboration may be supported by VCAA producing and promoting teacher support material including further resources within the DAL (VCAA, n.d.-c). There is precedence for this proposal given VCAA’s recent addition of the MAP platform to help PE teachers locate and undertake evidence-based, Fundamental Movement Skill (FMS) assessments within PE from F – 6 (VCAA, n.d.-h). In further supporting curriculum aligned assessment practice, the VCAA may promote adherence to the VC by auditing school-based assessment in Year Levels 7 - 10, as they do in the VCE (Year Levels 11 to 12) (VCAA, n.d.-a).

It was clear from the interviews, that participants’ assessment practices tacitly intersected several frameworks presented in the literature review in Chapter 2 and could benefit from strategies for implementing formative assessment (AfL). It was also evident that no participant articulated a process for designing their assessment tasks or tools. As key silences in participant responses, the

terms authentic assessment and AfL were not used by any respondent. Consequently, participants may benefit from critical consideration of the authentic assessment framework (Gulikers et al., 2004), strategies for implementing formative assessment or AfL (Black & Wiliam, 2009), elements of rubric design (Dawson, 2017), and the quality assessment framework (Chappuis et al., 2012). This might be supported by professional learning opportunities hosted by the VCAA, or any of the three professional organisations that supported the recruitment of participants for the cross-sectional, quantitative inquiry (Chapter 5).

To address students' limited awareness or understanding of their assessment, providing students access to rubrics prior to their assessment can promote transparency, reduce stress, improve feedback, and develop student self-efficacy (Panadero & Jonsson, 2013). Participants are encouraged to make their rubrics available to students, either electronically or through hard copy, and to schedule class time to explain each criterion (Black & Wiliam, 2009). Similarly, students should be made aware of when they are being assessed and provided with timely feedback of that assessment that might include where they are now, where they are heading, and how they might get there (Hattie & Timperley, 2007). Regarding the limited use of students in the assessment process, participants are encouraged to support sustainable feedback practices (Boud & Molloy, 2013), that support self-assessment and provide students opportunities to apply feedback/feed-forward in subsequent assessments. In doing so, students can develop student agency and build their capacity for self-regulation (Boud, 2000; Boud & Soler, 2016).

Chris's recognition that rubrics may constrain performance, while under-reported by the cohort, is widely recognised in the non-PE educational assessment literature (for example, Bennett, 2016; Panadero & Jonsson, 2020; Spandel, 2006). To address this, by adding a blank row to an existing rubric, users may add another criterion and their own qualitative descriptors of performance at the appropriate level(s). This blank row has been suggested in the context of writing as a "Wild card" (Turley & Gallagher, 2008, p. 90), that allowed students the opportunity to identify a criterion of their performance that sat outside of the existing rubric. Further, an extra column could be added to any existing rubric to allow for description of performance that exceeds the highest level within the rubric. Notwithstanding any negative impact on the validity and reliability of the assessment, this recommendation may be more useful for low stakes assessment, or assessment with a formative use.

Participants are also encouraged to avoid evaluative and frequency-based descriptors, and instead reframe their descriptors more qualitatively (Brookhart, 2018), to support student understanding of the next steps in their learning.

6.5 Strengths and Limitations

A major strength of this work was the construction of the four substantive themes that sought to explain and make meaning of findings from the preceding cross-sectional, quantitative inquiry (Chapter 5). Participant views were understood within the extant literature on assessment in PE (Chapter 2) and included findings that: assessment practices are inextricably linked to a broader social context (Hay & Penney, 2013); that traditional challenges like large class sizes and time constraints are still relevant today (Braga & Liversedge, 2017; Gallo et al., 2006; Veal, 1988; Veloo & Md Ali, 2016); that assessment was generally done by the teacher to the student (Boud, 2000); and that the pervasive use of scale language negatively impacted rubric utility (Brookhart, 2018).

A common limitation in qualitative work comprising interviews is establishing the completeness of the work, often through consideration of the sample size ($N = 8$) and the contested concept of saturation (Guest et al., 2006). Given this study's alignment to conceptual depth (Nelson, 2016), the degree of completeness of the work may be viewed as a limitation of the study, as the reader is charged with determining if the study is adequate in range, complexity, subtlety, resonance and validity (Nelson, 2016). Further, participant responses to questions and probes were based on recollections (O'Donoghue & Farrelly, 2022), as the inquiry did not occur at the same time as participants actioned their assessment practice. The intent was to have interviews closer to the time participants responded to the questionnaire, but due to the lockdowns in Melbourne as a consequence of COVID-19 (Tuffield, 2021), interviews were held later than originally planned. Thus, participant understandings and actions conveyed in the interviews that spanned November 2020 to March 2021 were based on recollections of events that happened across the school year of 2019 (January to December), and may be viewed as a limitation of the study due to the potential fallibility of participants' memories.

6.6 Conclusions and Further Research

This chapter described the interview design and analysis that utilised a reflexive thematic approach (Braun & Clarke, 2021a). This chapter described the means and methods applied in the practicing of rigor that supported the construction of four substantive themes in response to the research sub-question. This qualitative inquiry met its interpretive aim to further understanding as to how assessment of IGS in secondary school PE (Year Levels 7 - 10) was understood and practiced by a nested and purposive sample of participants from the preceding cross-sectional, quantitative inquiry (Chapter 5).

Key findings from the four themes included that sociocultural pressures (Hay & Penney, 2013) may have played a role in participants following school mandates that conflated assessment with reporting, thus supporting the assessment in PE literature (Moura et al., 2021; van der Mars et al., 2018a; 2018b). For some participants, there was less sociocultural pressure to follow the mandate from VCAA to implement and align assessment to the VC: HPE (VCAA, n.d.-j). Rubric use was widespread, but considered imperfect, due to the use of scale language (Brookhart, 2018); feedback and feed forward channels were under-developed; and students played a largely passive role in their assessment. Many participants identified deficiencies in their real-world assessment practice, and most were able to articulate idealised assessment strategies. These strategies included developing students' umpiring skills to free up time for teachers to assess, simplifying assessment tasks to capture evidence of performance, making use of more objective descriptors in rubrics, supporting assessment through video exemplars, improving the frequency and specificity of feedback and feed forward, and engaging students in self and peer assessment.

The findings in this chapter constitute an original contribution to knowledge in its capacity to explain and understand the practice of assessment by PE teachers in Victorian secondary schools in the context of IGS across Year Levels 7 - 10. Further research arising from this qualitative inquiry might involve interview studies within the same cohort to investigate their assessment practices in other game categories and Focus Areas. Field observations of the assessment practices of PE teachers in Victorian secondary schools, possibly using tools like the Systematic Observation of Formal Assessment of Students by Teachers (SOFAST) (van der Mars et al., 2018b), may also address gaps in the assessment literature. Given the participants' widespread use of "we" rather than "I" when asked

about their assessment practices, involving multiple PE teachers from the same school may provide insight into the consistency of assessment practice within schools. Lastly, field testing the proposed adaptation to rubrics that included an additional blank row and column for users to modify is also encouraged, and would serve to address criticisms that rubrics constraint performance (for example, Bennett, 2016; Cooper & Gargan, 2009; Spandel, 2006).

Due to COVID-19 restrictions in Melbourne (Tuffield, 2021), there was no option to undertake subsequent research on site in Victorian schools to complete this thesis. In pursuing lines of inquiry anchored in participant understandings (Kelly & Cordeiro, 2020), Chapter 7 presents a qualitative document analysis of rubrics used in the field. A document analysis was a logical extension of the analyses presented in this chapter that reported the widespread use of rubrics by participants, accompanied by reservations regarding their utility.

CHAPTER 7: A DOCUMENT ANALYSIS OF RUBRICS USED BY PHYSICAL EDUCATION TEACHERS IN INVASION GAMES AND SPORTS IN VICTORIAN SECONDARY SCHOOLS: A QUALITATIVE INQUIRY

7.1 Introduction

In Chapter 6, the use of rubrics intersected each of the four substantive themes with participants generally expressing the view that their rubrics could be better. The previous qualitative inquiry helped to explain the understanding and practice of assessment by PE teachers in Victorian secondary schools in the context of IGS in Year Levels 7 - 10. Participants' practices were dominated by the use of rubrics for reporting to parents/guardians, however, it remained unclear how rubrics were constructed, how well they aligned to a curriculum, and what use rubrics made of key performance criteria identified in the ScR (Chapter 4). I therefore decided to undertake the following qualitative document analysis of participants' assessment rubrics to fill gaps in the above data.

In this chapter I present the rationale for the study and describe the means and methods of the document analysis. Specifically, I identify the aim of the study and the research sub-question, before providing a brief overview of the potential benefits and limitations of assessment rubrics, as described in contemporary literature reviews (for example, Brookhart & Chen, 2015; Jonsson & Svingby, 2007; Reddy & Andrade, 2010). I describe the strategies employed to recruit participants and evaluate the sample size before justifying the use of content and thematic analysis as a suitable method (Bowen, 2009). The READ approach (Dalglish et al., 2021) is outlined before the four constructed themes that comprise the findings and discussion section are presented. The chapter offers some recommendations for the population, before concluding with an overview of the strengths and potential limitations of the study and possible areas for future research.

The document analysis of rubrics described in this chapter was considered an appropriate response to the findings in the preceding studies of this thesis. Specifically, findings of the ScR of evidence-based tools in the assessment of IGS (Chapter 4) provided some support for rubrics as valid assessment instruments. Findings of the cross-sectional, quantitative inquiry (Chapter 5) indicated that rubrics were the most widely used assessment tool, yet approximately half of the cohort identified another assessment tool as being the *most* useful. In support of the latter view that rubrics

may lack utility, in the preceding qualitative inquiry (Chapter 6), two participants encapsulated the views of the cohort when scoring rubrics as a modest 6 out of 10 and 6 ½ out of 10 for their usefulness.

Document analysis has been used to explore aspects of secondary school PE curriculum in Australia, Brazil, Canada, and New Zealand (for example, Araújo et al., 2021; DinanThompson & Penney, 2015; Kilborn et al., 2016; Whittle et al., 2017a). There is also support for document analysis complementing other qualitative study methods, namely interviews and field observations, to help strengthen overall findings (Bowen, 2009; Dalglish et al., 2021). As this document analysis sought to further clarify and understand results from earlier studies, it can be seen as having the aim of *complementarity* within the broader mixed methods study design (Schoonenboom & Johnson, 2017). As no similar inquiries investigating assessment rubrics used by secondary school PE teachers in an IGS context within Victoria were located in the assessment in PE literature review (Chapter 2), this study constitutes an original contribution to knowledge.

7.1.1 Aim

The aim of this inquiry was to describe and explain the assessment tools widely used by participants in the survey studies (Chapter 5 and 6). To meet this aim, a document analysis was employed to answer research sub-question 4: *How are Physical Education teachers' assessment tools constructed for invasion games and sports in Victorian secondary schools?* A reference point for the research sub-question included findings from the assessment tools located in the peer-reviewed literature described in the ScR (Chapter 4). In responding to the research sub-question, there was a pragmatic emphasis to create knowledge that could be useful to stakeholders in the field (Kelly & Cordeiro, 2020).

7.1.2 An Overview of Rubrics

There is no unified understanding of rubrics used in education, nor of the language used to describe their essential features. As an example of this lack of agreement, Popham (1997) suggested that a “rubric has three essential features: evaluative criteria, quality definitions and a scoring strategy” (p. 72), while Chappuis et al. (2012) suggested that rubrics have four features comprising criteria, indicators, levels and descriptors. More recently, Dawson (2017) has stated that rubrics are assessment tools that “usually include” (p. 349) the three elements described by Popham (1997). The

inconsistent use of language to describe key features of rubrics is exemplified by the term given to descriptions of performance that include: *descriptors* (Chappuis et al., 2012); *standards of attainment* (Jonsson & Svingby, 2007); *quality definitions* (Popham, 1997); *performance level descriptions* (Brookhart, 2018); and *performance criteria* (Fluckiger, 2010). With specific reference to rubrics designed for the assessment of IGS in Year Levels 7 - 10 in the VC: HPE, descriptions of performance are termed *quality criteria*, and the term *organising element* and *action* are used to represent criteria and sub-criteria respectively (VCAA, n.d.-g). There is limited use of the VC: HPE naming conventions in the corpus and the extant rubric literature (for example, Brookhart, 2018; Dawson, 2017; Popham, 1997). Throughout this chapter I use the term *criteria* to describe an overall element of performance and the term *descriptor* as the specific representation of a criterion across various levels of performance. As examples of both terms linked to the key performance criteria located in the ScR (Chapter 4), criteria might include passing a ball, while a high-level performance descriptor might indicate a pass was successful if it was received by a team mate.

Rubrics have been used in a wide range of PE contexts, including the assessment of personal and social learning (Gibbons & Robinson, 2004), FMS (Mohansen, 1998), generic sport skills (Hensley, 1997), specific sport skills (Chen et al., 2016), badminton (Casebolt & Zhang, 2020), basketball (Shaw, 2014), flag football (Robinson & Melnychuk, 2009), a generic invasion game (Harvey, 2007), racket games (Harvey & van der Mars, 2010) and tag rugby (Harvey & Hughes, 2009). The ScR described in Chapter 4 located two evidence-based rubrics that were used in the IGS contexts of basketball (Williams & Rink, 2003), rugby union, and soccer (Penney et al., 2012). While a body of evidence into rubric efficacy continues to grow, much of the corpus is based in a higher education setting and/or the skill of writing (Brookhart & Chen, 2015; Panadero & Jonsson, 2013), meaning that studies in the area of secondary school PE are under-represented or absent in many literature reviews (for example, Brookhart & Chen, 2015; Panadero & Jonsson, 2013, 2020).

There is generally modest to positive support for rubrics in the assessment literature across a broad educational setting. For example, a literature review examining the quality and effectiveness of rubrics by Brookhart and Chen (2015) did not make direct claims as to the efficacy of rubrics. Instead, the authors argued that it was the provision of clear learning goals and descriptions of performance embedded *within* rubrics that supported learning and student motivation (Brookhart & Chen, 2015).

Of the 38 studies included in their review none were based in PE. In another review of rubrics across multiple-disciplines, the authors suggested that rubrics “may have the potential” to promote student learning (Panadero & Jonsson, 2013, p. 129). While rubrics can promote transparency of criteria; reduce anxiety; enhance feedback; and improve student self-efficacy and self-regulation; these findings have been largely informed by student perception (Panadero & Jonsson, 2013). Of the 21 studies included in the review by Panadero and Jonsson (2013), none were in a PE context. The findings regarding the impact of rubrics in the assessment literature reviews are tentatively positive in their impact on student learning, but the findings may lack generalisability to a PE context given the omission of this subject from the corpus.

In a critical review of arguments *against* rubrics in a wider educational milieu, criticisms included claims that rubrics narrowed the curriculum, they promoted instrumentalism (tokenistic approaches to tasks), and they were inherently flawed in their use of pre-set and imprecise criteria (Panadero & Jonsson, 2020). Unlike the previous two reviews, this study did not make it clear if any studies were located in the context of PE. In addressing these criticisms of rubrics, the review’s authors argued that the underlying empirical support was “neither direct nor strong” (Panadero & Jonsson, 2020), suggesting that the above criticisms, including narrowing the curriculum and the use of inaccurate criteria, were ill-founded. Each of the three reviews covered in this chapter provide guarded support for the use of rubrics in education. However, the transferability of these findings to the secondary school PE context of this thesis is not assured because of the notable absence of studies in this subject in the above assessment rubric literature reviews.

7.2 Method

A document analysis was employed as a qualitative approach to answer the broadly descriptive research sub-question (Ramanadhan et al., 2021). From a pragmatic stance, this study was bound in a real-world context, it made use of rubrics from key stakeholders, and was able to identify several problems in the participants’ construction of rubrics. Document analysis was used in this study as a systematic process for evaluating documents in electronic form that involved locating, selecting, evaluating, and synthesising the data within the documents (Bowen, 2009). The largely interpretive process of document analysis undertaken was reflexive and recursive (Wood et al., 2020), as it involved the researcher moving between stages that comprised skimming, reading, and interpreting

(Bowen, 2009). In keeping with a qualitative inquiry that employs pattern-based analysis, the findings and discussion of this study are presented together to support links to the extant rubric literature and develop a fuller analysis (Braun & Clarke, 2013).

7.2.1 Participants and Sample

Following the cross-sectional, quantitative inquiry (Chapter 5), a total of 22 participants indicated their willingness to be contacted to participate in a follow-up study. Of these 22 participants, 11 agreed to submit assessment artefacts for the document analysis described in this chapter, including six of the eight participants from the previous qualitative inquiry (Chapter 6). The use of a nested volunteer sample is widely supported in sequential mixed methods studies (Onwuegbuzie & Collins, 2007; Tanner, 2023). As an alternative to the contested notion of saturation (Hennink et al., 2017; Rowlands et al., 2015), a consideration of the study's conceptual depth (Nelson, 2016) is discussed later in this chapter. In summary of the sample of this study, the 11 participants came from 11 different school sites and submitted a total of 24 documents constituting 24 unique rubrics. The documents used in the study were sufficient to generate four substantive themes that answered the research sub-question. Consistent with a functional (local) pragmatic imperative to construct knowledge to support improved action or practice (Goldkuhl, 2011, 2012), the findings and discussion section of this study led to a series of recommendations targeting the current population.

Each of the participants represented a unique school site and the number of participants from each school sector (Catholic, government, independent) approximated their equivalent representation in the cross-sectional, quantitative inquiry in Chapter 5. Data for school sector and curriculum were also drawn from participant responses in Chapter 5, while other school-based data were gathered from the My School website (<https://www.myschool.edu.au/>) in October 2022. The number of students in each school was rounded to the nearest 100 for consistency. An overview of the participants and their submitted artefacts is arranged alphabetically by participant school sector in Table 7.1.

Table 7.1*Document and Participant School Characteristics*

Pseudonym	Rubric Year Level(s)	Rubric n =	School sector	School name	Curriculum	Total student number	School Year Levels	School Region	Co-educational status
Betty	9	1	Catholic	A	VC: HPE	600	Prep to 12	Inner regional	Yes
Amy	8	4	Government	B	VC: HPE	1800	7 to 12	Major cities	Yes
Harry	8	4	Government	C	VC: HPE	700	7 to 12	Outer regional	Yes
Ian	7 to 9	1	Government	D	VC: HPE	1200	7 to 12	Inner regional	Yes
Kelly	7, 8, 9, 10	4	Government	E	VC: HPE	1100	7 to 12	Major cities	Yes
Pam	8, 9	2	Government	F	VC: HPE	700	7 to 12	Major cities	Yes
Carrie	8	1	Independent	G	VC: HPE	800	Prep to 12	Major cities	No
Ed	Not stated	1	Independent	H	VC: HPE	1200	K to 12	Major cities	Yes
Frank	7, 7 to 9	4	Independent	I	VC: HPE	1700	ELC to Year 12	Major cities	No
Louise	Not stated	1	Independent	J	VC: HPE	600	Prep to 12	Major cities	No
Sally	8	1	Independent	K	VC: HPE	1400	ELC to Year 12	Major cities	Yes

Note. The language used across schools to describe the youngest Years or Levels included the following terms: ELC – Early Learning Centre; K –

Kindergarten; and Prep- Preparatory school. In order to de-identify the 11 unique school sites the letters A to K have been used.

7.2.2 Ethics

Consistent with the standards of Social and Behavioural Research Ethics Committee (SBREC) at Flinders University, all 22 participants that expressed an interest to be contacted for a follow-up study after the cross-sectional, quantitative inquiry (Chapter 5) were invited to participate in this study via email. Further, participants were provided with an information sheet and consent form via email which were signed and returned to the researcher, again via email, before they could participate in this study. The email invitation asked participants to submit electronic copies of rubrics, exemplars of performance and/or any guidelines for the use of rubrics. Each of the documents describing the ethical requirements for this study can be found in Appendix M. Participants were reminded that that the context was student performance in IGS across Year Levels 7 - 10, and that there was no limit to the number of documents that they could submit. During this correspondence, participants were reminded that all documents would be de-identified upon receipt, before being numbered and stored on a password protected Flinders University account. While no student data (for example, images or names) was to be sent, one document, an exemplar containing student images, was submitted before being removed from the study due to ethics.

7.2.3 Data Collection and Analysis

The READ approach of Dalglish et al. (2021) provided me with the methodological scaffolding for the following qualitative document analysis. The four iterative steps involved reading material, extracting data, analysing data, and distilling findings (Dalglish et al., 2021). The stages in this process involved both content and thematic analysis to help shape my interpretation (Bowen, 2009), as neither analysis method was considered adequate to answer the research sub-question alone (Braun & Clarke, 2021a, 2021b). Content analysis involved coding the data in a systematic and replicable fashion (Mackieson et al., 2018) by organising information into categories to provide an overall picture of the material (Bowen, 2009). The thematic analysis involved recognising patterns in the data to construct themes that informed the analysis of the study (Bowen, 2009). Unlike reflexive thematic analysis, themes in the codebook analysis employed in this study “*may* consist of summaries of data domains” (Braun & Clarke, 2021a, p. 341).

In viewing document analysis as a qualitative research method (Bowen, 2009), the current study was considered an extension of the previous qualitative inquiry (Chapter 6) as it employed thematic

analysis and involved six of the same participants. Further, this study shared several codes from the inquiry described in Chapter 6, including several that were based on the influential work of Dawson (2017) that identified rubric design elements. The use of these established codes was consistent with document analysis methodology as:

Predefined codes may be used, especially if the document analysis is supplementary to other research methods employed in the study. The codes used in interview transcripts, for example, may be applied to the content of documents. Codes and the themes they generate serve to integrate data gathered by different methods. (Bowen, 2009, p. 32)

Use of deductive coding, considered necessary to respond to the research sub-question, involved the key performance criteria identified in the ScR of evidence-based assessment tools (Chapter 4) and relevant terminology from the VC: HPE (VCAA, n.d.-d). In addition, I was open to more data-driven or inductive coding (Braun & Clarke, 2021b), when coding novel items like *ordinal scale language*, *multiple indicators*, and *empty cells* in the submitted artefacts. This mix of inductive and deductive coding within the content analysis was considered abduction (Hennein & Lowe, 2020; Kovács et al., 2005) and consistent with my “pragmatist perspective” (Mitchell, 2018, p. 105). That is, abductive coding supported the wider pluralistic approach to the use of complementary research methods in the thesis.

Step one in applying the READ approach (Dalglish et al., 2021), *readying materials*, involved establishing parameters for the nature and number of documents to be collected to answer the research sub-question. This step involved determining what artefacts would be solicited and how many might be required to meet the aim of the study. The nature of the documents was clearly established from the previous cross-sectional, quantitative inquiry (Chapter 5) that indicated rubrics were the most widely used assessment tool by the population, and the previous qualitative inquiry (Chapter 6) where participant use and views on rubrics impacted all four themes. In addition to inviting participants to submit assessment rubrics for this document analysis, participants were also invited to submit any exemplars of performance and explanations or guidelines for the use of rubrics. These additional artefacts were based on several rubric design elements identified by Dawson (2017) that could respond to the research sub-question (see Appendix N). Thus, participants were invited to

submit electronic copies of rubrics, exemplars, and guidelines used in their assessment of student performance in IGS across Year Levels 7 - 10.

Given the diverse representation of rubrics in the literature described earlier in this chapter, no definition of rubrics was provided in the above invitation to participants. This allowed participants to show their understanding of what constituted a rubric, and for me to consider my understanding of rubrics based on the submitted artefacts in conjunction with the extant literature (for example, Brookhart, 2018; Chappuis et al., 2012; Dawson, 2017). The full script for the email invitations is included with other documents pertaining to ethical requirements for this study in Appendix M. After collecting participant submissions for this study, artefacts were classified as rubrics and included in the study if they comprised at least two of the following three elements:

1. At least one *criterion* - a criterion provides a broad indication of what skill, knowledge or understanding is to be assessed.
2. *Descriptors* of performance - these describe the quality of performance at a certain level based on the respective criterion. This requirement did not mean that every cell of a rubric in table format had to have a descriptor, or that the descriptor had to be text-based.
3. At least two *levels* of performance quality - these performance levels did not require labels or titles.

This flexible, rather than restrictive approach for inclusion, was adopted as “It is generally better to have access to a wide array of documents providing a preponderance of evidence” (Bowen, 2009, p. 33). In requiring artefacts to include at least two of the above three elements, all submitted artefacts were included in the corpus. The dates of inclusion for submission of artefacts were from the first email invitation November 2021 until the end of December 2021. Follow up email invitations were sent to the 22 participants that indicated their willingness to participate in further study, with 11 respondents responding affirmatively to the email invitations and participating in the document analysis. To manage the submitted artefacts, a file-naming system was employed that de-identified participants/schools, while allowing for the accurate tracking of rubrics, exemplars or guidelines. This naming system took the form of a capital letter to indicate the participant’s first initial (based on a pseudonym), lower case letters and numbers to indicate the Year Levels from 7 - 10, or a lower-case abbreviation if there were no Year Levels indicated (nyl), and a written number separated by a dash to

indicate multiple documents. For example, the third rubric from Amy for Year Level 8 was coded 'Ayr8 - 3', while the fourth rubric from Frank that targeted Year Levels 7 to 9 was coded as 'Fyr7to9 - 4'. All submitted documents were de-identified and ascribed unique codes and stored on a password protected network drive of Flinders University.

In keeping with the pragmatic approach to utilise the most appropriate quantitative and qualitative methods to answer the research sub-question (Essiet et al., 2022; Johnson & Onwuegbuzie, 2004), *extracting data* as step two of the READ approach (Dalglish et al., 2021) was supported by a codebook that I independently created in Microsoft Excel. When searching for patterns or trends in data, a codebook thematic analysis approach is positioned as a pragmatic middle ground between a reflexive thematic approach (qualitative) and a coding reliability approach (quantitative) (Braun & Clarke, 2021a; Byrne, 2022). This meant that the codebook had a number of tentative codes derived from the literature and the previous qualitative study, prior to any examination of the submitted documents. As the research sub-question focussed on *how* rubrics were constructed, and *how* they aligned to key performance criteria located in the ScR (Chapter 4) and elements of the VC: HPE, the items selected for data extraction were pivotal to the study. Given the modest volume of data to code, in contrast to the 62,000 words of verbatim transcript in the previous qualitative inquiry (Chapter 6), coding was done manually rather than electronically. The codes were identified abductively, as I moved between the related literature (deductive coding) and the artefacts (inductive coding) (Mitchell, 2018). As an example of related literature, some codes were based on the rubric design elements described in the influential work of Dawson (2017), for example *layout*, *type*, *generic* and *specific*, the original 14 elements of Dawson (2017) are presented in Appendix N. In addition, the final codebook comprising 47 discrete coding items is presented in Appendix O.

All categories/codes involved documenting low-inference items that were observable on the rubrics. I independently extracted all data before inviting a member of the supervisory team (SP) to a training session that included explaining the codes and having SP apply them to a sample rubric from outside the corpus (see Table 1.1). The use of a second coder and the construction of a codebook was primarily to test the robustness of the coding guidelines, promote data familiarity within the supervisory team, and activate SP as a critical friend in the data collection and analysis (Carlson et al., 2018; Deuchar, 2008). Consistent with the approach for reaching consensus in data charting in the ScR

data of Chapter 4, any disagreements in coding the extracted data were resolved via email and ZOOM™ meetings between both coders. Engaging a third member of the supervisory team to act as an arbitrator when the two coders could not reach consensus, as per the guidelines of Fink et al. (1984), was not required. As an example of initial disagreement in coding item 10 (number of criteria), several rubrics employed a double column format for listing criteria, potentially indicating their adherence to VCAA support material that employed the same format and terms *organising element* and *action* (VCAA, n.d.-g). Following a discussion of this format, it was agreed that criteria or sub-criteria may be presented in more than one row or column, and it was agreed these rubrics displayed eight criteria rather than four. The codebook was updated and all rubrics were reviewed by both coders considering this modification.

Where coding reliability approaches to thematic analysis reflect quantitative research values and reflexive thematic approaches reflect qualitative research sensitivities, the codebook thematic analysis approach in this study comprised a qualitative approach with pragmatic accommodations (Braun & Clarke, 2021a). In codebook thematic analysis approaches “consensus between coders and inter-rater reliability are not usually measures of quality” (Braun & Clarke, 2021a), as such, no statistical measures of inter-rater reliability were collected or presented as indicators of coding quality in this study (Braun & Clarke, 2021a). Instead, the reader is invited to assess the quality of the coding and thematic development according to the description of the specific means and methods presented in this chapter.

The final coding guidelines contribute to the study’s audit trail to support transparency and replicability (see Appendix O). Due to the study’s narrow focus and the relatively heterogenous nature of the submitted artefacts, the codebook comprised just 47 codes. This number is consistent with advice to be “parsimonious” when selecting coding categories and ensuring that each code responds to the study aim (Dalglish et al., 2021). This number of codes did not include the attribution of a unique identifier, rubric number or identifying the target Year Level(s) which would represent 50 codes in total. Of note, one exemplar document containing a series of videos was excluded for ethical reasons as it included images of students. No other exemplars or any explanations or guidelines for the use of rubrics were submitted for this study.

As step three, *analysing data* begins from the commencement of the study (Dalglish et al., 2021) through activities like memo writing and reflexive journaling. To support my journaling, I adapted six guiding questions from Dalglish et al. (2021) to support my analysis which included:

1. Was the rubric finished or a draft?
2. What was the likely purpose of the rubric?
3. Were there any internal contradictions in the criteria?
4. How did the documents compare across levels and schools?
5. How did rubric criteria compare to the criteria found in evidence-based tools?
6. What, if anything, was missing?

I was mindful that I was not the intended audience for these documents and that even gaps in the data may contribute to a rich interpretation (Rapley, 2018). The iterative process of data extraction and analysis was undertaken with the awareness that documents are not “necessarily precise, accurate or complete” (Bowen, 2009, p. 33). Underpinning the analysis was the construction of four substantive themes, each based on a unique central organising concept, to respond to the research sub-question (Braun & Clarke, 2021a). While the codebook approach made use of some pre-established codes from previous studies in the thesis and the extant literature, the themes were not constructed a priori.

In the final step of the READ approach, Dalglish et al. (2021) posited that *distilling findings* begins when researchers deem their review is complete. To support this determination the author suggests one of three things will indicate this: (a) every document fitting the criteria has been obtained; which is unlikely; (b) time runs out, indicating flaws in the document search criteria; or (c) when the researcher ‘sufficiently’ understands the phenomenon being investigated (Dalglish et al., 2021). In my case, I began distilling findings after point (c), when I was able to sufficiently understand the phenomenon under investigation. Readers are asked to judge the sufficiency of understanding conveyed in the findings and discussion section of this chapter by applying the criteria of conceptual depth (Nelson, 2016). This does not mean that the research reached a “final limit, beyond which it is impossible to achieve new insights, but it ... [has] reached a *sufficient depth* of understanding” to allow the development of theory. A fuller exposition of conceptual depth of the study is provided in the rigor section of this chapter.

In an extended metaphor that describes the first three steps of the READ approach (Dalglish et al., 2021) as being akin to strolling along a beach and collecting items of interest before sorting them into buckets, the final step involved me getting the specimens home, cleaning them off, and preparing them for display (Dalglish et al., 2021). The 'display' in this chapter includes a series of tables based on the extracted data and the four substantive themes. In linking the first and last study in this thesis, Table 7.7 provides a comparison of assessment characteristics and key performance criteria located in the evidence-based tools identified in the ScR (Chapter 4) and the current study. To preserve the integrity of design elements used by participants in the rubrics submitted for this document analysis, all visual references to rubrics are presented as screen clips. While all screen clips present partial or entire rubrics in tables rather than linear text, these visual references are treated as images and accordingly labelled as figures in the findings and discussion section of the chapter (see Figure 7.2 to Figure 7.9).

7.2.4 *Rigor*

Owing to the qualitative nature of this document analysis and the previous inquiry described in Chapter 6, a similar set of quality criteria were employed to demonstrate qualitative rigor. This view was underpinned by the belief that there is no widely accepted common set of criteria for mixed methods studies (Sparkes & Smith, 2009), or thematic analysis (Braun & Clarke, 2021a), and that quality criteria develop in response to the research (Smith & McGannon, 2018). Research criteria are historically and contextually located, making their broad application across different research contexts difficult (Halcomb, 2019). Rigor in this study begins by addressing the challenge of establishing what constitutes an adequate number of artefacts in document analysis, before describing various attributes of conceptual depth that support the completeness of the research and resulting themes (Nelson, 2016). As with all qualitative research criteria, the reader is invited to assess the degree to which this study displays methodological rigor (Elliott et al., 2023). This assessment begins with determining if the study constitutes a worthy topic (Tracy, 2010), in its analysis of *how* rubrics are constructed within the sample population. Similarly, readers are invited to consider the degree to which the study demonstrated rich rigor, credibility, and sincerity (Tracy, 2010).

Key to demonstrating rigor in document analysis is determining how many documents need to be included in the study (Bowen, 2009). Compounding this challenge is that document analysis

studies rarely state the number of documents collected and are more likely to state the number of *sites* from which they collect their artefacts (for example, Araújo et al., 2021; Backman & Larsson, 2014; Whittle et al., 2017a). In examining other document analyses within PE to try and determine an accepted standard for the number of documents or school sites, no consensus could be reached. As examples, international document analysis studies in the field of PE have determined that as few as three sites (Araújo et al., 2021) and five sites (Backman & Larsson, 2014) were sufficient to address their research questions investigating PE curricula. In contrast, an Australian document analysis of senior PE curriculum reported that eight sites generating 15 documents was an adequate sample for their research (Whittle et al., 2017a).

As 22 participants indicated a willingness to be contacted following the cross-sectional, quantitative Inquiry (Chapter 5), I established a provisional lower and upper range of 10 to 15 participants from different school sites, assuming that most participants would submit at least two rubrics. This range of participants might potentially yield a data set of between 20 to 30 rubrics, which was considered likely to be an adequate data set based on other document analysis studies in PE (for example, Backman & Pearson, 2016; Whittle et al., 2017a). Ultimately, this study comprised 11 participants from 11 unique school sites that submitted 24 documents describing 24 unique rubrics. This sample provided a richness of information (Roy et al., 2015), where none previously existed, and allowed for the construction of four substantive themes and a series of recommendations targeting the current population.

As an alternative to the contested nature of saturation (Smith & McGannon, 2018), the criteria associated with conceptual depth were used as a means to determine the degree of “completeness” of the data set (Nelson, 2016, p. 556). This qualitative study demonstrated a high degree of conceptual depth as it presented multiple excerpts from rubrics to support findings (range), was built upon a relatively broad code base to create four themes (complexity), made sense of often contradictory rubric terminology (subtlety), and was closely aligned to the wider educational literature on rubrics (resonance). When considering the number of artefacts in document analysis, “the concern should not be about ‘how many’; rather, it should be about the quality of the documents and the evidence they contain, given the purpose and design of the study” (Bowen, 2009,

p. 33). Thus, it was the quality, range and nuance within the submitted rubrics that was paramount in this study, rather than the number of documents gathered (Braun & Clarke, 2021b).

Several quality criteria as conceptualised by Tracy (2010), can be applied to the current study including a consideration of the nature of the inquiry as a *worthy topic*. This determination may be justified given the prevalent use of rubrics in the earlier cross-sectional, quantitative inquiry (Chapter 5) and the level of dissatisfaction with rubric utility conveyed in the previous qualitative inquiry (Chapter 6). Further, as rubrics featured prominently in another Australian investigation of GS assessment among secondary school PE teachers (Williams et al., 2020), it is possible that the findings of this study may be useful beyond the local sample. In judging *rich rigor*, the 24 documents from 11 unique school sites are widely referenced throughout the findings and discussion section and are closely linked to the literature on rubric design elements (Dawson, 2017). A detailed application of the READ approach (Dalglish et al., 2021), described in the method section of this chapter, began as rubrics were submitted in November 2021 and concluded with writing of the final report in June 2023, constituting a “sufficient, abundant, [and] appropriate...time in the field” (Tracy, 2010, p. 840).

The *credibility* and *sincerity* of the study (Tracy, 2010) were established when acknowledging my positionality and motivation for the broader thesis in Chapter 1. Details regarding participant school demographics, the transparency in the method, and the thick description (Tracy, 2010) in the findings and discussion section support the credibility of the study. Gathering additional data to explain rubric construction allowed me “to open up a more complex, in-depth, but still thoroughly partial, understanding of the issue” (Tracy, 2010, p. 844). This may be viewed as *crystallisation*, an aspect of credibility that is considered equivalent to the use of *triangulation* in quantitative research, but without the intent to pursue a single truth, and instead the aim of understanding the same phenomenon from a different perspective (Tracy, 2010). The study demonstrated sincerity as regular memo and journal writing provided an ongoing internal conversation for me to build, challenge and refine my interpretation of the data (Braun & Clarke, 2013). Comprising part of the audit trail for this study, a selection of memos is presented in Appendix L. In this study I drew on my chief supervisor (SP) as a critical friend to engage in dialogue via email and ZOOM™ meetings regarding the nature of the documents to solicit, what data to extract, and how to construct themes (Smith & McGannon, 2018). Given SP’s breadth of qualitative research experience, further support came via his

independent coding of the artefacts via a codebook I designed (see Appendix O). The choice of a codebook thematic analysis approach represented “qualitative pragmatism” (Braun & Clarke, 2021a) and was consistent with the pragmatic paradigm of the thesis.

7.3 Findings and Discussion

As with the previous qualitative inquiry described in Chapter 6, the following section of the chapter presents the findings and discussion together, to avoid repetition in the treatment of data that can occur in the reporting of these elements separately (Braun & Clarke, 2013). This integrated approach was supported by close reference to the extant literature in assessment in PE and the wider use of rubrics in education. This section of the chapter identified patterns and trends in the submitted rubrics, as well as anomalous or divergent cases. As the final study of the thesis, and consistent with the treatment of data and analysis in mixed methods research (Creswell & Creswell, 2018; Creswell & Plano Clark, 2018; Ivankova et al., 2016), relevant comparisons are made with the previous studies (Chapters 4, 5 and 6). As with other studies, the findings of this document analysis are followed by a series of recommendations for stakeholders. The following sections 7.3.1 through to 7.3.4 consider the four themes: A Narrow Conceptualisation of Rubrics; Rubrics are for Reporting; Appropriate Criteria are Key; and The Challenge of Describing Quality.

7.3.1 *A Narrow Conceptualisation of Rubrics*

The data comprising theme 1 presented in Table 7.2 centred on the fundamental aspects of rubric formatting. These design elements included the rubric layout, for example, if the content was presented in linear text or tabular form. This theme also included if the rubric made use of separate criteria (an analytical approach), or a single aggregated criterion (holistic approach). Another key design element considered if the rubric was suitable for assessing multiple-sports (generic), or a single sport (specific).

Table 7.2*Theme 1 - A Narrow Conceptualisation of Rubrics*

Identifier	Year Levels	Rubric No.	Layout	Type	Generic	Single sport
Fyr7 - 1	7	1	Table	Analytical	Generic	-
Fyr7 - 2	7	2	Table	Analytical	Generic	-
Kyr 7 - 1	7	3	Table	Analytical	Generic	-
Ayr8 - 1	8	4	Table	Analytical	Generic	-
Ayr8 - 2	8	5	Table	Analytical	Generic	-
Ayr8 - 3	8	6	Table	Analytical	Generic	-
Ayr8 - 4	8	7	Table	Analytical	Generic	-
Cyr8 - 1	8	8	Table	Analytical	Generic	-
Hyr8 - 1	8	9	Table	Analytical	-	Water polo
Hyr8 - 2	8	10	Table	Analytical	-	Football
Hyr8 - 3	8	11	Table	Analytical	-	Hockey
Hyr8 - 4	8	12	Table	Analytical	-	Netball
Kyr8 - 2	8	13	Table	Analytical	Generic	-
Pyr8 - 1	8	14	Table	Analytical	Generic	-
Syr8 - 1	8	15	Table	Analytical	Generic	-
Byr9 - 1	9	16	Table	Analytical	Generic	-
Kyr9 - 3	9	17	Table	Analytical	Generic	-
Pyr9 - 2	9	18	Table	Analytical	Generic	-
Kyr10 - 4	10	19	Table	Analytical	Generic	-
Fyr7to9 - 3	7 to 9	20	Table	Analytical	Generic	-
Fyr7to9 - 4	7 to 9	21	Table	Analytical	Generic	-
lyr7to9 - 1	7 to 9	22	Table	Analytical	Generic	-
Enyl - 1	-	23	Table	Analytical	Generic	-
Lnyl - 1	-	24	Table	Analytical	Generic	-
<i>N</i> =		24	24	24	20	4
Percentage		100.0	100.0	100.0	83.3	16.7

Note. The use of a dash (-) indicated absence of the relevant item.

The central organising concept of this theme (Braun & Clarke, 2013, 2021a) reframed a common criticism of rubrics, that they reduce the breadth of acceptable performance by offering a narrow range of criteria and performance levels (Cooper & Gargan, 2009; Martin et al., 2015; Panadero & Jonsson, 2020). Instead, this theme suggested that it may be the narrow *conceptualisation* of rubrics, as evidenced in their almost universal construction by the sample, that may be more constraining of student performance.

7.3.1.1 Rubric layout

With reference to the data presented in Table 7.2, there were clear patterns regarding rubric formatting that led to the naming of this theme. All rubrics were presented in a table layout as opposed to linear text, and analytical in nature, rather than holistic. As analytical rubrics assess a range of criteria separately, they may be better suited to provide feedback to students (Brookhart, 2013, 2018). In contrast, holistic rubrics that amalgamate criteria have been suggested as more appropriate for grading and reporting (Brookhart, 2018). While both rubrics located in the ScR (Chapter 4) were presented in linear text, one was analytical (Penney et al., 2012), and one was holistic (Williams & Rink, 2003). Unlike the assessment tools located in the ScR (Chapter 4), the psychometric properties of the submitted rubrics remain uncertain when considering a range of validity and/or reliability measures. While the intended or actual use of the rubrics could not be determined from this document analysis, many of the participants in the Chapter 6 study indicated that their rubrics served a summative purpose to support grading, which may be incongruent with the prevalent use of analytical rubrics in this study.

The following excerpt from the Observational Scoring Rubric located in the ScR (Chapter 4) is an example of a linear and holistic approach that was *not* found in any of the submitted rubrics, thus contributing to the naming of the theme, 'A Narrow Conceptualisation of Rubrics'. The excerpt describes the highest level of basketball performance described in the Observational Scoring Rubric (Williams & Rink, 2003):

Level 3:

- No observable errors in dribbling and passing
- Shoots proficiently and consistently
- Executes offensive strategies proficiently and consistently (runs patterns, avoids defenders, pass, catch, and pass/run decisions)
- Consistently applies defensive pressure with good technique
- Consistently executes defensive strategies (defensive pressure, defensive positioning) with good technique
- No observable errors in applying rules and rarely commits violations (e.g., walking, fouls)

(Williams & Rink, 2003, p. 606).


It is unclear why no rubrics in this study made use of linear text or a holistic format, however, participant views in the previous qualitative inquiry (Chapter 6) that school mandates required the use of analytical rubrics for reporting purposes may be partly responsible. As the contents of an analytical rubric can be easily organised into a holistic format by merging multiple criteria into a single criterion, this modest adaptation may support more efficacious grading by the sample (Brookhart, 2018).

