Bond University

DOCTORAL THESIS

Relationships between physiotherapy coursework assessments and student performance in clinical practice settings.

Terry, Rebecca

Award date:
2023

Licence:
CC BY-NC-ND

Relationships between physiotherapy coursework assessments and student performance in clinical practice settings

Rebecca Lynne Terry

Submitted in total fulfillment of the requirements of the degree of

Doctor of Philosophy

October 2022

Faculty of Health Sciences and Medicine

Professor Wayne Hing, Professor Robin Orr

and Associate Professor Nikki Milne

Abstract

Clinical placements are an essential component of the education of physiotherapy students and involve many stakeholders. As the demand for clinical placements continues to grow, it is important that students are adequately prepared to be successful in a clinical placement. If students who are less likely to succeed on a clinical placement could be identified prior to the commencement of a clinical placement, targeted guidance and support could be implemented with the aim of optimising their performance.

There is a wealth of literature based in the medical profession focused on the education and assessment of students. A smaller body of literature is focused on the physiotherapy profession. There is a tendency to translate research from medical education to physiotherapy student education. Evidence from the medical profession suggests that student performance on written examinations and Objective Structured Clinical Examinations (OSCEs) is related to their future performance in clinical practice. However, there are differences between the graduate standards expected of the two professions, and this suggests that research from medicine should be replicated in the physiotherapy profession.

This thesis investigated four cohorts of entry-level postgraduate physiotherapy students to determine the relationships between potential predictive measures and their future performance in clinical practice, as measured by the Assessment of Physiotherapy Practice (APP). The APP is a standardised instrument used by all physiotherapy education programs in Australia and New Zealand, as well as in Qatar and Iceland. Undergraduate GPA (UGPA) was identified to be the most utilised admission criterion into postgraduate physiotherapy programs in Australia. Analysis identified no significant relationship between UGPA and students' future performance in clinical practice.

Analysis of the relationships between students mean scores on summative assessments used in four pre-clinical coursework subjects identified significant relationships between future clinical performance and oral presentations, written examinations and OSCEs. Of these, the OSCE demonstrated the strongest relationship. The OSCE was then further investigated to explore differences in relationship strength between clinical performance

and individual OSCEs used in six pre-clinical coursework subjects. This highlighted that not all OSCEs demonstrate predictive validity and suggested that intrinsic properties of the OSCE influence their validity.

A content analysis of two OSCEs, one with a moderate predictive relationship with students' future clinical performance and one with no significant relationship, was conducted. This analysis identified differences between the two OSCEs and supports current recommendations for OSCE design from literature within the medical profession: mark-sheet items should be clear, unambiguous and be accompanied by an expected standard of performance; OSCE's content should focus on skills that cannot be included elsewhere in a program of assessment; and global rating scales should be considered for use with items or stations assessing holistic constructs such as communication and education.

Findings from this thesis question the use of UGPA as the sole eligibility criterion or selection measure for entry-level postgraduate physiotherapy education. This thesis provides support for the use of a range of assessment methods within physiotherapy entry-level education programs and supports the use of OSCEs to identify students at risk of poor performance in clinical practice. Practical recommendations are made to guide educators in the design of OSCEs with appropriate construct validity.

Keywords

Physiotherapy, student, clinical education, assessment, OSCE, written examination, grade-point average, predication, selection, validity.

Declaration by Author

This thesis is submitted to Bond University in fulfilment of the requirements of the degree of Doctor of Philosophy.

This thesis represents my own original work towards this research degree and contains no material that has previously been submitted for a degree or diploma at this University or any other institution, except where due acknowledgment is made.

Rebecca Terry

Date: 8th October 2022

Declaration of Author Contributions

Publication co-authored	Statement of contributions
Terry, R., Hing, W., Orr, R., Milne, N. (2017). Do coursework summative assessments predict clinical performance? A systematic review. BMC Medical Education, 17(1),40	RT 85%, WH 5%, RO 5%, NM 5%
Terry, R., Hing, W., Orr, R. M., & Milne, N. (2018). Undergraduate grade-point average as a selection criterion for a postgraduate entry-level physiotherapy program. *Australian Journal of Clinical Education*, *3*(1), 1 - 17.	RT 85%, WH 5%, RO 5%, NM 5%
Terry, R., Hing, W., Orr, R. M., & Milne, N. (2020). Relationships between pre-clinical summative assessment scores and the clinical performance of physiotherapy students. *Journal of Allied Health*, 49(1), E13 – E19.	RT 85%, WH 5%, RO 5%, NM 5%

Research Outputs

Peer-reviewed Publications

Terry. R., Hing, W., Orr, R., Milne, N. (2017) Do coursework summative assessments predict clinical performance? A systematic review. *BMC Medical Education*, 17(1),40

Terry, R., Hing, W., Orr, R. M., & Milne, N. (2018). Undergraduate grade-point average as a selection criterion for a postgraduate entry-level physiotherapy program. *Australian Journal of Clinical Education*, *3*(1), 1-17.

Terry, R., Hing, W., Orr, R. M., & Milne, N. (2020). Relationships between pre-clinical summative assessment scores and the clinical performance of physiotherapy students. *Journal of Allied Health*, 49(1), E13–E19.

Published and Presented Conference Abstracts

Terry. R., Hing, W., Orr, R., Milne, N. (2017) The relationship between pre-clinical summative assessment scores and student performance in clinical practice. Australian Physiotherapy Association Conference, Sydney, Australia.

Terry. R., Hing, W., Orr, R., Milne, N. (2019) Objective Structured Clinical Examination scores are strongly related to students' future performance in clinical practice. World Congress on Physical Therapy, Geneva, Sweden.

Terry. R., Hing, W., Orr, R., Milne, N. (2019) Undergraduate grade-point average is not a determinant of student's future performance in a postgraduate entry-level physiotherapy program. Australian Physiotherapy Association Conference, Adelaide, Australia.

Ethics Declaration

The research associated with this thesis received ethics approval from the Bond University Human Research Ethics Committee. Ethics application number RO1733.

Copyright Declaration

Copyright permission has been obtained from the publisher for the three manuscripts accepted for publication by the publishers. Copies of the permissions are referenced in the preface of the relevant chapters within the thesis.

Acknowledgements

Thank you to my supervisors Wayne, Rob, and Nikki. I am so grateful for all your guidance and support. You have been very patient with me, but we finally got here! Thank you for trusting me and giving me the freedom to work at a pace that helped keep my sanity intact. I am honoured to have had the chance to share in your ideas and learn from your experience.

To the Rando-Orrs and the Goodwins, thank you for being the best friends I could ask for. Your pep-talks always make me sound cleverer than I really am, but you do wonders for my confidence. I am so lucky to have friends who are so generous with their love, their time, and who build me up at every opportunity. Thank you for the thousand and one things you have done to help me along the way. You have been there whenever I needed you, and I can honestly say I would not have finished this without you.

Lisa, you are my research role model. You make everything look so easy; I aspire to be like you. Thank you for everything you have done over the years to help me, it has made such a difference. It is always a good day when we are working together!

To all my family, thank you for loving me, encouraging me, and helping me carve out time to work on my research. I am so lucky to have you.

To my beautiful and precious children Zac and Mini, I have learned the most from you. You have taught me to be brave, to try new things, and to never give up. You are always my inspiration and I hope I have made you proud. I love you both so, so much.

Finally, to my wonderful husband Mic. You are the best thing that has ever happened to me. There is no part of our life together that I am not grateful for. Everything I have managed to achieve is a credit to us. You have helped me become the best version of myself and I wouldn't have any of this without you. I love you with all my heart.

Table of Contents

List of Tables

List of Figures

Abbreviations

Advanced Pharmacy Practice Experiences	APPE
Association of Medical Education in Europe	AMEE
Assessment of Physiotherapy Practice	APP
Critical appraisal score	CAS
Cardiorespiratory	CR
Central Regional Dental Testing Service	CRDTS
Clinical performance examination	CPX
Competency Based Assessment in Speech Pathology	COMPASS®
Comprehensive Osteopathic Medical Licensing Examination of the United States Level 2-Performance Evaluation	COMLEX-USA Level 2-PE
Clinical performance examination	CPX
Extended matching questions	EMQ
Grade point average	GPA
Hong Kong University	HKU
Interdisciplinary Capability Assessment Tool	ICAT
Junior Doctor Assessment Tool	JDAT
Multiple choice question	MCQ
Modified essay questions	MEQ
Multiple Mini Interviews	MMI
Masters of Physiotherapy	MPT

Musculoskeletal	MSK
National Dental Hygiene Examination	NDHE
Neurological	Neuro
Objective Structured Clinical Examination	OSCE
Orthopaedic	Ortho
Problem-based learning	PBL
Physical Therapist Clinical Performance Instrument	PT CPI
Short answer questions	SAQ
Student Practice Evaluation Form	SPEF
Student Practice Evaluation Form – Revised Edition	SPEF-R
Standardised patient	SP
Undergraduate GPA	UGPA
United States of America	USA
United States Medical Licensing Exam Step 2 Clinical Skills	USMLE Step 2 CS

Chapter 1. Assessment in health profession education

1.1 Prelude

Clinical education is a necessary component of health professional education (Downing & Yudkowsky, 2009). It is important that students are prepared to maximise the learning opportunities available to them on a clinical placement, and that they are adequately equipped to be able to demonstrate the minimum standard required by the profession at the end of the experience. In the physiotherapy profession there is a global trend towards postgraduate entry-level education (Commission on Accreditation In Physical Therapy Education, 2018; Council of Canadian Physiotherapy University Programs, 2009). This requires students to have made the considerable financial and time commitment to complete pre-requisite undergraduate studies before entering the postgraduate physiotherapy program. Success on clinical placement is high stakes for students, given the time and financial investment made towards their chosen pathway of study.

Clinical education teams within physiotherapy programs play an important role in the support of both students and educators during clinical placement, with an aim to maximise students' success and foster a positive experience for all stakeholders. The structure of clinical education teams vary from institution to institution. Clinical education coordinators may be employed in either a professional or academic capacity with a focus on the organisation, administration, and quality control of the clinical education component of a program. Others, such as this author, can have dual roles that include both academic teaching in pre-clinical coursework as well as clinical education coordination responsibilities. Those who have academic interactions with students may form personal opinions on the students that they predict will encounter difficulty on clinical education, and can then act accordingly with pre-emptive support and closer monitoring during clinical placements. Those who do not are reliant on that information being passed on from other academics. Without some prior knowledge of students likely to encounter difficulty, the support role of the clinical education teams must be largely reactive, rather than proactive. This program of research was inspired by the question of how clinical education teams could more accurately predict students likely to encounter difficulty, so that timely, pre-emptive support for both students and educators could be implemented.

Students who do not demonstrate an adequate standard of competence on a clinical placement impact all stakeholders. These students tend to require a significant investment of time and energy by both university staff and the clinical educator during the unsuccessful placement (Davenport, Hewat, Ferguson, McAllister, & Lincoln, 2018). To maximise the likelihood that students are successful on clinical placement, education providers need to first ensure that students admitted into the program have the personal and academic attributes to succeed. They must also ensure that students are adequately prepared for the clinical practice environment before clinical placement commences, not only to maximise students' success, but to ensure public safety (Australian Physiotherapy Council, 2020).