7.3.1.2 Rubric type

Providing further support for the participants' narrow understanding of rubrics as evidenced in their relatively homogenous formatting, there was no use of single point rubrics in the corpus. The omission of this type of rubric is consistent with the literature review of rubrics used in higher education (Brookhart, 2018). This type of rubric is characterised by having several criteria in a column in a table, with a blank column(s) either side. These blank columns allow for bespoke descriptors to be written as evidence of performance below or above the standard described in the criteria by the student, teacher or peer (Fluckiger, 2010). Despite some disparate terminology found in the limited, single point, rubric literature (Brinson, 2022; Fluckiger, 2010), the assessable elements in Table 7.3 have been identified as criteria, allowing for the evidence of performance above or below the standard to be considered descriptors.

Table 7.3

Single Point Rubric

Descriptor for how to meet the criteria	Criteria	Descriptor of meeting the criteria	Descriptor of going beyond the criteria
			
	Criteria		
	Criteria		
	Criteria		
	Criteria		
	Criteria		

Single point rubrics allow for the provision of qualitative feedback and support student self-assessment and goal-setting (Fluckiger, 2010); they may also increase student attention to learning rather than grading (Brinson, 2022). Some examples of single point rubrics in the wider educational literature relate to music (Brinson, 2022), essay writing (Fluckiger, 2010) and computer programming (Estell et al., 2016). While the ScR (Chapter 4) and literature review of assessment in PE (Chapter 2) made no reference to single point rubrics used in the Focus Area of GS or the broader subject of PE respectively, their use in these contexts is consistent with AfL and assessment *as* learning strategies (Black & Wiliam, 2009; Gibbons & Kankkonen, 2011).

7.3.1.3 Generic versus single sport

Most rubrics were generic in nature, meaning they could be used widely as they did not identify a single sport. Two examples of a generic or multi-purpose rubric are provided in Figure 7.1 and Figure 7.2. Typical of most generic rubrics submitted by participants, the criteria and descriptors in rubric 22 (Figure 7.1) included references to games, team-mates, strategy, on-the-ball skills (for example, shooting, kicking and passing), and off-the-ball movement (for example, creating space and demonstrating game sense).

Figure 7.1

Rubric 22 – A Generic Rubric for Invasion Games and Sports

Descriptors for Y7-9 PE Practical Assessment

Criteria/Rating	1		2		3		4		5	
	NYC	NYC+	AC	AC+	C	C+	H	H+	A	A+
Effort/Improvement Are they in correct PE gear, organized for class and putting in the effort. Is this effort reflected in their improvement?	<ul style="list-style-type: none"> Active Participation -70% Prepared -70% Next to no improvement 	<ul style="list-style-type: none"> Active Participation 70-90% Prepared 70 – 90% Minimal improvement 	<ul style="list-style-type: none"> Active Participation 90% Prepared 90% Improvement evident 	<ul style="list-style-type: none"> Active Participation 90-100% Prepared 90-100% Clear Improvement 	<ul style="list-style-type: none"> Active Participation 100% Prepared 100% Proactive in their Improvement (challenged beyond norm) 					
Skills Sequences Hitting, shooting, passing, kicking type motors skills and linking in sequence.	<ul style="list-style-type: none"> Very basic skills No consistency Poor control and accuracy 	<ul style="list-style-type: none"> Basic skills Minimal consistency Control and accuracy low 	<ul style="list-style-type: none"> Competent skills to contribute Consistent in execution with control and some accuracy 	<ul style="list-style-type: none"> Competent skills Highly consistent Minimizes execution errors Consistent and accurate 	<ul style="list-style-type: none"> Advanced level skills Strategy-based application Highly consistent 					
Movement Patterns Where they run or move to, how they use the space and interact with other players.	<ul style="list-style-type: none"> Minimal movement Avoidance Stature 	<ul style="list-style-type: none"> Movement is restricted and random Limited movement patterns Reactive 	<ul style="list-style-type: none"> Some game contextual movement Appropriate use of space Some proactivity 	<ul style="list-style-type: none"> Movement shows game sense Uses space to advantage team Proactive 	<ul style="list-style-type: none"> Leads strategy based movement Creates space Highly proactive and adaptable 					
Teamwork Working cooperatively in team, bringing team mates into the game and providing opportunities to be involved. Inclusion mindset not exclusion	<ul style="list-style-type: none"> Demonstrates aspects of respectful behaviour 	<ul style="list-style-type: none"> Applies personal and social skills. Works effectively with others 	<ul style="list-style-type: none"> Applies elements of sportsmanship. Considers safety Respectful and Inclusive with other group members. 	<ul style="list-style-type: none"> Able to evaluate own and others contribution to group tasks 	<ul style="list-style-type: none"> Able to evaluate own and others contribution to group tasks. Demonstrates leadership 					
Feedback How they respond to teacher's improvement suggestions. Do they acknowledge and act or ignore and dismiss? Is it a positive response?	<ul style="list-style-type: none"> No response to feedback Struggles to make changes to performance 	<ul style="list-style-type: none"> Some use of feedback Limited ability to make changes to performance 	<ul style="list-style-type: none"> Responsive to feedback Can apply to improve performance 	<ul style="list-style-type: none"> Highly responsive to feedback. Ability to use alternative feedback sources Seeks areas to improve 	<ul style="list-style-type: none"> Highly responsive to feedback Uses feedback from a number of sources Self-evaluation to adapt and improve performance 					

*Can give half marks

Figure 7.2

Rubric 19 – A Generic Rubric for HPE

Assessment Rubric-Year 10 HPE Practical

Organising Element	Action	Not Shown (0)	Insufficient Evidence (1)	Quality Criteria (2-5)			
They explain the importance of cooperation, leadership and fair play across a range of health and movement contexts.	Reflect on how fairplay and ethical behaviour can influence the outcomes of movement activities.			Identify ways to include other members of the class to participate	Explain ways to others (refereeing/ coaching) to improve participation	Reflect on the importance of group participation in activities.	Analyse reasons why physical activity is beneficial for adolescents
They apply and transfer movement concepts and strategies to new and challenging movement situations.	Develop, implement and evaluate movement concepts and strategies for successful outcomes			Identify strategies that can be used in new game situations	Use tactics/strategies to overcome challenging movement situations	Use these tactics and strategies in multiple activities and movement situations	Evaluate and improve these tactics/strategies for a successful outcome
They apply criteria to make judgments about and refine their own and others' specialised movement skills and movement performances.	Perform and refine specialised movement skills in challenging movement situations			Use feedback to improve coordination when performing movement skills successfully	Perform movement skills successfully in challenging movement situations.	Create new movement skills in challenging movement situations	Lead others in creating new movement skills in challenging movement situations
Uniform	Wears uniform	Has not been changed for a practical class throughout the term	Has had uniform for 1-2 classes throughout the term	Has had uniform in 3-5 classes throughout the term	Has had uniform for 6-7 classes throughout the term	Has had uniform for all practical classes throughout the term	

Unlike the generic rubric for IGS in Figure 7.1, the rubric in Figure 7.2 made no reference to games and was generic to the broader Learning Area of HPE. While rubric 19 (Figure 7.2) refers to language found in the VC: HPE (VCAA, n.d.-d), terms like movement concepts and strategies, movement skills and performances, and tactics/strategies are not explicitly linked to games. In personal email correspondence that accompanied the submission of rubric 19, the participant indicated that the rubric was used at the end of each term, *regardless* of the content that was taught ('Kelly', personal communication, November 30, 2021). Thus, rubric 19 was used at the end of the term, presumably for reporting purposes, to assess any topic or unit that was taught in practical PE. In acknowledging that her rubrics lacked some utility, Kelly asked that if better rubrics were created as a result of this study that she would like to have a copy ('Kelly', personal communication, November 30, 2021). Specific to this thesis, it is unclear how useful rubric 19 was for teachers or students in the assessment context of IGS.

In considering the utility of generic rather than sport-specific assessment of IGS, pedagogical models like the TGFU (Bunker & Thorpe, 1986; Thorpe et al., 1984) and the Tactical Games Approach may be useful (Mitchell et al., 2013). Empirical support for general decision-making (tactical play) across IGS is supported by Memmert and Harvey (2010) in their identification of six non-specific tactical elements that included: (a) attacking the goal; (b) taking the ball near goal; (c) playing together; (d) Identification of gaps; (e) feinting; and (f) achieving an advantage through supporting, orienting and cooperating with partners. Similarly, in adapting various principles of play described by Ward and Griggs (2011) to the current research, generic assessment criteria could include: (a) attacking; (b) supporting; (c) creating space; (d) scoring and penetration; (e) defending; and (f) denying space and applying pressure. In coalescing any of the above conceptualisations with the view of general ball-handling ability (Rylander et al., 2019), the generation of empirically supported generic rubrics for IGS may be possible. Generic rubrics may be more congruent with GBAs applied in thematic games' units in a school setting and better suited to the assessment of transfer of movement concepts and strategies integral to the TGFU model (Bunker & Thorpe, 1986; Werner et al., 1996). Regarding the nature of the 15 school-developed tools in the ScR (Chapter 4), most tools were generic, which is similar to the nature of most rubrics submitted in this study.

7.3.1.4 Outliers

As an anomaly in the generally narrow understanding of rubrics, rubric 10 shown as Figure 7.3, is representative of the four sport-specific rubrics submitted by Harry. As one of two participants to submit rubrics designed in Microsoft Excel, these rubrics were the only artefacts that included drop down boxes for each of five criteria to enter a score of 0 (*poor*) or a 5 (*excellent*) to derive a total score out of 25 and generate a grade.

Figure 7.3

Rubric 10 – A Sport-Specific Rubric for Football (Australian)

Football						
Handballing	Kicking	Marrking	Gameplay attacking	Gameplay defending	Total ___/25	Grade
0	0	0	0	0	0	
0	0	0	0	0	0	
0	0	0	0	1	0	
0	0	0	0	2	0	
0	0	0	0	3	0	
0	0	0	0	4	0	
0	0	0	0	5	0	

Note. Australian Football (AF) is not synonymous to soccer. In the adult game of AF two sides comprising 18 players occupy a large field with the aim of moving the ball by running, handballing and kicking to score more goals than their opponent.

Contributing to the novelty of this format, these rubrics provided no text-based description of what students needed to demonstrate to achieve the various scores, meaning they might be more accurately categorised as rating scales (Brookhart, 2018). While these rubrics met the inclusion criteria for this study (see Step 1: Readying the material), these assessments seemed to be primarily

designed for efficacious scoring of students. Further, as they were designed to track class level data, privacy and confidentiality issues might arise if the document was shared with students. Conversely, if the rubric was solely for use by the teacher, it is not clear how students would be made aware of the assessment criteria or their results. Rating scales 'masquerading' as rubrics have been similarly found in higher education, with the research suggesting that some educators incorrectly consider any assessment with a scoring scheme to be a rubric (Brookhart, 2018). It is this narrow conceptualisation of rubrics as scoring guides (Siedentop et al., 2011) rather than instructional guides (Andrade, 2005) that informed the following theme.

7.3.2 Rubrics are for Reporting

Theme 2 focussed on the levels of performance described in the rubric, and the potential for the rubric to be used to score or grade student performance. The data that was coded included the total number of levels of performance, how any levels were labelled, the order with which quality of performance increased, and if the rubric generated an overall outcome (score, percentage, grade or reference to a standard). The latter code included any key or legend that complemented the rubric that could generate a performance outcome (see Table 7.4).

Table 7.4*Theme 2 - Rubrics are for Reporting*

Rubric No	Year Levels	Number of levels	Level label	Location of level label	Level order (low to high)	Outcome reported
1	7	5	Text	Top row	Right to left	-
2	7	5	Text	Top row	Right to left	-
3	7	3	Text/Number	Top row	Left to right	Score
4	8	5	Number	Top row	Right to left	Score/ Grade
5	8	5	Number	Top row	Right to left	Score/Grade
6	8	5	Number	Top row	Right to left	Score/Grade
7	8	5	Number	Top row	Right to left	Score/Grade
8	8	3	Text/Number	Top row	Right to left	Score
9	8	-	-	-	-	Score/Standard
10	8	-	-	-	-	Score/Standard
11	8	-	-	-	-	Score/Standard
12	8	-	-	-	-	Score/Standard
13	8	3	Text/Number	Top row	Left to right	Score
14	8	6	Text	Top row	Left to right	Standard
15	8	5	Number	Top row	Right to left	Score
16	9	4	Text/Number	Top row	Left to right	Score/Grade
17	9	3	Text/Number	Top row	Left to right	Score
18	9	6	Text	Top row	Left to right	Standard
19	10	6	Text/Number	Top row	Left to right	Score
20	7 to 9	4	Text	Top row	Right to left	-
21	7 to 9	5	Text	Left column	Bottom to top	-
22	7 to 9	5	Number	Top row	Left to right	Score
23	-	7	-	-	Left to right	-
24	-	4	Text/Number	Top row	Left to right	Score

Note. The use of a dash (-) indicated absence of the relevant item

7.3.2.1 Levels

Table 7.4 indicated that of the rubrics provided, 19 (79.2%) generated a score, grade, percentage or reference to a standard. Of those rubrics that generated a score or equivalent, the most prevalent total score of 25 marks comprised almost half of these rubrics. As the central organising concept of this theme (Braun & Clarke, 2013, 2021a), most artefacts provided further support that “physical education teachers continue to use assessment solely to grade students” (Moura et al., 2021, p. 388). The submitted rubrics assessed between three to seven levels of performance, with five levels of performance the most frequent configuration. The prevalence of five levels of performance and the generation of scores may indicate adherence to the Victorian government and Catholic school mandate to employ a five-point scale when reporting on student achievement and progress (DET, 2023c). The literature on rubrics is equivocal about the number of performance levels, however, in a primary and secondary school context four or less levels are advocated to support consistent judgement (Griffin, 2014). It is not clear if the participants submitting rubrics with five or more levels were able to make meaningful distinctions in student performance. Nor is it clear what empirical base supported the DET mandate to report student achievement across a five-point scale (DET, 2023c) that may have contributed to the formatting of the submitted rubrics.

A practical suggestion for improved rubric design is to include as many levels as “the number of reliable distinctions in student work that are possible and helpful” (Brookhart, 2018, p. 2). In expounding this view on the reliability of rubrics, two levels of performance can be reliably scored with minimal training (Hall et al., 2015; Williams & Rink, 2003), however, the information they provide to students and teachers may be limited. In contrast, rubrics with four or more levels of performance are more difficult to reach satisfactory levels of inter-rater agreement (Griffin, 2014; Hall et al., 2015; Williams & Rink, 2003). The implication for the submitted artefacts, given that most rubrics included four or more levels, is that they may have limited inter-rater reliability. Within a standards-based curriculum that describes a learning continuum in two-level bands like the VC: HPE (VCAA, n.d.-f), it may be manageable for practitioners to describe the standard of a criterion at Year Level 7 - 8, and adapt this description to indicate performance that is below the standard (Year Level 5 – 6) and above the standard (Year Level 9 – 10). A fourth level of performance might cater for students that were absent, injured or unable to demonstrate the lowest level of practical performance. Thus, in designing a rubric that requires three levels of performance-based judgement, more meaningful distinctions in

the quality of performance and increased inter-rater reliability may be realised (Hall et al., 2015; Williams & Rink, 2003).

7.3.2.2 Outcome reported

Consistent with the literature review of rubrics by Reddy and Andrade (2010), a detailed discussion of scoring strategies and the challenges of calculating final grades from rubrics is beyond the scope of this study. Instead, by removing scores, percentages and grades from rubrics, students may become more concerned with learning than with grading (Brinson, 2022; Panadero & Jonsson, 2020). This view is underpinned by having students focus on rubric descriptors, rather than scores, to better understand where they are in relation to their learning goal and their next steps in learning (Brookhart, 2018). Removing scores or grading information may be potentially challenging for participants, given the proclivity for rubrics to be aligned to school mandated reporting structures described in Chapter 6.

As shown in Figure 7.3, rubric 10 used the numbers 0 to 5 in rubric cells and this data item was coded as descriptors rather than levels. In contrast, most rubrics displayed level labels in the top row of their tables. The location for levels in the top row of table formatted rubrics has been similarly reported in the synthesis of the rubric literature by Dawson (2017). In labelling the levels of performance, text, numbers and a combination of the two were used at similar rates across the corpus. Given that the sample population acknowledged their use of the VC: HPE in this study, labelling levels of performance with reference to performance against the Achievement Standards may be more appropriate in their rubrics than numerical forms which link directly to grading and reporting.

No artefact had their performance levels mapped to Bloom's revised taxonomy (Krathwohl, 2002), or the Structure of Observed Learning Outcomes (SOLO) taxonomy (Dudley et al., 2016). This is not surprising given the dominance of the psychomotor domain in the assessment of IGS found in tools located in the ScR (Chapter 4). Beyond the scope of this document analysis, investigations into a psychomotor taxonomy, like the five-stage skill acquisition model described by Dreyfus (2016) may provide empirical support and guidance for the naming of performance levels in IGS rubrics. As examples of the text used to describe performance levels in the sample, rubric 16 used "N/A, low, medium, high", rubric 20 used "Beginning, developing, capable, exceptional" and rubric 18 used "Not

shown, emerging, working towards the standard, at the standard, above the standard, well above the standard”.

The phrasing used at the lower band of achievement in some rubrics was inconsistent with the strengths-based approach endorsed by the VC: HPE (VCAA, n.d.-e). For example, rubric 8 and 21 described the lowest level of achievement as “needs work” and “needs significant improvement”, respectively. Students receiving deficit-based messaging, like the above, may lack motivation to make progress (Tierney & Simon, 2004). In adopting a strengths-based perspective as described in the VC: HPE (VCAA, n.d.-e), teachers are encouraged to design rubrics that label the levels of performance positively. These performance levels should be based on what students *can* do, know, or understand (Lambert & O'Connor, 2023). Specific to this study, teachers of the VC: HPE are asked to “Think about the rubrics you use for assessment. Are they focused on what’s not happening? Or are they focused on what can be achieved and the way forward?” (Lambert & O'Connor, 2023). Where a student is unable to demonstrate a skill, a neutral statement like ‘not yet shown’, ‘not yet demonstrated’ or ‘insufficient evidence’ may be more consistent with the strengths-based perspective of the VC: HPE (VCAA, n.d.-e).

7.3.2.3 Outliers

As an outlier in rubric orientation and level labels, rubric 21, positioned each level label in the left-hand column and described the levels of quality as “Excellent, Good, Satisfactory, Needs improvement, and Needs significant improvement”. The extant literature on rubrics provides no advice for the location or ordering of performance levels in rubrics. Regardless of the orientation of rubrics presented in table format, consistent application of formatting options in each Learning Area or school may enhance their readability for stakeholders. Along with rubric 23 presented in Figure 7.4, the four rubrics from Frank were the only others in the corpus that did not generate a score or grade.

Figure 7.4

Rubric 23 – An Instructional Rubric

Student Name:			Class:			Teacher:		
I can... pass a ball	Insufficient evidence	Pass a ball towards a stationary target while stationary	Pass a ball towards a moving target while stationary	Pass a ball towards a moving team-mate while in motion	Pass a ball to the 'advantage' of a moving team-mate while in motion (by passing the ball within the receiving range of the teammate)	Pass a ball to the 'advantage' of a moving team-mate while in motion by using more than 1 type of pass	Pass a ball while moving to the 'advantage' of a moving team-mate using more than 1 type of pass, within a competitive situation	
I can... receive a ball	Insufficient evidence	Receive a pass whilst stationary	Receive a pass whilst moving	Receive a pass whilst in an open/advantageous position	Receive a pass in an open/advantageous position whilst under pressure from opposition			
I can... take up an attacking position	Insufficient evidence	When my team is in possession of the ball, I move	I take up a position to receive the ball	I take up a position to receive the ball when the ball carrier is ready to pass	I take up a position to receive the ball when the ball carrier is ready to pass and use body /verbal language to communicate this when required	I take up a position to draw or block a defender to the advantage my team	I take up a position to draw or block a defender to the advantage my team within a competitive situation	
I can... take up a defensive position	Insufficient evidence	When my team is not in possession of the ball, I defend an opponent/ space	I actively defend an opponent/ space (for example: by adjusting my position, being quick to respond to ball movement...)	I actively defend an opponent/ space closer to goal than my opponent	I actively defend an opponent/ space closer to goal than my opponent (for example: by adjusting my position, being quick to respond to ball movement...)			
I can... collaborate	Insufficient evidence	Work with others to achieve a common goal	Actively work with others to achieve a common goal (for example: by communicating, delegating, contributing ideas, encouraging others...)	Actively and respectfully work with others to achieve a common goal (for example: by accepting viewpoints that are different to my own)				
I can... demonstrate fair play	Insufficient evidence	Follow the referee's decision	Follow the referee's decision respectfully (by accepting viewpoints that are different to my own)					

In addition to not generating a score as an outlier in this theme, rubric 23 in Figure 7.4 made no use of level labels and assessed criteria at different levels (from 3 to 7). This rubric is presented above for readers to see the overall nuanced format rather than read any descriptors. Given the absence of a scoring scheme, this rubric may be classified more as an instructional guide than a scoring guide (Andrade, 2005; Siedentop et al., 2011). The inference is that rubrics that did not generate a score or grade may have been used as AfL, which is advocated by the AISEP position statement on assessment in PE (AISEP, 2020). As almost 80% of the artefacts generated a score, grade or reference to a standard, these findings potentially provide further support that assessment is synonymous with reporting in PE (Moura et al., 2021; Scanlon et al., 2023; Veal, 1988).

7.3.3 *Appropriate Criteria are Key*

Theme 3 focused on the total number of assessable criteria in the rubrics; their positioning in the table format; their alignment to aspects of the VC: HPE; and their use of key performance criteria identified in the ScR (Chapter 4). The naming of this theme acknowledged the literature review of rubrics in higher education by Brookhart (2018) aptly titled, “Appropriate Criteria: Key to Effective Rubrics”. An overview of rubric criteria presented in Table 7.5 provides the backdrop for this theme.

Table 7.5

Theme 3 - An Overview of Rubric Criteria

Rubric No	Year Levels	Number of criteria	Criteria location	Reference to the VC: HPE	Reference to the ScR criteria
1	7	6	Left column	-	X
2	7	4	Left column	-	X
3	7	8	Two left columns	X	X
4	8	5	Left column	X	X
5	8	5	Left column	X	X
6	8	5	Left column	X	X
7	8	5	Left column	X	X
8	8	4	Left column	-	X
9	8	5	Second row	-	X
10	8	5	Second row	-	X
11	8	5	Second row	-	X
12	8	5	Second row	-	X
13	8	8	Two left columns	X	X
14	8	6	Left column	-	X
15	8	6	Left column	-	X
16	9	8	Left column	X	X
17	9	8	Two left columns	X	X
18	9	5	Left column	-	X
19	10	8	Two left columns	X	X
20	7 to 9	2	Left column	-	X
21	7 to 9	2	Top row	-	X
22	7 to 9	5	Left column	-	X
23	-	6	Left column	-	X
24	-	8	Two left columns	-	X
<i>N</i> =				9	24
Percentage				37.5	100.0

Note. The use of a dash (-) indicated absence of the relevant item

7.3.3.1 Criteria number

Given the ability of this theme to respond to the research sub-question, and its capacity to make comparisons with data captured in the preceding studies of this thesis, it was the most substantive theme in the chapter and presents data in multiple tables. The rubrics comprised between two and eight criteria, with five criteria the most prevalent configuration. This is similar to findings with analytical rubrics used in higher education that reported three to five criteria were used most commonly (Brookhart, 2018). Most criteria were displayed in the left-hand column(s) which is consistent with the findings of the rubric synthesis of Dawson (2017). As there is no optimal number of rubric criteria suggested in the literature, rubric designers should only use as many criteria as required to indicate the desired learning outcome (Brookhart, 2018; Chappuis et al., 2012). This suggestion complements the previous suggestion in theme 2 to limit the number of levels of performance to what is needed to meet the learning goal(s), and what can be observed (Brookhart, 2018). Taken together, these suggestions indicate that rubrics should be more economical than expansive, and learning goals may need to be modest. A pragmatic issue for rubrics with an excessive number of criteria is that they may lack utility and are likely to “gather dust” (Popham, 1997, p. 74).

7.3.3.2 Links to curriculum

More than a third of the submitted rubrics made explicit reference to the VC: HPE, as determined by ‘word for word’ reference to relevant Achievement Standards or Content Descriptors from Year Levels 7 - 10 (VCAA,n.d.-d). Codes 30 to 47 in Appendix O provide excerpts from the curriculum that were identified in the artefacts. While it is possible that the VC: HPE had been adapted into student-friendly language in the rubrics, coding adaptations to the VC: HPE curricula statements was considered high inference and not used in this study. Similarly, unit plans or assessment policy documents were not requested for submission in this study, and it is possible that curriculum alignment may have been addressed in these documents. Cognisant of the limitations of these coding guidelines, the low level of curriculum alignment in the rubrics supported findings in the ScR (Chapter 4), where only 2 of the 15 school-developed tools were charted as having links to a curriculum. The low level of curriculum alignment in the submitted documents partially supported findings in the cross-sectional, quantitative inquiry (Chapter 5) that indicated less than half of the participants agreed that curriculum informed their assessment. The alignment of assessment in IGS to the VC: HPE indicated by the sample in Chapter 5 and this study is therefore considered low to

modest. This finding supports the historical view that assessment in PE, including GS, lacks alignment to curriculum (Lambert & Penney, 2020; Morrow, 1978; Williams et al., 2020). As an example of references to the VC: HPE from the corpus, the first two columns of the excerpt from rubric 3 presented in Figure 7.5 include verbatim statements from the Achievement Standards and Content Descriptors for Year Levels 7 - 8 (VCAA, n.d.-d).

Figure 7.5

Rubric 3 (Excerpt) - Criteria Linked to the VC: HPE

Organising Element	Action
Students explain personal and social skills required to establish and maintain respectful relationships and promote fair play and inclusivity.	Modify rules and scoring systems to allow fairplay, safety and inclusive participation.
Students demonstrate control and accuracy when performing specialised movement skills.	Use feedback to improve body control and coordination when performing specialised movement skills.
They apply and refine movement concepts and strategies to suit different movement situations.	Practice, apply and transfer movement concepts and strategies
Uniform	Wears uniform

Note. The first three criteria in both columns are linked to Year Levels 7 – 8 VC: HPE (VCAA, n.d.-d).

The top three criteria in the left-hand column were coded as items 30, 31 and 32 in Appendix O, while the first three criteria in the second column were coded as items 37, 33 and 34. As an example of non-performative criteria that are addressed later in this theme, the fourth criteria in both columns of the excerpt from rubric 3 are not found in the VC: HPE. The terminology used to describe the

criteria in rubric 3 is consistent with other documents from the same participant (Kelly) and is a rare example in the corpus that adopted the naming conventions and format found in support material from the VCAA (2019b) that is presented in Table 7.6.

Table 7.6

VC: HPE Invasion Game Rubric

Organising element	Action	Insufficient evidence	Quality criteria			
Movement skills	1 Performs movement skills	1.0 Insufficient evidence	1.1 Practise fundamental movement skills in isolation with prescribed technique.	1.2 Demonstrates movement skills in structured situations.	1.3 Apply movement skills to a modified game situation.	1.4 Perform movement skills with technique in pressure situations.
Movement concepts	2 Uses movement strategies	2.0 Insufficient evidence	2.1 Follows instructions to perform a set strategy in isolation.	2.2 Implement different movement strategies in structured situations.	2.3 Select and apply movement strategies in a game context.	2.4 Construct, apply and refine movement strategies in a game context.
Movement challenges	3 Solves movement challenges	3.0 Insufficient evidence	3.1 React to cues and select a strategy.	3.2 Attend to cues to solve structured movement challenges.	3.3 Select a solution to movement challenges using cues in a game situation.	3.4 Reflect on past game situations to solve movement challenges.

Note. From *Using Formative Assessment Rubrics in Health and Physical Education: Invasion, Net/Wall, Striking and Fielding Games - Levels 7 to 10*, by the VCAA, 2019b, (p. 7). Copyright 2019 by Victorian Curriculum and Assessment Authority. Reprinted by permission.

This ‘dual’ criteria structure of ‘organising element’ and ‘action’ was used in five rubrics from two participants and its utility in comparison to a single column of criteria remains unclear. It cannot be established how familiar participants in this study were with this VCAA source material, or what degree of guidance it provided to teachers that may have used it. For example, there is no advice about which quality criteria align to Year Levels 7 - 8 or Year Levels 9 – 10 and there are no specific on-the-ball skills or off-the-ball movements described. The rubric may lack internal consistency given that reflecting on past game situations (quality criteria 3.4) is likely to involve a written or verbal task

assessing declarative knowledge in the cognitive domain, in contrast to the lower levels of the same criterion that assess procedural knowledge in the cognitive and psychomotor domains. Given the potential lack of clarity, consistency and specificity within the rubric, reliable application in self or peer assessment is uncertain. It is possible that this sample rubric perpetuates a lack of sport and skill specificity in the assessment of GS reported with reference to the AC: HPE (Williams et al., 2020). As a potential support for users of the rubric, hyperlinks in the document to sample activities may assist assessment across the levels of quality (VCAA, n.d.-g).

7.3.3.3 Issues with clarity

Contributing to the naming of this theme was the finding that some criteria did not clearly indicate what was being assessed and may have been inappropriate. Single word criteria like “tactical” (rubric 20) and “teamwork” (rubric 7), do not describe what aspects of performance are being assessed. For example, while teamwork “sits comfortably within the vocabularies of most physical education teachers” (Barker & Rossi, 2012, p. 1), it is not clear how this term was understood by participants. Teamwork may represent various elements of the VC: HPE including personal and social skills, respectful relationships, fair play, inclusivity, cooperation, leadership and working collaboratively (VCAA, n.d.-f); but as a single word, it is not clear what meaning is conveyed to users. Other criteria were non-specific in nature; for example, “ability to perform relevant skills” (rubric 24) was not accompanied by any reference to what constituted *relevant* skills. Equally, the criterion “Offensive/Defensive movement” (rubric 15) provided no reference to what movement in either phase of play involved. While the rubric criteria were generally accompanied by descriptors to support their meaning, and this element of rubrics is discussed in theme 4 of this chapter, many criteria in the corpus may be considered ill-defined or inappropriate. An outlier in the corpus was rubric 8 as it assessed invasion game *design* rather than invasion game performance. Details of the specific key performance criteria assessed are provided in Table 7.7.

Table 7.7

Theme 3 - References to Key Performance Criteria

Rubric No	Year Levels	On-the-ball skills	Receive	Dribble	Pass	Score	Defend	Off-the-ball movement	Decision-making	Other criteria
1	7	X	X	-	X	-	X	X	X	-
2	7	X	X	X	X	-	X	X	X	-
3	7	X	-	-	-	-	-	-	X	X
4	8	X	-	-	-	-	X	X	X	X
5	8	X	-	-	-	-	X	X	X	X
6	8	X	-	-	-	-	X	X	X	X
7	8	X	-	-	-	-	X	X	X	X
8	8	-	-	-	-	-	-	-	-	X
9	8	X	X	X	X	-	X	-	-	-
10	8	X	X	-	X	-	X	-	-	-
11	8	X	X	X	-	-	X	-	-	-
12	8	X	-	-	X	-	X	-	-	X
13	8	X	-	-	-	-	-	-	X	X
14	8	X	-	X	X	X	X	X	X	X
15	8	X	X	-	X	-	X	X	X	X
16	9	X	-	-	X	-	-	X	-	X
17	9	X	-	-	-	-	-	-	X	X
18	9	X	-	X	X	X	X	X	X	X
19	10	X	-	-	-	-	-	-	X	X
20	7 to 9	X	X	X	X	X	X	X	X	-
21	7 to 9	X	X	X	X	X	X	X	X	-
22	7 to 9	X	-	-	X	X	-	X	X	X
23	-	X	X	-	X	-	X	X	-	X
24	-	X	-	-	-	-	-	X	X	X
<i>N</i> =		23	9	6	13	5	16	15	17	17
Percentage		95.8	37.5	25.0	54.2	20.8	66.7	62.5	70.8	70.8

Note. The use of a dash (-) indicated absence of the relevant item. The above key performance criteria were identified in the ScR (Chapter 4). Further to the totals provided in the bottom rows, more than half of all rubrics assessed on-the-ball skills, off-the-ball movement *and* decision-making.

7.3.3.4 Links to the scoping review

As most rubrics assessed decision-making (tactical play) and off-the-ball movement in authentic (game-based) situations, this is counter to the assessment criteria and conditions reported in the assessment literature (for example, Blomqvist et al., 2005; Borghouts et al., 2016). Some rubrics indicated that assessment tasks included practising FMS in isolated drills at the lowest level of achievement and executing complex skills in “game situations ... modified game situations ... [and] dynamic game situations” at higher levels of achievement (rubric 24). The latter rubric suggested students were assessed in playing form, that required an opposition and decision-making (Ford et al., 2010). Other references to playing form can be seen in rubrics with references to “invasion games” and “modified game” (for example, rubric 4), “opposition pressure” and “teammates” (rubric 15), and “fairplay” (for example, rubric 3), suggesting that most rubrics were used to assess students in authentic game-based play.

Of particular interest is that defending was the most assessed on-the-ball skill, while scoring was the least assessed. While both scoring and defending are key aspects of play in IGS (Memmert & Harvey, 2010; Ward & Griggs, 2011), it has been reported that few assessment tools assess both attack and defence, and that assessing attacking play is more prevalent (Arias & Castejón, 2012; Barquero-Ruiz et al., 2019). As an example, the TSAP (Gréhaigne et al., 1997) focusses on assessing attacking play with only one of six variables addressing defensive play (conquered ball). While some games like netball only allow certain players in a team to score, and FSG may be low scoring in comparison to SSG, the difference in the rates of assessment in scoring and defending as criteria is not easy to explain. One potential reason may be that the assessment of scoring may direct students to focus on winning and losing. Criticisms have been levelled at PE programs that emphasise competition as they can marginalise less skilled students (Bernstein et al., 2011; Bevans et al., 2010; Bryan et al., 2013; Garn et al., 2011). Overall, the criteria used in rubrics in this study supported findings in the ScR (Chapter 4) and the cross-sectional, quantitative inquiry (Chapter 5). Table 7.8 presents a comparison of key performance criteria across these studies.

Table 7.8*Key Performance Criteria Assessed in Relevant Studies Comprising the Thesis*

Key performance criteria	Scoping Review (Chapter 4)		Cross-sectional quantitative inquiry (Chapter 5)		Document analysis (Chapter 7)	
	<i>n</i> = 15 tools	%	<i>N</i> = 80	%	<i>N</i> = 24 rubrics	%
On-the-ball skills	14	93.3	60/72	83.3	23	95.8
Receive	11	73.3	50/70	71.4	9	37.5
Dribble	9	60.0	57/70	81.4	6	25.0
Pass	11	73.3	62/70	88.6	13	54.2
Score	10	66.7	34/70	48.6	5	20.8
Defend	7	46.7	Various*	55.0	16	66.7
Off-the-ball movement	12	80.0	57/68	83.8	15	62.5
Decision-making	11	73.3	69/71	97.2	17	70.8
Other	2	13.3	NA	NA	17	70.8

*Note**: Only the 15 tools designed in a school context have been included from the ScR. The maximum sample size for each item in the cross-sectional, quantitative inquiry (Chapter 5) was 80. Various* was derived by averaging the results for intercept/steal, block an attacking play, clear the ball, and tackle to generate an equivalent indicator for defend.

Like the charting in the ScR (Chapter 4), 'Other' criteria in the rubrics were not coded in any further detail. However, examples included teamwork or collaboration (for example, rubric 8), written work (for example, rubric 14), organisation (for example, rubric 24), participation (for example, rubric 15), effort (rubric 22) and wearing uniform (for example, rubric 19). Criteria like effort and wearing uniform do not inform the achievement standards of the VC: HPE and their inclusion on the submitted rubrics supports the wider literature that assessment in PE often includes criteria that are not aligned to a curriculum (for example, Baghurst, 2014; Blomqvist et al., 2005).

Despite the similarity in figures for the overall assessment of on-the-ball skills, most ball skills were assessed at much lower rates in the rubrics than the tools located in the ScR (Chapter 4) and the cross-sectional, quantitative inquiry (Chapter 5). This may be partially explained by the high rate of generic rubrics in the current study that did not identify specific sport skills. For example, generic rubrics preferred terms like "movement skills" (for example, rubric 17), "skills" (rubric 16), and "gameplay attacking" (for example, rubric 9). The prevalence of decision-making and off-the-ball skills across the rubrics suggests that authentic and holistic assessment of game performance was common practice within the sample. Of note, the assessment of 'Other' criteria in the submitted rubrics was much higher than in the tools located in the ScR (Chapter 4). This may be explained by assessment in PE needing to address other skills, knowledge and understandings beyond pure game performance.

7.3.4 *The Challenge of Describing Quality*

Closely linked to the importance of appropriate criteria, rubric descriptors provide the benchmark or anchor for assessing performance. As per the classification of rubric elements presented in the method section of this chapter, *descriptors* serve to describe the levels of quality within a criterion. In the typical table format, the descriptors constituted the text in each cell of a rubric. To support the following findings and discussion section, Table 7.9 defines five different types of descriptors based on the classification system of Brookhart (2018).

Table 7.9*Types of Rubric Descriptors*

Type of descriptor	Definition	Example from the rubric literature
Evaluative (scale)	Includes terms like excellent, good, satisfactory, competent, and beginning	“Excellent technical skills in passing, catching, shooting and dribbling” (Breed & Spittle, 2011, p. 183)
Frequency-based (scale)	Defines performance in terms of occurrences without referring to numbers, for example, consistently, usually or rarely	“Rarely demonstrates basketball skills, strategies, or knowledge of rules” (Williams & Rink, 2003, p. 606)
Numerical (scale)	Counts the number of times an action occurs or the number of visible performance elements, for example, once, two or more times, 50% of the time	“Demonstrates three of five components well” (Shaw, 2014, p. 36)
Ordinal (scale)	The use of a number to represent a level of performance	“Demonstrates all critical elements of the motor skills – 1, 2, 3, 4” (Wang & Rairigh, 2006, p. 39)
Qualitative	Describes an example of what the criterion looks like at varying levels of quality	“Receives ball on the move with head up” (Harvey, 2007, p. 49)

Note. The above types of descriptors are based on the classification system of Brookhart (2018). In an extension of the original classification system, ordinal (scale) has been added in response to the ‘descriptors’ in rubrics 9, 10, 11 and 12. The first four types of descriptor are followed by the term scale in parenthesis to indicate that these descriptors use rating scale language. Further, to promote clarity in the terminology employed by Brookhart (2018), the term *qualitative* replaces *descriptive* in the original to avoid the tautology ‘descriptive descriptor’. The term qualitative may also help to emphasise the need for the descriptor to indicate the *quality* of performance.

Underpinning this theme, the rubrics were coded based on the five types of descriptors presented in Table 7.9. In addition, descriptors that provided multiple indicator(s) of performance quality, and rubrics that included no descriptors (empty cells), were also coded. Collectively these seven data items contributed to the naming of the theme and indicated a potential barrier to rubric design and utility and are presented in Table 7.10, The Challenge of Describing Quality.

Table 7.10

Theme 4 - The Challenge of Describing Quality

Rubric No	Year Levels	Evaluative (scale)	Frequency-based (scale)	Numerical (scale)	Ordinal (scale)	Qualitative description	Multiple indicators	Empty cells
1	7	-	X	-	-	X	-	-
2	7	X	X	-	-	X	-	-
3	7	-	-	X	-	X	-	X
4	8	X	-	-	-	X	X	X
5	8	X	-	X	-	X	X	X
6	8	X	-	X	-	X	X	X
7	8	X	-	-	-	X	X	X
8	8	X	X	-	-	-	X	-
9	8	-	-	-	X	-	-	X
10	8	-	-	-	X	-	-	X
11	8	-	-	-	X	-	-	X
12	8	-	-	-	X	-	-	X
13	8	-	-	X	-	X	-	X
14	8	X	X	-	-	X	X	-
15	8	X	X	-	X	X	X	-
16	9	X	X	X	-	X	X	-
17	9	-	-	X	-	X	-	X
18	9	X	X	-	-	X	X	-
19	10	-	-	X	-	X	-	X
20	7 to 9	X	X	-	-	X	X	-
21	7 to 9	X	X	-	-	-	X	X
22	7 to 9	X	X	X	-	X	X	-
23	-	-	-	X	-	X	X	X
24	-	X	X	-	-	X	X	-
<i>N</i> =		14	11	9	5	18	14	14
Percentage		58.3	45.8	39.5	20.8	75.0	58.3	58.3

Note. The use of a dash (-) indicated absence of the relevant item.

7.3.4.1 Rating scales versus rubrics

In applying the classification of Brookhart (2018), no rubrics constituted a “true rubric” (p. 1). This type of rubric is characterised by the exclusive use of qualitative descriptors (termed descriptive in the original). Two rubrics were close to meeting this definition, as rubrics 4 and 7 were predominantly framed in qualitative language except for a single evaluative term (proficient). Figure 7.6 shows the first three descriptors within the first criterion of the excerpt from rubric 4. The descriptors were coded as qualitative language at level four and five, and evaluative language at level three due to the term “proficient”, as there was no description of what developing proficient space meant.

Figure 7.6

Rubric 4 (Excerpt) - Sample Qualitative and Evaluative Descriptors

Criteria	5	4	3
<i>Feedback to improve body control and coordination when performing specialised movement skills</i>	Student is able to use a qualitative analysis to give feedback to others and or self around technique	Is able to use ICT to assess and provide feedback on use of space	Develop proficient space based on feedback by teacher or peers

The use of evaluative terms provides no description of what performance looks like, and as with other scale language, such terms do not offer “students a description of the quality of their performance they can easily use to envision their next steps in learning” (Brookhart, 2018, p. 1). These terms have also been described as comparative language and may contribute to inconsistent marking in addition to providing limited guidance to students (Griffin, 2014). Ordinal descriptors were least prevalent, while qualitative descriptors were most prevalent. As most rubrics used a combination of two or more types of descriptor, and rating scale language was pervasive, this theme title captured the challenge of describing quality in performance. As an example, the subjective nature of evaluative language like sometimes, usually and often, provides no concrete direction for teacher instruction or student learning (Brookhart, 2018). Instead, the use of evaluative, frequency-based and numerical descriptors may be better suited for grading purposes rather than learning (Brookhart, 2018). The addition of ordinal scale to the classification of descriptors by Brookhart (2018) is encouraged.

The greater use of qualitative or rich descriptors (Bargainnier, 2003; Lund, 2006) in rubrics 4 and 7, in comparison to the corpus, does not necessarily suggest these rubrics were models of excellence, or that their criteria were better understood by users. For example, the first qualitative descriptor in the excerpt from rubric 4 in Figure 7.6 does not differentiate between any types of qualitative analysis, the nature of the feedback provided to peers, or what technique is critiqued. However, promoting the use of well-crafted qualitative descriptors is predicated on the belief that quality descriptions of performance are “the key to effective rubrics” (Brookhart, 2018, p. 10). A range of complementary strategies to support rubric utility include engaging students in the co-design of rubrics, providing students early access to rubrics to understand the learning goals, checking for student understanding of the criteria, and applying the rubrics to exemplars of performance to support transparency in assessment (Matshedisho, 2019).