The medical profession has been prolific in publishing literature focused on the education of medical students, however there is a much smaller body of research published with an allied health focus. Due to the paucity of research in this field, there is a tendency to translate findings from a medical student population and apply them to allied health education programs. However, medical graduates have an established framework of ongoing supervision from senior clinicians, whereas other graduates such as physiotherapists can immediately practice as autonomous, independent practitioners after registration. It is important to attempt to replicate the medical education findings and confirm or refute their applicability to an individual allied health profession. Measures taken early within students' education program that could possibly predict their future clinical performance give faculty the opportunity to work with those students and provide targeted enhancement strategies. This doctoral program of research has been undertaken to fill this need within the context of Australian postgraduate entry-level physiotherapy education.

This chapter aims to provide a review of key concepts relevant to the clinical education of health professional students. A broad exploration of assessment processes, with an emphasis on summative assessment, was conducted to anchor this thesis in sound pedagogy. Specific assessment tools used to measure clinical performance across a variety of health professions were reviewed, with an emphasis on physiotherapy education. This chapter aims to identify gaps within the literature, which will inform the program of doctoral research described in this thesis.

1.2 Introduction

Health profession education programs aim to produce graduates who are competent and ready to work as independent health professionals. Students must demonstrate they have a thorough knowledge base, are able to clinically reason, and have mastery of the practical skills required of a new graduate clinician. To be deemed competent they must be able to apply these skills in a real work environment, functioning alongside other professionals to assess and treat authentic clients.

In health profession education programs such as physiotherapy, students complete pre-clinical coursework that enables them to learn the knowledge, practical techniques, clinical reasoning skills and appropriate attitudes and behaviours that are required to be competent in clinical practice (Australian Physiotherapy Council, 2017; Physiotherapy Board of Australia & Physiotherapy Board of New Zealand, 2015). During this process of learning, multiple assessment items are used to assess students to ensure that the necessary learning has indeed taken place. Students must demonstrate that they meet minimum assessment standards before entering supervised clinical practice (Australian Physiotherapy Council, 2020). Minimum pre-clinical assessment standards aim to ensure that students possess the skills to be safe and effective learner practitioners in a clinical practice setting. This is essential to ensure public safety when students are enabled to implement their learning on authentic clients.

Physiotherapy is a registered health profession whose clinicians are first-contact practitioners (Australian Physiotherapy Association, 2022). Physiotherapists are experts in human movement and are uniquely qualified to support the development, restoration, and maintenance of a person's function (World Confederation for Physical Therapy, 2019). As an evidenced-based profession, physiotherapists are generally held in high regard by the community (Physiotherapy Board of Australia & Physiotherapy Board of New Zealand, 2015). Graduates work across all sectors of healthcare as independent autonomous practitioners, and as integral collaborators within multi-disciplinary healthcare teams (World Confederation for Physical Therapy, 2019). Physiotherapy practice is not only limited to the provision of clinical services, physiotherapists can work in a range of roles including "management, administration, education, research, advisory, regulatory or policy

development roles and any other roles that have an impact on safe, effective delivery of health services in physiotherapy" (Physiotherapy Board of Australia & Physiotherapy Board of New Zealand, 2015, pg. 9). Appropriate assessment of physiotherapy students is vital, given that graduates do not require ongoing supervision from a senior clinician, but can immediately practice autonomously as first-contact diagnosticians, case-managers, and advisors.

This chapter will focus on assessments as used in clinical education programs and discuss specific assessments used to measure academic and clinical performance. The current body of literature exploring the predictors of clinical performance will be reviewed and gaps in the literature and emerging trends identified for use as the basis for planning further research in this field. This chapter will encompass a range of health professions, before the thesis narrows to focus on the physiotherapy profession.

1.3 Clinical Education

Health profession education programs are specialised professional degree programs in which a defined component of the course usually includes practical training in a clinical workplace setting. Graduates of health profession education programs must be competent to function as health professionals across a wide range of settings. Professions that require a specific clinical health education program to qualify for practice include, but are not limited to, medicine, nursing, dentistry, psychology and allied health (Downing & Yudkowsky, 2009). These programs are delivered at tertiary institutions as undergraduate or postgraduate courses and produce entry-level health practitioners. Entry-level practitioners may be qualified to work independently and autonomously (Dental Board of Australia, 2020; Occupational Therapy Board of Australia, 2018; Physiotherapy Board of Australia & Physiotherapy Board of New Zealand, 2015), or under supervision from senior clinicians as is the case for junior doctors (Australian Medical Council Limited, 2012). Where it is relevant, Australian-trained graduates are eligible for registration with the Australian Health Practitioner Regulation Agency. Minimum standards of competence for new graduate practitioners are set by the professions themselves (Australian Medical Council Limited, 2012; Occupational Therapy Board of Australia, 2018; Physiotherapy

Board of Australia & Physiotherapy Board of New Zealand, 2015). These standards are required to control the quality of graduates and to ensure public safety.

Clinical placements are the structured learning components of a program that occur in an authentic workplace environment. Health Workforce Australia (2011) defines a clinical placement as occurring in a professional setting and providing opportunities for students to integrate theory into practice, familiarise themselves with workplace environments, and to build the knowledge, skills, and attributes essential for professional practice. Clinical placements are variously referred to as clinical practicums (Ho, Whitehill, & Ciocca, 2014), clerkships (Berg et al., 2008), work-integrated learning experiences (Smedts, Campbell, & Sweet, 2013) and clinical attachments (Carr, Celenza, Puddey, & Lake, 2014). Clinical placements may be embedded into the course structure or occur as a block at the end of the program. Programs may choose to include observational placements to facilitate students' learning and development. All students however, regardless of profession or program structure, must demonstrate they have learned and can apply the requisite theoretical knowledge, practical skills, clinical reasoning, and professional skills in coursework subjects developed and delivered by the academic institution.

Once in the professional workplace setting, students have the opportunity to apply their learning under the supervision and guidance of a qualified health practitioner who acts as the students' supervising clinician. The supervising clinician is responsible for ensuring safe and quality client care, while providing education, support and management of the student (Health Workforce Australia, 2011). They are also responsible for the assessment of student performance. The assessment result is provided to the academic institution, which is responsible for incorporating this result into the appropriate assessment framework in order to award the appropriate grade for the unit of study. The nomenclature associated with the role can vary between professions and may include terms such as clinical educator, clinical facilitator, clinical instructor, practice educator or program director. For this thesis, the person in the role of supervising clinician from this point forward will be referred to as clinical educator.

Throughout the clinical placement the clinical educator must evaluate the student's performance and rate this against a pre-determined set of competencies. Students are required to demonstrate to their clinical educator that they have reached the minimum standards of competency as dictated by their profession. The competencies required vary between professions. Individual professions have their own core set of professional competencies, but there are also generic competencies that are shared by all professions. These generic competencies include communication, professional behaviour, collaborative practice and client-centred care. Generic competencies can be assessed using not only profession-specific measures of clinical performance, but also multi-disciplinary assessment tools such as the Interprofessional Capability Assessment Tool (ICAT) (Brewer, Gribble, Robinson, Lloyd, & White, 2009).

1.4 Assessment

Assessment plays an integral role in all stages of education and serves many purposes. The key purposes of assessment have been identified by van der Vleuten et al. (2012) as:

- the promotion of learning
- the provision of robustness for high-stakes decisions such as the promotion of learners
- the provision of accountability for the quality of graduates
- the provision of feedback on the curriculum.

The focus of assessment in health education programs has progressed from the testing of pure factual content and technical skill (Pangaro & ten Cate, 2013). In recognition of the fact that knowledge alone does not make a successful clinician, modern assessment frameworks aim to require students to demonstrate performance at higher levels of function. In 1990, George Miller described a framework for the assessment of clinical competence which is now commonly known as Miller's pyramid (Figure 1).
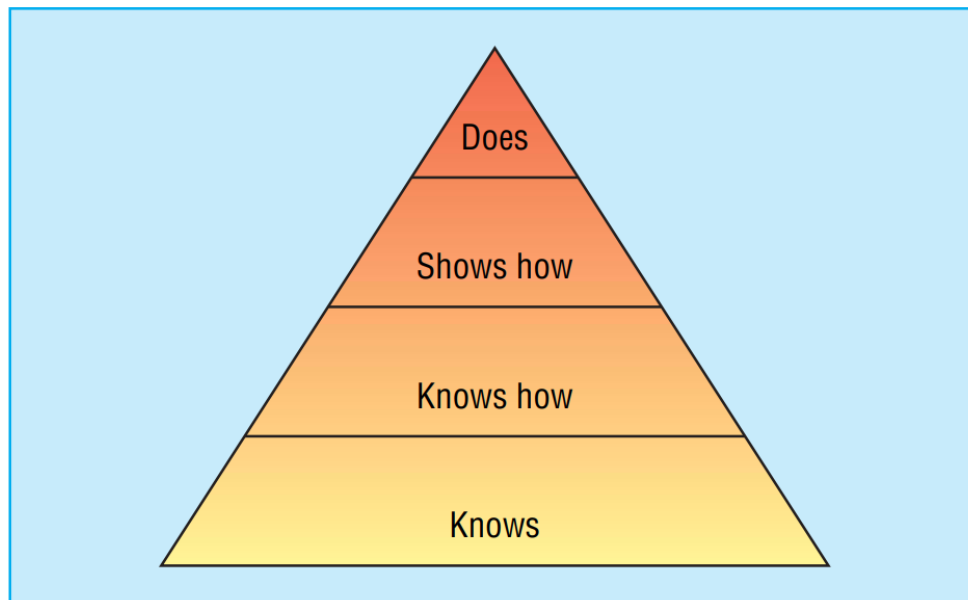
Figure 1. Miller's pyramid for assessing clinical competence

Reproduced from Work based assessment, Norcini, J., BMJ, 326(7392), 753-755, 2003. with permission from BMJ Publishing Group Ltd.

In Miller's framework, the assessment of more advanced levels of function sits closer to the apex of a pyramid. The assessment of knowledge forms the base of the pyramid and is labelled 'knows'. Above this lies the assessment of competence, i.e., 'knows how'. To demonstrate competence students must be able to acquire, analyse, and interpret information to create a rational diagnosis and/or management plan. Miller (1990) recognised that just because a student 'knows how', does not mean this will automatically translate into performance. Therefore, the next order of assessment is the evaluation of performance, i.e., 'shows'. At the peak of the pyramid is the assessment of what the student 'does' as an independent practitioner. Assessment frameworks should include items that evaluate a range of constructs at all levels of the pyramid, but particularly the upper levels.

Assessment can be formative, summative, or both. Assessment is an evaluation of student performance. Formative and summative assessments use essentially the same methodology. Scriven stated that "evaluation is itself a logical methodological activity which is essentially similar [irrespective of what is being evaluated]. The activity consists simply in the gathering and combining of performance data with a weighted set of goal scales" (Scriven, 1967, p. 2). The assessment process is therefore the same regardless of

whether the assessment is for formative or summative purposes (Taras, 2005), as the evaluation itself allows the assessor to make a judgment on the student's performance and compare it against a set standard. Formative assessment describes assessment that has been designed to provide useful feedback to the student (Wood, 2013). Formative feedback serves many functions and may be used to motivate students to learn, focus and direct student learning, and inform the teaching staff (Wood, 2013). Both student and teacher have opportunity to use the formative assessment feedback to modify, adapt or enhance student learning.

Summative assessment is a final measurement of student performance (Downing & Yudkowsky, 2009). In tertiary programs, summative assessments usually contribute heavily to students' formal academic grade (Downing & Yudkowsky, 2009). They provide information to facilitate high-stakes decisions regarding progression and graduation (van der Vleuten et al., 2012). In health education programs, coursework summative assessment has a societal responsibility to ensure only students demonstrating adequate competency progress into the clinical environment where they get to apply their learning in real-life situations (van der Vleuten et al., 2012). Summative coursework assessments must therefore measure the range of constructs required to prepare students to be successful in clinical practice. Assessment that is administered for summative purposes may also provide feedback to both student and teacher about student performance. In this instance, assessment is both formative and summative.

Considerable importance is placed on the reliability and validity of assessments. Examining bodies are required to collect evidence to support the hypothesis that assessment instruments achieve construct validity. The *Standards of Educational and Psychological Testing* discuss five sources of construct validity evidence (American Psychological Association, American Educational Research Association, & National Council on Measurement in Education, 1999). These include content validity, psychometric validity, predictive validity, analysis of response process, and the consequences of test use. There is a wealth of research commenting on the reliability and validity of assessment items such as the Objective Structured Clinical Examination (OSCE) (R. Cohen, Reznick, Taylor, Provan, & Rothman, 1990; Hodges, Regehr, Hanson, & McNaughton, 1998; Sloan, Donnelly, Schwartz, & Strodel, 1995). Most of this research pertains to the psychometric properties of the assessment. However, Hodges (1988) pointed out that validity should not

be limited to the psychometric properties of the assessment, but should also reflect contextual validity of the competencies it is attempting to assess.

Assessment frameworks used by health education programs share commonalities (Pangaro & ten Cate, 2013). Recently there has been a movement towards considering assessment as a program, rather than as a collection of individual items (Dijkstra et al., 2012; van der Vleuten & Schuwirth, 2005). The rationale for this approach is that no single assessment item can meet all the needs and purposes of assessment, but that a carefully considered program of assessment may (van der Vleuten et al., 2012). Health profession education programs routinely assess academic subjects using a variety of assessment items including written theory examinations, practical or clinical examinations including OSCE's and viva voces, written assignments including case studies, portfolios and reflective tasks, and oral presentations. This assessment framework targets not only different constructs of learning but aims to assess these constructs across the depth of Miller's pyramid. Figure 2 illustrates the level of Miller's pyramid targeted by individual assessment items.
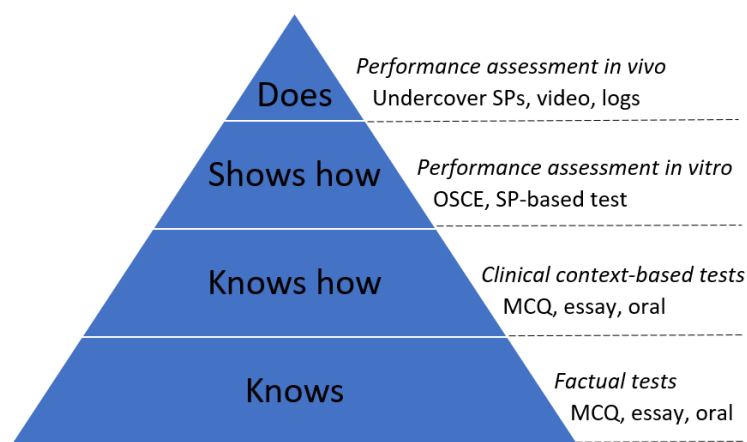
SP – Standardised patient; OSCE – Objective Structured Clinical Examination; MCQ – multiple choice question

Figure 2. The alignment of assessment items to Miller's pyramid of competence

Reprinted from The Lancet, Vol 357. Number 9260, Wass V, Van Der Vleuten C, Shatzer J, Jones R. Assessment of clinical competence, Pages 945-6., Copyright (2001), with permission from Elsevier.

1.5 Summative Assessments

1.5.1 Objective Structured Clinical Examinations (OSCE)

The use of the Objective Structured Clinical Examination (OSCE) is well established in medical and health education programs worldwide. It is used at both undergraduate and postgraduate level and in high-stakes licensing examinations (Reznick, Blackmore, Dauphinee, Rothman, & Smee, 1996) to assess student performance at the 'shows how' level of Miller's pyramid. Harden described the OCSE as "a 'focused' examination with each station focusing on one or two aspects of competence" (Harden, 1988, p. 19).

A modern definition of the OSCE from the Association of Medical Education in Europe (AMEE) is:

> An assessment tool based on the principles of objectivity and standardisation, in which the candidates move through a series of time-limited stations in a circuit for the purposes of assessment of professional performance in a simulated environment. At each station candidates are assessed and marked against standardised scoring rubrics by trained assessors (Khan, Ramachandran, Gaunt, & Pushkar, 2013, p. e1440).

Advantages of the OSCE are purported to be increased validity and reliability compared to traditional clinical examinations, and the ability to apply the examination to large numbers of students (Harden & Gleeson, 1979). Harden (1979) proposed that the standardisation of examiner and patient, as well as the use of checklists, increases reliability of the OSCE compared to traditional clinical examinations. A systematic review of the reliability of the OSCE in medical education programs identified that overall, scores on the OSCE are not always very reliable (Brannick, Erol-Korkmaz, & Prewett, 2011). Better reliability was associated with a greater number of stations, allowing for a wider sampling of cases. Some OSCEs were more reliable than others, suggesting that the structure of the OSCE stations contribute to its reliability (Brannick et al., 2011).

The OSCE in medical education programs has been extensively researched and has been demonstrated to be psychometrically valid (R. Cohen et al., 1990; Hodges et al., 1998;

Sloan et al., 1995). Psychometric validity relates to the internal structure of the examination such as the characteristics of the questions, the reproducibility and generalisabilty of the scale, and the model used to score the assessment (Downing & Haladyna, 2009). Reliability is also an aspect of validity which refers to the reproducibility of the assessment scores between raters, and also between tests (Downing & Haladyna, 2009). There is some evidence that the OSCE is valid in its use in other health education programs including physiotherapy (Wessel, Williams, Finch, & Gemus, 2003), pharmacy (Austin, O'Byrne, Pugsley, & Munoz, 2003), dentistry (Brown, Manogue, & Martin, 1999) and nursing (Rushforth, 2007). However, it has been suggested that while the format of the OSCE positively contributes to its reliability, it can be to the detriment of its context validity (Traska, 1988; Wass, van der Vleuten, Shatzer, & Jones, 2001). This is a result of the short station length (5-7 min) limiting the extent to which complex skills can be applied in an integrated way. Hodges (2003) suggests that qualitative OSCE research is required to determine its contextual validity.

Harden's (1979) original OSCE used checklists to evaluate student performance. However, OSCE checklists do not capture increasing levels of expertise (Hodges, Regehr, McNaughton, Tiberius, & Hanson, 1999). They can result in lower scores being awarded to more experienced candidates compared to novices (Hodges et al., 1999). Global rating scales allow an assessor to make an overall judgment of student performance. Global rating scales scored by experts have been demonstrated to have high inter-station reliability, better construct validity, and better concurrent validity than checklists (Regehr, MacRae, Reznick, & Szalay, 1998).

The first OSCE described by Harden included procedure stations requiring students to interpret data. No patients or examiners were required at these stations. Current recommendations from the Association of Medical Education in Europe are that OSCE stations should focus on assessing competencies or skills that cannot be adequately assessed using other tools such as the written assessment (Khan, Ramachandran, et al., 2013). The structure of modern OSCE stations do vary from program to program, which creates a challenge for researchers wanting to compare OSCEs between programs or institutions.

There are recognised challenges in the utilisation of OSCEs in health education programs. OSCEs require more time and resources to administer in comparison to other assessment methods (Khan, Ramachandran et al., 2013). Challenges with resourcing an OSCE include the amount of physical space required, shortage of facilities and equipment, and insufficient staff resources (Zamanzadeh et al., 2021). The financial cost of OSCEs can also be a limitation, although costs may vary considerably based on the structure of the OSCE (Carpenter, 1995). Students are also reported to experience increased levels of stress during an OSCE (Ferreira et al., 2020; Martínez-Pascual et al., 2022), more so than other assessment formats (Brand & Schoonheim-Klein, 2008).

A systematic review of physiotherapy OSCEs (Bobos et al., 2021) identified that OSCEs delivered in this profession tended to have a smaller number of stations than those original descriptions by Harden. Some OSCEs described in the literature have incorporated a smaller number of stations, with longer time spent at each station (Rushforth, 2007; Hawker et al., 2010; Terry et al., 2020).

1.5.2 Written Theory Examinations

Written theory assessments are widely used in health profession education programs. Written examinations test students' knowledge at the 'knows' and 'knows how' levels of Miller's pyramid. Assessment of the theoretical knowledge that underpins clinical decision making lends itself well to the written assessment format.

A discussion of the literature on written assessments by Hift (2014) proposed that written assessments can be placed into one of two categories: constructed-response questions and selected-response questions. Constructed response questions include essay questions, modified essay questions (MEQ), short answer questions (SAQ), and single word or sentence responses. Selected response questions include multiple choice questions (MCQ), one-best answer and extended matching items (Hift, 2014).

Selected response questions such as the MCQ are generally accepted to have greater reliability than constructed-response questions such as the essay (Hift, 2014). The large number of questions that can be included in an MCQ examination allow for a wide sampling of content. MEQ and essay question formats generally do not allow for such wide

sampling to occur in a realistic time frame; approximately 20 times more examination time is required using an essay question assessment to obtain the same information as a MCQ assessment (Lukhele, Thissen, & Wainer, 1994). The MCQ assessment also has a standardised, objective marking structure, whereas essay and MEQ assessments are traditionally more subjective and take longer to mark.

It has been suggested that constructed-response and selected-response assessments measure the same constructs (Norcini, Swanson, Grosso, & Webster, 1985). A meta-analysis by Rodriguez (2003) including 21 studies concluded that the difference in validity between the two assessment formats is trivial. As early as 1994 it was suggested that there was no justification for the continued use of MEQ or essay questions as selected-response questions are more efficient (Lukhele et al., 1994). While efficiency is an important consideration, so too is the influence that different response formats could have on student results. It has been demonstrated within a single cohort of students from the medical profession that students scored more highly on written examinations with a selected-response format compared to a constructed response format; the authors attributed the difference in scores to the impact of cuing in the selected-response format and recommended test developers consider this implication when designing written assessments (Desjardins, Touchie, Pugh, Wood, & Humphrey-Murto, 2014). The finding that students scored more highly on MCQ examinations has been replicated (Pham et al., 2018). This suggests that cut-scores should be considered if using a single-format assessment to determine progression within a program of study. However, MCQ examinations are vulnerable to item-writing flaws, and these can impact on students' achievement (Downing, 2005; Tarrant & Ware, 2008), both positively and negatively. The construction of an MCQ examination must be carefully considered.

1.5.3 Non-Traditional Assessments

1.5.3.1 Portfolios

A portfolio has been defined as "a focused purposeful collection of student work that documents evidence of traditional and non-traditional sources of student learning, progress, and achievement over time" (Gadbury-Amyot, Holt, Overman, & Schmidt, 2000, p. 375). The use of assessment portfolios in competency-based clinical education programs continues to increase with assessment portfolios now utilised in programs

including nursing (McCready, 2007), medicine (Tochel et al., 2009), dentistry, and dental hygiene (Gadbury-Amyot, McCracken, Woldt, & Brennan, 2014). While portfolios vary in structure and content, the documentation of key elements is usually required. Pitts (2010) identified the key elements as being: the experience, the learning that has occurred, the evidence of application of learning, and the identification of future learning needs. The use of the assessment portfolio as a measure of reflective competence has been suggested to be valid (Driessen, Overeem, van Tartwijk, van der Vleuten, & Muijtjens, 2006) and reliable (Gadbury-Amyot et al., 2005; Rees & Sheard, 2004) with high levels of inter-rater reliabilty (Driessen, van Tartwijk, van der Vleuten, & Wass, 2007). However, portfolios should be judged by multiple assessors as while they tend to reach the same overall judgement, there are differences in the way assessors reach their conclusions (Oudkerk Pool, Govaerts, Jaarsma, & Driessen, 2018). A systematic review (Driessen et al., 2007) on the use of portfolios in medical education identified that portfolios support the assessment of more general competencies required in clinical practice including reflective ability, communication and empathy. Driessen (2017) suggested that portfolios that are integrated into the curriculum, comprehensively chart learner progress, and are accompanied by explicit mentoring become a key instrument in a program of assessment.