7.3.4.2 Multiple indicators

Excluding the four rubrics that used an ordinal number scale in their descriptors, over two-thirds of the remaining rubrics included multiple text-based indicators of performance within their descriptors. The inherent problem of analytical rubrics including multiple assessable items within a single cell/descriptor is how an assessor recognises *partial* attainment of the descriptor (Griffin, 2014). For example, the second descriptor in the excerpt from rubric 4 in Figure 7.6 includes two items as the student is required to assess *and* provide feedback. Pragmatically, if the rubric is used in a hard copy format the assessor has the capacity to tick which indicators or sub-elements were met in each cell, but this functionality may not exist if the rubric is marked electronically. Further, if the score allocated to a cell containing the descriptor is one point, it is not clear how many indicators must be observed to score that one point. This issue is similar to the problem of two-part questions in questionnaires, when a respondent may not have a consistent response to both parts of a question (Ary et al., 2014). One solution for this practice is that assessment designers consider if the overarching criterion should be split into two or more separate criteria (Chappuis et al., 2012). In contrast to the challenge of assessing multiple indicators in cells, more than half the rubrics included empty cells that did not describe performance at all. Both aspects of multiple-indicators of performance and empty cells are evident in rubric 21 presented in Figure 7.7.

Figure 7.7

Rubric 21 – A Novel Rubric

ASSESSMENT	TECHNICAL	TACTICAL
Excellent	<ul style="list-style-type: none"> • Consistently maintains and control possession of the ball under pressure (when to run or pass, where to pass). • Consistently moves the ball to create scoring opportunities • Consistently executes the best type of pass in a given situation • Consistently demonstrates coordinated and fluent passing, dribbling and shooting skills • Consistently demonstrates correct body positioning in defensive situations 	<ul style="list-style-type: none"> • Student consistently knows where and when to pass, run or dribble when in possession (decision making) to maximise scoring opportunities • Student consistently creates space to gain advantage by using fakes and/or leads • Consistently demonstrates anticipatory skills (offense and defence) in game situations • Consistently moves into position which limit/ deny oppositions playing space • Consistently regains possession of the ball by defending a player or the ball
Good	Usually...	
Satisfactory	Sometimes...	
Needs improvement	Rarely...	
Needs significant improvement	Never...	

Figure 7.7 is an analytical rubric comprising two criteria described across five levels of performance from “excellent” to “needs significant improvement”. At the highest level of performance, there are multiple indicators (five) that comprise the descriptors. At the remaining levels of performance, frequency-based descriptors from “never” to “usually” are provided for the first criterion, while the cells in the second criterion remain entirely blank. Rubric 21 in Figure 7.7 was considered novel as it was the only rubric that presented levels of performance in rows, rather than in columns, and the only rubric that had more empty cells than descriptors. The literature is equivocal about which way to orient rubrics, although consistent orientation within Learning Areas, and possibly whole schools, may support the readability of rubrics by stakeholders. Regarding rubrics comprising empty cells, this may be preferable to the inclusion of arbitrary or unclear distinctions in student work (Brookhart, 2018).

7.3.4.3 Inconsistent descriptors

A further challenge in describing the level of performance in rubrics was found with inconsistency *across* descriptors representing a single criterion. Such inconsistency has been reported in the rubric literature as a threat to their utility (Tierney & Simon, 2004). To provide an illustrative example, rubric 24 presented in Figure 7.8 assesses multiple aspects of performance in the second criteria “Attitude & Effort” and the sub-criteria “Participation in class during unit”. Of note, attitude and effort do not feature in the VC: HPE at any Year Level (VCAA, n.d.-d); the term effort is only used to represent the biomechanical principle of force, rather than a student’s degree of energy or application (VCAA, n.d.-d).

Figure 7.8

Rubric 24 – Inconsistent and Incongruent Descriptors

capabilityMarkType: percent	useMarkRange: On	Beginner (1-4)	Developing (5-6)	Proficient (7-8)	Expert (9-10)
Skill Acquisition & Performance	Ability to perform relevant skills.	Students practices fundamental skills in isolation with prescribed technique.	Student demonstrates an ability to execute fundamental skills within game situations.	Student selects relevant skills and executes them in response to modified game situations.	Student selects and precisely executes complex skills in response to dynamic game situations.
10		4	6	8	10
Attitude & Effort	Participation in class during unit.	Attends classes.	Participates in a compliant manner, seeking small opportunities for improvement.	Enthusiastic participation coupled with the motivation to improve their ability. Actively supports peers and can accept team role responsibilities.	Participates with intensity and seeks feedback for improvement. Actively supports peers and acts as a role model in enacting fair play rules and team objectives.
10		4	6	8	10
Organisation	Organisation for class during unit.	Limited organisation of equipment for class.	Sometimes attends class with appropriate materials and uniform. Sometimes uses class time effectively.	Mostly attends class with appropriate materials and uniform. Mostly uses class time effectively.	Always attends class with appropriate materials and uniform. Consistently uses class time effectively.
10		4	6	8	10
Game Sense: Invasion Games	Ability to apply tactics during gameplay.	Student participates in a passive manner.	Student demonstrates a developing ability to apply decision making and movement to assist with maintaining possession, blocking space to protect goal and applying pressure to force opposition errors.	In basic gameplay and practice, student demonstrates decision making and movement to assist with maintaining possession, blocking space to protect goal and applying pressure to force opposition errors.	In complex gameplay, student consistently demonstrates precise decision making and movement to assist with maintaining possession, blocking space to protect goal and applying pressure to force opposition errors. They lead teammates to execute team objectives.

Within the criterion “Attitude & Effort”, the descriptor at the lowest level of performance states that the student “attends class”; this is followed by descriptors at increasing levels of quality that describe disparate elements of compliance, improvement, enthusiasm, motivation, peer support, accepting team roles, intensity, seeking feedback, role modelling and fair play. This lack of congruence and consistency in rubric descriptors *within* a criterion provides a key threat to effective rubric design (Tierney & Simon, 2004). Criticism may also be levelled at descriptors that assess similar attributes *across* criteria as they may reduce the breadth or scope of what is being assessed. For example, at the top level of performance for “Skill acquisition & Performance” and “Game Sense: Invasion games” the descriptors both refer to selecting complex skills and precise decision-making, effectively assessing the same attribute in both criteria and constraining the range of performance that can be assessed. The third criterion of “Organisation” assesses bringing equipment and wearing a uniform; these work habits support learning but are not identified in the VC: HPE Achievement Standards.

7.4 Recommendations

As per the knowledge created in the previous qualitative inquiry (Chapter 6), the observations made through the four themes in this chapter are not suggested as being universal to all PE teachers (Willis, 2012). As such, the following recommendations may not be generalisable beyond the local sample (McChesney & Aldridge, 2019). However, some of the recommendations may “meaningfully reverberate and affect an audience” (Tracy, 2010, p. 844) beyond PE teachers in Victorian secondary schools, making these recommendations contextually and analytically transferable (Smith, 2017; Tracy, 2010). It is incumbent upon other educators to critically examine any adoption of the following recommendations after reflecting on their assessment practice (Willis, 2012).

As a key recommendation from the narrow conceptualisation of rubrics, participants are encouraged to consider other rubric formatting options beyond a table layout and an analytical type. In embracing other formatting options, a relatively simple adjustment of amalgamating criteria in analytical rubrics into a single criterion to form a holistic rubric may support more time-efficient grading (Brookhart, 2018). Further, the use of single point rubrics may support AfL which is congruent with the AISEP position statement on assessment in PE (AISEP, 2020). The use of generic rubrics is consistent with GBAs like the TGfU model (Bunker & Thorpe, 1986; Thorpe et al., 1984; Werner et al., 1996), and reduces the need to create a large number of sport specific

versions. Generic rubrics also support guidance from the VCAA for students to transfer movement concepts/strategies and understanding from previous movement experiences across Year Levels 7 - 10 in the VC: HPE (VCAA, n.d.-d), as they allow for a range of sports to be assessed using the same assessment tool.

In broadening participants' conceptualisation of rubrics, embedding hyperlinked video or audio exemplars into digital rubrics may be useful in providing students with a greater understanding of assessment criteria at varying levels of performance (Ackermans, Rusman, Nadolski, et al., 2019). As an indication of partial support for the above innovation, one ancillary document comprising video exemplars of performance was submitted for this study. As this document included images of students it was excluded for ethical reasons. While no submitted rubrics made use of video or audio links, the embedding of multi-media in rubrics is evident in eRubrics (Company et al., 2019; Rivas et al., 2014), Video Enhanced Rubrics (VER), and the Viewbric (for example, Ackermans, Rusman, Brand-Gruwel, et al., 2019; Ackermans et al., 2021). Electronic rubrics may provide greater transparency for students in understanding the levels of performance and a benchmark for those completing the assessment to judge performance (for example, Lipnevich et al., 2013; Tierney & Simon, 2004). While the subject of PE is not specifically addressed in the above context, similar to the wider rubric literature discussed earlier in this chapter, the potential benefits of electronic rubrics may also apply to PE.

To support consistency in the naming of performance levels in rubrics used in PE and to avoid the pervasive use of scores and/or grades in level labels, adapting the naming conventions of the five-stage model of skill acquisition may be useful (Dreyfus, 2016). As an example, The terms of beginner, developing, proficient, and expert in rubric 24 align closely to the five levels identified by Dreyfus (2016). Adoption of this positively framed language is also consistent with guidance from the VCAA to adopt a strengths-based approach in the VC: HPE (VCAA, n.d.-e). Other recommendations from this study include designing more appropriate criteria and descriptors to support student learning (Brookhart, 2018). PE teachers and designers of rubrics in the field of education are encouraged to become familiar with the five types of descriptor identified in this chapter (see Table 7.9), understand the limits of scale type language, and privilege qualitative descriptors in their rubric design. Participants are also recommended that "For scoring rubrics to fulfil their educational ideal, they must first be designed or modified to reflect greater consistency in their performance criteria descriptors" (Tierney & Simon, 2004, p. 1). Thus, descriptors should

describe varying levels of quality within the same criterion (Griffin, 2014), and avoid assessing similar aspects of performance across criteria.

Given the lack of background information for the submitted artefacts in this study, it is suggested that some guidelines or protocols accompany each assessment rubric to help promote more reliable, fair and equitable assessment. As one of 14 rubric design elements identified by Dawson (2017), provision of an *explanation* might describe the nature and duration of the assessment task and offer advice on how to assess multiple indicators in a single descriptor of an analytical rubric. While determining the level of student involvement in the design and/or application of rubrics was not an aim of this study, the literature supports engaging students as assessors and co-designers of rubrics (Zheng et al., 2019). The use of self and peer assessment involving rubrics is also widely advocated (Andrade, 2007; Andrade & Valtcheva, 2009; Panadero & Romero, 2014; Panadero et al., 2012), and may support student self-regulation learning strategies (Fraile et al., 2017). While each of these recommendations is based outside of the subject of PE, the nature of their transferability remains unclear.

7.5 Strengths and Limitations

A major strength of this work was the combination of content and thematic analysis that led to the creation of four substantive themes that sought to explain and understand how the sample constructed their assessment rubrics for use in IGS. As the last of four studies comprising this thesis, another strength of the study was the integration of data based on key performance criteria found in evidence-based tools (Chapter 4) and comparisons to findings in the cross-sectional, quantitative inquiry (Chapter 5). As no similar study investigating assessment rubrics used by secondary school PE teachers in an IGS context within Victoria was located in the assessment in PE literature (Chapter 2), this study constitutes an original contribution to knowledge.

A common limitation in qualitative work is establishing the wholeness of the data, generally in respect to the sample size and saturation (Guest et al., 2006). Given a relatively modest sample of artefacts ($N = 24$), and the study's alignment to conceptual depth (Nelson, 2016) rather than saturation, the degree of comprehensiveness may be viewed as a limitation of the study. Claims of transferability and generalisability (Smith, 2017), as determined by the audience, may also be seen as a limitation of the study. The absence of contextual information regarding the purpose of the assessment, the designer(s), the user(s), student accessibility and validity or reliability measures of

the rubrics may also be seen as a limitation of the study. There was no capacity in this study to determine which IGS were used with the generic rubrics, how well students understood the criteria, or how consistently the rubrics were used within each participant's school. As this study was independent of any observations in the field, there was no information regarding the formative or summative use of the submitted rubrics. Without other background information, it is not clear which types of pedagogies were used, nor the conditions of assessment, such that the degree of instructional alignment cannot be established. Lastly, in choosing to code links to curriculum as verbatim references to Achievement Standards and Content Descriptors, it is possible that other curriculum references were not extracted in the coding protocol.

7.6 Conclusion and Further Research

The aim of this study was to explain and make meaning of how assessment tools (rubrics) were constructed by PE teachers in Victorian secondary schools in the assessment of IGS in Year Levels 7 - 10 within the VC: HPE. As with previous studies in this thesis, a reference point was the evidence-based assessment tools used in a school-age population located in the peer-reviewed literature (Chapter 4). To achieve this study's purpose, I followed the READ approach (Dalglish et al., 2021) and demonstrated conceptual depth (Nelson, 2016), rich rigor, credibility, and sincerity (Tracy, 2010). Patterns, anomalies and gaps across the artefacts were explained, which led to a series of recommendations for the sample population that may be transferable to the broader field of education.

The corpus established clear patterns in rubric design that included the use of a table layout, a generic rather than sport-specific focus, and an analytical rather than holistic format. Most rubrics generated a score, percentage or reference to a performance standard, suggesting their potential application in reporting student performance. The use of five criteria described at five levels of quality was the most prevalent configuration, which supported findings reported in the review of rubrics used in the context of higher education by Brookhart (2018) and suggested adherence to the mandated use of a five-point scale for reporting in government and Catholic schools in Victoria (DET, 2023c).

Many criteria did not adequately describe an attribute of performance, descriptors within a single criterion were not always clearly aligned, and in some cases, similar attributes in performance were assessed in multiple criteria. Despite this lack of clarity and consistency in criteria and descriptors, this study supported the prevalence of the three key performance criteria

identified in the ScR of Chapter 4 (on-the-ball skills, off-the-ball movement and decision-making). There was also support for the assessment of specific game skills identified in the ScR (Chapter 4), however, their rate of use in the submitted rubrics was lower for all skills except defend. There was modest alignment of the rubric criteria to the VC: HPE, as determined by verbatim references to the Achievement Standards and Content Descriptors. The widespread use of evaluative, frequency-based, and numerical rating scale language (Brookhart, 2018) may be one of the more significant findings of the study. It has been argued that rating scale language does not describe performance and therefore offers no information to students on how to progress their learning (Brookhart, 2018). Potentially contributing further to the limited utility of rubrics, most analytical rubrics had multiple indicators of assessment within their descriptors, which may reduce the efficacy of feedback to students if some indicators are met and others are not.

In proposing areas for further research, other document analysis studies in an IGS context across other Australian states and territories are encouraged. These studies may also collect and analyse data from unit plans, assessment policies, teacher interviews and student reports to help understand the complex phenomenon of assessment. Field observations of PE teachers' assessment practices, including the use of rubrics is also endorsed. As student views on assessment were not within the scope of this thesis, studies examining student perceptions of the use of rubrics in PE are supported. Future studies could aim to develop under-reported rubric formats, like holistic rubrics, single point rubrics, and VER rubrics that embed hyperlinks of exemplar performance (Ackermans, Rusman, Brand-Gruwel, et al., 2019). Investigating the levels of reliability and content validity of rubrics used in the field, including any of the rubrics submitted in this study, is also highly recommended. The findings and recommendations in this document analysis constitute an original contribution to knowledge in the discipline of PE. These contributions are offered as actionable knowledge (Kelly & Cordeiro, 2020) for educators within, and potentially beyond, a Victorian context.

CHAPTER 8: CONCLUSION

8.1 Introduction

This chapter summarises key findings from the four studies that comprise the explanatory, sequential, mixed methods research design (Creswell & Plano Clark, 2018). Following this summary, a translation of the overall findings is presented through the conceptualisation of two frameworks that provide guidelines for IGS assessment and rubric design, respectively. These frameworks integrate findings across the thesis to fulfil the pragmatic aim to make meaning from participant understandings (Kelly & Cordeiro, 2020) and create knowledge that contributes to improved action (Goldkuhl, 2011, 2012). The findings, translated assessment conceptualisations, and recommendations for researchers, policy makers and PE practitioners in this chapter make a significant and original contribution to knowledge in the discipline of PE, and potentially the wider field of educational assessment. The chapter concludes by addressing a number of strengths and limitations of the scope of studies, including the impact of COVID-19.

8.2 A Summary of Key Findings

This chapter provides a summation of key findings in response to the overarching research question: *How do Physical Education teachers understand and practice assessment in the curriculum focus area of Games and Sports in Victorian secondary schools?* Each of the four emergently designed studies addressed a research sub-question that contributed to meeting the aim of the thesis: *To describe, explain and make meaning of Physical Education teachers' understanding and practice of assessment in invasion games and sports in Victorian secondary schools (Year Levels 7 – 10)*. This phenomenon of assessment has been widely reported in the extant PE literature as problematic (for example, Hay & Penney, 2013; López-Pastor et al., 2013; Veal, 1988). The summary of key findings is presented by research sub-question, with close reference to the specific study chapter. Meaningful integration of data from other studies (chapters) was added, when it was likely to contribute to a fuller understanding of the research sub-question. This integration of data across studies (chapters) is consistent with the mixed methods design dimensions of *integration* and *complexity* (Schoonenboom & Johnson, 2017). Integrating the data in such a way led to the conceptualisation of the Invasion Games and Sports Assessment Framework and the Rubric Design Framework as a translation of the overall findings of the thesis.

8.2.1 *Research Sub-Question 1*

What does the extant literature say are the defining characteristics of assessment tools developed for invasion games and sports?

The ScR used to answer this question was described in Chapter 4. It employed a five-step framework (Arksey & O'Malley, 2005) that was complemented by guidelines in the PRISMA-ScR extension (Tricco et al., 2018) to locate 72 peer-reviewed studies that met the inclusion criteria. These studies applied 32, unique, evidence-based tools to substantively build on previous literature reviews in GS assessment of a youth population. For example, the review of the most frequently used tactical assessment instruments in PE and youth sports (Arias & Castejón, 2012) and the systematic review of assessment for tactical learning in games (Barquero-Ruiz et al., 2019) located six and four IGS assessment instruments, respectively. The findings of the ScR supported the prevalence of the GPAI and TSAP in other literature reviews (Arias & Castejón, 2012; Barquero-Ruiz et al., 2019). To provide a potential bank of assessment tools for secondary school PE teachers, Appendix E identifies and describes 15 tools that were designed, or validated in a school setting from the ScR. Identification of these tools, and making them available through professional associations like the VCAA, may help to address a challenge reported in the GS assessment literature that many PE teachers have difficulty locating appropriate assessment tools (Killian & Mays, 2021; Williams et al., 2020). In addition to providing PE teachers access to these tools, modification of these tools to suit a specific learning or assessment context is endorsed in the PE and performance analysis literature (for example, Brewer & Jones, 2002; Fernandes et al., 2019).

A key characteristic of the assessment tools was the prevalence of those that generated outcomes based on frequency-counts, rather than tools that rated performance holistically. This meant that rubrics, as a type of rating tool, were under-represented in the ScR. It was posited that the relatively modest identification of tools that rated performance may have been attributed to the eligibility criteria of validity and reliability that are more commonly associated with research and assessment tools based on quantitative (numerical) data (Ary et al., 2014). In keeping with the wide range of criteria that informs rigor in qualitative research (Smith & McGannon, 2018; Tracy, 2010), it is possible that other qualitatively-driven tools may have employed alternative criteria to validity and reliability and were subsequently not identified in the ScR. Further, most frequency-count tools in the ScR generated indices as indicators of overall game performance, which have been criticised in the field (Memmert & Harvey, 2008). For example, the use of indices may lead to

scores or indicators that mask results by aggregating data across multiple aspects of performance (Barquero-Ruiz et al., 2019).

In charting the data in the ScR (Chapter 4), three prevalent performance criteria were identified as on-the-ball skills, off-the-ball movement and decision-making. Subsequent coding of the most prevalent criterion, on-the-ball skills, identified five game skills as receiving, passing, dribbling, scoring and defending. There was no standardised definition for any of the above criteria, indicating the common practice of adapting or modifying tools in the field and the highly contextual nature of assessment. The three key performance criteria aligned closely to the GPAI, while the five game skills aligned closely to the TSAP. This finding is not surprising given that these two tools comprised one-third of the included studies. The utility of these findings is that researchers and PE teachers have empirical support for creating their own operational definitions of the above criteria when assessing game-based performance in IGS in a school-age population.

In terms of the application of tools across the studies, a sample comprising males and females was most prevalent, while exclusively female cohorts were under-represented. Approximately three-quarters of all studies employed a population that was 13 or under, equivalent to a primary school context in Victoria. This meant that studies involving students in Year Levels 7-10 were under-represented. Soccer and basketball were the most assessed sport contexts, and most tools were applied in the context they were developed (school or community), which indicated good alignment (Barquero-Ruiz et al., 2019). The pervasive use of SSG and evidence that four minutes may be adequate to assess student performance (da Costa et al., 2010) were positive findings that support the use of GBAs and efficacious assessment in the field. A range of challenges pertaining to the feasibility of assessment were also acknowledged; these included the expansive range of criteria employed in some tools, the widespread use of video capture, and the limited use of peer assessment. Given the complexity and pace of IGS, frequency-count tools like the GPAI and TSAP are more likely to be used by researchers than teachers (for example, Arias-Estero & Castejón, 2014; MacPhail et al., 2008; Williams et al., 2020).

In summary of the response to the sub-question, the evidence-based assessment tools located in the ScR were characterised by the use of frequency-based approaches, the calculation of performance indices, assessment in a single sport context, and the reporting of measures of reliability rather than measures of validity. The key performance criteria of the evidence-based assessment tools were based on actions on-the-ball, movement off-the-ball and decision-making.

The tools were generally applied in settings congruent with their development and comprised an all-male population. Most of the sample populations were aged 13 or under and the assessment context was SSG of soccer or basketball. The use of video capture and expert assessors was prevalent, while the length of observation to inform the assessment varied greatly. In addition to these findings, the ScR informed the design of the questionnaire used in the cross-sectional, quantitative inquiry (Chapter 5) and provided a reference point to describe, explain and make meaning of the assessment practices of the sample across the following three studies of the thesis.

8.2.2 *Research Sub-Question 2*

How do Physical Education teachers view the assessment of invasion games and sports in Victorian secondary schools?

To respond to the research sub-question, a sample of PE teachers in Victorian secondary schools ($N = 80$) completed a questionnaire that constituted the cross-sectional, quantitative inquiry described in Chapter 5. Participants identified rubrics as the most widely used, and most useful, assessment tool in IGS across Year Levels 7 - 10. In contrast, frequency-count tools were rarely used by teachers, students or peers. The limited use of frequency-count tools could be attributed to the limited familiarity participants had with the GPAI and the TSAP. The low level of familiarity and use of these tools is consistent with findings in other Australian studies in similar GS assessment contexts (Georgakis et al., 2015; Williams et al., 2020). Reasons for non-use reported by participants that were aware of these tools included the perceived complexity of the tools and the amount of time required to administer them. The latter point was endorsed by participants responding to the questionnaire item that indicated time-efficiency was the most widely supported aspect in their assessment practice.

The key performance criteria located in the ScR (Chapter 4) were widely used by participants in the cross-sectional, quantitative inquiry (Chapter 5). These findings indicate that despite the prevalence of frequency-count tools in the ScR, and the prevalence of rubrics reported in the questionnaire, the assessment criteria were congruent. In addition, various criteria that were provided as options from the state and national curricula, rather than the ScR, indicated the pervasive assessment of teamwork and the relatively modest assessment of creativity. As with the limited use of peer assessment in the tools located in the ScR (Chapter 4), self and peer assessment were not used as widely as teacher-led approaches for any assessment tool. The limited use of students in the process of assessment indicated in the questionnaire was explored

in the qualitative inquiry (Chapter 6) and led to the aptly named theme, 'Students - the missing ingredient'.

In establishing the degree of alignment between curriculum and assessment, less than half of all respondents in the cross-sectional, quantitative inquiry (Chapter 5) indicated that the curriculum informed their assessment of IGS. The challenge of curriculum alignment to GS assessment has been reported in other Australian studies (for example, Williams et al., 2020). The limited references to a curriculum in the tools developed in the school setting in the ScR (Chapter 4), and responses by participants in both qualitative inquiries (Chapter 6 and 7), reinforced the limited alignment of curriculum and assessment within the sample. For example, in the document analysis (Chapter 7), approximately two-thirds of the submitted rubrics made *no* explicit reference to the VC: HPE. The limited to modest curriculum alignment with assessment reported across the thesis presents a tension with the literature that describes this alignment as best practice in PE (Penney et al., 2009).

8.2.3 *Research Sub-Question 3*

How do Physical Education teachers practice and/or want to practice assessment of invasion games and sports in Victorian secondary schools?

The qualitative inquiry in Chapter 6 employed a series of semi-structured online interviews to help explain earlier findings by drawing on a nested, purposive, heterogenous sample ($n = 8$) from the cross-sectional, quantitative inquiry (Chapter 5). The use of reflexive thematic analysis (for example, Braun & Clarke, 2021a ; Braun et al., 2019) supported the construction of four themes that responded to the above research sub-question. Key findings addressing sociocultural influences on assessment (Hay & Penney, 2013) included that most participants felt compelled to follow school reporting guidelines, but not necessarily the state guidelines to assess students against the standards in the VC (VCAA, n.d.-j). Participants acknowledged a range of challenges in their assessment of IGS that included the perennial issues of large class sizes and time constraints (Braga & Liversedge, 2017; Gallo et al., 2006; Veloo & Md Ali, 2016). Participants recognised the limited role that students played in their assessment and were generally able to articulate a range of more useful strategies and novel assessment tools that might exist in an ideal or perfect world.

In describing the largely passive role of students, participants reported that students were not always aware that an assessment took place or the assessment criteria; in some cases, students did not receive any feedback until their end-of-semester report. While rubrics were used

widely by participants, many rubrics were located on the school's electronic Learning Management System (LMS), meaning that teachers were not aware what access students had to them. The significance of these findings is that, in most cases, assessment was something that was *done* to students by their teachers. Thus, students were largely ignorant of the assessment process, meaning that student agency was limited. Most participants acknowledged that a systematic use of self and peer assessment might be more useful in developing student agency.

Assessment for the purpose of driving student learning or teacher instruction received scant attention from the interview participants. Instead, assessment was persistently linked to the purpose of reporting to a parent audience. Participant views regarding the use of assessment supported findings in the historical and contemporary assessment in PE literature (Moura et al., 2021; Veal, 1988). Specifically, the term AfL was not used by any participant in the interview study, while formative assessment was mentioned by some. References to formative assessment were often linked to the admission that it was rarely done and that one-on-one feedback to students was infrequent. The limited use of AfL is contrary to guidance from the AISEP position statement on assessment in PE that AfL drives meaningful learning in PE and helps to legitimise the subject (AISEP, 2020).

Strategies suggested by participants to make assessment more useful included using rubrics formatively, teaching students to umpire games to allow teachers more time to assess, assessing fewer students in a single lesson, making use of video exemplars to describe levels of performance in rubrics, and developing student capacity to implement peer and self-assessment. In addition, a desire to provide students with more feedback was widely reported by participants as a means to improve their assessment practice. To support this, some participants reflected that their current multi-sport approach needed to be replaced by a thematic-based approach to extend the length of game-based units. In a secondary school context, the multi-sport or multi-activity approach has generally been criticised for providing insufficient time for student learning and its focus on directive teaching (Pill, 2011; Pill et al., 2017; SHAPE, 2014).

In addition to identifying various strategies to improve assessment, many adaptations to assessment tools were put forward by participants. Most interviewees recognised that the use of scale language limited the utility of their assessment rubrics, thus supporting the literature review of rubrics used in higher education (Brookhart, 2018). Participants spoke of the need to frame more objective descriptors for their rubrics, but were unclear how to do this. Further, the

vocabulary used to describe rubric elements was inconsistent, with several participants unclear about who, when and how their rubrics were constructed. As an example, one participant expressed a lack of interest and skill in rubric design. It was not clear what training participants had in rubric design or any guidelines they followed when constructing rubrics.

Most participants saw potential in the GPAI and TSAP, however, they were clear that these tools had to be more student-friendly (simple to understand and administer), and easy to convert into a rubric format for reporting purposes. Participant suggestions included combining elements of the GPAI and TSAP and adapting them into a generic rubric format. Participants expressed interest in designing more manageable assessment and feedback processes and developing new or modified assessment tools. This interest indicated a need for professional learning opportunities to promote assessment and feedback literacy models described in the PE literature (DinanThompson & Penney, 2015; Park, 2017). There was no indication by any participant that they would design any of the assessment tools they described, or enact any proposed refinements to their rubrics. Given the prevalence of rubrics across all four themes of the qualitative inquiry (Chapter 6), the final phase in the emergent mixed methods study design was a further qualitative inquiry that employed a document analysis of rubrics (Chapter 7).

8.2.4 *Research Sub-Question 4*

How are Physical Education teachers' assessment tools constructed for invasion games and sports in Victorian secondary schools?

This document analysis drew on a nested sample of 11 participants from the cross-sectional, quantitative inquiry detailed in Chapter 5. Six of these 11 participants had participated in the previous qualitative inquiry detailed in Chapter 6. The aim of the final study in the thesis was to explain how assessment tools (rubrics) were constructed for use in IGS. This included making references to the use of key performance criteria located in the ScR (Chapter 4), the language used in the VC: HPE (VCAA, n.d.-d), and various rubric design elements described in the extant literature (Brookhart, 2018; Dawson, 2017). In employing content and thematic analysis within the READ approach (Dalglish et al., 2021), the coding of documents ($N = 24$) was abductive, as it oscillated between the inductive and deductive coding of artefacts. In the former approach, coding was based on novel or unexpected data items, while in the latter coding approach, items were coded with reference to language used in the literature relating to rubric design (Brookhart, 2018; Dawson, 2017) and the VC: HPE (VCAA, n.d.-d).

As only two rubrics were identified in the ScR (Chapter 4), comparisons between the rubrics located in the ScR and the document analysis in the Chapter 7 are limited. Of the two rubrics in the ScR, one was analytical (Penney et al., 2012) and one was holistic (Williams & Rink, 2003). This contrasted with the document analysis study where all 24 rubrics were classified as analytical, as they assessed criteria separately. The literature suggests analytical rubrics may be better suited to provide students feedback (Brookhart, 2013, 2018), rather than the summative use described by most interview participants in the preceding qualitative inquiry in Chapter 6. The exclusive use of analytical rubrics in the corpus led to the theme, 'A narrow conceptualisation of rubrics', with recommendations to expand the repertoire of participants to include holistic rubrics, single point rubrics, and multi-media exemplars of performance in electronic rubrics.

Most rubrics were generic in nature, rather than sport-specific, which was incongruent with tools developed in school settings identified in the ScR (Chapter 4). The higher rate of generic assessment tools in the document analysis may be explained by participants' use of thematic GBAs to teach and assess games and/or the requirement for PE teachers to assess the *transfer* of movement skills, concepts and strategies within the VC: HPE (VCAA, n.d.-d). The most prevalent configuration for rubrics was to describe five criteria across five levels of performance in order to generate a score, grade or reference to a standard. The inference was that most rubrics were constructed to facilitate the reporting of student achievement on the state-mandated five-point scale (DET, 2023c). In removing scores, grades or percentages from rubrics, the literature on rubrics in education suggests that students may re-direct their focus towards learning and away from grading (Brinson, 2022; Panadero & Jonsson, 2020).

The key performance criteria identified in the ScR were used widely in the submitted rubrics. Further, as most rubrics assessed 'other' criteria, this indicated the potential requirement for PE teachers to report against a broad set of standards described in the VC: HPE (VCAA, n.d.-d). The rubrics also included criteria like effort, attitude and wearing a uniform that are widely criticised in the assessment in PE literature (Baghurst, 2014; Veal, 1988; Williams et al., 2020). As performance was commonly described using scale language in the submitted rubrics, none of the submitted artefacts constituted a "true rubric" (Brookhart, 2018, p. 1), characterised by the exclusive use of qualitative descriptors. The pervasive use of scale language identified in the rubrics supported participant views in the qualitative inquiry in Chapter 6, that the use of subjective language posed a challenge to reliable assessment.

In summary of the findings in response to the overarching research question and aim, the thesis indicated that most participants conflated assessment with reporting, were inconsistent in their alignment with assessment to a curriculum, and valued time efficiency (feasibility) in their assessment practice. Participants reported that students were rarely involved in their assessment and that rubrics were used widely, but lacked some utility. Rubrics were narrowly conceptualised, unhelpful scale language in rubric descriptors was pervasive, and there was no common language among participants to describe the essential elements of rubrics. In identifying 32 evidence-based tools in the ScR (Chapter 4), 15 of which were validated in a school context, the sample of PE teachers in Victorian schools typically assessed criteria similar to those found in the tools, but they avoided generating performance outcomes based on frequency of occurrence.

8.3 Translation of Findings

In keeping with the pragmatic aim of the thesis to create actionable knowledge (Kelly & Cordeiro, 2020) and the dimensions of integration and complexity within the mixed methods study design (Schoonenboom & Johnson, 2017), the following section of the chapter presents a translation of findings from the thesis. Putting together the findings and recommendations across the four emergently designed studies led to the conceptualisation of: (1) a flexible assessment framework for use in performance-based assessment of IGS and; (2) an expansive rubric design framework (both are presented in Appendix P). Each framework is briefly described with references to findings from relevant studies and the extant literature, before recommendations are made for researchers and practitioners to investigate their utility and robustness.

8.3.1 *Invasion Games and Sports Assessment Framework*

The key performance criteria located in the ScR (Chapter 4) have been arranged, refined and re-named to provide the foundation for the Invasion Games and Sports Assessment Framework. The idea of a potential assessment framework was initially proposed in Chapter 4 to address the issue of researchers and teachers locating, evaluating and selecting from the vast array of evidence-based tools in the literature. The proposal is conceptualised here as the Invasion Games and Sports Assessment Framework (see Table 2 in Appendix P). The framework is offered as a translation of the findings of the thesis and is positioned as a direct response to the myriad of descriptions of potentially “more useful” assessment tools described by interviewees in the qualitative inquiry (Chapter 6). In recognising that the design of the full range of tools was beyond

the scope of the thesis, a flexible assessment framework was conceptualised to support the creation of the multiple tools articulated by respondents.

Importantly, the criteria in the Invasion Games and Sports Assessment Framework align to the VC: HPE across Year Levels 7-10, with sample references to Year Levels 7-8 presented in Table P1 (VCAA, n.d.-d) in Appendix P. This was seen as a vital step in promoting curriculum alignment that was reported as being low to modest by participants (Chapters 5, 6 and 7), and supported in the assessment in PE literature (Chapter 2). To reflect participant views, that the language of assessment needed to be more “student-friendly” (Chapter 6), references to any complex language employed in the curriculum, including unique codes for content descriptors, may be better placed in unit plans than in any tools informed by the proposed framework.

Underpinning the framework are the three named key performance criteria and five specific on-the-ball skills identified in the ScR (Chapter 4). To honour participant requests to use “student-friendly” language in assessment, I chose a single word to represent each criterion; this led to the terms, *Move*, *Choose*, *Play*. Within the criterion of *Move* the terms support and cover have been included to reflect the importance of attacking and defensive positioning reported in the cross-sectional, quantitative inquiry (Chapter 5). Within the criterion of *Choose* users assess a player’s decision-making with reference to on-the-ball skills. Within the criterion of *Play* users assess the five game skills identified in the ScR. The framework encourages users to select from the three criteria and 12 sub-criteria according to their learning/assessment requirements. In selecting from these options, the use of inappropriate criteria commonly reported in the assessment in PE literature (Chapter 2) and participants in both qualitative inquiries (Chapter 6 and 7) may be reduced.

The flexible nature of the Invasion Games and Sports Assessment Framework was a pragmatic response to the adage, “Give a Man a Fish, and You Feed Him for a Day. Teach a Man To Fish, and You Feed Him for a Lifetime” (anon). In applying the framework, users can design a range of assessment tools including checklists, rating scales, rubrics and frequency-count approaches that support assessment of GS in PE (Mitchell et al., 2013). The proposed framework supports the TGfU model (Bunker & Thorpe, 1986; Werner et al., 1996) by providing users with the flexibility to assess decision-making (*Choose*) and skill-execution and performance (*Move* and *Play*). Further, the framework allows for ball skills to be assessed from a technique perspective, and an outcome perspective, that is also supported by the TGfU model (Bunker & Thorpe, 1986; Werner et al.,

1996). This flexibility in assessing skill execution (technique) and skill outcomes was endorsed by participants in the cross-sectional, quantitative inquiry (Chapter 5).

Given the focus of the framework is performance-based assessment, assessing declarative knowledge in the cognitive domain through written or verbal responses (García-Ceberino et al., 2020; Turner & Martinek, 1992), and aspects of the affective domain, like fairplay and collaboration (Mitchell et al., 2013), are not currently addressed by the proposed framework. There is potential for aspects of the affective domain to be added as a fourth criterion, with the earlier suggestion for further research in Chapter 4 to review ‘other’ criteria for this purpose. In adopting student friendly language, any potential affective criterion may be termed ‘*Feel*’ to capture the “feelings, attitudes, and dispositions that student have” (Lund & Veal, 2013, p. 134). In advocating for authentic assessment tasks, the framework assumes that within GS in the subject of PE “game performance improvement is the goal of instruction, [therefore] assessment should be conducted within the context of the game” (Mitchell et al., 2013, p. 42).

In addition to providing descriptions of the 12 sub-criteria based on the coding guidelines used in the ScR (Chapter 4), the three categories of the framework are deliberately ordered to represent a typical sequence of play. This sequence is characterised by a player moving into a position to gain possession of the ball (Move), selecting an on-the-ball skill (Choose), and then executing that ball skill (Play). Recognition of this playing sequence informed the development of the Game Play Observational Instrument (Turner & Martinek, 1999), and underpinned the Four Rs assessment instrument developed for use in net and wall games (Hopper, 2003). Given this sequencing, assessors might choose to assess student performance across the three criteria or to assess performance within a single criterion. Given the criteria and sub-criteria of the framework, it may also be viewed as a hybrid of the GPAI and TSAP. This hybrid view supports use of the framework to construct school-based rubrics as suggested by participants in the qualitative inquiry (Chapter 6). The following section of the chapter briefly addresses similarities and differences in the GPAI and TSAP to the conceptualised framework (see Table P2 in Appendix P).

Like the GPAI that identified seven observable and non-specific game components (Oslin et al., 1998), the Invasion Games and Sports Assessment Framework identifies 12 observable and non-specific aspects of game performance. In recognition that assessment in PE is highly contextual (Hay & Penney, 2013), users are directed to create their own definitions of successful performance based on the 12 sub-criteria. This design supports participant views in the qualitative

inquiry (Chapter 6) that creating their own definitions within the GPAI would be useful. Consistent with the GPAI, the framework assesses decision-making (Choose) in relation to on-the-ball skills (Play) rather than off-the-ball movement (Move) (Oslin et al., 1998). To provide support for practitioners in creating operational definitions that describe the sub-criteria within Choose, the generic tactical elements of IGS described by Memmert and Harvey (2010) and the adapted principles of play for IGS described by Ward and Griggs (2011), presented in Chapter 7 may be of use.

Unlike the GPAI, which was validated in three game categories (Oslin et al., 1998), the proposed framework is specifically designed for use in IGS. Further, the framework identifies five game skills, where the GPAI identifies none, and the framework reduces the assessment of movement or positional play from five aspects to two (Oslin et al., 1998). Specifically, the criteria of adjust and base in the GPAI have been omitted from the framework, while the term guard/mark has been absorbed into cover in an attempt to simplify assessment (Oslin et al., 1998). As attacking and defensive play were assessed by participants at high rates in the quantitative inquiry (Chapter 5), this led to the inclusion of support (attacking play) and cover (defensive play) within the criterion of Move in the framework. This design feature is consistent with a literature review of assessment for tactical learning in games that argued player roles can be generally classified as attacking (team with the ball) and defending (team without the ball) (Barquero-Ruiz et al., 2019).

Like the TSAP, the Invasion Games and Sports Assessment Framework includes ball-centred criteria that consider how players receive and dispose of the ball (Gréhaigne et al., 1997). For example, the framework has simplified the criteria of the TSAP by merging conquering and receiving the ball into receive, and merging playing a neutral ball and an offensive ball into pass (Gréhaigne et al., 1997). This supports participant views in the qualitative inquiry (Chapter 6) that they would change some of the abbreviations and terms used in the TSAP. Unlike the TSAP, the proposed framework includes a criterion for dribbling and omits the criteria of lost ball. The addition of dribbling to the framework was predicated on participant responses across the studies described in Chapters 5, 6, and 7, and was supported by findings in the ScR (Chapter 4). In omitting the criterion of lost balls (or turnovers), the framework adopts a strengths-based approach that is endorsed by the VC: HPE (VCAA, n.d.-e). As another departure from the TSAP, the framework makes no reference to a nomogram, which was rarely acknowledged by participants in the qualitative inquiry (Chapter 6), nor used widely in relevant studies located in the ScR (Chapter 4).

A major difference between both assessment instruments and the Invasion Games and Sports Assessment Framework is that the latter is not reliant on generating frequency-count outcomes, or using mathematical formulae to calculate indices of overall game performance indicators (Gréhaigne et al., 1997; Oslin et al., 1998). However, if a user determined that player involvement or efficiency ratios served the purpose of the assessment, then the framework can support this. Where users wish to generate frequency-counts based on the *Move* criterion, it is advised that support and cover are tallied with each change of possession between players (Nevett et al., 2001). The adjustments to the GPAI and TSAP that inform the proposed framework are consistent with participant views reported across the thesis and are supported by the GS assessment literature that promotes the adaptation of existing and validated tools (for example, Fernandes et al., 2019; Nadeau, Godbout, et al., 2008; Nadeau, Richard, et al., 2008). Consistent with recommendations from participants in the qualitative inquiry (Chapter 6), it is not suggested that all 12 sub-criteria are assessed at the same time. Given the prevalence of GS in most PE curricula (for example, Casey & Hastie, 2011; Gray et al., 2008; Ward & Griggs, 2011), the proposed framework may resonate with PE teachers beyond the local context, potentially making its use analytically generalisable and transferable (Smith, 2017; Tracy, 2010).

8.3.2 *The Rubric Design Framework*

A second conceptualisation that served to translate the overall findings of the thesis is presented in Table P3 (see Appendix P). The Rubric Design framework is an expansive adaptation of the 14 design elements of rubrics described by Dawson (2017) (see Appendix N). This conceptualisation was a response to the prevalent use of rubrics by the sample (Chapter 5), reports of their relatively modest utility (Chapter 5 and 6), the lack of clarity in rubric design and terminology (Chapter 6), the relatively narrow conceptualisation of rubrics that were often associated with reporting (Chapter 7), and the prevalence of rating scale language in descriptors (Chapter 7). The Rubric Design Framework aimed to clarify rubric terminology, to suggest a logical sequence in rubric design, and to make users aware of rubric design elements they might not have otherwise considered. In keeping with the original design framework of Dawson (2017), the proposed 27 decision-points include design processes and contextual information beyond the physical artefact of the rubric.

As with the work of Dawson (2017), the framework presented in Table P3 does not make claims of completeness, or advocate certain choices over others. However, the series of recommendations for rubric design and use, provided in the document analysis of Chapter 7, may

help to inform rubric design choices for the sample. The Rubric Design Framework is a significant extension of previous rubric design guidelines in the context of PE that include the three-step approach of Birky (2012), and the four-step approach of Wang and Rairigh (2006); neither of which were mentioned by participants in the qualitative inquiry (Chapter 6). Unlike the original arrangement by Dawson (2017), the decision-points are framed within six categories and follow a deliberate sequence to help scaffold rubric construction. While this proposed framework was designed to support individuals that formed the sample, a consistent approach within teaching teams may help stakeholders (students, teachers, parents) to read, understand and navigate rubrics. Each of the 27 decision-points in the framework are guided by a question(s) to help the designer(s) make an informed choice in their rubric design (see Table P3). Given the pervasive use of rubrics in the broader field of education and research (for example, Bearman & Ajjawi, 2019; Panadero & Jonsson, 2020; Zheng et al., 2019), the proposed Rubric Design Framework may be contextually generalisable, transferable and of interest to educators and researchers beyond the sample (Smith, 2017; Tracy, 2010).

8.4 Recommendations

The following section of the chapter is divided into recommendations for researchers in the field of GS assessment, policy makers at a curriculum and school level, and secondary school PE practitioners. As the findings of the ScR that informed the Invasion Games and Sports Assessment Framework were based in the international PE assessment literature, application of some of the following recommendations may be transferable across school sectors and geographical regions. Similarly, given the pervasive use of rubrics in education, the Rubric Design Framework may be of value to educators in subjects other than PE and educators in primary and tertiary sectors.

8.4.1 *For Researchers*

The ScR in Chapter 4 that lay the foundation for this thesis indicated that end-zone games (like touch rugby and ultimate Frisbee), exclusively female cohorts, and age groups 13 to 19 were under-represented. Only two of the 32 evidence-based tools were rubrics, suggesting investigations into the use, reliability, and validity of rubrics in a GS context are warranted. Alternative quality criteria for assessment, like feasibility, may also be used to inform other ScRs in GS assessment. It is suggested that ScRs of assessment in net and wall games, and striking and fielding games, may be of benefit to the discipline of PE. As a common outcome of ScRs is to identify systematic literature reviews that appraise the quality of included studies (Arksey &

O'Malley, 2005), systematic literature reviews are advocated for prevalent IGS assessment tools including the GPAI, TSAP, GPET and FUT-SAT.

As the ScR (Chapter 4) found that approximately one-third of tools did not report any measurement of validity, greater reporting of validity measures in future research in GS assessment tool design is encouraged (Arias & Castejón, 2012; Barquero-Ruiz et al., 2019). Of the eight different types of validity reported in tools located in the ScR (Chapter 4), the reporting of content, construct, ecological, and face validity are encouraged in further studies investigating IGS assessment. As there is no agreed 'gold standard' assessment for IGS, criterion validity that involves comparison of an assessment tool score to another tool (Terwee et al., 2007), remains problematic in this context. The inconsistent and unclear reporting of key features by researchers in the application of assessment tools was a challenge for data charting in the ScR. Researchers publishing in a GS assessment context are encouraged to provide their tool or procedure with a unique name, include a coding sheet, and clearly state the number and name of all criteria. This should include the provision of operational definitions for all criteria to allow for greater clarity in reporting.

Case studies or field observations in schools that examine how curriculum alignment and general assessment practice in secondary school PE settings is encouraged. Such field observations may be supported by evidence-based tools like the SOFAST (van der Mars et al., 2018b). Within these field observations, validation of various iterations of assessments informed by the Invasion Games and Sports Assessment Framework is advocated. Assessment tools theoretically supported by the framework and advocated for use in GS (Mitchell et al., 2013) may include checklists, rating scales, rubrics and frequency-count tools. Given the prevalence of decision-making and team-work as assessable criteria in the sample (for example, Chapter 5), a fuller examination of the operational definitions of these terms and their assessment in school contexts may be insightful. Further, focus groups, case studies or action research investigating the utility of the expansive Rubric Design Framework within secondary school PE departments and other educational settings is encouraged.

8.4.2 For Policy Makers

Within the field of curriculum design, the project indicated the need for greater clarity in documentation and support for curriculum alignment to assessment. The VCAA has recently introduced two-yearly online reviews of their F-10 programs across all learning areas, which may

be a suitable avenue for this to be addressed. Teachers must register online to access these reviews (VCAA, personal communication, August 15, 2022) and examples of contextually relevant questions in the 2022 online survey included asking participants if they would like to access “quality assessment tasks and marking rubrics/criteria” and the extent to which participants might need support in assessing “against the achievement standards in the curriculum”. These questions appear to be a proactive step in getting stakeholder contributions to inform curriculum and assessment alignment.

The establishment of a library of video exemplars to help teachers determine different levels of student performance was supported by participants (Chapter 6). The VCAA is currently expanding its online resources through the Digital Assessment Library (DAL), however, in the learning area of PE, this is yet to include assessment for the Focus Area of GS (VCAA, personal communication, August 15, 2022). The VCAA’s recent addition of the MAP platform to support teachers in locating and administering evidence-based FMS assessments (VCAA, n.d.-h) suggests the VCAA may be able to host a similar platform for the assessment of IGS. I am interested in pursuing this opportunity as part of the publication and dissemination of findings post thesis submission. The VCAA is a logical provider of exemplar videos and assessment resources due to the likelihood of stakeholders having free online access and the alignment of the resources to the VC: HPE. However, other peak bodies, such as those used in the recruitment for the cross-sectional, quantitative inquiry, may also be interested in the commercial development of such resources. Senior years PE within the VCE in Victorian schools (Year Levels 11 to 12) is high-stakes and tightly regulated by the VCAA. There may be merit in requiring greater accountability from teachers in government and Catholic schools to deliver and assess the VC: HPE across F to 10, as evidenced by some participants’ acknowledgement that the curriculum played a limited role in their assessment practice (Chapter 6).

Participants expressed some frustration with a ‘one size fits all’ reporting policy that was imposed by school leadership (Chapter 6). In contrast, there was no acknowledgment of any school-wide formative assessment policy (AfL); this may have been an unintended consequence of reporting policies that privileged summative assessment to report student achievement to parents. Greater consultation between school leaders and teachers in this sample may support greater autonomy and ownership in teachers’ assessment and reporting policies. Some participants expressed contradictory ICT guidelines at their schools that made the use of video capture problematic (Chapter 6). These ICT, policy-related barriers negatively impacted the

opportunity for some participants to create their own video libraries of performance in an effort to support student assessment.

8.4.3 For PE Practitioners

Pending engagement by the VCAA to support the development of an IGS assessment resource, much like that for the assessment of FMS through the MAP initiative (VCAA, n.d.-h), PE teachers are invited to access and adapt any of the 15 evidence-based assessment tools located in the ScR (Appendix E). Given the pervasive use of rubrics by participants in Chapter 5 and 6, the evidence-based Observational Scoring Rubric (Williams & Rink, 2003) and Standards-based rubric (Penney et al., 2012) may be useful starting points. PE teachers are invited to apply the Invasion Games and Sports Assessment Framework in teaching teams to help create assessments to suit the needs of their student cohort. PE teachers are also encouraged to make use of the Rubric Design Framework to review and design rubrics in any Focus Area, including those beyond GS. For the sample of PE teachers in Victorian secondary schools, both frameworks may be used concurrently and complemented by the recommendations for the design and use of rubrics provided in Chapter 7.

While this thesis focussed on assessment in the psychomotor and cognitive domains in performance-based assessment, PE teachers are encouraged to consider balancing their assessment across all three domains, including the affective (Hay & Penney, 2013; Lund & Veal, 2013; Mitchell et al., 2013). Participants are also advised to question their assessment of criteria, such as effort, organisation and uniform (Chapter 7) that is widely criticised in the literature (for example, Baghurst, 2014; Blomqvist et al., 2005; Lund, 1992). Further, to reduce the time burden of assessment and the challenge of assessing large classes (Braga & Liversedge, 2017; Gallo et al., 2006; Veloo & Md Ali, 2016), teaching students how to complete peer and self-assessment is recommended.

Where the aim of school-based IGS assessment is to determine the level of game involvement, choosing a frequency-count format may be appropriate. In contrast, if the aim of the assessment is to describe the quality of performance, then a rating scale or rubric format may be more appropriate. In terms of recommendations based on the submitted rubrics (Chapter 7), it may be of benefit for teachers to remove scores or grades so that students can focus on learning (Brinson, 2022). Further, three to five criteria (Brookhart, 2018) with a similar number of performance levels, assuming they can be reliably distinguished (Brookhart, 2018), is suggested.

The use of qualitative descriptors rather than evaluative, frequency-based or numerical descriptors may help assessors and performers better understand criteria and the next steps in learning (Brookhart, 2018).

8.5 Strengths and Limitations

Strengths of the thesis include the identification of an extensive collection of assessment tools in the peer-reviewed literature ($N = 32$) and the associated identification of key performance criteria. The three studies, drawing on a sample of PE teachers in Victorian secondary schools, provided considerable insights into their understanding and practice of assessment in IGS where none previously existed. The use of an explanatory, sequential, mixed methods study (Creswell & Plano Clark, 2018; Schoonenboom & Johnson, 2017) provided multiple data sources and allowed for the integration of findings that led to the conceptualisation of the Invasion Games and Sports Assessment Framework and the Rubric Design Framework in this chapter. Overall findings of the thesis tended to support the wider assessment literature regarding the challenges of assessment in PE, and of rubric design in a broader educational context. In addition, the final document analysis study provided information on rubric design to address a gap in the literature. The findings of the last study of the thesis included that rubrics were narrowly conceptualised, they were commonly associated with generating scores potentially for use in reporting, many lacked coherent criteria and all employed scale-type language which was viewed as an impediment to their utility.

Limitations of the thesis included that the overview of the assessment in PE literature (Chapter 2) and the ScR (Chapter 4) were both limited to studies in the English language. Other studies describing IGS assessment and evidence-based assessment tools were not located if published in other languages. Further, the decision to include validity and reliability terms in the search criteria may have inadvertently limited the number of assessment tools located. The questionnaire designed for use in the cross-sectional, quantitative inquiry (Chapter 5), may have benefited from more robust validation measures. As the maximum sample was 80 participants per question item, the response rate was under the estimated sample size of 91. Despite this sample size, the findings generally supported those reported in the literature for assessment in PE (Chapter 2). There was a lack of engagement with the pre-reading material in the qualitative inquiry (Chapter 6), which limited participant discussion of the potential strengths and limitations of the GPAI and TSAP. Equally, due to the limited use of AfL and provision of feedback reported by

most interview participants, several lines of inquiry could not be fully explored. The lack of contextual information regarding the design and use of the rubrics in the document analysis (Chapter 7) may also be viewed as a limitation of the final study.

As this thesis was undertaken by distance and in a part-time capacity, spanning 2016 to 2023, the challenges that faced the wider world in a time of COVID-19 were also experienced by me. Residing in a suburb of outer-Melbourne, I became a resident of the most locked down city in the world (Tuffield, 2021). The periods of lockdown and school closures also affected the timing of invitations for the online questionnaire. Research in schools within the state of Victoria was suspended from September 2021 until mid-2022, meaning that a number of avenues for a final study could not be realised (Victorian Government, 2022). Study options that were discussed with my supervisory team that could not be actioned in school settings included case studies, proof of concept studies testing novel or hybrid assessment instruments, and field observations of PE teachers' assessment practices.

8.6 Conclusion

Therefore, in conclusion this thesis has described the understandings and assessment practice of PE teachers in the context of IGS within Victorian secondary schools (Year Levels 7 – 10). Key findings included that assessment was commonly linked to reporting, inconsistently aligned to a curriculum, and largely overlooked the role of students. Participants indicated that while rubrics were the most commonly used assessment tool, their utility could be improved. These findings have been explained with reference to: the lack of accountability for some participants to adhere to a curriculum; school policies that link assessment to reporting rather than learning; and the pervasive view that peer and self-assessment did not support the participants' reporting agenda. The modest utility of rubrics was explained through the prevalent use of scale type language which promoted subjectivity and was a barrier for students to identify their next steps in learning. These findings are understood within a sociocultural perspective (Hay & Penney, 2013) that considers assessment practice as highly contextual and impacted by various systemic, social and cultural influences. Participants generally employed assessment criteria that were identified in evidence-based tools located in the ScR, however they privileged time-efficiency (feasibility) by preferring to generate assessment outcomes based on rating performance (through rubrics), rather than frequency-count approaches.

This explanatory, sequential, mixed methods study (Creswell & Plano Clark, 2018; Schoonenboom & Johnson, 2017) has achieved its aim of describing, explaining and making meaning of PE teachers' understandings and practice of assessment in IGS in Victorian secondary schools (Year Levels 7 – 10). In doing so, this thesis makes a significant and original contribution to the assessment of IGS within the discipline of PE. This population has rarely been canvassed in the current context in the literature, and has provided a wealth of information on their actual and idealised assessment practices. The findings and recommendations in this thesis serve the pragmatic aim of contributing knowledge to the discipline of PE that may make “a purposeful difference in practice” (Goldkuhl, 2012, p. 140). This research has also provided guidance to researchers and policy makers, and led to the conceptualisation of an Invasion Games and Sports Assessment Framework, and a Rubric Design Framework. Given the prevalence of IGS in international PE curricula, and the pervasive use of rubrics in education, it is possible that these frameworks may support improved assessment practice beyond the local context.

REFERENCES

- Aarskog, E. (2020). 'No assessment, no learning': Exploring student participation in assessment in Norwegian Physical Education (PE). *Sport, Education and Society*, 26(8), 875-888. <https://doi.org/10.1080/13573322.2020.1791064>
- Ackermans, K., Rusman, E., Brand-Gruwel, S., & Specht, M. (2019). Solving instructional design dilemmas to develop a video enhanced rubric with modeling examples to support mental model development of complex skills: The viewbrics-project use case. *Educational Technology Research and Development*, 67(4), 983-1002. <https://doi.org/10.1007/s11423-019-09668-1>
- Ackermans, K., Rusman, E., Nadolski, R., Brand - Gruwel, S., & Specht, M. (2021). Video-enhanced or textual rubrics: Does the viewbrics' formative assessment methodology support the mastery of complex (21st century) skills? *Journal of Computer Assisted Learning*, 37(3), 810-824. <https://doi.org/https://doi.org/10.1111/jcal.12525>
- Ackermans, K., Rusman, E., Nadolski, R., Specht, M., & Brand-Gruwel, S. (2019). Video-or text-based rubrics: What is most effective for mental model growth of complex skills within formative assessment in secondary schools? *Computers in Human Behavior*, 101, 248-258. <https://doi.org/10.1016/j.chb.2019.07.011>
- Alfrey, L., & Gard, M. (2017). Figuring out the prevalence of fitness testing in Physical Education: A figurational analysis. *European Physical Education Review*, 25(1), 187-202. <https://doi.org/10.1177/1356336x17715361>
- Allemang, B., Sitter, K., & Dimitropoulos, G. (2022). Pragmatism as a paradigm for patient-oriented research. *Health Expectations*, 25(1), 38-47. <https://doi.org/10.1111/hex.13384>
- Allen, I. E., & Seaman, C. A. (2007). Likert scales and data analyses. *Quality Progress*, 40(7), 64-65.
- Amatria, M., Lapresa, D., Arana, J., Anguera, M. T., & Garzon, B. (2016). Optimization of game formats in U-10 soccer using logistic regression analysis. *Journal of Human Kinetics*, 54, 163-171. <https://doi.org/10.1515/hukin-2016-0047>
- Andrade, H. (2007). Self-assessment through rubrics. *Educational Leadership*, 65(4), 60-63.
- Andrade, H., & Valtcheva, A. (2009). Promoting learning and achievement through self-assessment. *Theory into Practice*, 48(1), 12-19. <https://doi.org/10.1080/00405840802577544>
- Andrade, H. G. (2005). Teaching with rubrics: The good, the bad, and the ugly. *College Teaching*, 53(1), 27-31. <https://doi.org/10.3200/ctch.53.1.27-31>
- Anguera, M. T., Camerino, O., Castañer, M., Sánchez-Algarra, P., & Onwuegbuzie, A. J. (2017). The specificity of observational studies in physical activity and sports sciences: Moving forward in mixed methods research and proposals for achieving quantitative and qualitative symmetry. *Frontiers in Psychology*, 8(2196). <https://doi.org/10.3389/fpsyg.2017.02196>
- Annerstedt, C., & Larsson, S. (2010). 'I have my own picture of what the demands are ...': Grading in Swedish PEH - problems of validity, comparability and fairness. *European Physical Education Review*, 16(2), 97-115. <https://doi.org/https://doi.org/10.1177/1356336X10381299>
- Araújo, A. C. D., Knijnik, J., & Ovens, A. P. (2021). How does Physical Education and Health respond to the growing influence in media and digital technologies? An analysis of curriculum in Brazil, Australia and New Zealand. *Journal of Curriculum Studies*, 53(4), 563-577. <https://doi.org/10.1080/00220272.2020.1734664>
- Archibald, M. M., Ambagtsheer, R. C., Casey, M. G., & Lawless, M. (2019). Using Zoom videoconferencing for qualitative data collection: Perceptions and experiences of researchers and participants. *International Journal of Qualitative Methods*, 18, 1609406919874596. <https://doi.org/https://doi.org/10.1177/1609406919874596>
- Arias-Estero, J. (2013). Opportunities for and success in dribbling, passing, receiving, and shooting in youth basketball. *International Journal of Sports Science and Coaching*, 8(4), 703-711. <https://doi.org/https://doi.org/10.1260/1747-9541.8.4.703>
- Arias-Estero, J., & Castejón, F. (2014). Using instruments for tactical assessment in Physical Education and extra-curricular sports. *European Physical Education Review*, 20(4), 525-535. <https://doi.org/10.1177/1356336x14539214>

- Arias, J., & Castejón, F. (2012). Review of the instruments most frequently employed to assess tactics in Physical Education and youth sports. *Journal of Teaching in Physical Education*, 31(4), 381-391. <https://doi.org/https://doi.org/10.1123/jtpe.31.4.381>
- Arksey, H., & O'Malley, L. (2005). Scoping studies: Towards a methodological framework. *International Journal of Social Research Methodology*, 8(1), 19-32. <https://doi.org/10.1080/1364557032000119616>
- Arnold, P. J. (1979). *Meaning in movement, sport, and Physical Education*. Heinemann.
- Ary, D., Jacobs, L. C., Sorenson, C., & Walker, D. A. (2014). *Introduction to research in education* (9th ed.).
- Association Internationale des Écoles Supérieures d'Éducation Physique [AIESEP]. (2020). AIESEP position statement on PE assessment. Retrieved October 9, 2022, from <https://aiesep.org/wp-content/uploads/2020/06/AIESEP-Position-Statement-on-PE-Assessment-FINAL1.pdf>
- Australian Curriculum Assessment and Reporting Authority [ACARA]. (2023a). *The Australian Curriculum (version 9.0)*. ACARA. Retrieved February 10, 2023, from <https://v9.australiancurriculum.edu.au/>
- Australian Curriculum Assessment and Reporting Authority [ACARA]. (2023b). *F-10 curriculum overview*. ACARA. Retrieved March 2, 2023, from <https://v9.australiancurriculum.edu.au/f-10-curriculum/f-10-curriculum-overview#3>
- Australian Curriculum Assessment and Reporting Authority [ACARA]. (n.d.-a). *Australian Curriculum: Learning areas*. ACARA. Retrieved February 14, 2022, from <https://australiancurriculum.edu.au/f-10-curriculum/learning-areas/>
- Australian Curriculum Assessment and Reporting Authority [ACARA]. (n.d.-b). *Health and Physical Education (version 8.4)*. ACARA. Retrieved September 22, 2022, from <https://tinyurl.com/27k7c7r5>
- Australian Curriculum Assessment and Reporting Authority [ACARA]. (n.d.-c). *Health and Physical Education structure*. ACARA. Retrieved October 8, 2022, from <https://australiancurriculum.edu.au/f-10-curriculum/health-and-physical-education/structure/>
- Australian Curriculum Assessment and Reporting Authority [ACARA]. (n.d.-d). *Implementation of the Australian Curriculum*. ACARA. Retrieved April 14, 2022, from <https://australiancurriculum.edu.au/f-10-curriculum/implementation-of-the-australian-curriculum/>
- Australian Curriculum Assessment and Reporting Authority [ACARA]. (n.d.-e). *Implications for teaching, assessing and reporting*. ACARA. Retrieved April 16, 2022, from <https://australiancurriculum.edu.au/f-10-curriculum/implications-for-teaching-assessing-and-reporting/?searchTerm=reporting#dimension-content>
- Australian Institute for Teaching and School Leadership [AITSL]. (2017a). *Australian professional standards for teachers: All career stages*. AITSL. Retrieved October 16, 2022, from <https://www.aitsl.edu.au/standards>
- Australian Institute for Teaching and School Leadership [AITSL]. (2017b). *Spotlight - Reframing feedback to improve teaching and learning*. https://www.aitsl.edu.au/docs/default-source/research-evidence/spotlight/feedback.pdf?sfvrsn=cb2eec3c_14
- Ávila-Moreno, F. M., Chiroso-Ríos, L. J., Ureña-Espá, A., Lozano-Jarque, D., & Ulloa-Díaz, D. (2018). Evaluation of tactical performance in invasion team sports: A systematic review. *International Journal of Performance Analysis in Sport*, 18(2), 195-216. <https://doi.org/10.1080/24748668.2018.1460054>
- Backman, E., & Larsson, H. (2014). What should a Physical Education teacher know? An analysis of learning outcomes for future Physical Education teachers in Sweden. *Physical Education and Sport Pedagogy*, 21(2), 185-200. <https://doi.org/10.1080/17408989.2014.946007>
- Backman, E., & Pearson, P. (2016). 'We should assess the students in more authentic situations': Swedish PE teacher educators' views of the meaning of movement skills for future PE teachers. *European Physical Education Review*, 22(1), 47-64. <https://doi.org/10.1177/1356336x15589203>
- Baghurst, T. (2014). Assessment of effort and participation in Physical Education. *Physical Educator*, 71(3), 505.

- Bailey, R. (2005). Evaluating the relationship between Physical Education, sport and social inclusion. *Educational Review*, 57(1), 71-90.
<https://doi.org/10.1080/0013191042000274196>
- Bargainnier, S. (2003). Fundamentals of rubrics. *Pacific Crest*, 1-4. <https://www.pcrest.com/>
- Barker, D. M., & Rossi, A. (2012). The trouble with teamwork: A discursive investigation. *Physical Education and Sport Pedagogy*, 17(1), 1-19.
<https://doi.org/10.1080/17408989.2010.535198>
- Barquero-Ruiz, C., Arias-Estero, J. L., & Kirk, D. (2019). Assessment for tactical learning in games: A systematic review. *European Physical Education Review*, 1-21.
<https://doi.org/https://doi.org/10.1177/1356336X19889649>
- Barrientos Hernán, E. J., López-Pastor, V. M., Lorente-Catalán, E., & Kirk, D. (2022). Challenges with using formative and authentic assessment in Physical Education teaching from experienced teachers' perspectives. *Curriculum Studies in Health and Physical Education*, 14(2), 1-18. <https://doi.org/10.1080/25742981.2022.2060118>
- Basit, T. (2010). Manual or electronic? The role of coding in qualitative data analysis. *Educational Research*, 45(2), 143-154. <https://doi.org/10.1080/0013188032000133548>
- Bearman, M., & Ajjawi, R. (2019). Can a rubric do more than be transparent? Invitation as a new metaphor for assessment criteria. *Studies in Higher Education*, 46(2), 359-368.
<https://doi.org/10.1080/03075079.2019.1637842>
- Bearman, M., Dawson, P., Boud, D., Bennett, S., Hall, M., & Molloy, E. (2016). Support for assessment practice: Developing the Assessment Design Decisions Framework. *Teaching in Higher Education*, 21(5), 545-556. <https://doi.org/10.1080/13562517.2016.1160217>
- Bennett, C. (2016). Assessment rubrics: Thinking inside the boxes. *Learning and Teaching*, 9(1), 50-72. <https://doi.org/10.3167/latiss.2016.090104>
- Bennett, K. J. M., Novak, A. R., Pluss, M. A., Stevens, C. J., Coutts, A. J., & Fransen, J. (2017). The use of small-sided games to assess skill proficiency in youth soccer players: A talent identification tool. *Science and Medicine in Football*, 2(3), 231-236.
<https://doi.org/10.1080/24733938.2017.1413246>
- Bernstein, E., Phillips, S. R., & Silverman, S. (2011). Attitudes and perceptions of middle school students toward competitive activities in Physical Education. *Journal of Teaching in Physical Education*, 30(1), 69-83. <https://doi.org/https://doi.org/10.1123/jtpe.30.1.69>
- Bevans, K., Fitzpatrick, L.-A., Sanchez, B., & Forrest, C. B. (2010). Individual and instructional determinants of student engagement in Physical Education. *Journal of Teaching in Physical Education*, 29(4), 399-416. <https://doi.org/https://doi.org/10.1123/jtpe.29.4.399>
- Birky, B. (2012). Rubrics: A good solution for assessment. *Strategies*, 25(7), 19-21.
<https://doi.org/https://doi.org/10.1080/08924562.2012.10590978>
- Black, P., & William, D. (1998a). Assessment and classroom learning. *Assessment in Education: Principles, Policy and Practice*, 5(1), 7-74.
<https://doi.org/https://doi.org/10.1080/0969595980050102>
- Black, P., & William, D. (1998b). *Inside the black box: Raising standards through classroom assessment*. Granada Learning Group.
- Black, P., & William, D. (2009). Developing the theory of formative assessment. *Educational Assessment, Evaluation and Accountability*, 21(1), 5-31. <https://doi.org/10.1007/s11092-008-9068-5>
- Black, P., & William, D. (2018). Classroom assessment and pedagogy. *Assessment in Education: Principles, Policy and Practice*, 25(6), 551-575.
<https://doi.org/10.1080/0969594x.2018.1441807>
- Blomqvist, M., Vanttinen, T., & Luhtanen, P. (2005). Assessment of secondary school students' decision-making and game-play ability in soccer. *Physical Education and Sport Pedagogy*, 10(2), 107-119. <https://doi.org/10.1080/17408980500104992>
- Bonney, N., Berry, J., Ball, K., & Larkin, P. (2020). Validity and reliability of an Australian football small-sided game to assess kicking proficiency. *Journal of Sports Sciences*, 38(1), 79-85.
<https://doi.org/https://doi.org/10.1080/02640414.2019.1681864>
- Borghouts, L., Slingerland, M., & Haerens, L. (2016). Assessment quality and practices in secondary PE in the Netherlands. *Physical Education and Sport Pedagogy*, 22(5), 1-17.
<https://doi.org/https://doi.org/10.1080/17408989.2016.1241226>

- Boud, D. (2000). Sustainable assessment: Rethinking assessment for the learning society. *Studies in Continuing Education*, 22(2), 151-167. <https://doi.org/10.1080/713695728>
- Boud, D., & Molloy, E. (2013). Rethinking models of feedback for learning: The challenge of design. *Assessment and Evaluation in Higher Education*, 38(6), 698-712. <https://doi.org/10.1080/02602938.2012.691462>
- Boud, D., & Soler, R. (2016). Sustainable assessment revisited. *Assessment and Evaluation in Higher Education*, 41(3), 400-413. <https://doi.org/10.1080/02602938.2015.1018133>
- Bourke, B. (2014). Positionality: Reflecting on the research process. *Qualitative Report*, 19(33), 1-9. <https://doi.org/10.46743/2160-3715/2014.1026>
- Bowen, G. (2009). Document analysis as a qualitative research method. *Qualitative Research Journal*, 9(2), 27-40. <https://doi.org/https://doi.org/10.3316/QRJ0902027>
- Boyer, W. A. R., & Sweeting, T. (1996). Portfolio assessment in primary Physical Education program. *Canadian Children*, 21(1), 19-25. <http://ezproxy.flinders.edu.au/login?url=http://search.ebscohost.com/login.aspx?direct=true&db=s3h&AN=SPH411838&site=ehost-live>
- Braga, L., & Liversedge, P. (2017). Challenges and facilitators to the implementation of a sport education season: The voices of teacher candidates. *Physical Educator*, 74(1), 19. <https://doi.org/https://doi.org/10.18666/tpe-2017-v74-i1-6862>
- Braun, V., & Clarke, V. (2006). Using thematic analysis in psychology. *Qualitative Research in Psychology*, 3(2), 77-101. <https://doi.org/10.1191/1478088706qp063oa>
- Braun, V., & Clarke, V. (2012). Thematic analysis. In H. Cooper, P. M. Camic, D. L. Long, A. T. Panter, D. Rindskopf, & K. J. Sher (Eds.), *APA handbook of research methods in psychology, Vol 2: Research designs: Quantitative, qualitative, neuropsychological, and biological*. (Vol. 2, pp. 57-71). American Psychological Association. <https://doi.org/10.1037/13620-004>
- Braun, V., & Clarke, V. (2013). *Successful qualitative research: A practical guide for beginners*. SAGE.
- Braun, V., & Clarke, V. (2019). Reflecting on reflexive thematic analysis. *Qualitative Research in Sport, Exercise and Health*, 11(4), 589-597. <https://doi.org/10.1080/2159676x.2019.1628806>
- Braun, V., & Clarke, V. (2021a). One size fits all? What counts as quality practice in (reflexive) thematic analysis? *Qualitative Research in Psychology*, 18(3), 328-352. <https://doi.org/10.1080/14780887.2020.1769238>
- Braun, V., & Clarke, V. (2021b). To saturate or not to saturate? Questioning data saturation as a useful concept for thematic analysis and sample-size rationales. *Qualitative Research in Sport, Exercise and Health*, 13(2), 201-216. <https://doi.org/10.1080/2159676X.2019.1704846>
- Braun, V., Clarke, V., Hayfield, N., & Terry, G. (2019). Thematic analysis. In P. Liamputtong (Ed.), *Handbook of research methods in health social sciences* (pp. 843-860). https://doi.org/10.1007/978-981-10-5251-4_103
- Bredt, S., Praça, G., Figueiredo, L., Paula, L., Silva, P., Andrade, A., Greco, P., & Chagas, M. (2016). Reliability of physical, physiological and tactical measures in small-sided soccer games with numerical equality and numerical superiority. *Revista Brasileira de Cineantropometria and Desempenho Humano*, 18(5), 602-610. <https://doi.org/10.5007/1980-0037.2016v18n5p602>
- Breed, R., & Spittle, M. (2011). *Developing game sense through tactical learning: A resource for teachers and coaches*. Cambridge University Press.
- Brewer, C. J., & Jones, R. L. (2002). A five-stage process for establishing contextually valid systematic observation instruments: The case of rugby union. *The Sport Psychologist*, 16(2), 138-159. <https://doi.org/https://doi.org/10.1123/tsp.16.2.138>
- Brinson, J. (2022). Grading less or gradeless: Considerations for music educators. *Canadian Music Educator*, 63(3), 47-53.
- Broadfoot, P., Weeden, P., & Winter, J. (2002). *Assessment: What's in it for schools?* Routledge. <https://doi.org/https://doi-org.ezproxy.flinders.edu.au/10.4324/9780203468920>
- Brookhart, S. (2013). *How to create and use rubrics for formative assessment and grading*. ASCD.
- Brookhart, S. (2018). Appropriate criteria: Key to effective rubrics. *Frontiers in Education*, 3. <https://doi.org/10.3389/feduc.2018.00022>

- Brookhart, S., & Chen, F. (2015). The quality and effectiveness of descriptive rubrics. *Educational Review*, 67(3), 343-368. <https://doi.org/10.1080/00131911.2014.929565>
- Brookhart, S. M. (2003). Developing measurement theory for classroom assessment purposes and uses. *Educational Measurement: Issues and Practice*, 22(4), 5-12. <https://doi.org/10.1111/j.1745-3992.2003.tb00139.x>
- Brookhart, S. M. (2005). Assessment theory for college classrooms. *New Directions for Teaching and Learning*, Winter, (100), 5-14. <https://doi.org/10.1002/tl.165>
- Brown, T. D. (2013). 'In, through and about' movement: Is there a place for the Arnoldian dimensions in the new Australian Curriculum for Health and Physical Education? *Asia-Pacific Journal of Health, Sport and Physical Education*, 4(2), 143-157. <https://doi.org/10.1080/18377122.2013.801107>
- Bryan, C. L., Sims, S. K., Hester, D. J., & Dunaway, D. L. (2013). Fifteen years after the surgeon general's report: Challenges, changes, and future directions in Physical Education. *Quest*, 65(2), 139-150. <https://doi.org/https://doi.org/10.1080/00336297.2013.773526>
- Bryman, A. (2016). *Social research methods* (5th ed.). Oxford University Press.
- Bukamal, H. (2022). Deconstructing insider–outsider researcher positionality. *British Journal of Special Education*, 49(3), 327-349. <https://doi.org/10.1111/1467-8578.12426>
- Bunker, D., & Thorpe, R. (1986). The Curriculum Model. In D. Bunker, R. Thorpe, & L. Almond (Eds.), *Rethinking games teaching* (pp. 7-10). Loughborough University of Technology.
- Butler, S. A., & Hodge, S. (2001). Enhancing student trust through peer assessment in Physical Education. *Physical Educator*, 58(1), 30-41.
- Byrne, D. (2022). A worked example of Braun and Clarke's approach to reflexive thematic analysis. *Quality and Quantity*, 56(3), 1391-1412. <https://doi.org/https://doi.org/10.1007/s11135-021-01182-y>
- Camerino, O., Castañer, M., Anguera, T. M., & Nance, A. J. (2012). *Mixed methods research in the movement sciences: Case studies in sport, Physical Education and dance*. Routledge.
- Capel, S., & Blair, R. (2019). *Debates in Physical Education* (2nd ed.). Taylor & Francis Group. <http://ebookcentral.proquest.com/lib/flinders/detail.action?docID=5989297>
- Carlson, E., Nygren, F., & Wennick, A. (2018). Critical friends: Health professionals' experiences of collegial feedback in a clinical setting. *The Journal of Continuing Education in the Health Professions*, 38(3), 179-183. <https://doi.org/https://doi.org/10.1097/ceh.0000000000000210>
- Casebolt, K., & Zhang, P. (2020). An authentic badminton game performance assessment rubric. *Strategies*, 33(1), 8-13. <https://doi.org/https://doi.org/10.1080/08924562.2019.1680329>
- Casey, A., & Hastie, P. (2011). Students and teacher responses to a unit of student-designed games. *Physical Education and Sport Pedagogy*, 16(3), 295-312. <https://doi.org/https://doi.org/10.1080/17408989.2010.535253>
- Castelão, D., Garganta, J., Santos, R., & da Costa, I. (2014). Comparison of tactical behaviour and performance of youth soccer players in 3 v 3 and 5 v 5 small-sided games. *International Journal of Performance Analysis in Sport*, 14(3), 801-813. <https://doi.org/https://doi.org/10.1080/24748668.2014.11868759>
- Chan, K., Hay, P., & Tinning, R. (2011). Understanding the pedagogic discourse of assessment in Physical Education. *Asia-Pacific Journal of Health, Sport and Physical Education*, 2(1), 3-18. <https://doi.org/https://doi.org/10.1080/18377122.2011.9730340>
- Chappuis, J., Stiggins, R., Chappuis, S., & Arter, J. (2012). *Classroom assessment for student learning: Doing it right-using it well* (2nd ed.). Pearson.
- Charteris, J., & Thomas, E. (2016). Uncovering 'unwelcome truths' through student voice: Teacher inquiry into agency and student assessment literacy. *Teaching Education*, 28(2), 162-177. <https://doi.org/https://doi.org/10.1080/10476210.2016.1229291>
- Chatzopoulos, D., Drakou, A., Kotzamanidou, M., & Tsorbatzoudis, H. (2006). Girls' soccer performance and motivation: Games vs technique approach. *Perceptual and Motor Skills*, 103(2), 463-470. <https://doi.org/https://doi.org/10.2466/pms.103.2.463-470>
- Chen, W., Mason, S., Hypnar, A., & Bennett, A. (2016). Assessing motor skill competency in elementary school students: A three-year study. *Journal of Sports Science and Medicine*, 15(1), 102-110. <http://www.jssm.org>
- Chiles, M. (2020). *The craft of assessment: A whole school approach to assessment of learning* (1st ed.). John Catt Educational, Limited.

- Chng, L. S., & Lund, J. (2018). Assessment for learning in Physical Education: The what, why and how. *Journal of Physical Education, Recreation and Dance*, 89(8), 29-34. <https://doi.org/10.1080/07303084.2018.1503119>
- Cohen, S. A. (1987). Instructional alignment: Searching for a magic bullet. *Educational Researcher*, 16(8), 16-20. <https://doi.org/10.3102/0013189x016008016>
- Christensen, J. H. (2020). Enhancing mixed methods pragmatism with systems theory: Perspectives from educational research. *Systems Research and Behavioral Science*, 39(4), 104-115. <https://doi.org/10.1002/sres.2751>
- Clark, C. (2022). *Victorian Curriculum's achievement standards and assessment - Health and PE*, <https://vimeo.com/694690670/54a4f7a571>
- Collier, D. (2011). Increasing the value of Physical Education: The role of assessment. *Journal of Physical Education, Recreation and Dance*, 82(7), 38-41. <https://doi.org/https://doi.org/10.1080/07303084.2011.10598656>
- Colquhoun, H., Levac, D., O'Brien, K., Straus, S., Tricco, A., Perrier, L., Kastner, M., & Moher, D. (2014). Scoping reviews: Time for clarity in definition, methods, and reporting. *Journal of Clinical Epidemiology* 67(12), 1291-1294. <https://doi.org/http://dx.doi.org/10.1016/j.jclinepi.2014.03.013>
- Company, P., Otey, J., Agost, M.-J., Contero, M., & Camba, J. D. (2019). Teachers as designers of formative e-rubrics: A case study on the introduction and validation of go/no-go criteria. *Universal Access in the Information Society*, 18(3), 675-688. <https://doi.org/10.1007/s10209-019-00686-7>
- Contreras Jordán, O. R., García Lopez, L. M., & Cervelló Gimeno, E. (2005). Transfer of tactical knowledge from invasion games to floorball. *Journal of Human Movement Studies*, 49(3), 193-214.
- Cookson, C. J. (2018). Assessment terms half a century in the making and unmaking: From conceptual ingenuity to definitional anarchy. *Assessment and Evaluation in Higher Education*, 43(6), 930-942. <https://doi.org/https://doi.org/10.1080/02602938.2017.1420138>
- Cooper, B. S., & Gargan, A. (2009). Rubrics in education: Old term, new meanings. *Phi Delta Kappan*, 91(1), 54-55. <https://doi.org/https://doi.org/10.1177/003172170909100109>
- Creswell, J. W. (2012). *Educational research: Planning, conducting, and evaluating quantitative* (4th ed.). Pearson.
- Creswell, J. W., & Creswell, J. D. (2018). *Research design: Qualitative, quantitative, and mixed methods approaches* (5th ed.). SAGE.
- Creswell, J. W., & Plano Clark, V. L. (2018). *Designing and conducting mixed methods research* (3rd ed.). SAGE.
- Crotty, M. (1998). *The foundations of social research: Meaning and perspective in the research process*. SAGE.
- Currell, K., & Jeukendrup, A. E. (2008). Validity, reliability and sensitivity of measures of sporting performance. *Sports Medicine*, 38(4), 297-316.
- da Costa, I. T., Garganta, J., Greco, P., Mesquita, I., & Maia, J. (2011). System of Tactical Assessment in Soccer (FUT-SAT): Development and preliminary validation. *Motricidade*, 7(1), 69-84.
- da Costa, I. T., Garganta, J., Greco, P., Mesquita, I., & Afonso, J. (2010). Assessment of tactical principles in youth soccer players of different age groups. *Revista Portuguesa de Ciências do Desporto*, 10(1), 147-157. <https://doi.org/https://doi.org/10.5628/rpcd.10.01.147>
- Dalglis, S., Khalid, H., & McMahon, S. (2021). Document analysis in health policy research: The read approach. *Health Policy Plan*, 35(10), 1424-1431. <https://doi.org/10.1093/heapol/czaa064>
- Dann, R. (2014). Assessment as learning: Blurring the boundaries of assessment and learning for theory, policy and practice. *Assessment in Education: Principles, Policy and Practice*, 21(2), 149-166. <https://doi.org/https://doi.org/10.1080/0969594X.2014.898128>
- Darnis, F., & Lafont, L. (2013). Cooperative learning and dyadic interactions: Two modes of knowledge construction in socio-constructivist settings for team-sport teaching. *Physical Education and Sport Pedagogy*, 20(5), 459-473. <https://doi.org/https://doi.org/10.1080/17408989.2013.803528>
- Darr, C. (2005). A hitchhiker's guide to validity. *Set: Research Information for Teachers*(2), 55-56.
- Darst, P. W. (1989). *Analyzing Physical Education and sport instruction*. Human Kinetics.