1.5.3.2 Problem-Based Learning Evaluations

Problem-based learning (PBL) was introduced as an educational approach by McMaster University medical program in 1965, and this approach has since been adopted by other health education programs (Saarinen-Rahiika & Binkley, 1998). The PBL process and rationale for its use have previously been described by Schmidt (1983). Assessment of PBL performance is a form of non-traditional summative assessment that has been used to evaluate student academic performance in some programs featuring PBL in their curriculum. Van der Vleuten and Schuwirth (2019) discussed the challenges of assessment in the context of PBL, and the difficulty of achieving constructive alignment to the intended learning outcomes. Assessment may be self, peer, or tutor evaluation, or a combination of these. Intended learning outcomes of PBL that may be assessed include deep content learning, problem solving ability, and self-directed learning (Belland, French, & Ertmer, 2009).

There has been some suggestion that PBL evaluation forms can be used as reliable formative assessments (Valle et al., 1999). There has also been review of institution-

specific evaluation tools (Ho et al., 2014; Machado, Machado, Grec, Bollela, & Vieira, 2008) but further quality research is required to establish the validity and reliability of PBL evaluations as a form of summative assessment.

1.5.4 Oral Examinations

An oral examination, also referred to as a viva voce, is "characterised by a face-to-face interaction between an examinee and one or more examiners" (Tekian & Yudkowsky, 2009, p. 269). There is controversy over the ongoing use of traditional oral examinations due to concerns regarding their reliability and validity; factors such as limited content sampling, inconsistencies in cases and sub-sets of questions discussed, and potential biases of examiners threaten the validity of this unstructured assessment format (Tekian & Yudkowsky, 2009).

Under-sampling contributes to the unreliability of the oral examination (Tekian & Yudkowsky, 2009). At least five hours of testing time, or nine to 10 occasions, is required to achieve adequate reliability (Daelmans, Scherpbier, van der Vleuten, & Donker, 2001). Wass, Wakeford, Neighbour, and van der Vleuten (2003, p. 131) concluded that "provided an adequate length of testing time is given, and sufficient independent judgments are made on a wide range of topics, orals can be made psychometrically acceptable". To overcome the effects of under-sampling and content and examiner variability, there has been a move towards the use of a structured oral assessment. Structured oral assessments consist of predetermined clinically oriented scenarios with set questions and a specific marking scheme (Anastakis, Cohen, & Reznick, 1991). Memon, Joughin, and Memon (2010) have published a review of the literature relating to the validity and reliability of oral examinations in the context of postgraduate medical examinations. This review proposes a total of 15 conditions which must be met to ensure oral examinations are a reliable, valid, and fair method of assessing competence, particularly if used for high-stakes certification examinations (Memon et al., 2010). These conditions are:

- Examination items focus on the capabilities required for professional practice that are best assessed orally, namely clinical-reasoning and decision making.

- The specific capabilities for professional practice are established by a representative group of practitioners. The content of the examination is determined by a panel of experts based on these capabilities.
- Examination items are within the scope of professional practice.
- Where language capabilities are examined, this is done explicitly and at the level required of professional practice.
- An adequate sampling of questions are asked in order to provide sufficient coverage of the depth and breadth of practice and to ensure inter-item variability is at an acceptable level.
- Examiners are formally trained in oral examination issues and methods.
- Inter-examiner variations are monitored. Discrepancies are addressed.
- Items and implementation processes are standardised.
- Statistical methods are used to establish and monitor reliability.
- Consideration of bias is recognised by administrators as an essential element of good examination practice.
- Examination items are scrutinised by a representative panel to detect item bias.
- Result patterns are monitored to identify differential responses levels from identifiable sub-groups.
- Examinations are designed to minimise threats to their validity and reliability due to language differences of candidates.
- The language ability required in an examination should be commensurate with that required by professional practice.
- Where systematically lower or higher scores for particular groups of examinees occur, the possibility of bias should be considered.
(Memon et al., 2010, pp. 286-287)

1.6 Assessment in Clinical Practice

Assessment in the clinical setting serves similar purposes to those in the academic setting. Importantly, it provides the opportunity to assess specific competencies that students must possess to meet professional accreditation standards. In clinical practice, these competencies can be assessed with the added complexities of a clinical workplace and

authentic clients. This requires students to apply academic learning in a fluid, unpredictable environment where they must use reasoning and judgment to adapt to the circumstances at hand. Supervisors can observe students' performance and make judgment on their competence using evidence collected from discussion with the student, observation, documentation, and feedback from other staff and clients.

Supervised clinical practice provides the opportunity to assess students in the closest approximation to the 'does' level of Miller's pyramid, however, as students are aware they are being monitored and their performance assessed, it is not a true representation of what they may do as independent practitioners. Supervisors may allow a degree of independent practice to the student as trust is established, however insufficient observation time may jeopardise the reliability of the assessment.

There is a wide variety of methods used to measure competence and performance in the clinical practice setting. These vary not only between professions but also between programs.

1.7 Measures of Clinical Performance

1.7.1 Physiotherapy

A systematic review of clinical performance assessment tools utilised by the physiotherapy profession (O'Connor, McGarr, Cantillon, McCurtin, & Clifford, 2018) identified 14 tools that have been described in the literature: One tool from each of Australia, Malaysia, Japan, South Africa and the United Kingdom, two tools each from Ireland and Canada, and five assessment tools used within the United States of America (USA). The review by O'Connor et al. (2018) synthesised the published literature on the psychometric and edumetric properties of the 14 tools identified. No assessment tool was recognised as having Level A evidence for its use or was given a Class 1 recommendation.

Grading of the overall recommendations indicates the strength of the recommendation (Halperin, et al., 2106). Class 1 indicates that the clinical performance assessment tool is recommended for inclusion as a core component of a program of assessment, whereas

Class 2 indicates that the clinical performance assessment tool can be considered as one component of a program of assessment (Jelovsek, Kow, & Diwadkar, 2013). The three assessment tools with the highest recommendation for use (Class 2) in the review by O'Connor et al. (2018) were:

- the Assessment of Physiotherapy Practice (APP) used in Australia and New Zealand
- the University of Birmingham tool used within the UK
- the Physical Therapist Clinical Performance Instrument (PT CPI) (1997) version that was used in the USA until 2006 when superseded by an updated version (PT CPI: Version 2006).

The published literature regarding the University of Birmingham tool (Cross, Hicks, & Barwell, 2001) evaluates two alternative versions of the tool, but does not specify which form will be utilised going forward. Therefore, it will not be described in this thesis. Both the APP and the PT CPI are discussed in detail below.

1.7.1.1 The Assessment of Physiotherapy Practice

The Australian and New Zealand physiotherapy professions use a common instrument to assess student performance in clinical practice. The Assessment of Physiotherapy Practice (APP) instrument was developed by Dalton and colleagues in 2006 (Dalton, Keating, & Davidson, 2009). Support for the original work was provided by the Australian Learning and Teaching Council Ltd, an initiative of the Australian Government Department of Education, Employment and Workplace Relations (Dalton et al., 2009). The APP has now also been adopted in other countries including Qatar and Iceland, and is being piloted in China (Hu et al., 2020).

The development of the APP responded to a need identified by the physiotherapy profession. At the time of development there was a large variety of clinical practice assessment tools being used within physiotherapy education in Australia, as each entry-level program originally used its own unique clinical assessment form (Dalton et al., 2009). The need to be proficient in the use of multiple assessment tools was thought to burden clinical educators offering clinical placements to students from multiple institutions; this was recognised as being a barrier to the placement of students. Key stakeholders that

were considered or consulted in the development of the APP included national healthcare standards, the Australian Health Profession Regulation Agency, the national physiotherapy professional body the Australian Physiotherapy Association, universities, students, and clinical educators. Field testing of the APP occurred across a variety of clinical sites, with the majority occurring in a public hospital setting (53%). With the clinometric value of the APP established, universities and clinical educators have succeeded in standardising the assessment of clinical competence of physiotherapy students across Australia.

The APP is a criterion-based instrument covering seven domains of physiotherapy practice: professional behaviour, communication, assessment, analysis and planning, intervention, evidence-based practice, and risk management (see Figure 3). There are 20 items across the seven domains. Each item is accompanied by a list of performance indicators to be referred to as example behaviours that a student may display to demonstrate competency in that item. The standard of competence that each item is measured against is the minimally acceptable entry-level standards of physiotherapy practice. This is quantified as the standard of practice expected of a new graduate on their first day of practice. Students performing consistently at this minimum standard of competence are deemed to have reached the passing ('adequate') standard.

The APP uses a five-point scale (0-4) illustrated in Figure 3. Scores of '0' and '1' are considered 'inadequate'. A student is determined to have reached the 'adequate' passing standard for an item if they are scored a '2'. Scores of '3' and '4' are 'good' and 'excellent' respectively. To score a '2' a student must consistently demonstrate most of the performance indicators for that item to an adequate standard.

The APP is administered by the clinical educator and completed formatively mid-way through the placement, and summatively at the end of the placement. When the summative APP is administered it is accompanied by a global rating scale. The global rating scale is a four-point scale the clinical educator uses to make an overall judgement of the student's performance. The global ratings are 'inadequate', 'adequate', 'good' and 'excellent'. The global rating scale provides a second approach to assessment that can be utilised by the educational institution. It also provides opportunity for the clinical educator to reflect on their overall impression of the students and examine their scoring of individual

Assessment of Physiotherapy Practice (APP)

0 = Infrequently/rarely demonstrates performance indicators
1 = Demonstrates few performance indicators to an adequate standard
2 = Demonstrates most performance indicators to an adequate standard
3 = Demonstrates most performance indicators to a good standard
4 = Demonstrates most performance indicators to an excellent standard
n/a = (not assessed)
Note. a rating of 0 or 1 indicates that minimum acceptable competency has <u>not</u> been achieved

Professional Behaviour							Circle one number
1. Demonstrates an understanding of patient/client rights and consent	0	1	2	3	4	n/a
2. Demonstrates commitment to learning	0	1	2	3	4	n/a
3. Demonstrates ethical, legal & culturally sensitive practice	0	1	2	3	4	n/a
4. Demonstrates teamwork	0	1	2	3	4	n/a
Communication							
5. Communicates effectively and appropriately - Verbal/non-verbal	0	1	2	3	4	n/a
6. Demonstrates accurate record keeping skills	0	1	2	3	4	n/a
Assessment							
7. Conducts an appropriate patient/client interview	0	1	2	3	4	n/a
8. Selects appropriate methods for measurement of relevant health indicators	0	1	2	3	4	n/a
9. Performs appropriate physical assessment procedures	0	1	2	3	4	n/a
Analysis & Planning							
10. Appropriately interprets assessment findings	0	1	2	3	4	n/a
11. Identifies and prioritises patient's/client's problems	0	1	2	3	4	n/a
12. Sets realistic short and long term goals with the patient/client	0	1	2	3	4	n/a
13. Selects appropriate intervention in collaboration with patient/client	0	1	2	3	4	n/a
Intervention							
14. Performs interventions appropriately	0	1	2	3	4	n/a
15. Is an effective educator	0	1	2	3	4	n/a
16. Monitors the effect of intervention	0	1	2	3	4	n/a
17. Progresses intervention appropriately	0	1	2	3	4	n/a
18. Undertakes discharge planning	0	1	2	3	4	n/a
Evidence-based Practice							
19. Applies evidence based practice in patient care	0	1	2	3	4	n/a
Risk Management							
20. Identifies adverse events/near misses and minimises risk associated with assessment and interventions	0	1	2	3	4	n/a

In your opinion as a clinical educator, the overall performance of this student in the clinical unit was:

Not adequate ☐   Adequate ☐   Good ☐   Excellent ☐

Figure 3. The Assessment of Physiotherapy Practice

items for potential biases. The clinical educator is encouraged to give written feedback highlighting the student's strengths, and detailing areas to improve upon at both the mid-unit and end-unit assessment.