- Dawson, P. (2017). Assessment rubrics: Towards clearer and more replicable design, research and practice. *Assessment and Evaluation in Higher Education*, 42(3), 347-360. <https://doi.org/https://doi.org/10.1080/02602938.2015.1111294>
- den Duyn, N. (1996). Game sense: Why it makes sense to play games. *Sports Coach*, 19, 6-9.
- den Duyn, N. (1997). Game sense: It's time to play! *Sports Coach*, 19, 9-11.
- den Duyn, N. (2000). Games make sense out of sport. *The Sport Educator*, 12(2), 47-48.
- Deneen, C., & Boud, D. (2014). Patterns of resistance in managing assessment change. *Assessment and Evaluation in Higher Education*, 39(5), 577-591. <https://doi.org/https://doi.org/10.1080/02602938.2013.859654>
- Desrosiers, P., Genet-Volet, Y., & Godbout, P. (1997). Teachers' assessment practices viewed through the instruments used in Physical Education classes. *Journal of Teaching in Physical Education*, 16, 211-228. <https://doi.org/https://doi.org/10.1123/jtpe.16.2.211>
- Deuchar, R. (2008). Facilitator, director or critical friend?: Contradiction and congruence in doctoral supervision styles. *Teaching in Higher Education*, 13(4), 489-500. <https://doi.org/10.1080/13562510802193905>
- Diaz del Campo, D. G., González-Villora, S., Garcia Lopez, L. M., & Mitchell, S. (2011). Differences in decision-making development between expert and novice invasion game players. *Perceptual and Motor Skills*, 112(3), 871-888. <https://doi.org/10.2466/05.10.11.25.PMS.112.3.871-888>
- DinanThompson, M. (2013). Claiming 'educative outcomes' in HPE: The potential for 'pedagogic action'. *Asia-Pacific Journal of Health, Sport and Physical Education*, 4(2), 127-142.
- DinanThompson, M., & Penney, D. (2015). Assessment literacy in primary Physical Education. *European Physical Education Review*, 21(4), 485-503. <https://doi.org/https://doi.org/10.1177/1356336x15584087>
- Dodgson, J. E. (2019). Reflexivity in qualitative research. *Journal of Human Lactation*, 35(2), 220-222. <https://doi.org/10.1177/0890334419830990>
- Dreyfus, S. E. (2016). The Five-Stage Model of adult skill acquisition. *Bulletin of Science, Technology and Society*, 24(3), 177-181. <https://doi.org/10.1177/0270467604264992>
- Dudley, D., Goodyear, V., & Baxter, D. (2016). Quality and health-optimizing Physical Education: Using assessment at the health and education nexus. *Journal of Teaching in Physical Education*, 35(4), 324-336. <https://doi.org/10.1123/jtpe.2016-0075>
- Dudley, D., Mackenzie, E., Van Bergen, P., Cairney, J., & Barnett, L. (2022). What drives quality Physical Education? A systematic review and meta-analysis of learning and development effects from Physical Education-based interventions. *Frontiers in Psychology*, 13, 799330. <https://doi.org/10.3389/fpsyg.2022.799330>
- Earl, L. (2014). *Assessment as learning: Using classroom assessment to maximize student learning* (2nd ed.). Hawker Brownlow Education.
- Earl Rinehart, K. (2020). Abductive analysis in qualitative inquiry. *Qualitative Inquiry*, 27(2), 303-311. <https://doi.org/https://doi.org/10.1177/1077800420935912>
- Eddy, L. H., Bingham, D. D., Crossley, K. L., Shahid, N. F., Ellingham-Khan, M., Otteslev, A., Figueredo, N. S., Mon-Williams, M., & Hill, L. J. B. (2020). The validity and reliability of observational assessment tools available to measure fundamental movement skills in school-age children: A systematic review. *PLoS One*, 15(8), e0237919. <https://doi.org/10.1371/journal.pone.0237919>
- Elliott, S., Pankowiak, A., Eime, R., & Drummond, M. (2023). Sport parenting during the COVID-19 pandemic: Perceptions of parents and youth in Australia. *Psychology of Sport and Exercise*, 64, 102299. <https://doi.org/10.1016/j.psychsport.2022.102299>
- Emmanouilidou, K., Derri, V., Aggelousis, N., & Vassiliadou, O. (2012). Development and evaluation of a questionnaire to assess physical educators' knowledge of student assessment. *Physical Educator*, 69(2), 105.
- Ennis, C. D. (1999). Creating a culturally relevant curriculum for disengaged girls. *Sport, Education and Society*, 4(1), 31. <https://doi.org/https://doi.org/10.1080/1357332990040103>
- Essiet, I. A., Warner, E., Lander, N. J., Salmon, J., Duncan, M. J., Eyre, E. L. J., & Barnett, L. M. (2022). Exploring Australian teachers' perceptions of physical literacy: A mixed-methods study. *Physical Education and Sport Pedagogy*, 1-20. <https://doi.org/10.1080/17408989.2022.2028760>

- Estell, J. K., Sapp, H. M., & Reeping, D. (2016, June 26). *Work in progress: Developing single point rubrics for formative assessment*. 2016 American Society for Engineering Education (ASEE) Annual Conference and Exposition, Atlanta. <https://doi.org/10.18260/p.27221>
- Evangelio, C., Sierra-Díaz, M., González-Villora, S., & Clemente, F. (2019). 'Four goals for three players': Using 3 vs. 3 small-sided games at school. *Human Movement*, 20(4), 68-78. <https://doi.org/10.5114/hm.2019.85096>
- Fàbregues, S., & Molina-Azorín, J. F. (2017). Addressing quality in mixed methods research: A review and recommendations for a future agenda. *Quality and Quantity*, 51(6), 2847-2863. <https://doi.org/https://doi.org/10.1007/s11135-016-0449-4>
- Fardilha, F. d. S., & Allen, J. B. (2019). Defining, assessing, and developing creativity in sport: A systematic narrative review. *International Review of Sport and Exercise Psychology*, 13(1), 104-127. <https://doi.org/10.1080/1750984x.2019.1616315>
- Farias, C., Harvey, S., Hastie, P., & Mesquita, I. (2019). Effects of situational constraints on students' game-play development over three consecutive sport education seasons of invasion games. *Physical Education and Sport Pedagogy*, 24(3), 267-286. <https://doi.org/10.1080/17408989.2019.1571184>
- Farias, C., Mesquita, I., & Hastie, P. (2015). Game performance and understanding within a hybrid sport education season. *Journal of Teaching in Physical Education*, 34(3), 363-383. <https://doi.org/10.1123/jtpe.2013-0149>
- Farias, C., Mesquita, I., & Hastie, P. (2019). Student game-play performance in invasion games following three consecutive hybrid sport education seasons. *European Physical Education Review*, 25(3), 691-712. <https://doi.org/10.1177/1356336x18769220>
- Farias, C., Valério, C., & Mesquita, I. (2018). Sport education as a curriculum approach to student learning of invasion games: Effects on game performance and game involvement. *Journal of Sports Science and Medicine*, 17(1), 56-65. <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC5844209/pdf/jssm-17-56.pdf>
- Fenner, J. S., Iga, J., & Unnithan, V. (2016). The evaluation of small-sided games as a talent identification tool in highly trained prepubertal soccer players. *Journal of Sports Sciences*, 34(20), 1983-1990. <https://doi.org/https://doi.org/10.1080/02640414.2016.1149602>
- Fernandes, T., Camerino, O., Garganta, J., Pereira, R., & Barreira, D. (2019). Design and validation of an observational instrument for defence in soccer based on the dynamical systems theory. *International Journal of Sports Science and Coaching*, 14(2), 138-152. <https://doi.org/10.1177/1747954119827283>
- Fink, A., Kosecoff, J., Chassin, M., & Brook, R. H. (1984). Consensus methods: Characteristics and guidelines for use. *American Journal of Public Health* 74(9), 979-983. <https://doi.org/10.2105/AJPH.74.9.979>
- Fluckiger, J. (2010). Single point rubric: A tool for responsible student self-assessment. *The Delta Kappa Gamma Bulletin*, 76(4), 18-25.
- Ford, P. R., Yates, I., & Williams, A. M. (2010). An analysis of practice activities and instructional behaviours used by youth soccer coaches during practice: Exploring the link between science and application. *Journal of Sports Sciences*, 28(5), 483-495. <https://doi.org/10.1080/02640410903582750>
- Fraile, J., Panadero, E., & Pardo, R. (2017). Co-creating rubrics: The effects on self-regulated learning, self-efficacy and performance of establishing assessment criteria with students. *Studies in Educational Evaluation*, 53, 69-76. <https://doi.org/10.1016/j.stueduc.2017.03.003>
- Franklin, R., & Smith, J. (2015). Practical assessment on the run – iPads as an effective mobile and paperless tool in Physical Education and teaching. *Research in Learning Technology*, 23(1), 1-19. <https://doi.org/10.3402/rlt.v23.27986>
- French, K. E., & Thomas, J. R. (1987). The relation of knowledge development to children's basketball performance. *Journal of Sport Psychology*, 9(1), 15-32. <https://doi.org/https://doi.org/10.1123/jsp.9.1.15>
- Gallo, A., Sheehy, D., Patton, K., & Griffin, L. (2006). Assessment benefits and barriers: What are you committed to? *Journal of Physical Education, Recreation and Dance*, 77(8), 46-50. <https://doi.org/https://doi.org/10.1080/07303084.2006.10597926>
- García-López, L. M., González-Villora, S., Gutiérrez, D., & Serra, J. (2013). Development and validation of the Game Performance Evaluation Tool (GPET) in soccer. *Sport TK-Revista Euroamericana de Ciencias del Deporte*, 2(1), 89-99.

- García-Ceberino, J. M., Gamero, M. G., Feu, S., & Ibáñez, S. J. (2020). Experience as a determinant of declarative and procedural knowledge in school football. *International Journal of Environmental Research and Public Health*, 17(3).
<https://doi.org/10.3390/ijerph17031063>
- Garn, A. C., Ware, D. R., & Solmon, M. A. (2011). Student engagement in high school Physical Education: Do social motivation orientations matter? *Journal of Teaching in Physical Education*, 30(1), 84-98. <https://doi.org/https://doi.org/10.1123/jtpe.30.1.84>
- Gensemer, R. E. (1990). *Physical Education: Perspectives, inquiry, applications* (2nd ed.). Brown Publishers.
- Georgakis, S., Wilson, R., & Evans, J. (2015). Authentic assessment in Physical Education: A case study of game sense pedagogy. *Physical Educator*, 72(1), 67-86.
- Gibbons, S. L., & Kankkonen, B. (2011). Assessment as learning in Physical Education: Making assessment meaningful for secondary school students. *Physical and Health Education Journal*, 76(4), 6-12.
- Gibbons, S. L., & Robinson, B. A. (2004). Student-friendly rubrics for personal and social learning in Physical Education. *Physical and Health Education Journal*, 70(4), 4-9.
- Goldkuhl, G. (2011, October 14). Design research in search for a paradigm: Pragmatism is the answer. European Design Science Symposium (EDSS), Leixlip, Ireland.
https://doi.org/10.1007/978-3-642-33681-2_8
- Goldkuhl, G. (2012). Pragmatism vs interpretivism in qualitative information systems research. *European Journal of Information Systems*, 21(2), 135-146.
<https://doi.org/10.1057/ejis.2011.54>
- González-Villora, S., Serra-Olivares, J., Pastor-Vicedo, J. C., & da Costa, I. T. (2015). Review of the tactical evaluation tools for youth players, assessing the tactics in team sports: Football. *SpringerPlus*, 4, 1-17. <https://doi.org/10.1186/s40064-015-1462-0>
- Gouveia, É., Gouveia, B., Marques, A., Kliegel, M., Rodrigues, A., Prudente, J., Lopes, H., & Ihle, A. (2019). The effectiveness of a tactical games approach in the teaching of invasion games. *Journal of Physical Education and Sport*, 19(3), 962-970.
<https://doi.org/10.7752/jpes.2019.s3139>
- Gray, L. M., Wong-Wylie, G., Rempel, G. R., & Cook, K. (2020). Expanding qualitative research interviewing strategies: Zoom video communications. *Qualitative Report*, 25(5), 1292-1301.
<https://doi.org/https://doi.org/10.46743/2160-3715/2020.4212>
- Gray, S., & Sproule, J. (2011). Developing pupils' performance in team invasion games. *Physical Education and Sport Pedagogy*, 16(1), 15-32. <https://doi.org/10.1080/17408980903535792>
- Gray, S., Sproule, J., & Morgan, K. (2009). Teaching team invasion games and motivational climate. *European Physical Education Review*, 15(1), 65-89.
<https://doi.org/https://doi.org/10.1177/1356336x09105212>
- Gray, S., Sproule, J., & Wang, C. K. J. (2008). Pupils' perceptions of and experiences in team invasion games: A case study of a Scottish secondary school and its three feeder primary schools. *European Physical Education Review*, 14(2), 179-201.
<https://doi.org/10.1177/1356336x08090705>
- Green, K. (2002). Physical Education teachers in their figurations: A sociological analysis of everyday 'philosophies'. *Sport, Education and Society*, 7(1), 65-83.
<https://doi.org/10.1080/13573320120113585>
- Gréhaigne, J.-F., Bouthier, D., & Godbout, P. (1999). The foundations of tactics and strategy in team sports. *Journal of Teaching in Physical Education*, 18(2), 159-174.
<https://doi.org/https://doi.org/10.1123/jtpe.18.2.159>
- Gréhaigne, J.-F., Godbout, P., & Bouthier, D. (1997). Performance assessment in team sports. *Journal of Teaching in Physical Education*, 16(4), 500-516.
<https://doi.org/https://doi.org/10.1123/jtpe.16.4.500>
- Gréhaigne, J.-F., Richard, J.-F., & Griffin, L. (2005). *Teaching and learning team sports and games*. Psychology Press.
- Gréhaigne, J.-F., Wallian, N., & Godbout, P. (2005). Tactical-Decision Learning Model and students' practices. *Physical Education and Sport Pedagogy*, 10(3), 255-269.
<https://doi.org/https://doi.org/10.1080/17408980500340869>
- Griffin, L. L., & Butler, J. (2005). *Teaching games for understanding: Theory, research, and practice*. Human Kinetics.

- Griffin, P. (2014). *Assessment for teaching*. Cambridge University Press.
- Guest, G., Bunce, A., & Johnson, L. (2006). How many interviews are enough? *Field Methods*, 18(1), 59-82. <https://doi.org/10.1177/1525822x05279903>
- Gulikers, J. T. M., Bastiaens, T. J., & Kirschner, P. A. (2004). A five-dimensional framework for authentic assessment. *Educational Technology Research and Development*, 52(3), 67-86. <https://doi.org/https://doi.org/10.1007/BF02504676>
- Gutierrez, D., & García-López, L. M. (2012). Gender differences in game behaviour in invasion games. *Physical Education and Sport Pedagogy*, 17(3), 289-301. <https://doi.org/10.1080/17408989.2012.690379>
- Haerens, L., Krijgsman, C., Mouratidis, A., Borghouts, L., Cardon, G., & Aelterman, N. (2019). How does knowledge about the criteria for an upcoming test relate to adolescents' situational motivation in Physical Education? A self-determination theory approach. *European Physical Education Review*, 25(4), 983-1001. <https://doi.org/10.1177/1356336x18783983>
- Halcomb, E. J. (2019). Appraising mixed methods research. In P. Liamputtong (Ed.), *Handbook of Research Methods in Health Social Sciences* (pp. 1051-1067). Springer. https://doi.org/10.1007/978-981-10-5251-4_121
- Hall, T. J., Hicklin, L. K., & French, K. E. (2015). Teacher compliance and accuracy in state assessment of student motor skill performance. *Journal of Teaching in Physical Education*, 34(4), 588-602. <https://doi.org/10.1123/jtpe.2014-0115>
- Harvey, S. (2007). Using a generic invasion game for assessment. *Journal of Physical Education, Recreation and Dance*, 78(4), 19-50. <https://doi.org/https://doi.org/10.1080/07303084.2007.10598002>
- Harvey, S., Cushion, C., Wegis, H., & Massa-Gonzalez, A. (2010). Teaching games for understanding in American high-school soccer: A quantitative data analysis using the Game Performance Assessment Instrument. *Physical Education and Sport Pedagogy*, 15(1), 29-54. <https://doi.org/10.1080/17408980902729354>
- Harvey, S., & Hughes, C. (2009). Teaching and assessing tag rugby made simple. *Strategies*, 22(4), 17-28. <https://doi.org/https://doi.org/10.1080/08924562.2009.10590828>
- Harvey, S., & van der Mars, H. (2010). Teaching and assessing racquet games using "play practice" part 1: Designing the right games. *Journal of Physical Education, Recreation and Dance*, 81(4), 26-54. <https://doi.org/https://doi.org/10.1080/07303084.2010.10598461>
- Hastie, P. A., Layne, T. E., & Mesquita, I. (2013). Evaluating game performance from external experts in sport education: A case study of handball. *Revista Portuguesa de Ciências do Desporto*, 13(1), 23-37. <https://doi.org/https://doi.org/10.5628/rpcd.13.01.23>
- Hastie, P. A., Ward, J. K., & Brock, S. J. (2017). Effect of graded competition on student opportunities for participation and success rates during a season of sport education. *Physical Education and Sport Pedagogy*, 22(3), 316-327. <https://doi.org/http://dx.doi.org/10.1080/17408989.2016.1203888>
- Hattie, J., & Timperley, H. (2007). The power of feedback. *Review of Educational Research*, 77(1), 81-112. <https://doi.org/https://doi.org/10.3102/003465430298487>
- Hay, P., & Macdonald, D. (2008). (Mis) appropriations of criteria and standards-referenced assessment in a performance-based subject. *Assessment in Education: Principles, Policy and Practice*, 15(2), 153-168. <https://doi.org/https://doi.org/10.1080/09695940802164184>
- Hay, P., & Penney, D. (2009). Proposing conditions for assessment efficacy in Physical Education. *European Physical Education Review*, 15(3), 389-405. <https://doi.org/https://doi.org/10.1177/1356336X09364294>
- Hay, P., & Penney, D. (2013). *Assessment in Physical Education: A sociocultural perspective*. Routledge. <https://doi.org/10.4324/9780203133163>
- Hay, P., Tinning, R., & Engstrom, C. (2013). Assessment as pedagogy: A consideration of pedagogical work and the preparation of kinesiology professionals. *Physical Education and Sport Pedagogy*, 20(1), 31-44. <https://doi.org/10.1080/17408989.2013.788145>
- Hennein, R., & Lowe, S. (2020). A hybrid inductive-abductive analysis of health workers' experiences and wellbeing during the COVID-19 pandemic in the United States. *PLoS One*, 15(10), e0240646. <https://doi.org/10.1371/journal.pone.0240646>
- Hennink, M. M., Kaiser, B. N., & Marconi, V. C. (2017). Code saturation versus meaning saturation: How many interviews are enough? *Qualitative Health Research*, 27(4), 591-608. <https://doi.org/10.1177/1049732316665344>

- Henry, S. E., & Feuerstein, A. (2003). What pragmatism can offer educational reform debates: Method, means, and motive. *International Journal of Educational Reform*, 12(2), 146-164. <https://doi.org/10.1177/105678790301200205>
- Hensley, L. D. (1997). Alternative assessment for Physical Education. *Journal of Physical Education, Recreation and Dance*, 68(7), 19-24. <https://doi.org/https://doi.org/10.1080/07303084.1997.10604978>
- Herold, F. (2020). 'There is new wording, but there is no real change in what we deliver': Implementing the new national curriculum for Physical Education in England. *European Physical Education Review*, 26(4), 920-937. <https://doi.org/10.1177/1356336x19892649>
- Hesse-Biber, S., Rodriguez, D., & Frost, N. A. (2015). A qualitatively driven approach to multimethod and mixed methods research. In S. Hesse-Biber & R. B. Johnson (Eds.), *The Oxford Handbook of Multimethod and Mixed Methods Research Inquiry*. Oxford University Press. <https://doi.org/10.1093/oxfordhb/9780199933624.013.3>
- Hollingsworth, H., Heard, J., & Weldon, P. (2019). *Communicating student learning progress: A review of student reporting in Australia*. Australian Council for Educational Research. https://research.acer.edu.au/ar_misc/34
- Holt, J. E., Kinchin, G., & Clarke, G. (2012). Effects of peer-assessed feedback, goal setting and a group contingency on performance and learning by 10–12-year-old academy soccer players. *Physical Education and Sport Pedagogy*, 17(3), 231-250. <https://doi.org/10.1080/17408989.2012.690568>
- Hopper, T. (2003). Four Rs for tactical awareness: Applying game performance assessment in net/wall games. *Teaching Elementary Physical Education*, 14(2), 16-21.
- Hughes, M., Caudrelier, T., James, N., Donnelly, I., Kirkbride, A., & Duschesne, C. (2012). Moneyball and soccer - an analysis of the key performance indicators of elite male soccer players by position. *Journal of Human Sport and Exercise*, 7(2), 402-412. <https://doi.org/10.4100/jhse.2012.72.06>
- Hughes, M., & Franks, I. (2004). *Notational analysis of sport: Systems for better coaching and performance in sport*. Psychology Press.
- Hulteen, R. M., Barnett, L. M., True, L., Lander, N. J., Del Pozo Cruz, B., & Lonsdale, C. (2020). Validity and reliability evidence for motor competence assessments in children and adolescents: A systematic review. *Journal of Sports Sciences*, 38(15), 1717-1798. <https://doi.org/10.1080/02640414.2020.1756674>
- Hulteen, R. M., Lander, N. J., Morgan, P. J., Barnett, L. M., Robertson, S. J., & Lubans, D. R. (2015). Validity and reliability of field-based measures for assessing movement skill competency in lifelong physical activities: A systematic review. *Sports Medicine*, 45(10), 1443-1454. <https://doi.org/10.1007/s40279-015-0357-0>
- Hulteen, R. M., Smith, J. J., Morgan, P. J., Barnett, L. M., Hallal, P. C., Colyvas, K., & Lubans, D. R. (2017). Global participation in sport and leisure-time physical activities: A systematic review and meta-analysis. *Preventive Medicine*, 95, 14-25. <https://doi.org/10.1016/j.ypmed.2016.11.027>
- Ibanez, S. J., Martinez-Fernandez, S., Gonzalez-Espinosa, S., Garcia-Rubio, J., & Feu, S. (2019). Designing and validating a Basketball Learning and Performance Assessment Instrument (BALPAI). *Frontiers in Psychology*, 10, 1595. <https://doi.org/10.3389/fpsyg.2019.01595>
- Inns, J., Petancevski, E. L., Novak, A. R., & Fransen, J. (2023). Decision-making assessments in youth team invasion game athletes: A systematic scoping review. *International Journal of Sports Science and Coaching*. <https://doi.org/10.1177/17479541231185779>
- Ivankova, N. V., Creswell, J. W., & Stick, S. L. (2016). Using mixed-methods sequential explanatory design: From theory to practice. *Field Methods*, 18(1), 3-20. <https://doi.org/10.1177/1525822x05282260>
- Jadeera, P. G. C., Lay, B., & Rizal, R. (2016). Investigating the contextual interference effect using combination sports skills in open and closed skill environments. *Journal of Sports Science and Medicine*, 15(1), 167-175. <https://www.proquest.com/scholarly-journals/investigating-contextual-interference-effect/docview/2295522210/se-2>
- Jenkinson, K., & Benson, A. (2010). Physical Education, sport education and physical activity policies: Teacher knowledge and implementation in their Victorian state secondary school. *European Physical Education Review*, 15(3), 365-388. <https://doi.org/10.1177/1356336x09364456>

- Jiménez-Salas, J., Morillo-Baro, J. P., Reigal, R. E., Morales-Sánchez, V., & Hernández-Mendo, A. (2020). Polar coordinate analysis to study counterattacks in senior and under-16 men's handball. *Cuadernos de Psicología del Deporte*, 20(1), 48-61. <https://doi.org/10.6018/cpd.396521>
- Johnson, R. B. (2017). Dialectical pluralism: A metaparadigm whose time has come. *Journal of Mixed Methods Research*, 11(2), 156-173. <https://doi.org/10.1177/1558689815607692>
- Johnson, R. B., & Onwuegbuzie, A. J. (2004). Mixed methods research: A research paradigm whose time has come. *Educational Researcher*, 33(7), 14-26. <https://doi.org/https://doi.org/10.3102/0013189x033007014>
- Jonsson, A., & Svingby, G. (2007). The use of scoring rubrics: Reliability, validity and educational consequences. *Educational Research Review*, 2(2), 130-144. <https://doi.org/10.1016/j.edurev.2007.05.002>
- Kelly, L. M., & Cordeiro, M. (2020). Three principles of pragmatism for research on organizational processes. *Methodological Innovations*, 13(2). <https://doi.org/10.1177/2059799120937242>
- Kervin, L. K., Vialle, W., Howard, S. J., Herrington, J. A., & Okely, A. D. (2016). *Research for educators* (2nd ed.). Cengage Learning.
- Kilborn, M., Lorusso, J., & Francis, N. (2016). An analysis of Canadian Physical Education curricula. *European Physical Education Review*, 22(1), 23-46. <https://doi.org/10.1177/1356336x15586909>
- Killian, C., & Mays, W. (2021). Assessment practices in K-12 Physical Education in the United States: A scoping review of research, 2000-2020. *Research Quarterly for Exercise and Sport*, 92(2), 248-258. <https://doi.org/10.1080/02701367.2021.1894315>
- Kirk, D. (2006). Sport education, critical pedagogy, and learning theory: Toward an intrinsic justification for Physical Education and youth sport. *Quest*, 58(2), 255-264. <https://doi.org/10.1080/00336297.2006.10491882>
- Kirk, D. (2010). Defining Physical Education: Nature, purposes and future/s. *Physical Education Matters*, 5(3), 30-31.
- Kitts, M. (2003). Assessment strategies for Physical Education. *Strategies*, 16(6), 29-32. <https://doi.org/https://doi.org/10.1080/08924562.2003.10591057>
- Koekoek, J., Van Der Mars, H., van der Kamp, J., Walinga, W., & van Hilvoorde, I. (2018). Aligning digital video technology with game pedagogy in Physical Education. *Journal of Physical Education, Recreation and Dance*, 89(1), 12-22. <https://doi.org/https://doi.org/10.1080/07303084.2017.1390504>
- Koopmann, T., Faber, I., Baker, J., & Schorer, J. (2020). Assessing technical skills in talented youth athletes: A systematic review. *Sports Medicine*, 50(9), 1593-1611. <https://doi.org/10.1007/s40279-020-01299-4>
- Kovács, G., van Hoek, R., & Spens, K. M. (2005). Abductive reasoning in logistics research. *International Journal of Physical Distribution and Logistics Management*, 35(2), 132-144. <https://doi.org/10.1108/09600030510590318>
- Krathwohl, D. R. (2002). A revision of Bloom's taxonomy: An overview. *Theory into Practice*, 41(4), 212-218. https://doi.org/10.1207/s15430421tip4104_2
- Lambert, K., & O'Connor, J. (2023). *Why PE teachers should flip their thinking when adopting a strengths-based approach*. Monash University. Retrieved April 27, 2023, from <https://www.monash.edu/education/teachspace/articles/why-pe-teachers-should-flip-their-thinking-when-adopting-a-strengths-based-approach>
- Lambert, K., & Penney, D. (2020). Curriculum interpretation and policy enactment in Health and Physical Education: Researching teacher educators as policy actors. *Sport, Education and Society*, 25(4), 378-394. <https://doi.org/10.1080/13573322.2019.1613636>
- Lander, N., Barnett, L., Brown, H., & Telford, A. (2015). Physical Education teacher training in fundamental movement skills makes a difference to instruction and assessment practices. *Journal of Teaching in Physical Education*, 34(3), 548-556. <https://doi.org/https://doi.org/10.1123/jtpe.2014-0043>
- Lander, N., Morgan, P. J., Salmon, J., & Barnett, L. M. (2016). Teachers' perceptions of a Fundamental Movement Skill (FMS) assessment battery in a school setting. *Measurement in Physical Education and Exercise Science*, 20(1), 50-62. <https://doi.org/10.1080/1091367x.2015.1095758>

- Lauder, A. (2001). *Play practice: The games approach to teaching and coaching sports*. Human Kinetics.
- Lauder, A., & Piltz, W. (2013a). *Play practice: Engaging and developing skilled players from beginner to elite*. Human Kinetics.
- Lauder, A., & Piltz, W. (2013b). *Play practice: The games approach to teaching and coaching sports*. Human Kinetics.
- Layne, T. E., & Hastie, P. A. (2014). Development of game performance by novice learners participating in a sport education Physical Education unit. *Journal of Physical Education and Sport*, 14(3), 317. <https://doi.org/https://doi.org/10.7752/jpes.2014.03048>
- Lee, M.-A., & Hare, M. K. (2007). A web-based learning portfolio system. *Journal of Physical Education, Recreation and Dance*, 78(9), 45-48. <https://doi.org/https://doi.org/10.1080/07303084.2007.10598097>
- Lee, M.-A., & Ward, P. (2009). Generalization of tactics in tag rugby from practice to games in middle school Physical Education. *Physical Education and Sport Pedagogy*, 14(2), 189-207. <https://doi.org/http://dx.doi.org/10.1080/17408980801974937>
- Leirhaug, P. E., & Annerstedt, C. (2015). Assessing with new eyes? Assessment for learning in Norwegian Physical Education. *Physical Education and Sport Pedagogy*, 21(6), 616-631. <https://doi.org/10.1080/17408989.2015.1095871>
- Levac, D., Colquhoun, H., & O'Brien, K. (2010). Scoping studies: Advancing the methodology. *Implementation Science*, 5(1), 1-9. <https://doi.org/https://doi.org/10.1186/1748-5908-5-69>
- Lindberg-Sand, Å., & Olsson, T. (2008). Sustainable assessment? *International Journal of Educational Research*, 47(3), 165-174. <https://doi.org/10.1016/j.ijer.2008.01.004>
- Lipnevich, A. A., McCallen, L. N., Miles, K. P., & Smith, J. K. (2013). Mind the gap! Students' use of exemplars and detailed rubrics as formative assessment. *Instructional Science*, 42(4), 539-559. <https://doi.org/10.1007/s11251-013-9299-9>
- Llobet-Martí, B., López-Ros, V., Barrera-Gómez, J., & Comino-Ruiz, J. (2016). Assessing novices' game performance in rugby union: The Rugby Attack Assessment Instrument (RAAI). *Journal of Teaching in Physical Education*, 35(2), 181-186. <https://doi.org/10.1123/jtpe.2014-0134>
- López-Pastor, V. M., Kirk, D., Lorente-Catalán, E., MacPhail, A., & Macdonald, D. (2013). Alternative assessment in Physical Education: A review of international literature. *Sport, Education and Society*, 18(1), 57-76. <https://doi.org/https://doi.org/10.1080/13573322.2012.713860>
- Lorente, E., & Kirk, D. (2013). Alternative democratic assessment in PETE: An action-research study exploring risks, challenges and solutions. *Sport, Education and Society*, 18(1), 77-96. <http://ezproxy.flinders.edu.au/login?url=http://search.ebscohost.com/login.aspx?direct=true&db=s3h&AN=84221155&site=ehost-live>
- Low, J. (2019). A pragmatic definition of the concept of theoretical saturation. *Sociological Focus*, 52(2), 131-139. <https://doi.org/10.1080/00380237.2018.1544514>
- Lund, J. (1992). Assessment and accountability in secondary Physical Education. *Quest*, 44(3), 352-360. <https://doi.org/https://doi.org/10.1080/00336297.1992.10484061>
- Lund, J. (1997). Authentic assessment: Its development and applications. *Journal of Physical Education, Recreation and Dance*, 68(7), 25-28. <https://doi.org/https://doi.org/10.1080/07303084.1997.10604979>
- Lund, J. (2006). NASPE/NCATE report preparation for the accreditation process. *Journal of Physical Education, Recreation and Dance*, 77(3), 13-31. <https://doi.org/10.1080/07303084.2006.10597839>
- Lund, J., & Veal, M. L. (2008). Chapter 4: Measuring pupil learning - How do student teachers assess within instructional models? *Journal of Teaching in Physical Education*, 27(4), 487-511. <https://doi.org/doi:10.1123/jtpe.27.4.487>
- Lund, J., & Veal, M. L. (2013). *Assessment-driven instruction in Physical Education with web resource: A standards-based approach to promoting and documenting learning*. Human Kinetics.
- Lynch, T. (2014). Australian Curriculum reform II: Health and Physical Education. *European Physical Education Review*, 20(4), 508-524. <https://doi.org/https://doi.org/10.1177/1356336X14535166>

- Lynch, T., Soukup, G. J., & Usher, W. (2016). "Physical Education", "Health and Physical Education", "Physical Literacy" and "Health Literacy": Global nomenclature confusion. *Cogent Education*, 3(1). <https://doi.org/10.1080/2331186x.2016.1217820>
- Maarouf, H. (2019). Pragmatism as a supportive paradigm for the mixed research approach: Conceptualizing the ontological, epistemological, and axiological stances of pragmatism. *International Business Research*, 12(9). <https://doi.org/10.5539/ibr.v12n9p1>
- Mackieson, P., Shlonsky, A., & Connolly, M. (2018). Increasing rigor and reducing bias in qualitative research: A document analysis of parliamentary debates using applied thematic analysis. *Qualitative Social Work*, 18(6), 965-980. <https://doi.org/10.1177/1473325018786996>
- MacPhail, A., Kirk, D., & Griffin, L. (2008). Throwing and catching as relational skills in game play: Situated learning in a modified game unit. *Journal of Teaching in Physical Education*, 27(1), 100-115. <https://doi.org/https://doi.org/10.1123/jtpe.27.1.100>
- MacPhail, A., & Murphy, F. (2017). Too much freedom and autonomy in the enactment of assessment? Assessment in Physical Education in Ireland. *Irish Educational Studies*, 36(2), 1-16. <https://doi.org/https://doi.org/10.1080/03323315.2017.1327365>
- MacPhail, A., Tannehill, D., Leirhaug, P. E., & Borghouts, L. (2023). Promoting instructional alignment in Physical Education teacher education. *Physical Education and Sport Pedagogy*, 28(2), 153-164. <https://doi.org/10.1080/17408989.2021.1958177>
- Maguire, M., & Delahunt, B. (2017). Doing a thematic analysis: A practical, step-by-step guide for learning and teaching scholars. *All Ireland Journal of Higher Education*, 9(3).
- Maher, C., Hadfield, M., Hutchings, M., & De Eyto, A. (2018). Ensuring rigor in qualitative data analysis: A design research approach to coding combining NVivo with traditional material methods. *International Journal of Qualitative Methods*, 17(1), 1-13. <https://doi.org/https://doi.org/10.1177/1609406918786362>
- Malterud, K., Siersma, V. D., & Guassora, A. D. (2016). Sample size in qualitative interview studies: Guided by information power. *Qualitative Health Research*, 26(13), 1753-1760. <https://doi.org/10.1177/1049732315617444>
- Martin, C., Horton, M. L., & Tarr, S. J. (2015). Building assessment tools aligned with grade-level outcomes. *Journal of Physical Education, Recreation and Dance*, 86(7), 28-34. <https://doi.org/https://doi.org/10.1080/07303084.2015.1064713>
- Masters, G. (2013). *Reforming educational assessment: Imperatives, principles and challenges*. Australian Council for Educational Research (ACER). <http://research.acer.edu.au/aer/12/>
- Masters, G. (2014). Assessment: Getting to the essence. *Designing the Future*(1), 1-6.
- Matshedisho, K. R. (2019). Straddling rows and columns: Students' (mis)conceptions of an assessment rubric. *Assessment and Evaluation in Higher Education*, 45(2), 169-179. <https://doi.org/10.1080/02602938.2019.1616671>
- McChesney, K., & Aldridge, J. (2019). Weaving an interpretivist stance throughout mixed methods research. *International Journal of Research and Method in Education*, 42(3), 225-238. <https://doi.org/10.1080/1743727x.2019.1590811>
- McGannon, K. R., Smith, B., Kendellen, K., & Gonsalves, C. A. (2021). Qualitative research in six sport and exercise psychology journals between 2010 and 2017: An updated and expanded review of trends and interpretations. *International Journal of Sport and Exercise Psychology*, 19(3), 359-379. <https://doi.org/10.1080/1612197x.2019.1655779>
- McLennan, J., & Thompson, J. (2015). *Quality Physical Education (QPE): Guidelines for policy makers* [eBook]. United Nations Educational, Scientific and Cultural Organization <https://unesdoc.unesco.org/ark:/48223/pf0000231101>
- Melograno, V. J. (1997). Integrating assessment into Physical Education teaching. *Journal of Physical Education, Recreation and Dance*, 68(7), 34-37. <https://doi.org/10.1080/07303084.1997.10604981>
- Memmert, D. (2010). Testing of tactical performance in youth elite soccer. *Journal of Sports Science and Medicine*, 9(2), 199-205.
- Memmert, D., & Harvey, S. (2008). The Game Performance Assessment Instrument (GPAI): Some concerns and solutions for further development. *Journal of Teaching in Physical Education*, 27(2), 220-240. <https://doi.org/https://doi.org/10.1123/jtpe.27.2.220>

- Memmert, D., & Harvey, S. (2010). Identification of non-specific tactical tasks in invasion games. *Physical Education and Sport Pedagogy*, 15(3), 287-305. <https://doi.org/10.1080/17408980903273121>
- Mesquita, I., Farias, C., & Hastie, P. (2012). The impact of a hybrid Sport Education-Invasion Games Competence Model soccer unit on students' decision making, skill execution and overall game performance. *European Physical Education Review*, 18(2), 205-219. <https://doi.org/10.1177/1356336x12440027>
- Metzler, M. W. (2011). *Instructional models for Physical Education* (3rd ed.). Holcomb Hathaway.
- Miller, A., Eather, N., Duncan, M., & Lubans, D. (2019). Associations of object control motor skill proficiency, game play competence, physical activity and cardiorespiratory fitness among primary school children. *Journal of Sports Sciences*, 37(2), 173-179. <https://doi.org/10.1080/02640414.2018.1488384>
- Mitchell, A. (2018). A review of mixed methods, pragmatism and abduction techniques. *Electronic Journal of Business Research Methods*, 16(3), 103-116. <https://academic-publishing.org/index.php/ejbrm/article/view/1378>
- Mitchell, S., Griffin, L., & Oslin, J. (1994). Tactical awareness as a developmentally appropriate focus for the teaching of games in elementary and secondary Physical Education. *Physical Educator*, 51(1), 21-28.
- Mitchell, S., Oslin, J., & Griffin, L. (2013). *Teaching sport concepts and skills: A tactical games approach for ages 7 to 18*. Human Kinetics.
- Mohler, E., & Rudman, D. (2022). Negotiating the insider/outsider researcher position within qualitative disability studies research. *Qualitative Report*, 27(6), 1511-1521 <https://doi.org/10.46743/2160-3715/2022.5047>
- Mohnsen, B. (1998). Assessing and grading middle school students. *Teaching Elementary Physical Education*, 9(6), 13-15.
- Mohnsen, B. (2006). Assessment and grading in Physical Education. *Strategies*, 20(2), 24-28.
- Morgan, P. J., & Hansen, V. (2008). Classroom teachers' perceptions of the impact of barriers to teaching Physical Education on the quality of Physical Education programs. *Research Quarterly for Exercise and Sport*, 79(4), 506-516. <https://doi.org/10.1080/02701367.2008.10599517>
- Morrow, J. R. (1978). Measurement techniques- Who uses them? *Journal of Physical Education and Recreation*, 49(9), 66-67.
- Mosston, M. (1966). *Teaching Physical Education from command to discovery*. C. E. Merrill Books.
- Mosston, M. (2002). *Teaching Physical Education* (5th ed.). B. Cummings.
- Moura, A., Graça, A., MacPhail, A., & Batista, P. (2021). Aligning the principles of assessment for learning to learning in Physical Education: A review of literature. *Physical Education and Sport Pedagogy*, 26(4), 388-401. <https://doi.org/10.1080/17408989.2020.1834528>
- Nadeau, L., Godbout, P., & Richard, J. F. (2008). Assessment of ice hockey performance in real-game conditions. *European Journal of Sport Science*, 8(6), 379-388. <https://doi.org/10.1080/17461390802284456>
- Nadeau, L., Richard, J. F., & Godbout, P. (2008). The validity and reliability of a performance assessment procedure in ice hockey. *Physical Education and Sport Pedagogy*, 13(1), 65-83. <https://doi.org/10.1080/17408980701444718>
- Nelson, J. (2016). Using conceptual depth criteria: Addressing the challenge of reaching saturation in qualitative research. *Qualitative Research*, 17(5), 554-570. <https://doi.org/10.1177/1468794116679873>
- Nevett, M., Rovegno, I., Babiarez, M., & McCaughtry, N. (2001). Chapter 6: Changes in basic tactics and motor skills in an invasion-type game after a 12-lesson unit of instruction. *Journal of Teaching in Physical Education*, 20(4), 352-369. <https://doi.org/https://doi.org/10.1123/jtpe.20.4.352>
- Nevill, A., Atkinson, G., & Hughes, M. (2008). Twenty-five years of sport performance research in the Journal of Sports Sciences. *Journal of Sports Sciences*, 26(4), 413-426. <https://doi.org/10.1080/02640410701714589>
- Newton, P. E. (2007). Clarifying the purposes of educational assessment. *Assessment in Education: Principles, Policy and Practice*, 14(2), 149-170. <https://doi.org/10.1080/09695940701478321>

- Nieminen, J. H., & Tuohilampi, L. (2020). 'Finally studying for myself' – examining student agency in summative and formative self-assessment models. *Assessment and Evaluation in Higher Education*, 45(7), 1031-1045. <https://doi.org/10.1080/02602938.2020.1720595>
- O'Connor, J., Alfrey, L., & Payne, P. (2012). Beyond games and sports: A socio-ecological approach to Physical Education. *Sport, Education and Society*, 17(3), 365-380. <https://doi.org/10.1080/13573322.2011.608940>
- O'Connor, J., & Penney, D. (2020). Informal sport and curriculum futures: An investigation of the knowledge, skills and understandings for participation and the possibilities for Physical Education. *European Physical Education Review*, 27(1), 3-26. <https://doi.org/10.1177/1356336x20915937>
- O'Donoghue, T., & Farrelly, T. (2022). Mixed methods or mixed up: A critical exposition on interpretive research undertaken within mixed methods' studies. *Irish Educational Studies*, 1-16. <https://doi.org/10.1080/03323315.2022.2148265>
- Onwuegbuzie, A. J., & Collins, K. M. T. (2007). A typology of mixed methods sampling designs in social science research [Report]. *The Qualitative Report*, 12, 281+. <https://link.gale.com/apps/doc/A172525650/AONE?u=flinders&sid=bookmark-AONE&xid=5c978dc9>
- Ortega-Toro, E., Garcia-Angulo, A., Gimenez-Egido, J. M., Garcia-Angulo, F. J., & Palao, J. M. (2019). Design, validation, and reliability of an observation Instrument for technical and tactical actions of the offense phase in soccer. *Frontiers in Psychology*, 10, 22. <https://doi.org/10.3389/fpsyg.2019.00022>
- Oslin, J., Mitchell, S., & Griffin, L. (1998). The Game Performance Assessment Instrument (GPAI): Development and preliminary validation. *Journal of Teaching in Physical Education*, 17(2), 231-243. <https://doi.org/https://doi.org/10.1123/jtpe.17.2.231>
- Otero-Saborido, F. M., & González-Jurado, J. A. (2015). Design and validation of a tool for the formative assessment of invasion games. *Journal of Physical Education and Sport*, 15(2), 254-263. <https://doi.org/https://doi.org/10.7752/jpes.2015.02039>
- Panadero, E., & Jonsson, A. (2013). The use of scoring rubrics for formative assessment purposes revisited: A review. *Educational Research Review*, 9, 129-144. <https://doi.org/10.1016/j.edurev.2013.01.002>
- Panadero, E., & Jonsson, A. (2020). A critical review of the arguments against the use of rubrics. *Educational Research Review*, 30, 1-19. <https://doi.org/10.1016/j.edurev.2020.100329>
- Panadero, E., & Romero, M. (2014). To rubric or not to rubric? The effects of self-assessment on self-regulation, performance and self-efficacy. *Assessment in Education: Principles, Policy and Practice*, 21(2), 133-148. <https://doi.org/10.1080/0969594x.2013.877872>
- Panadero, E., Tapia, J. A., & Huertas, J. A. (2012). Rubrics and self-assessment scripts effects on self-regulation, learning and self-efficacy in secondary education. *Learning and Individual Differences*, 22(6), 806-813. <https://doi.org/10.1016/j.lindif.2012.04.007>
- Park, Y. (2017). Examining South Korea's elementary Physical Education performance assessment using assessment literacy perspectives. *International Electronic Journal of Elementary Education*, 10(2), 207-213. <https://doi.org/10.26822/iejee.2017236116>
- Penney, D., Brooker, R., Hay, P., & Gillespie, L. (2009). Curriculum, pedagogy and assessment: Three message systems of schooling and dimensions of quality Physical Education. *Sport, Education and Society*, 14(4), 421-442. <https://doi.org/https://doi.org/10.1080/13573320903217125>
- Penney, D., Jones, A., Newhouse, P., & Cambell, A. (2012). Developing a digital assessment in senior secondary Physical Education. *Physical Education and Sport Pedagogy*, 17(4), 383-410. <https://doi.org/10.1080/17408989.2011.582490>
- Pérez-Morales, J. C., Greco, P. J., Ferreira-Lopes, B., Estevão, B. J., & Ibañez, S. J. (2018). Development and preliminary validation of a new Procedural Tactical Knowledge Test for basketball using 3vs.3 situation. *Revista Internacional de Ciencias del Deporte*, 14(53), 256-267. <https://doi.org/10.5232/ricyde2018.05306>
- Perlman, D. J., & Forrest, G. (2015). An examination of in-class physical activity across games classifications. *Physical Educator*, 72(2), 266-277.
- Pesce, C., Faigenbaum, A., Crova, C., Marchetti, R., & Bellucci, M. (2013). Benefits of multi-sports Physical Education in the elementary school context. *Health Education Journal*, 72(3), 326-336. <https://doi.org/https://doi.org/10.1177/0017896912444176>

- Peters, M., Marnie, C., Tricco, A., Pollock, D., Munn, Z., Alexander, L., McInerney, P., Godfrey, C., & Khalil, H. (2020). Updated methodological guidance for the conduct of scoping reviews. *JBI Evid Synth*, 18(10), 2119-2126. <https://doi.org/10.11124/JBIES-20-00167>
- Peters, M. D. J., Marnie, C., Tricco, A. C., Pollock, D., Munn, Z., Alexander, L., McInerney, P., Godfrey, C. M., & Khalil, H. (2021). Updated methodological guidance for the conduct of scoping reviews. *JBI Evidence Implementation*, 19(1), 3-10. <https://doi.org/10.1097/XEB.0000000000000277>
- Pham, M. T., Rajić, A., Greig, J. D., Sargeant, J. M., Papadopoulos, A., & McEwen, S. A. (2014). A scoping review of scoping reviews: Advancing the approach and enhancing the consistency. *Research Synthesis Methods*, 5(4), 371-385. <https://doi.org/10.1002/jrsm.1123>
- Pill, S. (2007a). Physical Education-what's in a name?: A praxis model for holistic learning in Physical Education. *ACHPER Australia Healthy Lifestyles Journal*, 54(1), 5-10.
- Pill, S. (2007b). *Play with purpose: A resource to support teachers in the implementation of the game-centred approach to Physical Education*. ACHPER
- Pill, S. (2011). Seizing the moment: Can game sense further inform sport teaching in Australian Physical Education? *Phenex Journal*, 3(1), 1-15. <https://ojs.acadiau.ca/index.php/phenex/article/view/1327>
- Pill, S. (2012). *Play with purpose: Developing game sense in afl footballers*. ACHPER.
- Pill, S. (2013). *Play with purpose: Game sense to sport literacy: A resource to support teachers and coaches in the implementation of a game-centred approach to sport teaching*. ACHPER.
- Pill, S., & Stolz, S. (2017). Exploring Australian secondary Physical Education teachers' understanding of Physical Education in the context of new curriculum familiarisation. *Asia-Pacific Journal of Health, Sport and Physical Education*, 8(1), 67-79. <https://doi.org/10.1080/18377122.2016.1272425>
- Pill, S., Swabey, K., & Penney, D. (2017). Investigating PE teacher use of models based practice in Australian secondary PE. *Phenex Journal*, 9(1), 1-17.
- Pinheiro, V. (1994). Diagnosing motor skills—A practical approach. *Journal of Physical Education, Recreation and Dance*, 65(2), 49-54. <https://doi.org/https://doi.org/10.1080/07303084.1994.10606857>
- Piotrowski, S., & Capel, S. (2000). Formal and informal modes of assessment in Physical Education. In S. Piotrowski & S. Capel (Eds.), *Issues in Physical Education* (pp. 99-114). Taylor & Francis Group. <https://doi.org/https://doi.org/10.4324/9780203135716>
- Placek, J. (1983). Conceptions of success in teaching: Busy, happy and good. *Teaching in Physical Education*, 14, 46-56.
- Popham, J. (1997). What's wrong-and what's right - with rubrics. *Educational Leadership*, 55(2), 72-75.
- Potrac, P., Jones, R., & Nelson, L. (2014). Interpretivism. In L. Nelson, R. Groom, & P. Potrac (Eds.), *Research Methods in Sports Coaching* (pp. 31-41). Routledge.
- Praxedes, A., Moreno, A., Gil-Arias, A., Claver, F., & Del Villar, F. (2018). The effect of small-sided games with different levels of opposition on the tactical behaviour of young footballers with different levels of sport expertise. *PLoS One*, 13(1), e0190157. <https://doi.org/10.1371/journal.pone.0190157>
- Ramanadhan, S., Revette, A. C., Lee, R. M., & Aveling, E. L. (2021). Pragmatic approaches to analyzing qualitative data for implementation science: An introduction. *Implementation Science Communications*, 2(1), 70. <https://doi.org/10.1186/s43058-021-00174-1>
- Rapley, T. (2018). *Doing conversation, discourse and document analysis* (2nd ed.). SAGE <https://doi.org/10.4135/9781526441843>
- Reddy, Y. M., & Andrade, H. (2010). A review of rubric use in higher education. *Assessment and Evaluation in Higher Education*, 35(4), 435-448. <https://doi.org/10.1080/02602930902862859>
- Redelius, K., & Hay, P. J. (2012). Student views on criterion-referenced assessment and grading in Swedish Physical Education. *Physical Education and Sport Pedagogy*, 17(2), 211-225. <https://doi.org/10.1080/17408989.2010.548064>
- Reid, A., & Price, D. (2018). *The Australian Curriculum: Promises, problems and possibilities*. Australian Curriculum Studies Association.

- Richard, J.-F., Godbout, P., & Gréhaigne, J.-F. (1998). The establishment of team-sport performance norms for grade 5 to 8 students. *Avante-Ontario*, 4, 1-19.
- Richard, J.-F., Godbout, P., & Gréhaigne, J.-F. (2000). Students' precision and interobserver reliability of performance assessment in team sports. *Research Quarterly for Exercise and Sport*, 71(1), 85-91. <https://doi.org/10.1080/02701367.2000.10608885>
- Rivas, M. R., De La Serna, M. C., & Martínez-Figueira, E. (2014). Electronic rubrics to assess competences in ICT subjects. *European Educational Research Journal*, 13(5), 584-594. <https://doi.org/10.2304/eerj.2014.13.5.584>
- Robertson, S. J., Burnett, A. F., & Cochrane, J. (2014). Tests examining skill outcomes in sport: A systematic review of measurement properties and feasibility. *Sports Medicine*, 44(4), 501-518. <https://doi.org/10.1007/s40279-013-0131-0>
- Robinson, D., & Melnychuk, N. (2009). Assessing games understanding. *Strategies*, 22(5), 25-28. <https://doi.org/https://doi.org/10.1080/08924562.2009.10590837>
- Robinson, D., & Randall, L. (2017). Gadgets in the gymnasium: Physical educators' use of digital technologies. *Canadian Journal of Learning and Technology*, 43(1). <https://doi.org/10.21432/T24C82>
- Rowat, O., Fenner, J., & Unnithan, V. (2017). Technical and physical determinants of soccer match-play performance in elite youth soccer players. *Journal of Sports Medicine and Physical Fitness*, 57(4), 369-379. <https://doi.org/10.23736/S0022-4707.16.06093-X>
- Rowlands, T., Waddell, N., & McKenna, B. (2015). Are we there yet? A technique to determine theoretical saturation. *Journal of Computer Information Systems*, 56(1), 40-47. <https://doi.org/10.1080/08874417.2015.11645799>
- Roy, K., Zvonkovic, A., Goldberg, A., Sharp, E., & LaRossa, R. (2015). Sampling richness and qualitative integrity: Challenges for research with families. *Journal of Marriage and Family*, 77(1), 243-260. <https://doi.org/10.1111/jomf.12147>
- Ryan, G. (2018). Introduction to positivism, interpretivism and critical theory. *Nurse Researcher*, 25(4), 14-20. <https://doi.org/10.7748/nr.2018.e1466>
- Rylander, P., Karlsteen, M., Kougioumtzis, K., & Gustafsson, J. E. (2019). The specificity versus generality of ball-handling skills - Empirical evidence for a general ball-handling ability. *Human Movement Science*, 66, 477-486. <https://doi.org/10.1016/j.humov.2019.05.024>
- Scanlon, D., MacPhail, A., Walsh, C., & Tannehill, D. (2023). Embedding assessment in learning experiences: Enacting the principles of instructional alignment in Physical Education teacher education. *Curriculum Studies in Health and Physical Education*, 14(1), 3-20. <https://doi.org/10.1080/25742981.2022.2039074>
- Schoonenboom, J., & Johnson, R. B. (2017). How to construct a mixed methods research design. *Kolner Zeitschrift für Soziologie und Sozialpsychologie*, 69(Suppl 2), 107-131. <https://doi.org/10.1007/s11577-017-0454-1>
- Sgrò, F., Schembri, R., Nicolosi, S., Manzo, G., & Lipoma, M. (2013). A mixed-method approach for the assessment of fundamental movement skills in Physical Education. *Procedia - Social and Behavioral Sciences*, 106, 102-111. <https://doi.org/10.1016/j.sbspro.2013.12.013>
- Shan, Y. (2021). Philosophical foundations of mixed methods research. *Philosophy Compass*, 17(1). <https://doi.org/10.1111/phc3.12804>
- SHAPE. (2014). *National standards and grade-level outcomes for K-12 Physical Education*. Human Kinetics.
- Shaw, G. F. (2014). Introducing rubrics to Physical Education teacher candidates. *Journal of Physical Education, Recreation and Dance*, 85(6), 31-37. <https://doi.org/https://doi.org/10.1080/07303084.2014.926846>
- Shaw, G. F. (2015). Introducing iPad applications to teacher education students in a fencing course: A journey of self discovery. *Journal of Physical Education, Recreation and Dance*, 86(5), 21-26. <https://doi.org/https://doi.org/10.1080/07303084.2015.1022671>
- Shusterman, R. (2010). What pragmatism means to me: Ten principles. *Revue Française D'études Américaines*, 124, 59-65. <https://doi.org/https://doi.org/10.3917/rfea.124.0059>
- Siedentop, D. (1987). High school Physical Education: Still an endangered species. *Journal of Physical Education, Recreation and Dance*, 58(2), 24-25. <https://doi.org/10.1080/07303084.1987.10609502>

- Siedentop, D. (1998). What is sport education and how does it work? *Journal of Physical Education, Recreation and Dance*, 69(4), 18-20.
<https://doi.org/10.1080/07303084.1998.10605528>
- Siedentop, D., Hastie, P., & van der Mars, H. (2011). *Complete guide to sport education* (2nd ed.). Human Kinetics.
- Simon, S., Collins, L., & Collins, D. (2017). Observational heuristics in a group of high level paddle sports coaches. *International Sport Coaching Journal*, 4(2), 235-245.
<https://doi.org/10.1123/iscj.2017-0012>
- Slade, D. G. (2010). *Transforming play: Teaching tactics and game sense*. Human Kinetics.
- Slingerland, M., Haerens, L., Cardon, G., & Borghouts, L. (2014). Differences in perceived competence and physical activity levels during single-gender modified basketball game play in middle school Physical Education. *European Physical Education Review*, 20(1), 20-35. <https://doi.org/https://doi.org/10.1177/1356336X13496000>
- Smith, B. (2017). Generalizability in qualitative research: Misunderstandings, opportunities and recommendations for the sport and exercise sciences. *Qualitative Research in Sport, Exercise and Health*, 10(1), 137-149. <https://doi.org/10.1080/2159676x.2017.1393221>
- Smith, B., & McGannon, K. R. (2018). Developing rigor in qualitative research: Problems and opportunities within sport and exercise psychology. *International Review of Sport and Exercise Psychology*, 11(1), 101-121. <https://doi.org/10.1080/1750984X.2017.1317357>
- Smith, C. D., Worsfold, K., Davies, L., Fisher, R., & McPhail, R. (2013). Assessment literacy and student learning: The case for explicitly developing students 'assessment literacy'. *Assessment and Evaluation in Higher Education*, 38(1), 44-60.
<https://doi.org/10.1080/02602938.2011.598636>
- Smith, D. (1996). Peer coaches' problems with videotape recording for teacher observation. *Action in Teacher Education*, 17(4), 18-27.
<https://doi.org/https://doi.org/10.1080/01626620.1996.10463349>
- Smith, T. K. (1997). Authentic assessment: Using a portfolio card in Physical Education. *Journal of Physical Education, Recreation and Dance*, 68(4), 46-52.
<https://doi.org/https://doi.org/10.1080/07303084.1997.10604927>
- Spandel, V. (2006). In defense of rubrics. *English Journal*, 96(1), 19-22.
- Sparkes, A. C., & Smith, B. (2009). Judging the quality of qualitative inquiry: Criteriology and relativism in action. *Psychology of Sport and Exercise*, 10(5), 491-497.
<https://doi.org/10.1016/j.psychsport.2009.02.006>
- Stanley, T. (2019). *Using rubrics for performance-based assessment: A practical guide to evaluating student work*. Taylor & Francis Group.
<http://ebookcentral.proquest.com/lib/flinders/detail.action?docID=6736778>
- Starck, J. R., Richards, K. A. R., & O'Neil, K. (2018). A conceptual framework for assessment literacy: Opportunities for Physical Education teacher education. *Quest*, 70(4), 519-535.
<https://doi.org/10.1080/00336297.2018.1465830>
- State Government of Victoria. (2018). *All schools FTE enrolments - February 2018*. DET. Retrieved July 5, 2023, from <https://discover.data.vic.gov.au/dataset/all-schools-fte-enrolments-feb-2018>
- Stiggins, R. (2014). Improve assessment literacy outside of schools too. *Phi Delta Kappan*, 96(2), 67-72.
- Stiggins, R. (2018). Better assessments require better assessment literacy. *Educational Leadership*, 75(5), 18-19.
- Stiggins, R. J. (1991). Assessment literacy. *Phi Delta Kappan International*, 72(7), 534-539.
<https://www.jstor.org/stable/20404455>
- SueSee, B., Edwards, K., Pill, S., & Cuddihy, T. (2018). Self-reported teaching styles of Australian senior Physical Education teachers. *Curriculum Perspectives*, 38(1), 41-54.
<https://doi.org/10.1007/s41297-018-0041-2>
- Sutton, A., Clowes, M., Preston, L., & Booth, A. (2019). Meeting the review family: Exploring review types and associated information retrieval requirements. *Health Information and Libraries Journal*, 36(3), 202-222. <https://doi.org/https://doi.org/10.1111/hir.12276>
- Svennberg, L. (2017). Swedish PE teachers' understandings of legitimate movement in a criterion-referenced grading system. *Physical Education and Sport Pedagogy*, 22(3), 257-269.
<https://doi.org/http://dx.doi.org/10.1080/17408989.2016.1176132>

- Svennberg, L., Meckbach, J., & Redelius, K. (2014). Exploring PE teachers' 'gut feelings': An attempt to verbalise and discuss teachers' internalised grading criteria. *European Physical Education Review*, 20(2), 199-214. <https://doi.org/https://doi.org/10.1177/1356336X13517437>
- Tallir, I. B., Lenoir, M., Valcke, M., & Musch, E. (2007). Do alternative instructional approaches result in different game performance learning outcomes? Authentic assessment in varying game conditions. *International Journal of Sport Psychology*, 38(3), 263-282. <https://tinyurl.com/Tallir2007>
- Tangalos, C., Robertson, S. J., Spittle, M., & Gastin, P. B. (2015). Predictors of individual player match performance in junior Australian football. *International Journal of Sports Physiology and Performance*, 10(7), 853-859. <https://doi.org/10.1123/ijspp.2014-0428>
- Tanner, P. (2023). Nested sampling in sequential mixed methods research designs: Comparing recipe and result. *Forum, qualitative social research*, 24(1). <https://doi.org/10.17169/fqs-24.1.4004>
- Taylor, W. (2014). Analysis of qualitative data. In L. Nelson, R. Groom, & P. Potrac (Eds.), *Research methods in sports coaching* (pp. 181-191). Routledge.
- Teaching Games for Understanding Special Interest Group [TGfU SIG] (2021). *Game-based consensus statement*. <http://www.tgfu.info/game-based-consensus-statement.html>
- Teddlie, C., & Tashakkori, A. (2009). *Foundations of mixed methods research: Integrating quantitative and qualitative approaches in the social and behavioral sciences*. SAGE
- Teddlie, C., & Tashakkori, A. (2012). Common "core" characteristics of mixed methods research. *American Behavioral Scientist*, 56(6), 774-788. <https://doi.org/10.1177/0002764211433795>
- Terry, G., Hayield, N., Braun, V., & Clarke, V. (2017). Thematic analysis. In C. Willig & W. S. Rogers (Eds.), *The SAGE handbook of qualitative research in psychology* (pp. 17-37). <https://doi.org/10.4135/9781526405555>
- Terwee, C. B., Bot, S. D. M., de Boer, M. R., van der Windt, D. A. W. M., Knol, D. L., Dekker, J., Bouter, L. M., & de Vet, H. C. W. (2007). Quality criteria were proposed for measurement properties of health status questionnaires. *Journal of Clinical Epidemiology* 60(1), 34-42. <https://doi.org/10.1016/j.jclinepi.2006.03.012>
- Thomas, J. R., Silverman, S., & Nelson, J. (2015). *Research methods in physical activity* (7th ed.). Human Kinetics.
- Thorpe, R., Bunker, D., & Almond, L. (1984). Four fundamentals for planning a games curriculum. *Bulletin of Physical Education*, 20(1), 24-29.
- Tierney, R., & Simon, M. (2004). What's still wrong with rubrics: Focusing on the consistency of performance criteria across scale levels. *Practical Assessment, Research, and Evaluation*, 9(1), 2.
- Tillema, H. (2014). Student involvement in assessment of their learning. In C. Wyatt-Smith, V. Klenowski, & P. Colbert (Eds.), *Designing assessment for quality learning*, (pp. 39-53). https://doi.org/10.1007/978-94-007-5902-2_3
- Tinning, R. (1980, August n.d.). *Recent innovations in Victorian primary school Physical Education*. In R. Crawford (Ed.), *Influences and innovators*, (240-260), School of Physical Education and Leisure Studies, Preston Institute of Technology.
- Tolgfors, B. (2018a). Different versions of assessment for learning in the subject of Physical Education. *Physical Education and Sport Pedagogy*, 23(3), 311-327. <https://doi.org/10.1080/17408989.2018.1429589>
- Tolgfors, B. (2018b). Transformative assessment in Physical Education. *European Physical Education Review*, 25(4), 1211-1225. <https://doi.org/10.1177/1356336x18814863>
- Tolgfors, B., & Öhman, M. (2015). The implications of assessment for learning in Physical Education and Health. *European Physical Education Review*, 22(2), 150-166. <https://doi.org/10.1177/1356336x15595006>
- Tracy, S. J. (2010). Qualitative quality: Eight "big-tent" criteria for excellent qualitative research. *Qualitative inquiry*, 16(10), 837-851. <https://doi.org/10.1177/1077800410383121>
- Tricco, A. C., Lillie, E., Zarin, W., O'Brien, K., Colquhoun, H., Kastner, M., Levac, D., Ng, C., Sharpe, J. P., Wilson, K., Kenny, M., Warren, R., Wilson, C., Stelfox, H. T., & Straus, S. E. (2016). A scoping review on the conduct and reporting of scoping reviews. *BMC Medical Research Methodology*, 16(15), 1-10. <https://doi.org/10.1186/s12874-016-0116-4>

- Tricco, A. C., Lillie, E., Zarin, W., O'Brien, K. K., Colquhoun, H., Levac, D., Moher, D., Peters, M. D. J., Horsley, T., Weeks, L., Hempel, S., Akl, E. A., Chang, C., McGowan, J., Stewart, L., Hartling, L., Aldcroft, A., Wilson, M. G., Garritty, C., . . . Straus, S. E. (2018). PRISMA extension for scoping reviews (PRISMA-ScR): Checklist and explanation. *Annals of Internal Medicine*, 169(7), 467-473. <https://doi.org/doi:10.7326/M18-0850>
- Tuffield, R. (2021). *Melbourne records 246 days in lockdown, becoming world's most locked down city*. Retrieved October 29, 2022, from <https://tinyurl.com/Tuffield2021>
- Turley, E. D., & Gallager, C. W. (2008). On the "uses" of rubrics: Reframing the great rubric debate. *English Journal*, 97(4), 87-92. <https://www.jstor.org/stable/30047253>
- Turner, A., & Martinek, T. J. (1992). A comparative analysis of two models for teaching games (technique approach and game-centered (tactical focus) approach). *International Journal of Physical Education*, 29(4), 15-31.
- Turner, A. P., & Martinek, T. J. (1999). An investigation into teaching games for understanding: Effects on skill, knowledge, and game play. *Research Quarterly for Exercise and Sport*, 70(3), 286-296. <https://doi.org/10.1080/02701367.1999.10608047>
- Vaismoradi, M., Turunen, H., & Bondas, T. (2013). Content analysis and thematic analysis: Implications for conducting a qualitative descriptive study. *Nursing and Health sciences*, 15(3), 398-405. <https://doi.org/10.1111/nhs.12048>
- van der Mars, H., McNamee, J., & Timken, G. (2018a). Physical Education meets teacher evaluation: Supporting physical educators in formal assessment of student learning outcomes. *Physical Educator*, 75(4), 582-616. <https://doi.org/10.18666/tpe-2018-v75-i4-8471>
- van der Mars, H., Timken, G., & McNamee, J. (2018b). Systematic Observation of Formal Assessment of Students by Teachers (SOFAST). *Physical Educator*, 75(3), 341-373. <https://doi.org/10.18666/tpe-2018-v75-i3-8113>
- van Maarseveen, M., Oudejans, R. R. D., & Savelsbergh, G. J. P. (2017). System for notational analysis in small-sided soccer games. *International Journal of Sports Science and Coaching*, 12(2), 194-206. <https://doi.org/10.1177/1747954117694922>
- Vaughn, M. (2020). What is student agency and why is it needed now more than ever? *Theory into Practice*, 59(2), 109-118. <https://doi.org/10.1080/00405841.2019.1702393>
- Veal, M. L. (1988). Pupil assessment issues: A teacher educator's perspective. *Quest*, 40(2), 151-161. <https://doi.org/10.1080/00336297.1988.10483896>
- Veloo, A., & Md Ali, R. (2016). Physical Education teachers challenges in implementing school based assessment. *International Review of Management and Marketing*, 6(8S), 48-53. <https://core.ac.uk/download/pdf/83553275.pdf>
- Viciano, J., Mayorga-Vega, D., Guijarro-Romero, S., & Martínez-Baena, A. (2017). Effect of two alternated teaching units of invasion team sports on the tactical learning in primary school children. *International Journal of Performance Analysis in Sport*, 17(3), 256-270. <https://doi.org/10.1080/24748668.2017.1331575>
- Victoria University. (2023). *Assessment and feedback*. Victoria University. Retrieved May 17, 2023, from <https://www.vu.edu.au/learning-teaching/teaching-for-learning-at-vu/learning-teaching-approaches/curriculum-design/assessment-feedback>
- Victorian Curriculum and Assessment Authority [VCAA]. (2019a). *Develop your own formative assessment rubrics*. VCAA. Retrieved February 18, 2022, from <https://www.vcaa.vic.edu.au/assessment/f-10assessment/formative-assessment/DevelopFormativeAssessmentRubrics/Pages/default.aspx>
- Victorian Curriculum and Assessment Authority [VCAA]. (2019b). *Using formative assessment rubrics in Health and Physical Education: Invasion, net/wall, striking and fielding games - levels 7 to 10*. <https://www.vcaa.vic.edu.au/assessment/f-10assessment/formative-assessment/formative-assessment-rubric-samples/Pages/HealthPhysicalEducationSamples.aspx>
- Victorian Curriculum and Assessment Authority [VCAA]. (n.d.-a). *Administrative information: VCAA*. Retrieved September 15, 2022, from <https://www.vcaa.vic.edu.au/administration/vce-vcal-handbook/sections/Pages/05AdminInformationVCAA.aspx>
- Victorian Curriculum and Assessment Authority [VCAA]. (n.d.-b). *Content description VCHPEM139*. VCAA. Retrieved October 22, 2022, from <https://victoriancurriculum.vcaa.vic.edu.au/Curriculum/ContentDescription/VCHPEM139>

- Victorian Curriculum and Assessment Authority [VCAA]. (n.d.-c). *DAL: Digital Assessment Library*. VCAA. Retrieved September 5, 2022, from <https://www.vcaa.vic.edu.au/assessment/f-10assessment/digital-assessment-library/Pages/Index.aspx>
- Victorian Curriculum and Assessment Authority [VCAA]. (n.d.-d). *Health and Physical Education (7-10)*. VCAA. Retrieved October 28, 2022, from <https://tinyurl.com/msa4nd5y>
- Victorian Curriculum and Assessment Authority [VCAA]. (n.d.-e). *Health and Physical Education rationale and aims*. VCAA. Retrieved October 8, 2022, from <https://victoriancurriculum.vcaa.vic.edu.au/health-and-physical-education/introduction/rationale-and-aims>
- Victorian Curriculum and Assessment Authority [VCAA]. (n.d.-f). *Health and Physical Education structure*. VCAA. Retrieved October 22, 2022, from <https://victoriancurriculum.vcaa.vic.edu.au/health-and-physical-education/introduction/structure>
- Victorian Curriculum and Assessment Authority [VCAA]. (n.d.-g). *Health and Physical Education: Put formative assessment rubrics into practice in Health and Physical Education*. VCAA. Retrieved February 21, 2022, from <https://www.vcaa.vic.edu.au/assessment/f-10assessment/formative-assessment/formative-assessment-rubric-samples/Pages/HealthPhysicalEducationSamples.aspx>
- Victorian Curriculum and Assessment Authority [VCAA]. (n.d.-h). *Movement Assessment in Practice (MAP)*. VCAA. Retrieved July 20, 2023, from <https://vcaa.vic.edu.au/curriculum/foundation-10/resources/health-and-physical-education/map/Pages/Index.aspx>
- Victorian Curriculum and Assessment Authority [VCAA]. (n.d.-i). *Navigation and terminology*. VCAA. Retrieved April 15, 2023, from <https://victoriancurriculum.vcaa.vic.edu.au/overview/navigation-and-terminology>
- Victorian Curriculum and Assessment Authority [VCAA]. (n.d.-j). *Overview*. VCAA. Retrieved October 9, 2022, from <https://victoriancurriculum.vcaa.vic.edu.au/overview/about>
- Victorian Curriculum and Assessment Authority [VCAA]. (n.d.-k). *Reporting results: Score aggregation*. VCAA. Retrieved March 21, 2023, from <https://www.vcaa.vic.edu.au/administration/vce-vcal-handbook/sections/Pages/12ReportingResultsScoreAggregation.aspx>
- Victorian Curriculum and Assessment Authority [VCAA]. (n.d.-l). *VCE curriculum*. VCAA. Retrieved August 15, 2022, from <https://www.vcaa.vic.edu.au/curriculum/vce/Pages/Index.aspx>
- Victorian Curriculum and Assessment Authority [VCAA]. (n.d.-m). *The Victorian Curriculum F–10*. VCAA. Retrieved October 8, 2022, from <https://victoriancurriculum.vcaa.vic.edu.au/>
- Victorian Department of Education [DET]. (2021a). *Assessment*. DET. Retrieved May 20, 2022, from <https://www.education.vic.gov.au/school/teachers/teachingresources/practice/Pages/assessment.aspx>
- Victorian Department of Education [DET]. (2021b). *Quality Physical Education*. DET. Retrieved March 8, 2022, from <https://www.education.vic.gov.au/school/teachers/teachingresources/discipline/physed/Pages/qualityphysed.aspx>
- Victorian Department of Education [DET]. (2023a). *High Impact Teaching Strategies (HITS)*. DET. Retrieved May 2, 2023, from <https://www.education.vic.gov.au/school/teachers/teachingresources/practice/improve/Pages/hits.aspx>
- Victorian Department of Education [DET]. (2023b). *School operations Framework for Improving Student Outcomes (FISO 2.0)*. DET. Retrieved June 9, 2023, from <https://www2.education.vic.gov.au/pal/fiso/policy>
- Victorian Department of Education [DET]. (2023c). *School operations: Reporting student achievement and progress foundation to 10*. DET. Retrieved September 10, 2023, from <https://www2.education.vic.gov.au/pal/reporting-student-achievement/policy>
- Victorian Department of Education [DET]. (2023d). *Victorian Teaching and Learning Model (VTLM)*. DET. Retrieved September 7, 2023, from <https://www.education.vic.gov.au/school/teachers/teachingresources/practice/improve/Pages/Victorianteachingandlearningmodel.aspx>

- Victorian Government. (2022). *Conduct research with the department of education and training*. State Government of Victoria. Retrieved March 25, 2022, from <https://www.vic.gov.au/conduct-research-department-education-and-training>
- Vogt, W. P., Gardner, D. C., & Haeffele, L. M. (2012). *When to use what research design* (1st ed.). Guilford Press.
- Wang, J., & Rairigh, R. M. (2006). Using instructional rubrics in Physical Education. *Teaching Elementary Physical Education*, 17(3), 37-41.
- Ward, G., & Griggs, G. (2011). Principles of play: A proposed framework towards a holistic overview of games in primary Physical Education. *Education 3-13*, 39(5), 499-516. <https://doi.org/https://doi.org/10.1080/03004279.2010.480945>
- Webb, C. (1992). The use of the first person in academic writing: Objectivity, language and gatekeeping. *Journal of Advanced Nursing*, 17(6), 747-752.
- Weeldenburg, G., Borghouts, L. B., Slingerland, M., & Vos, S. (2020). Similar but different: Profiling secondary school students based on their perceived motivational climate and psychological need-based experiences in physical education. *PLoS One*, 15(2), 1-17. <https://doi.org/10.1371/journal.pone.0228859>
- Weeldenburg, G., Borghouts, L. B., Slingerland, M., & Vos, S. (2021). Through students' eyes: Preferred instructional strategies for a motivating learning climate in secondary school Physical Education. *Curriculum Studies in Health and Physical Education*, 268-286. <https://doi.org/10.1080/25742981.2021.1889383>
- Weir, T., & Connor, S. (2009). The use of digital video in Physical Education. *Technology, Pedagogy and Education*, 18(2), 155-171. <https://doi.org/https://doi.org/10.1080/14759390902992642>
- Weller, S. C., Vickers, B., Bernard, H. R., Blackburn, A. M., Borgatti, S., Gravlee, C. C., & Johnson, J. C. (2018). Open-ended interview questions and saturation. *PLoS One*, 13(6), 1-18. <https://doi.org/https://doi.org/10.1371/journal.pone.0198606>
- Wen, D., Robertson, S., Hu, G., Song, B., & Chen, H. (2018). Measurement properties and feasibility of the Loughborough soccer passing test: A systematic review. *Journal of Sports Science*, 36(15), 1682-1694. <https://doi.org/10.1080/02640414.2017.1409611>
- Werner, P., & Almond, L. (1990). Models of games education. *Journal of Physical Education, Recreation and Dance*, 61(4), 23-30. <https://doi.org/http://dx.doi.org/10.1080/07303084.1990.10606501>
- Werner, P., Thorpe, R., & Bunker, D. (1996). Teaching games for understanding: Evolution of a model. *Journal of Physical Education, Recreation and Dance*, 67(1), 28-33. <https://doi.org/doi:10.1080/07303084.2017.1271266>
- Whittle, R., Benson, A., & Telford, A. (2017a). Enrolment, content and assessment: A review of examinable senior secondary (16–19 year olds) Physical Education courses: An international perspective. *The Curriculum Journal*, 28(4), 598-625. <https://doi.org/10.1080/09585176.2017.1318770>
- Whittle, R., Benson, A., & Telford, A. (2017b). Exploring context specific teacher efficacy in senior secondary (VCE) Physical Education teachers. *Teaching and Teacher Education*, 68, 21-29. <https://doi.org/https://doi.org/10.1016/j.tate.2017.08.006>
- Wiggins, G. (1993a). *Assessing student performance: Exploring the purpose and limits of testing*. Jossey-Bass.
- Wiggins, G. (1993b). Assessment: Authenticity, context, and validity. *Phi Delta Kappan*, 75(3), 200-213. <https://www.proquest.com/scholarly-journals/assessment-authenticity-context-validity/docview/218496931/se-2?accountid=10910>
- William, D. (2011). What is assessment for learning? *Studies in Educational Evaluation*, 37(1), 3-14. <https://doi.org/10.1016/j.stueduc.2011.03.001>
- William, D., & Leahy, S. (2015). *Embedding formative assessment: Practical techniques for K-12 classrooms*. Hawker Brownlow Education.
- Williams, J. (2016). Invented tradition and how Physical Education curricula in the Australian Capital Territory has resisted Indigenous mention. *Asia-Pacific Journal of Health, Sport and Physical Education*, 7(3), 219-234. <https://doi.org/10.1080/18377122.2016.1233803>
- Williams, J., Pill, S., & Hewitt, M. (2020). 'I think everyone is on board with changing how we do things, but we are yet to find a best fit model': A figurational study of assessing games and

- sport in Physical Education. *Sport, Education and Society*, 26(3), 1-14.
<https://doi.org/10.1080/13573322.2020.1716212>
- Williams, J. J. (2012). Operational definitions in performance analysis and the need for consensus. *International Journal of Performance Analysis in Sport*, 12(1), 52-63.
<https://doi.org/10.1080/24748668.2012.11868582>
- Williams, L., & Rink, J. (2003). Teacher competency using observational scoring rubrics. *Journal of Teaching in Physical Education*, 22(5), 552.
<https://doi.org/https://doi.org/10.1123/jtpe.22.5.552>
- Willis, J. (2012). Foundations of qualitative research: Interpretive and critical approaches. In J. Willis, M. Jost, & R. Nilakanta (Eds.), *SAGE Research Methods* (pp. 1-52). SAGE.
<https://doi.org/10.4135/9781452230108>
- Wood, L. M., Sebar, B., & Vecchio, N. (2020). Application of rigour and credibility in qualitative document analysis: Lessons learnt from a case study. *Qualitative Report*, 25(2), 456-470.
<https://www.proquest.com/scholarly-journals/application-rigour-credibility-qualitative/docview/2375475754/se-2?accountid=10910>
- Wyatt-Smith, C., Klenowski, V., & Colbert, P. (2014). *Designing assessment for quality learning* (Vol. 1). Springer Science and Business Media.
- Young, L., O'Connor, J., & Alfrey, L. (2020). Physical literacy: A concept analysis. *Sport, Education and Society*, 25(8), 946-959.
<https://doi.org/https://doi.org/10.1080/13573322.2019.1677586>
- Young, S. (2011). A survey of student assessment practice in Physical Education: Recommendations for grading. *Strategies*, 24(6), 24-26.
<https://doi.org/10.1080/08924562.2011.10590959>
- Zessoules, R., & Gardner, H. (1991). Authentic assessment: Beyond the buzzword and into the classroom. *Expanding Student Assessment*, 47-71.
- Zheng, H., Ding, L., Lu, Z., & Branch, R. M. (2019). The motivational effects of involving students in rubric development on animation instruction. *TechTrends*, 64(1), 137-149.
<https://doi.org/https://doi.org/10.1007/s11528-019-00443-w>

APPENDIX A: Ethics Approval and Modification

8434 ETHICS approval notice (6 November 2019)

This message was sent with High importance.

HE

Human Research Ethics

Wed 11/6/2019 1:18 PM

To: David Gow; Shane Pill <shane.pill@flinders.edu.au>; Kate Ridley <kate.ridley@flinders.edu.au>; Sam Elliott <sam.elliott@flinders.edu.au>

Dear David,

Your conditional approval response for project 8434 was reviewed by the Deputy Chair of the Social and Behavioural Research Ethics Committee (SBREC) and was **approved**.
notice can be found below.

APPROVAL NOTICE

Project No.:	8434		
Project Title:	Assessment practices of Victorian Secondary Physical Education (PE) teachers regarding individual student performance in invasion games and sports		
Principal Researcher:	Mr David Gow		
Email:	gow0012@flinders.edu.au		
Approval Date:	6 November 2019	Ethics Approval Expiry Date:	1 April 2021

The above proposed project has been **approved** on the basis of the information contained in the application, its attachments and the information subsequently provided.

D

donotreply@infonetica.net

To: David Gow

Cc: David Gow

Dear Mr David Gow,

We are pleased to advise that the requested modifications to the below transfer project have been approved on 11 November 2021.

Project ID: 8434

Project Title: Assessment of invasion games

Chief Investigator: Mr David Gow

Expiry Date: 09/08/2022

Application Link: <https://researchnow-ethics-forms.flinders.edu.au/Project/Index/2115>

You can access the application in the ResearchNow Ethics & Biosafety system via the Application Link above.

Please don't hesitate to contact the **Human** Ethics Executive Officer if you have any questions.

Regards,

Ms Camilla Dorian

Research Development and Support
human_researchethics@flinders.edu.au

Flinders University
Sturt Road, Bedford Park, South Australia, 5042
GPO Box 2100, Adelaide, South Australia, 5001

http://www.flinders.edu.au/research/researcher-support/ebi/human-ethics/human-ethics_home.cfm

APPENDIX B: Charting Guidelines for ScR (Chapter 4)

Data item	Definition and charting guidelines (reference numbers for studies are provided)
<i>Assessment tool characteristics</i>	
1. Tool name	The name of the tool. Abbreviations or acronyms may be used if presented in studies.
	Where no name was provided, and to avoid a large number of un-named tools, the following options were applied in order. All names were capitalised as proper nouns.
	Option 1: Where the name for the same tool was provided in another reference, that name was coded. For example, reference 26 does not identify the TSAP by name, but multiple studies do (9, 29, 30, 33, 44, 45, 52, 60, 61, 66).
	Option 2: Where another tool is acknowledged as the base of the current tool, that name will be coded. For example, reference 25 refers to the tool used in reference 4. In the case of reference 23, the observation instrument acknowledged reference 19, the observational instrument, and was coded as the same tool despite the different suffix.
	Option 3 - The name is derived from the study title. For example, the Australian Football Small-Sided Game Kicking Proficiency Assessment (5) and Polar Coordinate Analysis (32).
2. Original context	The intended context for the tool's application (community or school). With tools used in multiple studies this was derived from the earliest reference in the corpus.
	As an anomaly, reference 63 drew on a sample of players from youth clubs (the community) but classified them into school-age groups and assessed them in a school. This was coded as community.
3. Developed for use in: (IGS)	The intended GS the tool was developed for. With tools used in multiple studies this was derived from the earliest reference in the corpus.
4. Tool outcome: Rating scale	Any reference to the tool rating performance, in contrast to counting actions/behaviours. This may include the use of Likert scales or rubric formats. This does not include the myriad frequency-tools that use a binary or ternary coding system to tally player behaviours.
	Two tools applied both rating scales and frequency-count approaches (5 and 14).
5. Tool outcome: Frequency-count	Any reference to the tool counting actions/choices. This is in contrast to the act of rating performance.
6. Tool outcome: Indices	Any reference to a frequency-count tool generating a quantifiable performance index (ratio) as a summary of performance. For example, indices are used in the GPAI and TSAP.

Data item	Definition and charting guidelines (reference numbers for studies are provided)
7. Reported validity	<p>Any reference to the type(s) of validity data of the tool analysed using the current study sample.</p> <p>This includes the actual type of validity type. As exceptions, reference 63 referred to five experts developing assessment criteria using the Delphi method (content validity), reference 2 referred to six experts determining the criteria conjointly (content validity), and reference 58 referred to ecological dynamics in assessing a match (ecological validity). These 3 instances were coded as validity.</p>
8. Reported reliability	<p>Any reference to intra-rater or inter-rater reliability data of the tool analysed using the current study sample. This includes any reference to the percentage agreement but excludes any reference to the performance stability of the tool.</p> <p>Further, the use of an ICC coefficient between observers was coded as inter-rater reliability. For example, see references 24 and 37.</p>

Assessment tool criteria

All assessment criteria were reviewed to identify the following.

- | | |
|-----------------------|--|
| 9. On-the-ball skills | <p>Any reference to on-the-ball skills within the assessment criteria that include actions of receiving, dribbling, passing, scoring and defending. This criterion may be assessed independent of any other criterion, as in the GPAI (50), or within the criterion of decision-making, as in the Coding Instrument (4).</p> <p>Of note, the five named on-the-ball skills below are only coded when assessed from a skill execution perspective (that is, not if they are named within decision-making).</p> |
| 10. Receive | <p>Any reference to the action of a catch, reception, possession or ball control that comes from a team-mate that does not create a score. This includes balls received in open play where the ball was not in possession of either team (termed a 'loose ball' in reference 10).</p> <p>It excludes any act of stealing or intercepting which is coded as defend. This is best demonstrated in the TSAP that codes a received ball (from own team) and conquered ball (from opposition) as two separate criteria. As with all five named on-the-ball skills, this is not reported if assessed within decision-making.</p> |
| 11. Dribble | <p>Any reference to the action of dribbling, carrying, maintaining possession or running with the ball that does not create a score. As with all five named on-the-ball skills, this is not reported if assessed within decision-making.</p> |
| 12. Pass | <p>Any reference to the action of a pass, kick or throw that does not create a score. As with all five named on-the-ball skills, this is not reported if assessed within decision-making.</p> |
| 13. Score | <p>Any reference to the action of a score/shot. As with all five named on-the-ball skills, this is not reported if assessed within decision-making.</p> |

Data item	Definition and charting guidelines (reference numbers for studies are provided)
14. Defend	Any reference to defensive actions, for example, tackle, deflect, intercept, save goal or defensive pressure. As with all five named on-the-ball skills, this is not reported if assessed within decision-making.
15. Off-the-ball movement	Any reference to off-the-ball behaviours, movement patterns/skills, positioning, support or cover. This may be assessed independent of any other criterion, as in the GPAI (50), or within the criterion of decision-making, as in the Coding Instrument (4).
16. Decision-making	<p>Any reference to decision-making/selection/or making choices within the assessment criteria.</p> <p>As an example of not coding DM, reference 5 suggested players/coaches felt the tool assessed decision-making similar to match play (p.81), but on p.80 only kicking total and efficiency are indicated as criteria.</p> <p>Decision-making is invariably based on on-the-ball skills and/or off-the ball movement. In both cases, on-the-ball skills and off-the ball movement are also coded.</p>
17. 'Other' criteria	Any reference to criteria that cannot be classified as on-the-ball skills, off-the-ball movement or decision-making.

Assessment tool applications

18. Study aim (abbreviated)	A concise statement that captured the most relevant study aim pertaining to the assessment of game performance.
19. Sample size	<p>The total number of participants used to generate the assessment data for the GS reported. This may be different to a larger total <i>n</i> reported.</p> <p>Where an <i>n</i> was not provided separately for the test and any control group the total <i>n</i> was reported. For example, reference 11.</p>
20. Gender	Gender of participants (where indicated).
21. Age (years)	<p>The age of all participants in years between ages 5 to 19, or grade/ year/ school level. Where an age or mean age is reported, this is preferred over an age band. For example, reference 64 is reported as age 9 rather than U/10. The mean will be rounded up.</p> <p>Where more than two ages or age bands are reported, the range (top and bottom ages) will be reported, for example reference 15 provides mean ages for four bands, so the bottom and top mean age is reported, and reference 12 provides no ages and is reported as U/11 to U/20.</p> <p>Where a grade, year or school type is mentioned without any accompanying ages that will be reported. For example, reference 69 was reported as Grade 6 to 7, reference 72 was reported as High school and reference 53 was reported as Year 11.</p>
22. Applied context	The context the tool was applied in (community or school).