The APP Clinical Educator Resource Manual (Dalton et al., 2009) typifies minimum competence as the ability to:

> manage a variety of patients with relatively uncomplicated needs, such that the patient/clients' major problems are identified, major goals established, and treatment is completed safely and effectively within a reasonable time frame. While achieving this, the student is aware of their limitations and where to seek assistance (Dalton et al., 2009, p. 12).

A student is deemed to have reached minimum competence in a clinical practice area if their global rating scale on the end-unit APP is 'adequate' or higher and their total score is at least 40/80.

The validity of the APP has been established using Rasch analysis (Dalton, Davidson, & Keating, 2011). The APP was confirmed to be a unidimensional scale that could discriminate four levels of performance. The 'written documentation' item of the APP did not fit the model, however, removal of the item made only small improvement in the overall model fit. As written documentation was part of the Australian Standards for Physiotherapy (Australian Physiotherapy Council, 2006) the item was retained in the final product (Dalton et al., 2011). Further research investigating the psychometric properties of the APP identified that a two-factor representation of the APP is superior to the original unidimensional scale (Reubenson, Ng, & Gucciardi, 2020). Reubenson et al. (2020) suggest that an updated interpretation and scoring protocol reflective of the two factors identified is appropriate, where items 1-4 represent professional competencies and items 5-20 represent clinical competencies.

A cross-sectional reliability study of the APP evaluating 33 pairs of clinical educators demonstrated a high level of consistency between raters (Dalton, Davidson, & Keating, 2012). In this study, 80% of the rating pairs were in complete agreement on the global rating scale, and the remainder agreed within one point of each other. The level of

agreement between raters, established using Pearson's correlation, was strong ($r = 0.92$). It should be noted that all clinical educators involved in the study had access to training on the use of the APP, either through workshop attendance or via the APP manual. Further research has found that assessor consistency was achieved when rating students as 'not adequate' (97%) and 'good/excellent' (89%) levels of performance (Kirwan et al., 2019). Consensus was notably less when rating students at an 'adequate' level (43%), however raters were able to consistently differentiate between a 'not adequate' level performance and an 'adequate' or above (89%) (Kirwan et al., 2019).

In the systematic review by O'Connor et al. (2018) the APP was identified as scoring Level A evidence for validity and Level B evidence for reliability. The Level A score for the validity of the APP indicates that there is strong evidence from a methodologically sound study evaluating the tool across multiple settings (Jelovsek et al., 2013). It was scored as having Level B evidence for edumetric properties including ease of use, ease of interpretation, and educational impact. Level B score indicates that the evidence is acceptable and suggests that the APP could be usable by multiple programs (Jelovsek et al., 2013). Since this systematic review was published, further psychometric evidence was published in support of an updated scoring protocol for the APP as described above (Reubenson et al., 2020).

The APP was originally developed to align to the Australian Standards for Physiotherapy (Australian Physiotherapy Council, 2006). These standards were:
1. Demonstrate professional behaviour appropriate to physiotherapy
2. Communicate effectively
3. Access, interpret and apply information to continuously improve practice
4. Assess the client
5. Interpret and analysis the assessment findings
6. Develop a physiotherapy intervention plan
7. Implement safe and effective physiotherapy intervention(s)
8. Evaluate the effectiveness and efficient of physiotherapy intervention(s)
9. Operate effectively across a range of settings

(Australian Physiotherapy Council, 2006)

Each standard was accompanied by elements and criteria which provided further detail and context. However, during the development of the construct map for the APP, a variety of relevant sources were considered, and potential items were also drawn from:

- all existing instruments in use in Australia and New Zealand
- Australian Physiotherapy competency standards 2002
- National Patient Safety Framework
- National OT competency assessment document
- National Speech Pathology competency-based assessment tool, COMPASS™
- The Australian Council on Healthcare Standards EQuIP Standards 2002. (Dalton et al., 2009)

In 2015 the Australian Standards for Physiotherapy were replaced by the Physiotherapy Practice Thresholds in Australia and Aotearoa New Zealand (Physiotherapy Board of Australia & Physiotherapy Board of New Zealand). The 2015 Practice Thresholds, which remains the current version as of Dec 30 2022, outlines the seven key roles of a physiotherapist. These roles are:

1. Physiotherapy practitioner
2. Professional and ethical practitioner
3. Communicator
4. Reflective practitioner and self-directed learner
5. Collaborative practitioner
6. Educator
7. Manager/leader

(Physiotherapy Board of Australia & Physiotherapy Board of New Zealand 2015)

Like the previous standards, each role is accompanied by further descriptive elements called key competencies and enabling components. The APP was updated in 2016 to realign to the language used in the new practice thresholds. The update resulted in changes to the wording of some APP items; these changes are outlined in the table below.

Table 1. Changes to wording of APP items in update

Original item	Item wording after 2016 Update
Item 1: Demonstrates an understanding of patient/client rights and consent	Demonstrates an understanding of client rights and consent
Item 4: Demonstrates teamwork	Demonstrates collaborative practice
Item 6. Demonstrates accurate record keeping skills	Demonstrates clear and accurate documentation
Item 7: Conducts an appropriate patient/client interview	Conducts an appropriate client-centred interview
Item 8: Selects appropriate methods of measurement of relevant health indicators	Selects and measures relevant health indicators and outcomes
Item 11. Identifies and prioritises patient/client's problems	Identifies and prioritises client's problems
Item 12. Sets realistic short and long term goals with the patient/client	Sets realistic short and long term client-centred goals
Item 13. Selects appropriate intervention in collaboration with the patient/client	Selects appropriate intervention in collaboration with the client
1tem 19. Applies evidenced based practice in patient care	Applies evidenced based practice in client-centred care

(Dalton et al., 2009; APPLinkup, 2022)

The APP does not have a stand-alone item that explicitly aligns to the role of 'manager/leader' as per the current practice thresholds. However, this role has two key competencies, one of which is *competency 7.1: 'organise and prioritise their workload and resources to provide safe, effective and efficient physiotherapy autonomously and, where relevant, as a team member'* (Physiotherapy Board of Australia & Physiotherapy Board of New Zealand, 2015). This competency is reflected in the APP performance indicators for *Item 14: Performs interventions appropriately,* which includes the behaviours:

- considers the scheduling of treatment in relation to other procedures e.g. medication for pain, wound care
- demonstrates skill in case management
- recognises when to enlist assistance of others to complete workload
- completes intervention in acceptable time
- refers client to other professional/s when physiotherapy intervention is not appropriate or requires a multi-disciplinary approach.
(Dalton et al., 2009)

These performance indicators align with the enabling components of key competency 7.1,

7.1A: use appropriate strategies to manage their workload safely, effectively and efficiently

7.1B: use appropriate strategies to effectively manage and supervise individuals and groups in their work environment
(Physiotherapy Board of Australia & Physiotherapy Board of New Zealand 2015)

1.7.1.2 Clinical Performance Instrument

The American Physical Therapy Association commissioned the development of the Physical Therapist Clinical Performance Instrument (PT CPI) in 1994 (Roach et al., 2002). The fourth and final draft of the PT CPI was approved for use in 1997. The PT CPI contained 20 criteria that were scored using a 100mm visual analogue scale. The scale represented the spectrum of student performance from novice at the lower anchor and entry-level at the upper anchor. No further reference points were provided. Sample behaviours were provided for each criterion. There were separate check boxes to identify student performance as 'with distinction' or 'significant concerns/at risk'.

A field study examined the psychometric properties of the third draft of the PT CPI (Roach et al., 2002). A sample of 319 physical therapy students were each observed and rated by two clinical educators. The internal consistency of this draft of the PT CPI was high (α = 0.97). Inter-class correlation coefficients ranged from 0.21 - 0.76. Construct validity was tested using the hypothesis that performance on the CPI and the amount of clinical experience should be related. Pearson's correlation between performance on the PT CPI and number of days clinical experience ranged between 0.12 - 0.40, with all but two items

achieving statistical significance (p < 0.001). The results indicated acceptable psychometric properties of the third draft of the PT CPI. However, it was the subsequent fourth draft of the PT CPI that was approved for use and was utilised until 2006.

A retrospective analysis of the validity of the PT CPI conducted in 2003 showed that raters were unable to distinguish between the 100 possible responses available on the 100mm visual analogue scale (Straube & Campbell, 2003). The scale only demonstrated the qualities of ordered step calibration and average measure values that are required of a valid scale when it consisted of six categories. The analysis also showed that raters were not using the 'with distinction' box unless students reached entry-level performance. This was of note as the 'with distinction' box was intended to indicate students who had exceeded the expectations of the clinical educator, not students who had exceed entry-level performance.

In a review of the PT CPI in 2006, the visual analogue scale was removed from the instrument and replaced with an ordered categorical scale with six defined anchors and five distinct intervals (Roach et al., 2012). Raters could mark the scale either on an anchor if all the defined conditions for that anchor were met, or between anchors if only some of the upper anchor conditions were fulfilled. To address the inconsistent use of the 'with distinction' box it was removed; in its place the designation of the upper anchor was adjusted to read 'beyond entry-level performance'. The revised rating scale, as published by Roach et al. (2012), is shown in Figure 4. The revised version, PT CPI: Version 2006, was shown to be valid and to demonstrate a very high level of internal consistency, although inter-rater reliability has not been established (Roach et al., 2012).

The PT CPI: 2006 Version of the PT CPI was made available for use in 2008. It is administered at mid-term and end-term by the student's clinical educator. There is space provided for clinical educators to provide written feedback on student strengths, and areas to improve at the summative end-term assessment.
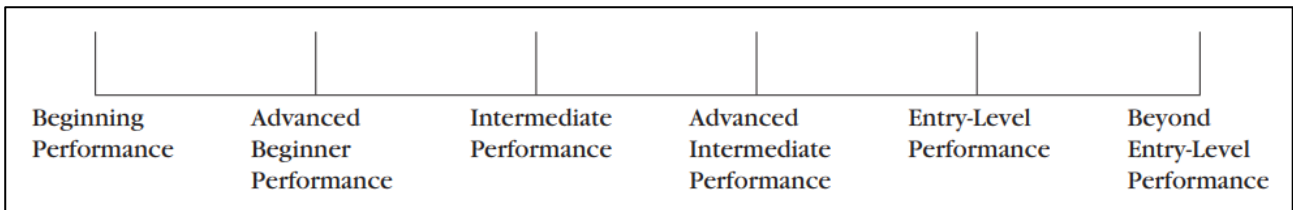
Figure 4. PT CPI: Version 2006 Rating Scale

Roach, KE et al., Validation of the Revised Physical Therapist Clinical Performance Instrument (PT CPI): Version 2006, Physical Therapy, 2012, volume *92,* issue 3, pages 416-428, by permission of Oxford University Press.