Data item	Definition and charting guidelines (reference numbers for studies are provided)
23. Applied GS	The GS that reports data collected by the tool. For simplicity, any subsequent 'study' using the same tool within the same reference was excluded. The two instances were reference 37 and reference 21.
24. Game format	This involves coding from two options of small-sided game (SSG) or full-sided game (FSG), based on the adult form of the game.
25. Observation period	The total time (in minutes) that a single player/team is assessed. Where time in minutes cannot be calculated, the figure will be based on the number of plays or matches. The observation period is reported for a single game, game format and reporting period, rather than aggregating any of these data points. In some cases, where the observation period cannot be established for a single player/team, the figure represents the observation period for assessment of the entire population.
26. Video capture	Any reference to the use of video capture during player assessment. This excludes references to the use of video capture outside of player assessment (for example reference 20 uses video to examine construct validity, but then assesses the game in real time, this was coded as not using video capture).
27. Peer assessor	Any reference to players completing the assessment.

Note. The abbreviated study aim was the only item that was independently charted by the researcher.

To reduce confusion in reporting data, where an article applied the same assessment tool in a second study, only the first study was charted (for example, French & Thomas, 1987; Memmert, 2010).

APPENDIX C: Bibliographic Details for Included Studies in ScR (Chapter 4)

Reference Number	Authors	Year of publication	Title of study	Tool name	Journal
1	Amatria, M., Lapresa, D., Arana, J., Anguera, M., & Garzon, B.	2016	Optimization of game formats in u-10 soccer using logistic regression analysis	Football Observation System (SOF)	Journal of Human Kinetics
2	Arias-Estero, J.	2013	Opportunities for and success in dribbling, passing, receiving, and shooting in youth basketball.	Opportunities for and Success in Dribbling, Passing, Receiving, and Shooting in Youth Basketball	International Journal of Sports Science & Coaching
3	Bennett, K., Novak, A., Pluss, M., Stevens, C., Coutts, A., & Fransen, J.	2018	The use of small-sided games to assess skill proficiency in youth soccer players: A talent identification tool	Talent Identification Tool	Science and Medicine in Football
4	Blomqvist, M., Vääntinen, T., & Luhtanen, P.	2005	Assessment of secondary school students' decision-making and game-play ability in soccer	Coding Instrument	Physical Education and Sport Pedagogy
5	Bonney, N., Berry, J., Ball, K., & Larkin, P.	2020	Validity and reliability of an Australian football small-sided game to assess kicking proficiency	Australian Football Small-Sided Game Kicking Proficiency Assessment	Journal of Sports Sciences
6	Bredt, S., Praça, G., Figueiredo, L., Paula, L., Silva, P., Andrade, A., Greco, P., Chagas, M.	2016	Reliability of physical, physiological and tactical measures in small-sided soccer games with numerical equality and numerical superiority	System of Tactical Assessment in Soccer (FUT-SAT)	Revista Brasileira de Cineantropometria e Desempenho Humano
7	Castelão, D., Garganta, J., Santos, R., & da Costa, I.	2014	Comparison of tactical behaviour and performance of youth soccer players in 3 v 3 and 5 v 5 small-sided games	System of Tactical Assessment in Soccer (FUT-SAT)	International Journal of Performance Analysis in Sport
8	Chatzopoulos, D., Drakou, A., Kotzamanidou, M., & Tsorbatzoudis, H.	2006	Girls' soccer performance and motivation: Games vs technique approach	Game Performance Assessment Instrument (GPAI)	Perceptual and Motor Skills
9	Clemente, F., Martins, F., Lourenço, M., Mendes, R.	2015	Technical accuracy it is associated with prominence levels in basketball?	Team Sport Assessment Procedure (TSAP)	Journal of Physical Education and Sport
10	Cobb, N., Unnithan, V., & McRobert, A.	2018	The validity, objectivity, and reliability of a soccer specific behaviour measurement tool	Soccer Specific Behaviour Measurement Tool (S-SBMT)	Science and Medicine in Football
11	Contreras-Jordan, O., García-López, L., & Cervelló-Gimeno, E	2005	Transfer of tactical knowledge: From invasion games to floorball.	Game Play Observational Instrument	Journal of Human Movement Studies
12	da Costa, I. T., Garganta, J., Greco, P., Mesquita, I., & Afonso, J.	2010	Assessment of tactical principles in youth soccer players of different age groups	System of Tactical Assessment in Soccer (FUT-SAT)	Revista Portuguesa de Ciências do Desporto
13	da Costa, I., Garganta, J., Greco, P., Mesquita, I., & Maia, J.	2011	System of Tactical Assessment in Soccer (FUT-SAT): Development and preliminary validation	System of Tactical Assessment in Soccer (FUT-SAT)	Motricidade
14	Darnis, F., & Lafont, L.	2015	Cooperative learning and dyadic interactions: Two modes of knowledge construction in socio-constructivist settings for team-sport teaching	Collective Game Efficacy and Individual Skill Level Tool	Physical Education and Sport Pedagogy
15	Evangelio, C., Sierra-Díaz, M., González-Víllora, S., & Clemente, F.	2019	Four goals for three players: using 3 vs. 3 small-sided games at school	Game Performance Assessment Instrument (GPAI)	Human Movement
16	Farias, C., Harvey, S., Hastie, P., & Mesquita, I.	2019a	Effects of situational constraints on students' game-play development over three consecutive sport education seasons of invasion games	Game Performance Assessment Instrument (GPAI)	Physical Education and Sport Pedagogy
17	Farias, C., Mesquita, I., & Hastie, P.	2015	Game performance and understanding within a hybrid sport education season	Game Performance Observation Instrument	Journal of Teaching in Physical Education
18	Farias, C., Mesquita, I., & Hastie, P.	2019b	Student game-play performance in invasion games following three consecutive hybrid sport education seasons.	Game Performance Assessment Instrument (GPAI)	European Physical Education Review

Reference Number	Authors	Year of publication	Title of study	Tool name	Journal
19	Farias, C., Valério, C., & Mesquita, I.	2018	Sport education as a curriculum approach to student learning of invasion games: Effects on game performance and game involvement	Game Performance Assessment Instrument (GPAl)	Journal of Sports Science and Medicine
20	Fenner, J., Iga, J., & Unnithan, V.	2016	The evaluation of small-sided games as a talent identification tool in highly trained prepubertal soccer players	Game Technical Scoring Chart (GTSC)	Journal of Sports Sciences
21	French, K., & Thomas, J.	1987	The relation of knowledge development to children's basketball performance	Observational Instrument	Journal of Sport Psychology
22	García-López, L., González-Villora, S., Gutiérrez, D., & Serra, J.	2013	Development and validation of the Game Performance Evaluation Tool (GPET) in soccer	Game Performance Evaluation Tool (GPET)	Sport TK-Revista Euroamericana de Ciencias del Deporte
23	Gil-Arias, A., Garcia-Gonzalez, L., Del Villar Alvarez, F., & Iglesias Gallego, D.	2019	Developing sport expertise in youth sport: A decision training program in basketball	Observation Instrument	The Journal of Life & Environmental Sciences (PeerJ)
24	Gouveia, É., Gouveia, B., Marques, A., Kliegel, M., Rodrigues, A., Prudente, J., Lopes, H., & Ihle, A.	2019	The effectiveness of a tactical games approach in the teaching of invasion games	Game Performance Assessment Instrument (GPAl)	Journal of Physical Education and Sport
25	Gray, S., & Sproule, J.	2011	Developing pupils' performance in team invasion games	Coding Instrument	Physical Education and Sport Pedagogy
26	Gréhaigne, JF., Godbout, P., Bouthier, D.	1997	Performance assessment in team sports	Team Sport Assessment Procedure (TSAP)	Journal of Teaching in Physical Education
27	Gutierrez, D., & García-López, L.	2012	Gender differences in game behaviour in invasion games	Game Performance Evaluation Tool (GPET)	Physical Education and Sport Pedagogy
28	Harvey, S., Cushion, C., Wegis, H., & Massa-Gonzalez, A.	2010	Teaching games for understanding in American high-school soccer: A quantitative data analysis using the game performance assessment instrument	Game Performance Assessment Instrument (GPAl)	Physical Education and Sport Pedagogy
29	Hastie, P.	1998	Skill and tactical development during a sport education season	Team Sport Assessment Procedure (TSAP)	Research Quarterly for Exercise and Sport
30	Hastie, P., Ward, J., Brock, S.	2017	Effect of graded competition on student opportunities for participation and success rates during a season of sport education	Team Sport Assessment Procedure (TSAP)	Physical Education and Sport Pedagogy
31	Ibanez, S., Martinez-Fernandez, S., Gonzalez-Espinosa, S., Garcia-Rubio, J., & Feu, S.	2019	Designing and validating a Basketball Learning and Performance Assessment Instrument (BALPAI)	Basketball Learning and Performance Assessment Instrument (BALPAI)	Frontiers in Psychology
32	Jiménez-Salas, J., Morillo-Baro, J. P., Reigal, R. E., Morales-Sánchez, V., & Hernández-Mendo, A.	2020	Polar coordinate analysis to study counterattacks in senior and Under-16 men's handball	Polar Coordinate Analysis	Cuadernos de Psicología del Deporte
33	Layne, T., Hastie, P.	2014	Development of game performance by novice learners participating in a sport education physical education unit	Team Sport Assessment Procedure (TSAP)	Journal of Physical Education and Sport
34	Lee, M., & Ward, P.	2009	Generalization of tactics in tag rugby from practice to games in middle school physical education	Supporting Movement Tool	Physical Education and Sport Pedagogy
35	Llobet-Martí, B., López-Ros, V., Barrera-Gómez, J., & Comino-Ruiz, J.	2016	Assessing novices' game performance in rugby union: The Rugby Attack Assessment Instrument (RAAI)	Rugby Attack Assessment Instrument (RAAI)	Journal of Teaching in Physical Education
36	MacPhail, A., Kirk, D., & Griffin, L.	2008	Throwing and catching as relational skills in game play: Situated learning in a modified game unit	Game Performance Assessment Instrument (GPAl)	Journal of Teaching in Physical Education

Reference Number	Authors	Year of publication	Title of study	Tool name	Journal
37	Memmert, D.	2010	Testing of tactical performance in youth elite soccer	Game Test Situation 1: Taking Advantage of Openings and Game Test Situation 2: Offering and Orienting	Journal of Sports Science and Medicine
38	Mesquita, I., Farias, C., & Hastie, P.	2012	The impact of a hybrid Sport Education–Invasion Games Competence Model soccer unit on students' decision-making, skill execution and overall game performance	Coding Instrument	European Physical Education Review
39	Miller, A., Christensen, E., Eather, N., Gray, S., Sproule, J., Keay, J., & Lubans, D.	2016	Can physical education and physical activity outcomes be developed simultaneously using a game-centered approach?	Coding Instrument	European Physical Education Review
40	Miller, A., Eather, N., Duncan, M., & Lubans, D.	2019	Associations of object control motor skill proficiency, game play competence, physical activity and cardiorespiratory fitness among primary school children	Game Performance Assessment Instrument (GPAI)	Journal of Sports Sciences
41	Miller, A., Harvey, S., Morley, D., Nemes, R., Janes, M., & Eather, N.	2016	Exposing athletes to playing form activity: Outcomes of a randomised control trial among community netball teams using a game-centred approach	Coding Instrument	Journal of Sports Sciences
42	Morales-Belando, M., Calderón, A., & Arias-Estero, J.	2018	Improvement in game performance and adherence after an aligned tgf floorball unit in physical education	Game Performance Assessment Instrument (GPAI)	Physical Education and Sport Pedagogy
43	Moreno, D., García-López, L., Del Valle Díaz, M., & Martínez, I.	2011	Spanish primary school students' knowledge of invasion games	Game Performance Evaluation Tool (GPET)	Physical Education and Sport Pedagogy
44	Nadeau, L., Godbout, P., Richard, JF.	2008a	Assessment of ice hockey performance in real-game conditions	Team Sport Assessment Procedure (TSAP)	European Journal of Sport Science
45	Nadeau, L., Richard, JF., Godbout, P.	2008b	The validity and reliability of a performance assessment procedure in ice hockey	Team Sport Assessment Procedure (TSAP)	Physical Education and Sport Pedagogy
46	Nathan, S.	2015	Coaching school hockey in Malaysia: A exploratory analysis and effect of improvised tgf pedagogical model on small sided game play	Game Play Observational Instrument	Journal of Physical Education and Sport
47	Nathan, S., & Haynes, J.	2013	A move to an innovative game teaching model: Style E Tactical (SET)	Game Play Observational Instrument	Asia-Pacific Journal of Health, Sport and Physical Education
48	Nevett, M., Rovegno, I., Babiarez, M., & McCaughtry, N.	2001	Changes in basic tactics and motor skills in an invasion-type game after a 12-lesson unit of instruction	Cutting or Off-the-ball-Actions Coding Instrument and Passing Decision Coding Instrument	Journal of Teaching in Physical Education
49	Ortega-Toro, E., Garcia-Angulo, A., Gimenez-Egido, J., Garcia-Angulo, F., & Palao, J.	2019	Design, validation, and reliability of an observation instrument for technical and tactical actions of the offense phase in soccer	Observation Instrument for Technical and Tactical Actions of the Offense Phase in Soccer	Frontiers in Psychology
50	Oslin, J., Mitchell, S., & Griffin, L.	1998	The Game Performance Assessment Instrument (GPAI): Development and preliminary validation	Game Performance Assessment Instrument (GPAI)	Journal of Teaching in Physical Education
51	Otero-Saborido, F., & González-Jurado, J.	2015	Design and validation of a tool for the formative assessment of invasion games	Tool for Assessment and Learning of an Invasion Situation (TALIS)	Journal of Physical Education and Sport
52	Otero-Saborido, F., Lluch, Á., & Gonzalez-Jurado, J.	2015	Student precision and reliability of the team sport assessment in basketball: A primary education case study.	Team Sport Assessment Procedure (TSAP)	South African Journal for Research in Sport, Physical Education and Recreation
53	Penney, D., Jones, A., Newhouse, P., & Cambell, A.	2012	Developing a digital assessment in senior secondary physical education	Standards Based Rubric	Physical Education and Sport Pedagogy
54	Pérez-Morales, J., Greco, P., Ferreira-Lopes, B., Estevão, B., & Ibañez, S	2018	Development and preliminary validation of a new procedural tactical knowledge test for basketball using 3 vs.3 situation	Procedural Tactical Knowledge Test for Basketball (PTKT: Bb)	Revista Internacional de Ciencias del Deporte

Reference Number	Authors	Year of publication	Title of study	Tool name	Journal
55	Pizarro, D., Práxedes, A., Travassos, B., del Villar, F., & Moreno, A.	2019	The effects of a nonlinear pedagogy training program in the technical-tactical behaviour of youth futsal players	Game Performance Evaluation Tool (GPET)	International Journal of Sports Science & Coaching
56	Praça, G., Morales, J., Bredt, S., Sousa, R., Andrade, A., & Greco, P.	2017	The development of tactical skills in u-14 and u-15 soccer players throughout a season: a comparative analysis	System of Tactical Assessment in Soccer (FUT-SAT)	Human Movement
57	Praxedes, A., Moreno, A., Gil-Arias, A., Claver, F., & Del Villar, F.	2018	The effect of small-sided games with different levels of opposition on the tactical behaviour of young footballers with different levels of sport expertise	Game Performance Evaluation Tool (GPET)	PLOS ONE
58	Prieto-Ayuso, A., Pastor-Vicedo, J., Gonzalez-Villora, S., & Contreras-Jordan, O.	2019	Observation criteria for physical education teachers to identify gifted children through invasion games	Game Performance Evaluation Tool (GPET)	International Journal of Environmental Research and Public Health
59	Pritchard, T., McCollum, S., Sundal, J., & Colquit, G.	2014	Effect of the Sport Education Tactical Model on coeducational and single gender game performance	Game Performance Assessment Instrument (GPAI)	The Physical Educator
60	Richard, J., Godbout, P., & Gréhaigine, J.	1998	The establishment of team-sport performance norms for grade 5 to 8 students	Team Sport Assessment Procedure (TSAP)	Avante-Ontario
61	Richard, J., Godbout, P., & Gréhaigine, J.	2000	Students' precision and interobserver reliability of performance assessment in team sports	Team Sport Assessment Procedure (TSAP)	Research Quarterly for Exercise and Sport
62	Rowat, O., Fenner, J., & Unnithan, V.	2017	Technical and physical determinants of soccer match-play performance in elite youth soccer players	Game Technical Scoring Chart (GTSC)	The Journal of Sports Medicine and Physical Fitness
63	Rubajczyk, K., & Rokita, A.	2015	Relationships between results of soccer-specific skill tests and game-related soccer skill assessment in young players aged 12 and 15 years	Observation sheet	TRENDS in Sport Sciences
64	Serra-Olivares, J., García-López, L., & Calderón, A.	2016	Game-based approaches, pedagogical principles and tactical constraints: Examining games modification	Game Performance Evaluation Tool (GPET)	Journal of Teaching in Physical Education
65	Serra-Olivares, J., Gonzalez-Villora, S., & Garcia-Lopez, L.	2015	Effects of modification of task constrains in 3-versus-3 small-sided soccer games	Game Performance Evaluation Tool (GPET)	South African Journal for Research in Sport, Physical Education and Recreation
66	Sgrò, F., Pignato, S., Lipoma, M.	2018	Assessing the impact of gender and sport practice on students' performance required in team games	Team Sport Assessment Procedure (TSAP)	Journal of Physical Education and Sport
67	Tallir, I., Lenoir, M., Valcke, M., & Musch, E.	2007	Do alternative instructional approaches result in different game performance learning outcomes? Authentic assessment in varying game conditions	Game performance Coding Instrument	International Journal of Sport Psychology
68	Tangalos, C., Robertson, S., Spittle, M., & Gustin, P.	2015	Predictors of individual player match performance in junior Australian football	Coach Rating Tool	International Journal of Sports Physiology and Performance
69	Turner, A., & Martinek, T.	1999	An Investigation into teaching games for understanding: Effects on skill, knowledge, and game play	Game Play Observational Instrument	Research Quarterly for Exercise and Sport
70	van Maarseveen, M., Oudejans, R., & Savelsbergh, G.	2017	System for notational analysis in small-sided soccer games	System for Notational Soccer Analysis	International Journal of Sports Science & Coaching
71	Viciano, J., Mayorga-Vega, D., Guijarro-Romero, S., & Martínez-Baena, A.	2017	Effect of two alternated teaching units of invasion team sports on the tactical learning in primary schoolchildren	Game Performance Assessment Instrument (GPAI)	International Journal of Performance Analysis In Sport
72	Williams, L., & Rink, J	2003	Teacher competency using observational scoring rubrics	Observational Scoring Rubric	Journal of Teaching in Physical Education

APPENDIX D: PRISMA – ScR Extension Checklist (Chapter 4)

SECTION	ITEM	PRISMA-ScR CHECKLIST ITEM	REPORTED
TITLE			
Title	1	Identify the report as a scoping review.	Chapter 4 title
ABSTRACT			
Structured summary	2	Provide a structured summary that includes (as applicable): background, objectives, eligibility criteria, sources of evidence, charting methods, results, and conclusions that relate to the review questions and objectives.	Preliminary pp. v - vi
INTRODUCTION			
Rationale	3	Describe the rationale for the review in the context of what is already known. Explain why the review questions/objectives lend themselves to a scoping review approach.	4.1
Objectives	4	Provide an explicit statement of the questions and objectives being addressed with reference to their key elements (e.g., population or participants, concepts, and context) or other relevant key elements used to conceptualize the review questions and/or objectives.	4.1.1
METHODS			
Protocol and registration	5	Indicate whether a review protocol exists; state if and where it can be accessed (e.g., a Web address); and if available, provide registration information, including the registration number.	NA
Eligibility criteria	6	Specify characteristics of the sources of evidence used as eligibility criteria (e.g., years considered, language, and publication status), and provide a rationale.	4.2.1
Information sources*	7	Describe all information sources in the search (e.g., databases with dates of coverage and contact with authors to identify additional sources), as well as the date the most recent search was executed.	4.2.2
Search	8	Present the full electronic search strategy for at least 1 database, including any limits used, such that it could be repeated.	4.2.2
Selection of sources of evidencet	9	State the process for selecting sources of evidence (i.e., screening and eligibility) included in the scoping review.	4.2.2
Data charting process‡	10	Describe the methods of charting data from the included sources of evidence (e.g., calibrated forms or forms that have been tested by the team before their use, and whether data charting was done independently or in duplicate) and any processes for obtaining and confirming data from investigators.	4.2.3
Data items	11	List and define all variables for which data were sought and any assumptions and simplifications made.	4.2.3
Critical appraisal of individual sources of evidence§	12	If done, provide a rationale for conducting a critical appraisal of included sources of evidence; describe the methods used and how this information was used in any data synthesis (if appropriate).	NA

SECTION	ITEM	PRISMA-ScR CHECKLIST ITEM	REPORTED
Synthesis of results	13	Describe the methods of handling and summarizing the data that were charted.	4.2.3
RESULTS			
Selection of sources of evidence	14	Give numbers of sources of evidence screened, assessed for eligibility, and included in the review, with reasons for exclusions at each stage, ideally using a flow diagram.	4.2.2 (including Figure 4.1)
Characteristics of sources of evidence	15	For each source of evidence, present characteristics for which data were charted and provide the citations.	4.3 (including Appendix C)
Critical appraisal within sources of evidence	16	If done, present data on critical appraisal of included sources of evidence (see item 12).	NA
Results of individual sources of evidence	17	For each included source of evidence, present the relevant data that were charted that relate to the review questions and objectives.	4.3
Synthesis of results	18	Summarize and/or present the charting results as they relate to the review questions and objectives.	4.3
DISCUSSION			
Summary of evidence	19	Summarize the main results (including an overview of concepts, themes, and types of evidence available), link to the review questions and objectives, and consider the relevance to key groups.	4.4
Limitations	20	Discuss the limitations of the scoping review process.	4.5
Conclusions	21	Provide a general interpretation of the results with respect to the review questions and objectives, as well as potential implications and/or next steps.	4.7
FUNDING			
Funding	22	Describe sources of funding for the included sources of evidence, as well as sources of funding for the scoping review. Describe the role of the funders of the scoping review.	Funding was provided by Flinders University, College of Education, Psychology & Social Work. These funds allowed for a research assistant to verify the coding of the researcher and was given the project code of 13577.

APPENDIX E: An Overview of Evidence-Based Assessment Tools Developed in a School Context (Chapter 4)

Tool No	Assessment tool name	Studies (n)	Reference number(s) in Scoping Review	Intended game(s) or sport(s)	Rating scale	Frequency count	On-the-ball skills	Off-the-ball movement	Decision-making	Other criteria
1.	Basketball Learning and Performance Assessment Instrument (BALPAI)	1	31	Basketball	-	Yes	Yes	Yes	Yes	-
2.	Coding Instrument	5	4, 25, 38, 39, 41	Soccer	-	Yes	Yes	Yes	Yes	-
3.	Collective Game Efficacy and Individual Skill Level Tool	1	14	Basketball	Yes	Yes	Yes	Yes	-	Yes
4.	Cutting or Off-the-Ball-Actions Coding Instrument	1	48	Invasion games	-	Yes	Yes	Yes	-	-
5.	Game Performance Assessment Instrument (GPAL)	13	8, 15, 16, 18, 19, 24, 28, 36, 40, 42, 50, 59, 71	All game categories	-	Yes	Yes	Yes	Yes	-
6.	Game Performance Coding Instrument	1	67	Basketball	-	Yes	Yes	Yes	Yes	-
7.	Game Performance Evaluation Tool (GPET)	8	22, 27, 43, 55, 57, 58, 64, 65	Modified invasion game	-	Yes	Yes	Yes	Yes	-
8.	Game Performance Observation Instrument	1	17	Soccer	-	Yes	Yes	Yes	Yes	-
9.	Game Play Observational Instrument	4	46, 47, 69, 11	Field hockey	Yes	Yes	Yes	-	Yes	-
10.	Observational Scoring Rubric	1	72	Movement forms	Yes	-	Yes	Yes	Yes	-
11.	Passing Decision Coding Instrument	1	48	Invasion games	-	Yes	Yes	-	Yes	-
12.	Standards Based Rubric	1	53	All game categories	Yes	-	Yes	Yes	Yes	Yes
13.	Supporting Movement Tool	1	34	Tag rugby	-	Yes	-	Yes	-	-
14.	Team Sport Assessment Procedure (TSAP)	11	9, 26, 29, 30, 33, 44, 45, 52, 60, 61, 66	Invasion games, net and wall games	-	Yes	Yes	-	Yes	-
15.	Tool for Assessment and Learning of an Invasion Situation (TALIS)	1	51	Invasion games	-	Yes	Yes	Yes	-	-
	<i>n</i> =				4	13	14	12	11	2
	Percentage				26.7	86.7	93.3	80.0	73.3	13.3

Note. The use of a dash (-) indicates the absence of the relevant item. Total *n* = 15 tools.

APPENDIX F: Questionnaire (Chapter 5)

The following Appendix was downloaded from the online platform of Qualtrics. Minor adjustments to paragraphing, page breaks and item spacing have been made to enhance readability in this document. The text in the shaded boxes indicates the use of skip logic to ensure participants met the inclusion criteria and were not asked redundant questions. For example, participants that indicated they were not familiar with the GPAI in question 14a), moved to the next level of questions relating to the TSAP beginning at 15a).

Invasion Games Assessment - 2019

Front page of questionnaire

My name is Dr Shane Pill, Associate Professor in Physical Education and Sport at Flinders University. I am supervising Mr David Gow who is completing a research higher degree (Doctor of Philosophy) at Flinders University in the College of Education, Psychology and Social Work. David is exploring Victorian Physical Education (PE) teachers' assessment practices regarding Invasion games and sports.

Title: PE teachers' assessment of student performance in invasion games

The purpose of this study is to establish the current practices of PE teachers in the field and what they find most useful when assessing invasion game performance. Responses from the voluntary survey will be anonymized and analysed using descriptive statistics (e.g. group averages and measures of variability) to report the results. As this survey is focused on PE in years / levels 7-10, if you have not taught any of these years / levels in a Victorian school in 2019, please do not proceed with the survey. The online survey will take approximately 10 minutes. Your completion of the survey will indicate your consent to participate in the study. If you have any enquiries about this project, please contact me directly by telephone on +61 8201 2277 or e-mail at shane.pill@flinders.edu.au

Please open and read the following information sheet prior to undertaking the survey, which begins at the bottom of this page, and thank you for your consideration of this invitation.

<https://docs.google.com/document/d/1i25O0e-LYsKJh3ekZmy84vR6Y-IHUIF-uroby5uJMQk/edit?usp=sharing>

yours sincerely Associate Professor Shane Pill

College of Education, Psychology and Social Work Flinders University

This thesis project has been approved by the Flinders University Social and Behavioral Research Ethics Committee (Project Number 8434). For more information regarding ethics approval of the project the Executive Officer of the Committee can be contacted by telephone on 8201 3116, by fax on 8201 2035 or by email human.researchethics@flinders.edu.au

In this questionnaire you will not be asked to identify yourself nor your school by name. You will be anonymous, and your comments will not be linked directly to you.

- I have read the above information and give consent to participate
- I do not give consent to participate

Skip To: End of Survey If In this questionnaire you will not be asked to identify yourself nor your school by name. You will ... = I do not give consent to participate

Q 1. The following survey is aimed at Physical Education (PE) teachers in Victorian schools for the school year of 2019. Please indicate your school category for this time.

- Catholic
- Government
- Independent
- I did not teach in Victoria
- Other

Skip To: End of Survey If 1. The following survey is aimed at Physical Education (PE) teachers in Victorian schools for the ... = I did not teach in Victoria

Start of Block: Title: Development of an invasion game performance assessment tool

Q 2. Do you have a teaching / education degree in PE?

- Yes
 - No
-

Q 3. What is the highest level of education you have completed?

- Bachelor's degree
 - Diploma of teaching
 - Masters
 - PhD
 - Other
-

Q 4. Indicate the number of years you have taught PE.

- Less than 5
 - Between 5 and less than 10
 - Between 10 and less than 15
 - Between 15 and less than 20
 - More than 20
-

Q 5. Please identify your gender:

- Male
 - Female
 - I do not identify with either of the above
-

Q 6. Assessment involves gathering data in order to provide feedback to students. Assessment can be provided by the teacher, a peer or the student themselves.

Select the curriculum that was most influential in your assessment of PE across years / levels 7-10 in 2019.

- Australian Curriculum
 - Victorian Curriculum
 - International Baccalaureate - Middle Years Program (MYP)
 - School - developed
 - Other
-

Q 7. Both the Australian and Victorian Curricula refer to the following Focus Areas for practical PE across years / levels 7-10. Please indicate the Focus Area that occupies most curriculum time across these years / levels.

- Challenge and Adventure Activities
 - Games and Sports
 - Lifelong Physical Activities
 - Rhythmic and Expressive Activities
 - Other curriculum area
-

Q 8. Within the Focus Area of Games and sports are four different game categories. Please indicate the game category that occupies most curriculum time across these years / levels.

- Invasion
- Net and wall
- Striking/ fielding
- Target
- Games and Sports are not taught

Skip To: End of Survey If 8. Within the Focus Area of Games and Sports are four different game categories. Please indicate ... = Games and Sports are not taught

Q 9. Please respond to the following statements with reference to Invasion Games and Sports.

A curriculum sets out what students should learn. The Victorian and Australian Curricula include level descriptors, achievement standards, content descriptors and elaborations.

	Strongly agree	Agree	Neutral	Disagree	Strongly disagree
The curriculum clearly sets out learning goals for invasion games and sports	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
The curriculum clearly outlines student progress in invasion games and sports from year to year (or level to level)	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
The curriculum directly informs my assessment practice of invasion games and sports	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>

Q 10. With reference to individual student performance in invasion games and sports indicate any use of the following assessment forms and users. You may choose multiple options.

	Teacher	Peer	Self	Not used
Checklists	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Frequency count tools (like the GPAI and TSAP)	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Journal	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Portfolio	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Rating scales	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Rubrics	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Skill tests	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Verbal responses	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Written test/ assignment	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Other	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>

Q 11. Which of the following assessment forms do you consider most useful when assessing an individual student's performance in Invasion games and sports.

- Checklists
 - Frequency count tools (like the GPAI and TSAP)
 - Journal
 - Portfolio
 - Rating scales
 - Rubrics
 - Skill tests
 - Verbal responses
 - Written test/ assignment
 - Other
-

Q 12. Please respond to the following statements.

My assessment of an individual student's performance in invasion games and sports is based on:

	Strongly agree	Agree	Neutral	Disagree	Strongly disagree
Attacking play	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Creativity	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Decision-making	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Defensive play	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Positioning 'off the ball'	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Skill execution = result of performance	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Skill execution = technique	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Team work	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>

Q 13. Please indicate which of the following game skills you assess in Invasion games and sports. You may choose multiple options.

- Block an attacking play
- Catch or trap the ball
- Clear the ball
- Dribble or carry ball
- Intercept or steal
- Pass
- Score
- Tackle
- Other
- None of the above

Q 14a. The Game Performance Assessment Instrument (GPAI) involves an observer coding individual performance behaviours like support, cover, guard, skill execution and decisions made. Please indicate any use of the GPAI (including in modified form) in regard to invasion games and sports in 2019.

- I am not familiar with the GPAI
- I am familiar with the GPAI and have not used it
- I have used the GPAI

Display This Question:

If 14a. The Game Performance Assessment Instrument (GPAI) involves an observer coding individual = I am familiar with the GPAI and have not used it

Q 14b. Please indicate which of the following reasons affected your choice not to use the GPAI. You may choose multiple options.

- Complicated
- Misses important aspects of game performance
- Not linked to the curriculum
- Peer assessment lacks accuracy
- Takes too much time
- None of the above
- Other

Display This Question:

If 14a. The Game Performance Assessment Instrument (GPAI) involves an observer coding individual = I have used the GPAI

Q 14c. Please indicate which assessment form you find most useful when assessing with the GPAI.

- Frequency counts/ coding behaviours
- Modified Rubric
- Other

Display This Question:

If 14a. The Game Performance Assessment Instrument (GPAI) involves an observer coding individual = I have used the GPAI

Q 14d. Please indicate which GPAI performance indicators you made use of in 2019. You may choose multiple options.

- Adjust
- Any 'position' based index
- Base
- Cover
- Decisions made
- Decisions made index
- Game involvement
- Game performance
- Guard/ mark
- Skill execution
- Skill execution index
- Support
- None of the above
- Other

Display This Question:

If 14a. The Game Performance Assessment Instrument (GPAI) involves an observer coding individual = I have used the GPAI

Q 14e. Please indicate which of the following skills you assess with the GPAI. You may choose multiple options.

- Block an attacking play
- Catch or trap the ball
- Clear the ball
- Dribble or carry ball
- Intercept or steal
- Pass
- Score
- Tackle
- None of the above
- Other

Q 15a. The Team Sport Assessment Procedure (TSAP) involves an observer coding individual performance behaviours, like received ball, lost ball and successful shot. Please indicate any use of the TSAP (including in a modified form) in regard to invasion games and sports in 2019.

- I am not familiar with the TSAP
- I am familiar with the TSAP and have not used it
- I have used the TSAP

Display This Question:

If 15a. The Team Sport Assessment Procedure (TSAP) involves an observer coding individual ... = I am familiar with the TSAP and have not used it

Q 15b. Please indicate which of the following reasons affected your choice not to use the TSAP. You may choose multiple options.

- Complicated
- Misses important aspects of game performance
- Not linked to the curriculum
- Peer assessment lacks accuracy
- Takes too much time
- None of the above
- Other

Display This Question:

If 15a. The Team Sport Assessment Procedure (TSAP) involves an observer coding individual ... = I have used the TSAP

Q 15c. Please indicate which assessment form you find most useful when assessing with the TSAP.

- Frequency counts/ coding behaviours with the nomogram
- Frequency counts/ coding behaviours without the nomogram
- Modified Rubric
- Other

Display This Question:

If 15a. The Team Sport Assessment Procedure (TSAP) involves an observer coding individual ... = I have used the TSAP

Q 15d. Please indicate which of the following TSAP performance indicators you made use of in 2019. You may choose multiple options.

- Attack ball
 - Conquered ball
 - Received ball
 - Lost ball
 - Neutral ball
 - Offensive ball
 - Performance score (using the nomogram)
 - Played ball
 - Successful shot
 - Efficiency index
 - Volume of play
 - None of the above
 - Other
-

Q 16. Please respond to the following statements.

When assessing an individual student's performance in invasion games and sports it is useful if the tool:

	Strongly agree	Agree	Neutral	Disagree	Strongly disagree
Is accurate in indicating the level of performance	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Is linked to the curriculum	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Is time efficient to use	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Can be completed electronically	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Suggests the next steps in learning	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Can be completed by peers	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>

Q 17. Would you be interested in participating in a follow up interview at a time and place convenient to you?

By entering your email in the space below you may be invited to participate in a follow up interview. You will be provided with further information outlining steps taken to support confidentiality and anonymity prior to accepting any invitation for interview.

APPENDIX G: Supporting Documentation (Chapter 5)

Peak Body Recruitment

Dear X (ACHPER Victoria branch))

As per previous correspondence regarding my PhD through Flinders University under the supervision of Associate Professor Shane Pill, would you be able to run the following message and links in your newsletter from September 1st until November 1st, 2019? After completion of the survey and subsequent analysis of the data I would provide you with a summary document that would interpret the results and can be published in your newsletter.

In brief, it is an invitation to Victorian PE teachers to complete an online questionnaire with the aim of exploring current assessment practices in IGS with a view to developing a new assessment tool to be used in the field. The ethics approval number is 8434 and the contact details of Assoc. Professor Shane Pill by telephone is +61 8201 2277 or e-mail shane.pill@flinders.edu.au

Regards

(signature removed under advice from the Office of Graduate Research)

David Gow,
Flinders University HDR student 2172775

Dear X (Peak Phys Ed)

As per previous correspondence regarding my PhD through Flinders University under the supervision of Associate Professor Shane Pill, would you be able to run the following message and links in your newsletter from September 1st until November 1st, 2019? After completion of the survey and subsequent analysis of the data I would provide you with a summary document that would interpret the results and can be published in your newsletter.

In brief, it is an invitation to Victorian PE teachers to complete an online questionnaire with the aim of exploring current assessment practices in IGS with a view to developing a new assessment tool to be used in the field. The ethics approval number is 8434 and the contact details of Assoc. Professor Shane Pill by telephone is +61 8201 2277 or e-mail shane.pill@flinders.edu.au

Regards

(signature removed under advice from the Office of Graduate Research)

David Gow
Flinders University HDR student 2172775

Peak Body Recruitment Continued

Dear X (ConnectedPE)

As per previous correspondence regarding my PhD through Flinders University under the supervision of Associate Professor Shane Pill, would you be able to run the following message and links in your newsletter from September 1st until November 1st, 2019? After completion of the survey and subsequent analysis of the data I would provide you with a summary document that would interpret the results and can be published in your newsletter.

In brief, it is an invitation to Victorian PE teachers to complete an online questionnaire with the aim of exploring current assessment practices in IGS with a view to developing a new assessment tool to be used in the field. The ethics approval number is 8434 and the contact details of Assoc. Professor Shane Pill by telephone is +61 8201 2277 or e-mail shane.pill@flinders.edu.au

Regards

(signature removed under advice from the Office of Graduate Research)

David Gow

Flinders University HDR student 2172775

Participant Recruitment

My name is David Gow. I am a research higher degree student at Flinders University undertaking a study to explore Victorian PE teachers' assessment practices regarding invasion games/ sports.

The following survey looks at PE across years 7-10 and will take approximately 10 min to complete. If you have not taught any of these Year Levels in the state of Victoria in 2019, please do not proceed with the survey.

Any enquiries you may have concerning this project should be directed to Associate Professor Shane Pill by telephone on +61 8201 2277 or e-mail at shane.pill@flinders.edu.au

Please read the following **letter of introduction** prior to undertaking the survey

<https://drive.google.com/open?id=1jll7-YaK90ZgeF0yfONL0rkIW4OuOQEF>

thank you for your consideration of this invite

Regards

(signature removed under advice from the Office of Graduate Research)

David Gow
Flinders University RHD student 2172775

Letter of Introduction

Title: PE teachers' assessment of student performance in invasion games

My name is David Gow. I am a research higher degree student at Flinders University undertaking a study into game assessment in secondary physical education.

Description of the study:

The literature indicates that "assessment is one of the most fraught and troublesome issues physical educators have had to deal with over the past 40 years" (Lopez-Pastor, Kirk, Lorente- Catalan, MacPhail, & Macdonald, 2013, p.57). However, little is known about the assessment practices of Australian physical education (PE) teachers.

This study aims to explore Victorian PE teachers' assessment practices regarding invasion games/ sports across Years 7-10. While there are several reliable and validated assessment tools in the literature, their use is more prevalent in elite athlete and teacher education settings. The purpose of this study is to establish the current practices of PE teachers in the field and what they find most useful when assessing invasion game performance. Responses from the survey will be anonymized and analysed using descriptive statistics (e.g. group averages and measures of variability) to report the results.

As this survey is focused on PE across Years 7-10, if you have not taught any of these levels in a Victorian school in **2019**, please do not proceed with the survey.

The survey will take approximately 10 minutes

Any enquiries you may have concerning this project should be directed to Associate Professor Shane Pill by telephone on +61 8201 2277 or e-mail at shane.pill@flinders.edu.au

Please open and read the following **information sheet** prior to undertaking the survey

<https://drive.google.com/file/d/1bN8J0N70mBnH8hlmMakunCoLxt9NlmAi/view?ths=true>

Click here to take the **survey**

https://flindersuw.au1.qualtrics.com/jfe/preview/SV_55PVUTVkJEdCU61?Q_SurveyVersionID=current&Q_CHL=preview

thank you for your consideration of this invite

Regards

(signature removed under advice from the Office of Graduate Research)

David Gow
Flinders University HDR student 2172775

Information Sheet



Mr David Gow
PhD candidate
College of Education, Psychology & Social
Work (EPSW)
Flinders University
Sturt Road
Bedford Park SA 5042
GPO Box 2100
Adelaide SA 5001
Tel: +61 03 9787 7735
gow0012@flinders.edu.au

CRIOS Provider No. 00114A

INFORMATION SHEET (for 'Questionnaire')

Title: Assessment practices of Victorian Secondary Physical Education (PE) teachers regarding individual student performance in invasion games and sports.

Researcher

Mr David Gow
College of Education, Psychology &
Social Work (EPSW)
Flinders University
Tel: +61 3 9787 7735
david.gow@flinders.edu.au

Associate Supervisor

Dr. Kate Ridley
College of Education, Psychology &
Social Work (EPSW)
Flinders University
Tel: +61 8 8201 3355
kate.ridley@flinders.edu.au

Principal Supervisor

Associate Professor Shane Pill
College of Education, Psychology &
Social Work (EPSW)
Flinders University
Tel: +61 8 8201 2277
shane.pill@flinders.edu.au

Associate Supervisor

Dr. Sam Elliott
College of Education, Psychology &
Social Work (EPSW)
Flinders University
Tel: +61 8 8201 3495
sam.elliott@flinders.edu.au

Description of the study

This study is part of the project titled, 'Development of an invasion game performance assessment tool: An exploratory sequential mixed methods approach.' This project will investigate assessment practices of Victorian Physical Education (PE) teachers across levels (years) 7 to 10 regarding invasion games. This project is supported by Flinders University, College of Education, Psychology & Social Work (EPSW).

Purpose of the study

This project aims to explore which tools, protocols and criteria Victorian PE teachers use to assess individual student performance in invasion games and sports across levels (years) 7 to 10. Furthermore, to investigate which aspects of performance in invasion games and sports Victorian PE teachers suggest as being most useful in assessment.

inspiring
achievement

What will I be asked to do?

You are invited to complete the on-line questionnaire which will ask you questions regarding your practices and views about assessment of student performance in invasion games across levels (years) 7 - 10. Participation is entirely voluntary. The questionnaire will take about 15 minutes.

What benefit will I gain from being involved in this study?

The sharing of your experiences will help the development of a new assessment tool for use with student performance in invasion games across levels (years) 7 - 10. The completion of the questionnaire may help you to reflect on and develop your assessment practices.

Will I be identifiable by being involved in this study?

In accordance with usual practice, study results (questionnaire responses) become property of the researchers and will be published later. We do not need your name, you will be anonymous, and your comments will not be linked directly to you. All information and results obtained in this study will be stored in a secure way, with access restricted to relevant researchers.

Are there any risks or discomforts if I am involved?

The researcher anticipates minimal risks or discomfort from your involvement in this study. Discomfort may occur regarding questions regarding your professional practice, at which time you may skip the specific question or exit the questionnaire. If you have any concerns regarding anticipated or actual risks or discomforts, please raise them with the researcher directly or via email on gow0012@flinders.edu.au

How do I agree to participate?

Participation is voluntary. You may answer 'no comment' or refuse to answer any questions, and you are free to withdraw from the questionnaire at any time without effect or consequences. If you are willing to participate, you can provide consent by checking the appropriate box on the first page of the questionnaire.

How will I receive feedback?

On project completion, outcomes of the project will be distributed through the ACHPER e-bulletin and Peak Phys Ed on-line newsletters.

Thank you for taking the time to read this information sheet, and we hope that you will accept our invitation to be involved.

This thesis project has been approved by the Flinders University Social and Behavioural Research Ethics Committee (8434). For more information regarding ethical approval of the project only, the Executive Officer of the Committee can be contacted by telephone on (08) 8201 3116, by fax on (08) 8201 2035, or by email to human.researchethics@flinders.edu.au

APPENDIX H: Interview Schedule (Chapter 6)

Subject: _____ Date: _____

- a) Which **Year Levels** (from 7-10) did you teach practical PE in 2019? 7, 8, 9, 10
- b) What curriculum do you follow?
- c) Do you have any leadership role within your school? HOD/ HOT&L/ HOY/ HOH/ No

The following questions relate to the assessment of individual student performance in IGS (IGS) across Year Levels 7-10 (game examples include: basketball, soccer, football, rugby, handball, hockey and lacrosse). In particular, the questions focus on what approaches you use, how you use them and if anything, else could be more useful.

HIT RECORD ON ZOOM AND IPHONE

Assessment process (including links to curriculum)

- How do you define assessment (*involves gathering data in order to provide feedback to students, teachers and parents*)
- Can you tell me how, if at all, you use technology in your assessment of IGS - useful
- How do you think your approach to assessment of IGS compares to other topics within PE?
- How do you think your approach to assessment of IGS compares across Years 7-10? Assessment in other subjects?
- Can you describe how, if at all, you use your written curriculum (e.g. VC: HPE) in assessment of IGS? – useful

Assessment strategies (including links to feedback and reporting)

- Can you tell me about any assessment forms or strategies you use in IGS? - useful
- Do students have any input into their assessment of IGS? – can you describe it
- What can you tell me about any feedback students receive from their assessment of IGS – nature of feedback - team/ individual – written/ verbal - How well they are going - Progress made - Next steps in learning
- Do you moderate assessment of IGS? How – how does this process compare to other topics/ subjects
- Can you tell me how, if at all, your assessment of IGS links to reporting? – useful
- How do you keep track of any assessment of IGS data gathered – useful – more useful
- How do you use any data gathered?

AfL or formative assessment

- What is your understanding of student awareness of *when* they are being assessed?
- What is your understanding of student awareness of *what* they are being assessed on? - how
- Do you think your students understand what *excellence* looks like? – how – more useful
- Can you describe the tasks students do while during assessment of IGS? Is there a particular teaching model you follow?
- Who does the assessment of IGS in your classes? Can you describe how, if at all, you use self/peer - useful

Rubrics and frequency-count tools

- Can you tell me about any use of rubrics (specific/ generic, number of criteria/ levels) – useful +/-
- Can you tell me about any use of frequency count tools (tactical instruments) – useful +/-
- GPAI +/- - TSAP +/- - more useful

How consistent do you think assessment practices are across your department?

The survey indicated time efficiency was important in assessment – how do you think this can be achieved?

Is there anything else that you think could be useful when assessing students in IGS?

APPENDIX I: Stimulus Material (Chapter 6)

Thank you for agreeing to take part in phase two of this study exploring assessment practices in invasion games and sports. The following material provides an outline of two assessment tools that were referred to in the questionnaire. It would be greatly appreciated if you could read through this material prior to your interview to help support your responses to a number of related questions.

Regards David Gow

THE GAME PERFORMANCE ASSESSMENT INSTRUMENT (GPAI) GPAI - CRITERIA

Users can choose from 7 'open' criteria and tally their occurrence for a single player in a game. They are called 'open' as specific definitions need to be created by the users. The three highlighted criteria have been popular in invasion games studies, with users choosing particular ball skills within decisions made and skill execution as shown in the sample table at the bottom of the page

1. Base: Appropriate return of performer to a "home" or "recovery" position between skill attempts
2. Adjust: Movement of performer, either offensively or defensively, as required by the flow of the game
3. Decisions made: Making appropriate choices about what to do with the ball (or projectile) during the game
4. Skill execution: Efficient performance of selected skills
5. Support: Off-the-ball movement to a position to receive a pass (or throw)
6. Cover: Defensive support for player making a play on-the-ball, or moving to the ball (or projectile)
7. Guard/ mark: defending an opponent who may or may not have the ball (or projectile)

GPAI – Sample definitions for soccer and coding table over page

1. Decisions made: Player chooses to pass to an open teammate
Player chooses to shoot when appropriate
2. Skill execution: Reception--Control of pass and set up of the ball
Passing- Ball reaches target
Shooting- Ball stays below head height and is on target
3. Support: The player appeared to attempt to support the ball carrier by being in/ moving to an appropriate position to receive a pass

Decisions made		Skill execution		Support	
A	IA	E	IE	A	IA

Key: A – Appropriate, IA – Inappropriate, E – Efficient, IE – Inefficient

GPAI - CALCULATIONS

After counting the frequencies of the above criteria there are some calculations that can be made to create five scores:

- a) Decisions Made Index (DMI) = number of appropriate decisions made / (number of appropriate + number of inappropriate decisions made)
- b) Skill Execution Index (SEI) = number of efficient skill executions / (number of efficient skill executions + number of inefficient skill executions)
- c) Support Index (SI) = number of appropriate supporting movements / (number of appropriate supporting movements + number of inappropriate supporting movements)
- d) Game Performance = (DMI + SEI + SI) / 3
- e) Game Involvement = total appropriate responses + number of efficient skill executions + number of inefficient skill executions + number of inappropriate decisions made

Oslin, J. L., Mitchell, S. A., & Griffin, L. L. (1998). The Game Performance Assessment Instrument (GPAI): Development and Preliminary Validation. *Journal of Teaching in Physical Education, 17*(2), 231-243.

THE TEAM SPORT ASSESSMENT PROCEDURE (TSAP)

TSAP - CRITERIA

Users tally each of the six 'closed' criteria for a single player in a game. They are called closed as specific definitions are provided. The first two variables consider how the player gains possession of the ball.

1. Conquering the ball (CB): A player is considered having conquered the ball if he or she intercepted it, stole it from an opponent, or recaptured it after an unsuccessful shot on goal or after a near-loss to the other team.
2. Receiving the ball (RB): The player receives the ball from a partner and does not immediately lose control of it.

After gaining possession of the ball, the following consider how the player disposes of the ball in one of four ways:

3. Playing a neutral ball (NB): A routine pass to a partner or any pass which does not truly put the other team in jeopardy is considered a neutral ball.
4. Losing the ball (LB): A player is considered having lost the ball when he or she loses it to the other team without having scored a goal.
5. Playing an offensive ball (OB): An offensive ball is a pass to a partner which puts pressure on the other team and, most often, leads to a shot on goal.
6. Executing a successful shot (SS): A shot is considered successful when it scores or possession of the ball is retained by one's team.

TSAP – Sample coding table

CB	RB	NB	LB	OB	SS

TSAP - CALCULATIONS

After counting the frequencies of the above there are some calculations that can be made to create three further variables:

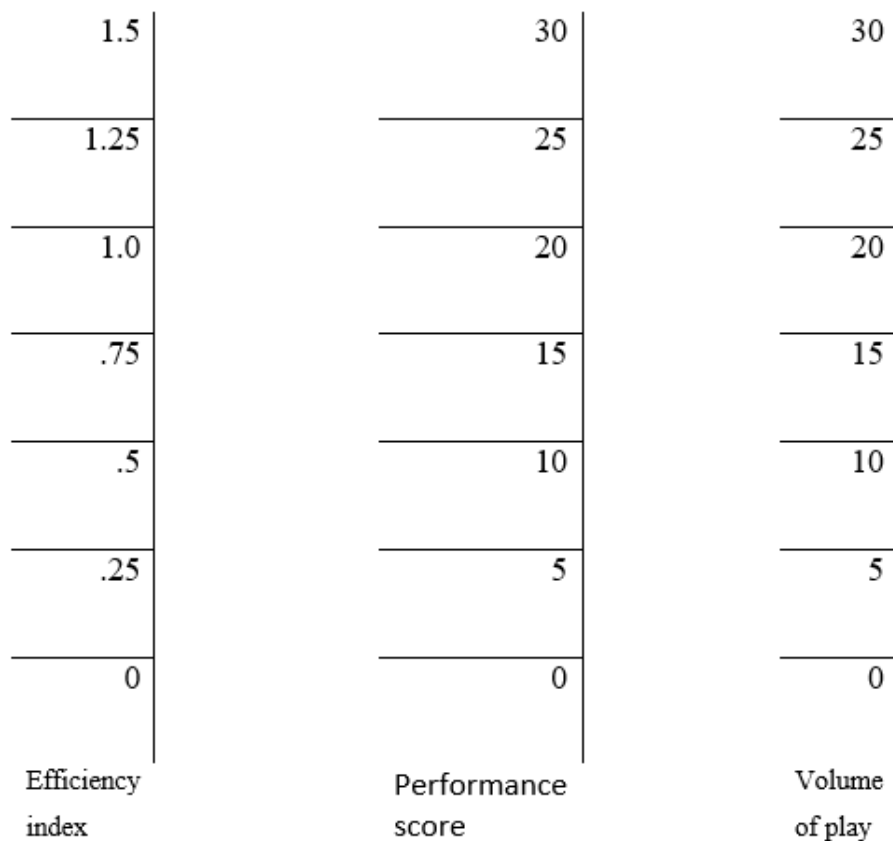
a) Attack balls (AB) = OB + SS

b) The volume of play (PB) = CB + RB.

c) The efficiency index (EI) = (CB + AB) / (10 + LB)

TSAP - PERFORMANCE SCORE/ NOMOGRAM

After calculating the efficiency index and volume of play, a line can be drawn on the nomogram to indicate a performance score.



Gréhaigne, J. F., Godbout, P., & Bouthier, D. (1997). Performance Assessment in Team Sports. *Journal of Teaching in Physical Education*, 16(4), 500-516.

APPENDIX J: Sample Memos (Chapter 6)

Text in *italics* are Amy's responses from the interview transcript, the highlighted text led to the researcher's memos that have been numbered.

They would be assessed on not every lesson, not formally every lesson. But maybe you know, informally, yes, but our summative assessment. We would do that across a number of lessons looking at different skills, yeah?

1. Is assessment still skills dominant?

The students are getting a little bit of feedback based on what we decided was going to be the learning intention that day.

2. Who is we?

Uh, we give it to them, and so we write on the white board and we do say it and we do give examples. I think I'm talking a summative rubric and so the students get that rubric

3. We - is it just this teacher or all teachers?
4. Earlier talks of formative rubric - now summative - how are they used?

Um, I think it would be helpful if we were able to measure it a lot easier.

5. It - game performance/skill?

We've tried our best to turn them into, well, you know, we've got - worked with Melbourne Uni around developmental rubrics,

6. Links to tertiary/academic support – how many schools do this?

Um, if we were going across the - so we try to have a rubric from Grade 6, to, or level 6 sorry to year 9, level 9. If we were talking a year 7 and 8 students or 10, so we could have somewhere between like you know 8 and 10 columns. I don't know if its columns or rows but across -

7. What is the optimum rubric size?

And that's not, well, that's not what our school wants either. It doesn't match up with our school, so you're trying to scale that back.

8. How much autonomy do teachers have if schools direct their assessment?

APPENDIX K: Supporting Documentation (Chapter 6)

Participant invitation

Dear X

My name is David Gow and I am a research higher degree (RHD) student at Flinders University undertaking a study to explore Victorian PE teachers' assessment practices. I would like to thank you for completing my survey that investigated invasion game assessment across years/ levels 7-10.

In that survey you indicated your willingness to participate in a one-on-one follow up interview and that is why I am contacting you now. I would like to invite you to take part in an online interview at a time that is convenient to you, using the Zoom conferencing platform. The aim of the interview is to investigate how Victorian PE teachers use assessment, and what they find most useful, when assessing invasion games and sports.

This interview will last for approximately 35 minutes. In all follow up written work any identifying information will be removed, and your comments will not be linked directly to you or your school. I have attached an information sheet and consent form to this invitation. Should you decide to participate please sign the consent form and attach it in reply to this email.

Any enquiries you may have concerning this project should be directed to Associate Professor Shane Pill by telephone on +61 8201 2277 or e-mail at shane.pill@flinders.edu.au

Thank you for your consideration of this invite.

Regards

(signature removed under advice from the Office of Graduate Research)

David Gow
Flinders University RHD student 2172775

Consent form



CONSENT FORM FOR PARTICIPATION IN RESEARCH - (Interview)

Assessment of invasion games in Physical Education

I

Being over the age of 18 years hereby consent to participate as requested in an interview for the research project with the title listed above.

1. I have read the information provided and agree to participation
2. Details of procedures and any risks have been explained to my satisfaction.
3. I agree to the recording of my information (and have the option of turning my video off)
4. I am aware that I should retain a copy of the Information Sheet and Consent Form for future reference.
5. I understand that:
 - I may not directly benefit from taking part in this research.
 - Participation is entirely voluntary, and I am free to withdraw from the project at any time; and can decline to answer particular questions.
 - The information gained in this study will be published as explained, and my participation will be confidential. On the basis that the interview will be undertaken in my place of employment, anonymity cannot be guaranteed.
 - I may ask that the recording be stopped at any time, and that I may withdraw at any time from the session or the research without disadvantage.

Participant's name.....

Participant's signature..... Date.....

I certify that I have explained the study to the volunteer and consider that she/he understands what is involved and freely consents to participation.

Researcher's name – David Gow

Date: 13/11/20

Researcher's signature

(signature removed under advice from the Office of Graduate Research)

Date: 13/11/20

This research project has been approved by the Flinders University Social and Behavioural Research Ethics Committee (Project number 8434). For more information regarding ethical approval of the project please contact the Executive Officer on (08) 8201-3110 or human_researchethics@flinders.edu.au

Information sheet



Mr David Gow
PhD candidate
College of Education, Psychology & Social
Work (EPSW)
Flinders University
Sturt Road
Bedford Park SA 5042
GPO Box 2100
Adelaide SA 5001
Tel: +61 03 9787 7735
dgo012@flinders.edu.au

CR2005 Provider No. 90114A

INFORMATION SHEET (for *Interview*)

Title: Assessment practices of Victorian Secondary Physical Education (PE) teachers regarding individual student performance in invasion games and sports.

Researcher

Mr David Gow
College of Education, Psychology &
Social Work (EPSW)
Flinders University
Tel: +61 3 9787 7735
david.gow@flinders.edu.au

Associate Supervisor

Dr. Kate Ridley
College of Education, Psychology &
Social Work (EPSW)
Flinders University
Tel: +61 8 8201 3355
kate.ridley@flinders.edu.au

Principal Supervisor

Associate Professor Shane Pill
College of Education, Psychology &
Social Work (EPSW)
Flinders University
Tel: +61 8 8201 2277
shane.pill@flinders.edu.au

Associate Supervisor

Dr. Sam Elliott
College of Education, Psychology &
Social Work (EPSW)
Flinders University
Tel: +61 8 8201 3495
sam.elliott@flinders.edu.au

Description of the study

This study is part of the project titled, 'Development of an invasion game performance assessment tool: An exploratory sequential mixed methods approach.' This project will investigate assessment practices of Victorian Physical Education (PE) teachers across levels (years) 7 to 10 regarding invasion games. This project is supported by Flinders University, College of Education, Psychology & Social Work (EPSW).

Purpose of the study

This project aims to explore how Victorian PE teachers use, and what they find most useful, when assessing individual student performance in invasion games and sports across levels (years) 7 to 10.

What will I be asked to do?

You will be invited to participate in a one-on-one interview with the chief researcher using Microsoft teams or Zoom. You will be asked mostly open-ended questions regarding your practices and views about assessment of student performance in invasion games across levels (years) 7 - 10. Participation is entirely voluntary. The interview will be conducted at a time that is convenient to you, at the researcher's cost. The interview (video and audio) will be recorded using Microsoft Teams or Zoom. You will have the option of turning off your camera should you not want your image recorded. The interview will be transcribed (typed-up) and stored, with password protection, as a computer file on the network drive of Flinders University. This transcription may be supported by the automatic transcription function on Microsoft Teams. The interview will take approximately 45 minutes.

What benefit will I gain from being involved in this study?

The sharing of your experiences will help us to articulate what teachers find useful when assessing invasion games and sports. It is possible that this will lead to the development of a new assessment tool for use with student performance in invasion games across levels (years) 7 - 10. The sharing of your experiences may help you to explore, reflect on and develop your understanding of assessment practices.

Will I be identifiable by being involved in this study?

In accordance with usual practice, study results (interview recordings and transcripts) become property of the researchers and will be published later. In all written work your name will not be recorded and any data relating to your school will be de-identified. Any identifying information will be removed, and your comments will not be linked directly to you. All information and results obtained in this study will be stored in a secure way, with access restricted to relevant researchers.

Are there any risks or discomforts if I am involved?

The researcher anticipates minimal risks or discomfort from your involvement in this study. Discomfort may occur regarding questions regarding your professional practice, or inconvenience, at which time you may skip the specific question or withdraw from the interview. If you have any concerns regarding anticipated or actual risks or discomforts, please raise them with the researcher.

How do I agree to participate?

Participation is voluntary. You may answer 'no comment' or refuse to answer any questions, and you are free to withdraw from the interview at any time without effect or consequences. A consent form accompanies this information sheet. If you are willing to participate, simply read and sign the form and send it back to the chief researcher at gow0012@flinders.edu.au

How will I receive feedback?

On project completion, outcomes of the project will be distributed through the ACHPER e-bulletin and Peak Phys Ed on-line newsletters.

Thank you for taking the time to read this information sheet, and we hope that you will accept our invitation to be involved.

This thesis project has been approved by the Flinders University Social and Behavioural Research Ethics Committee (8434). For more information regarding ethical approval of the project only, the Executive Officer of the Committee can be contacted by telephone on (08) 8201 3116, by fax on (08) 8201 2035, or by email to human.researchethics@flinders.edu.au

APPENDIX L: Sample Memos (Chapter 7)

In keeping with Dawsons framework the 27 elements are offered as possible decision points - not a checklist

Add text form (or similar)

Other than any numbers or symbols used to generate a score or indicate levels are there any non-text....

Emoji's, images

In exemplars- still images or hyperlinks to moving images

Go over VER

VIEWBRIC

FOR criteria and descriptors - is language consistent with users

Check Dawson aim - is mine same or different?

How will assessment results be moderated validity and reliability?

What is a decision matrix or decision tree?

Do The descriptor aligned to the criteria, Do, How

One set of criteria- not 2,

Refer to VC exemplars

.....

1 Clarify what a single point rubric is

2 address student involvement as a central element in discussion

3 find something about the value of transparent criteria

4 get more on student involvement in assessment design

.....

Trubric

Non-scalar? Qualitative language

Match rather than judge?

Qualitative, observable, measurable and objective description

Aim to use parallel language within each criterion

Issue for practical based rubrics is it the verbs do not discriminate between the performance - for example if you're looking at passing - passing is the verb so the context around this must change to suggest increases in sophistication or quality of performance

.....
Should Likert and rubric criteria have even or odd numbers

Some lit suggest an odd number will make the middle band a dumping ground

Degrees of deficiency = evaluative language
vs strengths based

Checklist = tick list/ circle
Just have criteria and binary choices (are these non-descriptive)
Yes/ no, present/ absent
No levels, no descriptors
What then if terms are used

Rating scale
Criteria
At least two levels indicating frequency or quality
Uses scale language (rarely to consistently/ poor to excellent) -
This language is more subjective/ abstract, but offers no qualitative descriptors

The above also fits mould of old rubric

Old rubric, The catch
Not all have independent criteria
not all have named/ numbered levels
not all have descriptors (but numbers 1-5 do indicate quality differences, but they don't describe the difference) just as the above frequency and quality scale language does not actually describe performance in concrete terms

Trubric - All of the above and exclusively qualitative descriptors, more objective/ concrete - They describe performance at different levels of accomplishment

Levels offers shorthand way to indicate the level of quality - e.g. at standard/ level

Ultimately more than two qualitative descriptors equals a Trubric
Advocate independent criteria and named levels

Excluding any evaluative, numerical or frequency based language. Two latter types are better suited to tally tools, where things are counted as rubrics are qualitative, not quantitative tools.

Count (quantity) = TT
Describe (quality) = trubric

APPENDIX M: Supporting Documentation (Chapter 7)

Returning Participant Invitation

Dear X

My name is David Gow and I am a research higher degree (RHD) student at Flinders University undertaking a study to explore Victorian PE teachers' assessment practices. I would like to thank you for completing my survey on the assessment of IGS in years seven to ten and for considering my previous invitation to participate in an interview.

I am contacting you again to invite you to participate in the final phase of this study, a document analysis of assessment related documents in the context of invasion games and sports. The aim of the final study is to critique assessment in the field. It is possible that this study will lead to the development of an assessment design framework and/ or exemplar tools for use in evaluating student performance. All submitted documents will have any identifying information removed prior to being numbered and stored on a password protected Flinders University account. Any document that cannot be de-identified will be removed from the study.

There are no minimum or maximum limits on the number of documents you submit which may include: Office documents like Word, Excel or PowerPoint and Google documents like Docs, Sheets or Slides. You may also send relevant screen shots from mobile devices or your school's Learning Management System (LMS).

I am inviting electronic submissions of documentation in the following categories:

1. Rubrics
2. Guidelines for the use of the rubrics
3. Exemplars or samples of performance (excluding any images of students)

I have attached an information sheet and consent form to this invitation. Should you choose to participate please sign the consent form and attach it in reply to this email address. From then on you are encouraged to send any of the above documents at your convenience to: david.gow@flinders.edu.au

Any enquiries you may have concerning this project should be directed to Associate Professor Shane Pill by telephone on +61 8201 2277 or e-mail at shane.pill@flinders.edu.au

Thank you for your consideration of this invite.

Regards

(signature removed under advice from the Office of Graduate Research)

David Gow, Flinders University RHD student 2172775

First Time Participant Invitation

Dear FIRST TIME PARTICIPANT

My name is David Gow and I am a research higher degree (RHD) student at Flinders University undertaking a study to explore Victorian PE teachers' assessment practices. I would like to thank you for completing my survey on the assessment of invasion games/ sports in years seven to ten and for considering my previous invitation to participate in a follow-up study.

I am contacting you to invite you to participate in the final phase of this study, a document analysis of rubrics used in the field. The survey revealed that rubrics were the most widely used tool, used by almost 90% of participants, yet only 50% said that they were the most useful. The aim of the final study is to describe these rubrics which may also lead to a series of recommendations for rubric design and/ or sample rubrics for use in the field.

All documents that are submitted will have any identifying information removed prior to being numbered and stored on a password protected Flinders University account.

There is no limit on the number or type of documents you submit. Documents may include, but are not limited to: Text files e.g. Word, Docs; Spreadsheet files e.g. Excel, Sheets; Audio files e.g. MP3; Video/ image files e.g. MP4, JPEG; Web files e.g. HTML. You may also send screen shots from mobile devices or your school's Learning Management System (LMS).

In the context of assessing student performance in year (level) seven to ten invasion games/ sports, I am inviting electronic submissions of:

1. Rubrics
2. Guidelines for the use of the rubrics
3. Exemplars or samples of performance (excluding any images of students)

I have attached an information sheet and consent form to this invitation. Should you choose to participate please sign the consent form and attach it in reply to this email. From then on you are invited to send any of the above documents to:

gow0012@flinders.edu.au

Any enquiries you may have concerning this project should be directed to Associate Professor Shane Pill by telephone on +61 8201 2277 or e-mail at shane.pill@flinders.edu.au

Thank you for your consideration of this invite.

Regards

(signature removed under advice from the Office of Graduate Research)

David Gow
Flinders University RHD student 2172775

Consent Form



CONSENT FORM FOR PARTICIPATION IN RESEARCH - (Document analysis)

Assessment of invasion games in Physical Education

I

Being over the age of 18 years hereby consent to participate as requested in submitting relevant documents for the research project with the title listed above.

1. I have read the information provided and agree to participation
2. Details of procedures and any risks have been explained to my satisfaction.
3. I am aware that I should retain a copy of the Information Sheet and Consent Form for future reference.
4. I understand that:
 - I may not directly benefit from taking part in this research.
 - Participation is entirely voluntary
 - The information gained in this study will be published in aggregate form, all documents will be de-identified and my participation will be confidential.

Participant's name.....

Participant's signature.....Date.....

I certify that I have explained the study to the volunteer and consider that she/he understands what is involved and freely consents to participation.

Researcher's name – David Gow

Date: 12/10/21

Researcher's signature

(signature removed under advice from the Office of Graduate Research)

Date: 12/10/21

This research project has been approved by the Flinders University Social and Behavioural Research Ethics Committee (Project number 8434). For more information regarding ethical approval of the project please contact the Executive Officer on (08) 8201-3116 or human.researchethics@flinders.edu.au

Information Sheet



Mr David Gow
PhD candidate
College of Education, Psychology & Social
Work (EPSW)
Flinders University
Sturt Road
Bedford Park SA 5042
GPO Box 2100
Adelaide SA 5001
Tel: +61 03 9787 7735
gow0012@flinders.edu.au

CRICOS Provider No. 00114A

INFORMATION SHEET (For Document analysis)

Title: Assessment practices of Victorian Secondary Physical Education (PE) teachers regarding individual student performance in invasion games and sports.

Researcher

Mr David Gow
College of Education, Psychology &
Social Work (EPSW)
Flinders University
Tel: 0407 324944
david.gow@flinders.edu.au

Associate Supervisor

Dr. Kate Ridley
College of Education, Psychology &
Social Work (EPSW)
Flinders University
Tel: +61 8 8201 3355
kate.ridley@flinders.edu.au

Principal Supervisor

Associate Professor Shane Pill
College of Education, Psychology &
Social Work (EPSW)
Flinders University
Tel: +61 8 8201 2277
shane.pill@flinders.edu.au

Associate Supervisor

Dr. Sam Elliott
College of Education, Psychology &
Social Work (EPSW)
Flinders University
Tel: +61 8 8201 3495
sam.elliott@flinders.edu.au

Description of the study

This study is part of the project titled, 'Exploring secondary Physical Education teachers' assessment of invasion games and sports.' This project will investigate assessment practices of Victorian Physical Education (PE) teachers across levels (years) 7 to 10 regarding invasion games. This project is supported by Flinders University, College of Education, Psychology & Social Work (EPSW).

Purpose of the study

This project aims to explore what Victorian PE teachers use, and what they find most useful, when assessing individual student performance in invasion games and sports across levels (years) 7 to 10. This includes a consideration of the types of assessment used and the nature of any criteria. In the final phase of the study this involves a description of rubrics used in the field via document analysis.

inspiring
achievement

What will I be asked to do?

You will be invited to submit relevant documents associated with your assessment of students in IGS across levels (years) 7 - 10. Documents may include, but are not limited to: Text files e.g. Word, Docs; Spreadsheet files e.g. Excel, Sheets; Audio files e.g. MP3; Video/ image files e.g. MP4, JPEG; Web files e.g. HTML. You may also send screen shots from mobile devices or your school's Learning Management System (LMS). You may make electronic submissions of the following:

1. Rubrics
2. Guidelines for the use of the rubrics
3. Exemplars or samples of performance (excluding any images of students)

Participation is entirely voluntary. You may send documents to the chief researcher's email at a time that is convenient to you. The documents will be de-identified and stored on a password protected network drive of Flinders University.

What benefit will I gain from being involved in this study?

The sharing of your documentation will help us to describe rubrics used in the field and it is possible that this may also lead to a series of recommendations for rubric design and/ or exemplar rubrics for use in assessing student performance in IGS across levels (years) 7 - 10. The sharing of your documents may help you to explore, reflect on and develop your understanding of assessment practices.

Will I be identifiable by being involved in this study?

In accordance with usual practice, study results (drawing on submitted documents) become property of the researchers and will be published later. In all written work your name will not be recorded and any data relating to your school or students will be de-identified. All information and results obtained in this study will be stored in a secure way, with access restricted to relevant researchers.

Are there any risks or discomforts if I am involved?

The researcher anticipates minimal risks or discomfort from your involvement in this study. Discomfort may occur regarding reflection on your professional practice, or inconvenience in locating and sending documents, at which time you may cease the emailing of documents. If you have any concerns regarding anticipated or actual risks or discomforts, please raise them with the researcher.

How do I agree to participate?

Participation is voluntary. A consent form accompanies this information sheet. If you are willing to participate, simply read and sign the consent form and send it back to the chief researcher at:

gow0012@flinders.edu.au

How will I receive feedback?

Upon project completion, outcomes of the project will be distributed through the ACHPER Victoria e-bulletin and Peak Phys Ed on-line newsletters.

Thank you for taking the time to read this information sheet, and we hope that you will accept our invitation to be involved.

This thesis project has been approved by the Flinders University Social and Behavioural Research Ethics Committee (8434). For more information regarding ethical approval of the project only, the Executive Officer of the Committee can be contacted by telephone on (08) 8201 3116, by fax on (08) 8201 2035, or by email to human.researchethics@flinders.edu.au

APPENDIX N: Design Elements of Rubrics (Chapter 7)

The 14 elements of rubric design (Dawson, 2017)

Elements in original order	Definition
Specificity	Rubrics can be 'generic', where they can be used across multiple tasks assessing similar skills, knowledge and understandings, or 'specific', where they have a narrower set of criteria aligned to a single task for single use. The latter format may also be termed 'task-specific'.
Secrecy	Some rubrics are not shared with students as they may 'give away' answers. Other rubrics are shared with students as part of the assessment process.
Exemplars	Exemplars of performance at a variety of levels may accompany a rubric to help users understand the definitions of quality. These exemplars can include products or performances and may also be referred to as work samples or anchors.
Scoring strategy	Rubrics can be holistic, where all assessable criteria come together to represent a single level of quality, or analytical, where separate criteria are described in varying levels of quality.
Evaluative criteria	Typically displayed in the leftmost column of a table, these criteria indicate the key qualities to be assessed.
Quality levels	Typically displayed in the top row of a table, these terms describe the different levels of quality. For example, there may be references to levels in a learning taxonomy like Bloom's revised taxonomy or the SOLO taxonomy. The levels may also be expressed through grade levels, numbers or percentages. There may also be references to levels of competency like beginner to proficient or novice to expert. There may be various combinations of the above.
Quality definitions	These definitions sit within each of the cells of a typical table. These definitions require judgement from the assessor. Typically, there are uniform definitions of attributes that vary only in the increasing levels of sophistication/ quality.
Judgement complexity	This describes the level of inference or judgement required to decide between the quality definitions.
Users and uses	Rubrics may be used by teachers or students for both self and peer assessment. Rubrics may be used formatively, to improve learning, or summatively, to indicate the level of learning.
Creators	Rubrics may be design by single teachers, teacher teams, third parties or students. This process can also be shared between teachers and students.
Quality processes	Rubrics can be validated and reliability tested. This may include consulting existing rubrics and experts along with using a range of statistical measures including inter-rater reliability.
Accompanying feedback information	Rubrics can be accompanied by feedback beyond that provided by the checking of cells.
Presentation	Rubrics typically come in the form of a grid, table or a matrix that is occupied by written text. Information can also be presented in non-text images, symbols or emojis. They may be paper based or electronic.
Explanation	Rubrics may come with instructions or protocols for use.

APPENDIX O: Coding Guidelines (Chapter 7)

Code	Coding guideline for rubrics
1. Layout	Is the layout presented in a standard table, linear text or another format?
2. Type	Is the rubric holistic (each level of performance has multiple assessment items) or analytical (criteria are assessed separately)?
3. Generic	Is the rubric generic (designed to assess a range of tasks or GS)?
4. Specific	Is the rubric specific (designed to assess a single IGS)?
5. Number of levels	How many levels of performance are there?
6. Level labels	If level labels are used, do they comprise text, images or numbers?
7. Location of level labels	If level labels are used, where are they located?
8. Level order (low to high)	How are the increasing levels of quality in performance ordered?
9. Outcome reported	Does the rubric generate a total score, percentage, grade or refer to a standard? This may include the use of a key or legend that sits outside the rubric.
10. Number of criteria	How many criteria are there? This may involve criterias being presented in more than one column or row.
11. Criteria location	Where are the criteria located?
12. Reference to the VC: HPE	Any word for word reference to the VC: HPE Achievement Standards or Content Descriptors for Year levels 7 - 8.
13. Reference to ScR criteria	Any reference to the three named key performance criteria from the ScR (on-the-ball skills, off-the-ball movement and decision-making).
14. 'Other' criteria	Any reference to criteria that could not be classified as on-the-ball skills, off-the-ball movement or decision-making.
15. On-the-ball skills	Any reference to on-the-ball skills, movement skills or gameplay. This may include references to receiving, dribbling, passing, scoring, or defending. This criterion may be assessed independently or within the criterion of decision-making.
16. Receive	Any reference to the action of a catch, trap, reception or ball control that does not create a score. It excludes any act of stealing or intercepting the ball from the opposition which is coded as defend.

Code	Coding guideline for rubrics
17. Dribble	Any reference to the action of dribbling or maintaining possession of the ball that does not create a score.
18. Pass	Any reference to the action of a pass, kick or throw that does not create a score.
19. Score	Any reference to the action of a score/shot.
20. Defend	Any reference to on-the-ball defensive actions or gameplay. This may include references to tackling, deflecting, intercepting, saving a goal or applying defensive pressure.
21. Off-the-ball movement	Any reference to off-the-ball behaviours, running, positioning, movement patterns, movement into/ leading into/ creation of/use of space. This criterion may be assessed independently or within the criterion of decision-making.
22. Decision-making	Any reference to making decisions in gameplay. This may include demonstrating/ applying/ using/ following/ implementing tactics or strategies.
23. Evaluative scale	Terms that refer to different levels of quality that provide no description of what these different levels look like. A common evaluative scale includes terms like excellent, very good, good and poor.
24. Frequency-based scale	<p>Terms that describe the frequency of a performance that are not based on numbers.</p> <p>Frequency-based descriptors located in the corpus included terms like always, consistently, inconsistently, sometimes, regularly, mostly, always, limited, usually and rarely.</p>
25. Numerical scale	Terms that describe the frequency of a performance that are based on numbers. Numerical descriptors located in the corpus included once, more than one, one to two, at least three, 70%, 90-100% and 100%.
26. Ordinal scale	A number representing a ranked order of performance. For example, from 1 (poor) to 5 (excellent).
27. Qualitative description	Terms that describe performance without any use of evaluative, frequency-based or numerical descriptors. For example, 'Student is able to...perform an interception or deflection'
28. Multiple indicators	Within analytical rubrics this refers to the assessment of more than one item in any descriptor. While bullet points or multiple sentences are more obvious for coding this, multiple assessment indicators may be found in a single bullet or sentence that assesses different actions/ behaviours.
29. Empty cells	Any empty descriptor cells in a standard table format rubric

Coding links to VC: HPE

Code	Aspect of VC	Coding guideline – reference to the excerpts from the VC: HPE
Year Levels 7 - 8		
30.	Achievement standard	Students explain personal and social skills required to establish and maintain respectful relationships and promote fair play and inclusivity
31.		Students demonstrate control and accuracy when performing specialised movement skills.
32.		They apply and refine movement concepts and strategies to suit different movement situations.
33.	Content descriptor	Use feedback to improve body control and coordination when performing specialised movement skills (VCHPEM133)
34.		Practise, apply and transfer movement concepts and strategies (VCHPEM135)
35.		Practise and apply personal and social skills when undertaking a range of roles in physical activities (VCHPEM139)
36.		Evaluate and justify reasons for decisions and choices of action when solving movement challenges (VCHPEM140)
37.		Modify rules and scoring systems to allow for fair play, safety and inclusive participation (VCHPEM141)
Year Levels 9 - 10		
38.	Achievement standard	They explain the importance of cooperation, leadership and fair play across a range of health and movement contexts.
39.		They apply and transfer movement concepts and strategies to new and challenging movement situations.
40.		They apply criteria to make judgments about and refine their own and others' specialised movement skills and movement performances.
41.	Content descriptor	Perform and refine specialised movement skills in challenging movement situations (VCHPEM152)
42.		Evaluate own and others' movement compositions, and provide and apply feedback in order to enhance performance situations (VCHPEM153)
43.		Develop, implement and evaluate movement concepts and strategies for successful outcomes (VCHPEM154)
44.		Analyse the impact of effort, space, time, objects and people when composing and performing movement sequences (VCHPEM156)
45.		Devise, implement and refine strategies demonstrating leadership and collaboration skills when working in groups or teams (VCHPEM158)
46.		Transfer understanding from previous movement experiences to create solutions to movement challenges (VCHPEM159)
47.		Reflect on how fair play and ethical behaviour can influence the outcomes of movement activities (VCHPEM160)

Additional coding notes:

1. To be consistent with the coding of the tools in the ScR, criteria and descriptors were scanned for the coding of variables 13 to 22.
2. The five named on-the-ball skills were only coded when assessed from a skill execution perspective (not if they were named within decision-making). The basis for this was to identify any rubrics that assessed these on-the-ball skills independent of decision-making (as was the case with rubric 3, 13, 17, 19 and 24).
3. The first four types of descriptor are based on rating scales as they do not describe performance, and make use of term adapted from Brookhart (2018).

Researcher's note

In supporting the second coder, the following descriptors located in the corpus were considered **evaluative**: proficient, excellent, very good, good, satisfactory, competent, beginning, strong, basic, complex, sound, competent, advanced, confidently, the best X, the correct option/ decision, appropriate, effectively, under pressure, no pressure, and depending on the flow of the game.




APPENDIX P: Translation of Findings (Chapter 8)

Table P1. Sample references to Games and Sports in Year Levels 7 - 8 VC: HPE and their alignment to the key performance criteria identified in the ScR (Chapter 4)

Movement term	Achievement Standard	Content Descriptor	Elaborations in the VC: HPE categorised within key performance criteria from the ScR (Chapter 4)		
			Off-the-ball movement	Decision-making	On-the-ball skills
Movement skills	Students demonstrate control and accuracy when performing specialised movement skills.	Use feedback to improve body control and coordination when performing specialised movement skills (VCHPEM133)			Using visual and kinaesthetic feedback when coordinating eye–hand and eye–foot movements to control different pieces of equipment (GS)
Movement sequences	They apply the elements of movement to compose and perform movement sequences.	Compose and perform movement sequences for specific purposes in a variety of contexts (VCHPEM134)	Performing movement sequences to create, use and defend space (GS)		Marking and intercepting to achieve and retain possession (GS)
Movement concepts and strategies	They apply and refine movement concepts and strategies to suit different movement situations.	Practise, apply and transfer movement concepts and strategies (VCHPEM135)		Selecting previously successful strategies and applying the most appropriate when solving new movement challenges (GS)	

Note. The above text is based on the VC: HPE Year Levels 7 – 8 as described by the VCAA (n.d.-d). The letters and numbers after each content descriptor are unique identifiers found in the VC: HPE. The abbreviation of GS within the elaborations indicates their links to the Focus Area of Games and Sports, including the category of IGS. The key performance criteria located in the ScR (Chapter 4) are framed in more ‘student-friendly’ terms as Move, Choose, and Play in the Invasion Games and Sports Assessment Framework.

Table P2. Invasion Games and Sports Assessment Framework

 MOVE		 CHOOSE		 PLAY	
Support	Player provides a passing option, blocks an opponent, or draws a defender out of play	Receive	Player chooses to receive	Receive	The player's act of controlling, catching, receiving, or intercepting the ball (not in the possession of an opponent)
Cover	Player covers an opponent or opposition scoring area	Dribble	Player chooses to dribble	Dribble	The player's act of dribbling, carrying, maintaining possession or running with the ball
		Pass	Player chooses to pass	Pass	The player's act of passing, throwing, striking or kicking the ball to a team-mate
		Score	Player chooses to score	Score	The player's act of scoring/shooting
		Defend	Player chooses to defend	Defend	The player's act of clearing the ball from their defensive area or away from an opponent, tackling, deflecting a pass/ score attempt, or stealing the ball (in the possession of an opponent)

Note. The criterion of Choose is linked exclusively to the selection of Play actions as per the GPAI (Oslin et al., 1998). Within the criterion of Play, score is recorded if an act to receive, dribble, pass or defend generates a score, thus making the five sub-criteria mutually exclusive. As with the GPAI, this framework requires users to create operational definitions for the above 12 sub-criteria.

Table P3. Rubric Design Framework

Category	Decision-point	Questions to inform rubric design
1. Getting started - Prior to design	1. Purpose(s)	What is the purpose of the rubric? For example, will the rubric determine what a student knows, understands or can do?
	2. Use(s)	How will the rubric be used? For example, will the rubric be used to improve learning, indicate the level of learning, provide certification or course progression?
	3. Audience(s)	Who is the intended audience for the rubric? For example, students, parents/guardians, teachers or other stakeholders?
	4. Designer(s)	Who will design the rubric? If students are involved, what elements will they contribute to?
	5. User(s)	Will the rubric be administered by the teacher and/ or the student? If students are involved, will the rubric be used for peer or self-assessment?
2. Format - Key structural elements	6. Layout	Is the layout presented in a table (grid/matrix), as linear/non-linear text, or in another format?
	7. Type	Is the rubric holistic (criteria are aggregated), analytical (criteria are separate), or single point (one descriptor for each criterion)? If analytical, will blank rows or columns be included for users to add criteria/ descriptors/ levels of performance?
	8. General or specific	Is the rubric general (able to suit a range of similar tasks) or specific (for a single task)?
	9. Text form(s)	What text forms will the rubric make use of? For example, written text, diagrams, still images, emoji's, hyperlinked audio or video files.
	10. Access	How will users access the rubric? For example, will it be online or in hard copy?
3. Criteria - The broad assessable component(s)	11. Criteria description	What are the criteria? Do they adequately cover the learning goal?
	12. Criteria weighting	Do the criteria need to be weighted to reflect their level of significance? If so, how is this conveyed on the rubric?
	13. Links to curriculum	Do the criteria link to the curriculum? If so, which part(s) of the curriculum do they link to? Do references make use of student-friendly language?
	14. Criteria number	How many criteria are there?

Category	Decision-point	Questions to inform rubric design
	15. Criteria location	Where are the criteria located? In the case of an analytical rubric this refers to a portrait or landscape orientation.
4. Levels - The range of performance	16. Level number	How many levels of performance are there? Will all criteria be assessed across the same number of levels?
	17. Level labels	How are the increasing levels of quality in performance labelled? For example, through the use of text, number or image?
	18. Level order	How are the increasing levels of quality in performance ordered? For example, from left to right or top to bottom.
	19. Level location	Where are any level labels located? In the case of an analytical rubric this refers to a portrait or landscape orientation.
	20. Performance outcome	Does the rubric allow the generation of an outcome like a total score, percentage, grade or reference to a standard? This may include the use of a key or legend that sits outside a typical table format.
5. Descriptors - The specific descriptions of performance quality	21. Descriptors	Is performance described using qualitative language or is rating scale language also used?
	22. Descriptor items	In analytical rubrics, are there multiple indicators of performance in the descriptors?
	23. Descriptor gaps	Are there any empty cells instead of descriptors?
6. Support - Wider considerations	24. Instructions for use	Are there any accompanying instructions for the use of the rubric? For example, any task constraints including the timing of any observation.
	25. Additional feedback	Will there be space to provide further feedback?
	26. Exemplars	Are any exemplars of performance provided to help describe different levels of quality? This may include still images, diagrams or hyperlinked audio or video files within the rubric, or ancillary documents.
	27. Quality control	Will any steps be taken to determine the validity (accuracy) and reliability (consistency) of the rubric? How will any process of moderation be managed? If students are involved in peer or self-assessment will they have any training to apply the rubric?

Note. The 27 items in the framework are considered an extension of the 14 design elements originally reported by Dawson (2017).