The 1997 Version of the PT CPI was awarded Level B evidence for validity in the systematic review by O'Conner et al (2018). This review identified there was not enough information contained within the published literature to judge the reliability or edumetric properties of the PT CPI (1997 version), however overall, it was awarded a Class 2 recommendation. O'Connor et al. (2018) awarded the PT CPT (2006 Version) an overall Class 3 recommendation, indicating that it can be used provisionally as a component of the program of assessment, given the lack of additional evidence for inter-rater reliability.

1.7.2 Other Allied Health Professions

1.7.2.1 Speech Pathology

The Speech Pathology profession also has a common assessment tool used by all institutions across Australia, New Zealand, Singapore and Hong Kong, the Competency Based Assessment in Speech Pathology (COMPASS®) (McAllister, Lincoln, Ferguson, & McAllister, 2011). COMPASS® was developed in consultation with key stakeholders including students, clinical educators, universities and the professional body Speech Pathology Australia (McAllister, Lincoln, Ferguson, & McAllister, 2010). Development of COMPASS® began in 2001 and field testing commenced in 2003.

COMPASS® is comprised of two sets of competencies that include four professional or generic competencies, and seven occupational competencies. The occupational competencies are in line with the Competency-based Occupational Standards for Speech Pathologists developed by Speech Pathologists Australia. COMPASS® uses a visual analogue scale where the lower anchor represents novice performance, and the upper anchor represents entry-level performance. There is a box that can be checked to identify

that student performance is above entry-level standard. Student performance on each of the 11 competencies is rated by the clinical educator and marked on the visual analogue scale. Each competency is accompanied by a set of behavioural descriptors to assist the clinical educator in their decision-making. COMPASS® is completed formatively at mid-unit and summatively at end-unit.

COMPASS® has been validated using a Rasch model which showed a unidimensional scale that discriminated seven levels of performance (McAllister et al., 2010). COMPASS® demonstrates high internal consistency with a Pearson Reliability statistic of 0.98 (S. McAllister, 2005). Student performance as measured by COMPASS® was demonstrated to increase according to hours of experience, as well over consecutive clinical placements, which supports the validity of the tool (S. McAllister, 2005).

1.7.2.2 Occupational Therapy

The occupational therapy profession in Australia has a national assessment tool, the Student Practice Evaluation Form – Revised Edition (SPEF-R). The original Student Practice Evaluation Form (SPEF) was developed and implemented in 1998 by the University of Queensland as a competency-based assessment tool, which was then licensed to institutions across Australia (Allison & Turpin, 2004). The SPEF is a flexible tool that utilises banks of assessment items that can be grouped together to suit the diverse range of practice settings in which a student can be placed. In 2005, the SPEF underwent a revision funded by the Australian Award for University Teaching and the SPEF-R was developed in consultation with academics, clinical educators and students (Turpin, Fitzgerald, & Rodger, 2011).

The SPEF-R encompasses eight domains, each containing a bank of learning objectives referred to as items: professional behaviour, self-management skills, co-worker communication, communication skills, documentation, information gathering, service provision and service evaluation (Turpin et al., 2011). As described by Rodger et al. (2014), clinical educators can select from two streams depending on the nature of the clinical placement: a Direct Client Contact stream (38 items) or a Consultancy stream (37 items). Professional behaviour, self-management skills, co-worker communication and documentation are considered core domains and include the same items for both streams. For the remaining domains, the streams include a different set of items. For each item,

clinical educators must rate the student's performance on a five-point rating scale. On the rating scale the lower anchor, '1', represents 'performs unacceptably' and the upper anchor, '5', represents 'performs with distinction'. At the end of the placement, students are awarded an overall result of a pass or fail.

The SPEF-R has been investigated for content validity (Turpin et al., 2011) and construct validity (Rodger et al., 2016). A reliability study using video vignettes to portray student performances in clinical practice scenarios found that most raters demonstrated rater consistency (Rodger et al., 2014). The authors suggested that while only one of the three clinical practice scenarios investigated achieved satisfactory test-retest agreement, the SPEF-R could be considered reliable if targeted training of clinical educators was implemented. A further revision, the SPEF-R (Second Edition) is underway and is currently being piloted (Caine, Copley, Turpin, Fleming, & Herd, 2021).

1.7.2.3 Dental Hygiene

In the United States of America, the Central Dental Testing Regional Service (CRDTS) uses the National Dental Hygiene Examination (NDHE) to judge if dental hygiene graduates are competent to practice as dental hygienists in its member states. The NDHE is a clinical performance test composed of five subtests. The five subtests are extra/intra oral assessment, periodontal probing, scaling, supraginval deposit removal and tissue management. As described by Haladyna (2011, pp. 3-4) each subtest consists of six or more items that are dichotomously scored. Penalty points are deducted for critical issues in areas of tissue management, violation of standards, improper professional demeanour and submitting unqualified patients.

Haladyna (2011) stated that the NDHE had good reliability ($\alpha = 0.897$) with a small margin of error around the cut-score. Four of the five subtests had low alpha coefficients (0.423 - 0.654), which he argued was due to the high performance of candidates and restricted scoring range. The 'Scaling' subtest, which has the largest number of items and was the most heavily weighted, had an alpha coefficient of 0.904. However, the internal consistency of this subtest was the lowest with examiner consistency between 67.6% - 74.4% for the 14 items in this subtest. The examiners who conducted the NDHE are subject-matter experts in their field (Haladyna, 2011). These examiners received extensive training on the evaluation of candidate performance, including calibration of their precision

and accuracy of judgement when scoring items of the NDHE. The evaluation by Haladyna (2011) of the NDHE asserts that the test scores are highly valid. The NDHE also demonstrated content validity as its content was based on a practice analysis comprising of a survey of more than 1500 dental hygienists that was conducted in 2008.

1.7.3 Medicine

In Australia, students who graduate from allied health education programs can practice their profession independently and autonomously once they are registered with the Australian Health Practitioner Regulation Agency. Medical students however, progress from their undergraduate medical studies into a workplace-based training program where they continue to practice under supervision and undergo regular assessment.

Doctors may be referred to as 'junior' or 'senior' doctors. These terms refer to whether a doctor is still in training and can practice independently without the need for supervision (British Medical Association, 2017). Doctors in their first two postgraduate years enter into the prevocational stage of their training and are referred to as 'junior doctors' (Confederation of Postgraduate Medical Education Councils, 2012). Junior doctors require supervision from a senior doctor who has completed their training program and can practice independently (British Medical Association, 2017). Doctors may still be considered junior doctors after completing their two years prevocational training program, as they require supervision until they have completed their vocational training program, which can take up to eight years.

1.7.3.1 Global Ratings

Wilkinson and Frampton (2003) published an adaptation of a global rating instrument used by the Royal Australasian College of Physicians (Paget, Newble, Saunders, & Du, 1996) to assess practicing doctors. The instrument was altered to reflect the difference in expectation between students and practicing doctors. The adapted instrument comprises of 12 items that are scored on a seven-point scale. The lower anchor score of '1' indicates the student is 'clearly incompetent' and the upper anchor score of '7' indicates the student is 'extremely competent'. A score of '4' defines the 'minimally acceptable level of competence'. The global rating instrument is scored by a senior doctor. In defined clinical

This Trainee Intern was:
1   Clearly incompetent on this item
2   Unacceptable level of competence on this item
3   Marginally unacceptable level of competence
4   Minimally acceptable level of competence
5   Satisfactory level of competence
6   High level of competence
7   Extremely competent

A   **Interpersonal/Communication Skills**
Ability to relate to patients and colleagues, ability to communicate with patients, their families and other professionals | 1 | 2 | 3 | 4 | 5 | 6 | 7 |

B   **Practical/Technical Skills**
Ability to perform practical/technical procedures | 1 | 2 | 3 | 4 | 5 | 6 | 7 |

C   **Diagnostic Skills**
Critically assesses information, identifies major issues and makes timely decisions | 1 | 2 | 3 | 4 | 5 | 6 | 7 |

D   **Patient Management Skills**
Shows wisdom in selecting treatment; adapts management to different circumstances | 1 | 2 | 3 | 4 | 5 | 6 | 7 |

E   **Management of Psychosocial Aspects of Illness**
Ability to recognise and/or respond to psychosocial aspects of illness | 1 | 2 | 3 | 4 | 5 | 6 | 7 |

F   **Management of Multiple Complex Problems**
Ability to manage patients with multiple complex problems | 1 | 2 | 3 | 4 | 5 | 6 | 7 |

G   **Care Skills**
Ability to treat patients and coordinate care | 1 | 2 | 3 | 4 | 5 | 6 | 7 |

H   **Theoretical Knowledge**
Knowledge of theoretical aspects of medical care | 1 | 2 | 3 | 4 | 5 | 6 | 7 |

I   **Compassion**
Appreciates patients' and families special needs for comfort and help | 1 | 2 | 3 | 4 | 5 | 6 | 7 |

J   **Respect**
Shows personal commitment to honouring the choices and rights of other persons | 1 | 2 | 3 | 4 | 5 | 6 | 7 |

K   **Examination skills**
Ability to examine patients with care and elicit signs accurately | 1 | 2 | 3 | 4 | 5 | 6 | 7 |

L   **Responsibility**
Accepts responsibility for own actions and decisions | 1 | 2 | 3 | 4 | 5 | 6 | 7 |

Figure 5. Trainee Intern Rating Instrument

Used with permission of John Wiley & Sons - Books, from 'Assessing performance in final year medical students. Can a postgraduate measure be used in an undergraduate setting?', Wilkinson, TJ., & Frampton, CM., Medical Education, volume 37, 2003; permission conveyed through Copyright Clearance Center, Inc.

placements, the form can also be scored by a nominated junior doctor, nurse, or health professional. The instrument, as published by Wilkinson and Frampton (2003), is shown in Figure 5.

Wilkinson and Frampton (2003) performed reliability and validity studies on the assessments of 62 students over two years of their training. A total of 856 global rating forms were returned and psychometric testing performed. The global rating form had high internal consistency (α = 0.973). Three factors were identified: a clinical subset (α = 0.99), a humanistic subset (α = 0.803) and an overall score (α = 0.757). There was a high level of correlation of subset and overall scores between raters (r = 0.43 - 0.64). The correlation of scores between attachments was not as high but still significant (r = 0.15 - 0.66). These results demonstrated that the global rating instrument was highly reliable.

Construct validity of the instrument was demonstrated by comparing student performance on the global rating instrument with traditional assessments used to measure student performance. There were correlations of 0.40 - 0.76, indicating that the global rating instrument was a valid assessment.

1.7.3.2 The United States Medical Licensing Examination Step 2 Clinical Skills
The United States Medical Licensing Examination (USMLE) is a three-step examination process that candidates are required pass to be licensed to practice medicine in the United States of America. According to the USMLE website www.usmle.org, Step 1 assesses the concepts of science fundamental to the practice of medicine, Step 2 assesses the clinical knowledge and patient-centred skills required to provide patient care under supervision, and Step 3 assesses the general concepts required to provide unsupervised general health care to a patient. In 2004, Step 2 of the USMLE was expanded to include a clinical skills component (Hawkins et al., 2005). The USMLE Step 2 Clinical Skills (CS) comprised of 12 stations at which candidates had 15-minute encounters with a standardised patient. Candidate performance was rated by the standardised patient using checklists that represent appropriate client history and physical examination content. The standardised patient also rated the candidate's communication, interpersonal skills and spoken English proficiency. At the completion of the standardised patient encounter, the candidate had 10 minutes to document a summary of the encounter which was marked by a trained physician.

The USMLE Step 2 CS was structurally valid as demonstrated by a factor analysis of the checklist responses of 387 candidates which identified the current scoring rubric as appropriate for use in a high-stakes assessment (De Champlain, Swygert, Swanson, & Boulet, 2006). In January 2021 it was announced that the USMLE Step 2 CS would be discontinued with no plan for it to be reinstated, after it's suspension in May 2020 during the COVID-19 pandemic (United States Medical Licensing Examination ®, 2021).

1.7.3.3 The Junior Doctor Assessment Tool

The Junior Doctor Assessment Tool (JDAT) was developed by the Postgraduate Medical Council of Western Australia and has been utilised in three tertiary hospitals since 2008 (Carr, Celenza, & Lake, 2013). The JDAT was designed to be applied by the clinical educator at the end of a 10-week clinical rotation undertaken by junior doctors in their first two postgraduate years. The tool is comprised of 10 items that are rated on a five-point Likert Scale. The lowest score of '0' represents 'not observed' and the highest score of '4' represents 'better than expected' performance. A score of '3' represents the junior doctor is performing 'as expected'. The individual item scores are summed to give a total score out of 40. Clinical educators are provided space to provide written feedback on the examinee's strengths and weaknesses. The JDAT, as published by Carr et al. (2013), is shown in Figure 6.

The validity study including 302 5th and 6th Year medical students demonstrated the JDAT to be a valid and reliable method of measuring junior doctor clinical performance (Carr et al., 2013). A total of 688 individual assessments were included in an analysis of the psychometric properties of the JDAT identifying two subscales: a Clinical Management subscale and a Communication subscale. The JDAT demonstrated high internal consistency, both as an overall score (α = 0.883) and as individual subscales (Clinical Management α = 0.829 and Communication α = 0.834).

	Not observed	Below expected level Requires substantial assistance	Borderline Requires assistance	At expected level	Better than expected
CLINICAL MANAGEMENT					
1. Clinical Assessment and Patient Management					
2. Procedural skills					
3. Emergency management					
4. Adverse event identification and risk management					
		Please support these ratings with comments overleaf			
COMMUNICATION					
5. Interpersonal skills with Patients					
6. Teamwork/Interpersonal skills with others in the healthcare team					
7. Written communication/record keeping					
		Please support these ratings with comments overleaf			
PROFESSIONALISM					
8. Professional behaviour *(responsive/reflective/ethical)*					
9. Scholarly practice *(learning/critical thinking)*					
10. Doctor's role in society *(manager/role model)*		*Please support these ratings with comments overleaf*			

Figure 6. Replica of a junior doctor assessment

Reproduced from Carr, S. E., Celenza, A., & Lake, F. (2013). Assessment of Junior Doctor performance: a validation study. *BMC Med Educ, 13*(129), 129. Under CC BY 4.0 licence.

1.7.3.4 Entrustable Professional Activities

An emerging form of assessment being utilised by not only the medical profession, but also other health professions (Shorey, Lau, Lau & Ang., 2019; Bramley & McKenna, 2021), is the Entrustable Professional Activity (EPA). An EPA is a unit of professional practice (i.e. taking a blood pressure or conducting a history) that can be fully entrusted to a student once they have demonstrated the required competencies (ten Cate & Taylor, 2021). An EPA can be utilised as a form of formative and summative assessment.

Ten Cate et al. (2015) described the development process of an EPA in AMEE Guide No. 99, which included identifying appropriate EPAs, curriculum mapping to the number of EPAs required, competency mapping for individual EPAs, and the design of an EPA matrix. Assessment using an EPA determines if a student is competent to a perform professional task unsupervised (Hauer et al., 2013). The student's performance is

evaluated against the EPA matrix based on the level of supervision that they require to perform the EPA. The suggested supervision milestones against which a postgraduate student's competence is assessed are:

- Be present and observe
- Act with direct, pro-active supervision
- Act with indirect, reactive supervision
- Act with supervision not readily available
- Provide supervision to junior trainees
  (ten Cate et al., 2015).

Alternative supervision milestones appropriate for the undergraduate learner have been proposed:

- Act in co-activity with supervision
- Act alone, but with a supervisor in room ready to step in if needed
- Act with supervisor immediately available, all findings being double checked
- Act with supervisor immediately available, key findings only being double checked
- Act with supervisor distantly available (e.g. by phone), findings being reviewed.
  (Chen, van den Broek, & ten Cate (2015).

An EPA is completed at regular timepoints throughout the clinical practice experience, providing formative feedback to the learner. Guidance is limited as to how many occasions of observation are required to support a summative entrustment decision; the number of observations will vary based on the breadth of the individual EPA. There is also limited information currently available in the reliability or validity of EPAs. However, sufficient data is required for a clinical educator to make the decision that a student is sufficiently competent to be trusted to practice independently (ten Cate & Taylor, 2021).

A systematic review by O'Dowd, Lydon, O'Connor, Madden, & Byrne (2019) identified significant variability in the quality of EPAs and is a limitation in research exploring the validity of EPAs. The review identified that the recommended template for the development of an EPAs was used by four of the 49 included studies. The variable quality of EPAs being utilised and investigated limits the ability to draw general conclusions on the

validity of EPA as a summative assessment of clinical performance. Taylor, Park, Smith, ten Cate & Tekian (2021) posit that construct validity of EPAs is dependent on a rigorous development process resulting in an EPA that is an accurate reflection of the work of the profession. These authors acknowledge that a lack of validity evidence for some EPAs is a concern. The review by O'Dowd et al. (2019) also highlighted a lack of reporting on the implementation and assessment of EPAs. Further research on the utility and validity of EPAs is needed to support the use of this novel form of summative assessment.

1.8 Factors Predicting Clinical Performance of Health Professional Students

There has been research related to identifying predictors of student clinical performance within the medical profession, particularly to identify predictive criteria for success in residency (Hamdy et al., 2006; Harfmann & Zirwas, 2011). Hamdy et al. (2006) conducted a systematic review of the predictive values of measurements obtained in medical schools for future clinical performance. Of 38 studies that met inclusion criteria, 19 could be included in the meta-analysis due to variability in the both the outcomes and the method of measurement. Hamdy et al. (2006) synthesised the research around predictor variables including USMLE Step 1, 2 and 3, OSCE, and pre-clinical grade-point average (GPA). The OSCE was found to be a predictor variable for clinical educator ratings in five studies, with a summary correlation coefficient of 0.37 (95% CI 0.22 - 0.50). This correlation was statistically significant but low. Pre-clinical GPA was also a predictor variable for clinical educator ratings in five studies, and again there was a significant but low summary correlation (0.25, 95% CI 0.19 - 0.31). The results of this review indicated that the variables of OSCE and pre-clinical GPA have limited ability to predict clinical performance given they describe 13.7% and 6.3% of the variation in students' clinical performance respectively (Hamdy et al., 2006).

A compilation and review of correlative studies by Harfmann and Zirwas (2011) looked to answer whether performance in medical school could predict performance in residency. The OSCE was the only specific summative assessment that had been investigated as a predictor of residency performance and in this the authors referenced Hamdy et al's.

(2006) review. A variety of other performance measures were reviewed, including pre-clinical GPA, clinical GPA, research experience, honour society membership, letters of recommendation and faculty interviews. The authors identified inconsistent results for most of the performance measures used but noted that medical student pre-clinical GPA was one of the indicators that correlated most strongly with performance on in-training examinations in residency (Harfmann & Zirwas, 2011). Several factors were identified that could have contributed to the inconsistency of results: coursework and grading schemes varied greatly from school to school making it difficult to compare between institutions, and the lack of a widely used measure of success in residency that was validated and reliable.

The two reviews discussed above focused on research in the medical profession, meaning studies involving other health profession disciplines were not evaluated. While the reviews looked at a range of predictor variables, the only summative assessment evaluated was the OSCE. Given the limited number of review papers identified discussing predictors of clinical performance, the focus on the medical profession, and the lack of information regarding specific summative assessments other than the OSCE, a more systematic and critical review of the literature investigating the ability of academic coursework summative assessments is warranted.

1.9 Gaps in the literature

The current body of literature reporting on education practices within the health professions is heavily skewed towards the medical profession. The medical and physiotherapy profession do share similarities in the domains of competence required of their practitioners: professional and ethical behaviour, communication and interpersonal skills, knowledge, safety and quality, leadership and management, and collaborative practice. Like medicine, physiotherapy education programs use similar assessment frameworks (Pangaro & ten Cate, 2013). Unlike medicine however, physiotherapy graduates are considered autonomous practitioners and do not undergo further supervised practice after completing their entry-level university programs. This is a significant difference in the standard expected of students at the point of graduation.

While justification can be made that evidence supporting assessment practices from the medical profession could be applied to the physiotherapy profession given the similarities in professional competencies, findings should be verified within the appropriate population. It should be considered that the standard of performance expected to be demonstrated by physiotherapy students in the clinical practice components of their programs may be higher than that of medical students, given the expectation of professional autonomy by the point of graduation. This is further impetus to test the assumption that evidence from the medical profession can be generalised to the physiotherapy profession. This thesis will aim to address this need by investigating the coursework summative assessments used within a physiotherapy education institution.

1.10 Research Rationale

The initial review of the literature discussed in this chapter identified a paucity of information regarding predictors of student's future clinical performance in both the medical, and allied health populations. Objective measures taken prior to, or during students' professional education that are predictive of their future clinical performance would provide opportunity for educators to maximise a student's preparedness for clinical placement. It would also provide opportunity to plan and implement supports for the benefit of both the student and the supervising educator of a clinical placement.

Justifications are needed to support the utilisation of resources within education programs. Should measures with demonstrated relationships with students' future clinical performance exist, not only could these be used to more accurately predict students at risk of poor performance on clinical placement, but such a measure could be utilised by faculties to support the allocation of resources to upskill and support these students.

1.11 Research Aims

The aims of this program of doctoral research are:

- To conduct a narrative review of the literature on assessment in health profession education, systematically review the literature reporting on relationships between pre-clinical coursework assessment and students' future performance in clinical practice across the health professions, and identify gaps in the literature;

- To investigate and evaluate the relationships between academic measures of physiotherapy students' performance on entry, and in the pre-clinical stages of their education program, and their future clinical performance;

- To identify the content and explore the properties of a summative assessment with demonstrated context validity in a physiotherapy education program.

1.12 Thesis Overview

Ch 1 • Narrative review of the literature surrounding summative assessment and clinical performance

Ch 2 • Study 1: Systematic review investigating relationships between summative assessments and student performance in clinical practice

Ch 3 • Study 2: Evaluate the relationship between students' undergraduate grade-point average and their future clinical performance in a postgraduate entry-level physiotherapy program

Ch 4 • Study 3: Evaluate the relationship between students' performance on coursework summative assessments and their future clinical performance in a postgraduate entry-level physiotherapy program

Ch 5 • Study 4: Identify the summative assessment within a postgraduate entry-level physiotherapy program with the largest relationship to student's future clinical performance

Ch 6 • Study 5: explore the content of the summative assessment most related to physiotherapy students' future clinical performance and compare to the assessment used measure clinical performance

Ch 7 • Study 6: explore the content of the summative assessment least related to physiotherapy students' future clinical performance

Ch 8 • Discussion and recommendations

Figure 7. Thesis overview

Chapter 2. Do coursework summative assessments predict clinical performance? A systematic review.

The formatting of this chapter is in accordance with the guidelines of *BMC Medical Education.*

2.1 Prelude

This chapter describes a systematic review conducted to collate and synthesise the existing research reporting on the relationships between pre-clinical summative assessments used in health profession education and the future clinical performance of students. An initial search of the literature identified two reviews focused within the medical profession, but no published reviews containing studies from the wider allied health professions. There are similarities in assessment frameworks between health profession programs and research findings from medical education are often used to inform practice in other health profession education programs. However, it is important to also acknowledge the differences between professions and ensure that research findings can be replicated in specific populations to confirm the practices are appropriate for the needs of each profession.

This review identifies the paucity of research conducted within the allied health professions on this topic and highlights a need for further investigation. Clinical performance measures used by studies included in this review are described in Chapter 1 (pp. 17-34).

2.2 Abstract

Background: Two goals of summative assessment in health profession education programs are to ensure the robustness of high stakes decisions such as progression and licensing, and predict future performance. This systematic and critical review aims to investigate the ability of specific modes of summative assessment to predict the clinical performance of health profession education students.

Methods: PubMed, CINAHL, SPORTDiscus, ERIC and EMBASE databases were searched using key terms with articles collected subjected to dedicated inclusion criteria. Rigorous exclusion criteria were applied to ensure a consistent interpretation of 'summative assessment' and 'clinical performance'. Data were extracted using a pre-determined format and papers were critically appraised by two independent reviewers using a modified Downs and Black checklist with level of agreement between reviewers determined through a Kappa analysis.

Results: Of the 4783 studies retrieved from the search strategy, 18 studies were included in the final review. Twelve were from the medical profession and there was one from each of physiotherapy, pharmacy, dietetics, speech pathology, dentistry and dental hygiene. Objective Structured Clinical Examinations featured in 15 papers, written assessments in four and problem-based learning evaluations, case-based learning evaluations and student portfolios each featured in one paper. Sixteen different measures of clinical performance were used. Two papers were identified as 'poor' quality and the remainder categorised as 'fair' with an almost perfect (k = 0.852) level of agreement between raters. Objective Structured Clinical Examination scores accounted for 1.4 - 39.7% of the variance in student performance; multiple choice/extended matching questions and short answer written examinations accounted for 3.2 - 29.2%; problem-based or case-based learning evaluations accounted for 4.4 - 16.6%; and student portfolios accounted for 12.1%.

Conclusions: Objective Structured Clinical Examinations and written examinations consisting of multiple choice/extended matching questions and short answer questions do have significant relationships with the clinical performance of health professional students. However, caution should be applied if using these assessments as predictive measures for clinical performance due to a small body of evidence and large variations in the predictive

strength of the relationships identified. Based on the current evidence, the Objective Structured Clinical Examination may be the most appropriate summative assessment for educators to use to identify students that may be at risk of poor performance in a clinical workplace environment. Further research on this topic is needed to improve the strength of the predictive relationship.

2.3 Background

Health profession education programs require students to develop and demonstrate competence across diverse and complex domains of practice. The curriculums delivered across the medical, nursing, and allied health professions vary in the attitudes, knowledge and skills required of their graduates. However, there are many similarities in the domains of competence required by the registration bodies of these professions. To be a licenced medical, nursing or allied health professional, graduates must demonstrate competence across domains of practice such as: professional and ethical behaviour, communication and interpersonal skills, knowledge, safety and quality, leadership and management, and collaborative practice (General Medical Council, 2013; Nursing & Midwifery Council, 2010; Physiotherapy Board of Australia & Physiotherapy Board of New Zealand, 2015). Educators must ensure that only students meeting the required standards of competence become eligible for licensing (van der Vleuten et al., 2012).

As the domains of practice required by the different health professions share similarities, so to do the assessment frameworks used by their education programs (Pangaro & ten Cate, 2013). No single mode of assessment can adequately measure performance across all domains of practice, but a well-considered program of assessment may (van der Vleuten et al., 2012). Formative assessment plays an important role in the promotion of learning, but it is summative assessment that provides a final measure of student performance (Downing & Yudkowsky, 2009; Wood, 2010). Summative assessment in health profession education has three main goals: (i) the promotion of future learning, (ii) to ensure that high-stakes decisions such as progression, graduation and licensing are robust so the public is protected from incompetent practitioners, (iii) and to provide a basis for choosing applicants for advanced training (Epstein, 2007). To achieve the goals of providing robust evidence of competence, and the identification of appropriateness for advanced training, summative assessments scores must necessarily be predictive of students' future performance. However, there is limited evidence to support this assumption.

A systematic review by Hamdy et al. (2006) of predictors of future clinical performance in medical students found Objective Structured Clinical Examinations (OSCEs) and pre-clinical grade point average (GPA) to be significant predictor variables for clinical

performance, however the predictive relationships were limited. Additionally, a compilation and review of correlative studies by Harfmann and Zirwas (2011) looked to answer whether performance in medical school could predict performance in residency. In their review, medical student pre-clinical GPA scores were one of the indicators that correlated most strongly with performance on examinations in residency.

While the reviews by Hamdy et al. (2006) and Harfmann and Zirwas (2011) looked at a range of predictor variables, the only specific mode of summative assessment common to all health professions evaluated was the OSCE and this was limited only to medical education programs. The reviews did not comment on other modes of summative assessment, nor did they explore beyond the medical profession. On this basis, the ability of a variety of modes of assessment to predict future clinical performance has yet to be investigated in detail.

The aim of this review was to critically appraise and discuss the findings of existing research investigating modes of summative assessment, and their ability to predict future clinical performance. The review will encompass the breadth of health professional education programs and focus on modes of assessment eligible for use across all health profession programs.

2.4 Methods

2.4.1 Search Strategy

Peer reviewed research papers were gathered using a search of the PubMed, CINAHL, SPORTDiscus, ERIC and EMBASE databases. Key search terms were chosen to capture the breadth of assessments commonly used within the non-clinical components of health profession programs, as well as the variety of terms used to describe performance in a clinical setting. These search terms were generated following consultation with educators from health professions and are outlined in Table 2.

Table 2. Systematic review databases and search terms

Database	Search Terms	
PubMed		student*
CINAHL	AND	predict* OR
SPORTDiscus		associat* OR
EMBASE		correlat* OR
ERIC		relat*
	AND	clinical performance OR
		clinical practice OR
		work* performance OR
	AND	summative assess* OR
		OSCE OR
		objective structured clinical examination OR
		practical exam* OR
		practical assess* OR
		written exam* OR
		written assess* OR
		theory exam* OR
		theory assess* OR
		oral exam* OR
		oral assess* OR
		oral presentation OR
		VIVA OR
		viva voce OR
		clinical exam* OR
		clinical assess* OR

2.4.2 Screening and Selection

Title and abstracts of all papers identified by the initial database searches were screened and assessed against the following inclusion criteria:

a. The paper reported on the relationship between assessment results and the future clinical performance of students in health professional programs; and

b. The paper was published in the English language; and

c. The paper was published after 1996.

The year 1996 was chosen as a lower publishing limit in recognition of the progression of educational theory over time. This date allows for the capture of 20 years of literature following on from the papers by Harden (Harden & Gleeson, 1979) regarding the

development of the OSCE and Miller's framework for the assessment of clinical competence (Miller, 1990).
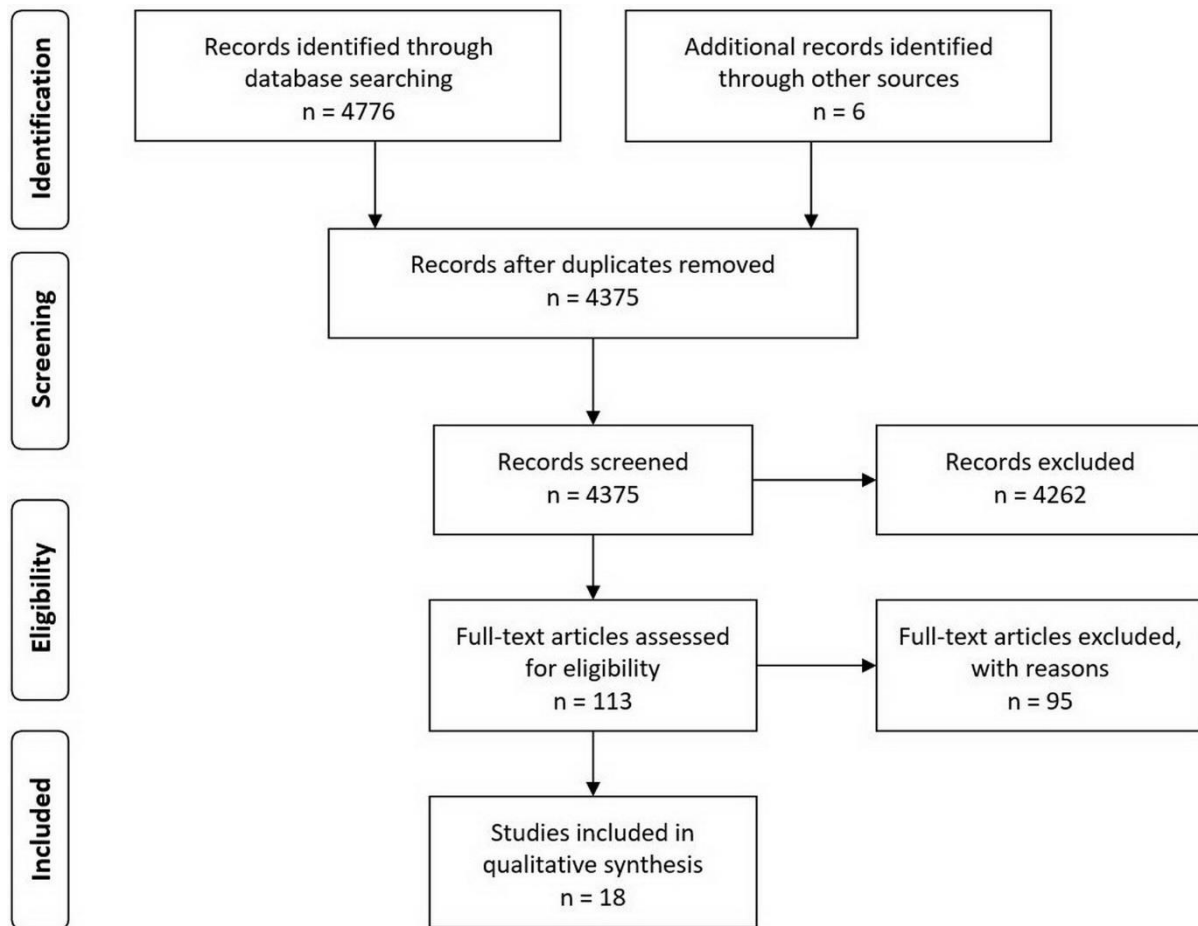
Figure 8. Preferred reporting items for systematic reviews and meta-analyses (PRISMA)

Papers selected for inclusion from the initial database searches were then subject to the application of rigorous exclusion criteria:

a. The independent variable was a formative assessment;
b. Individual modes of summative assessment were not specified (e.g. used overall GPA);
c. The independent variable was a standardised assessment limited to use by a single health profession (e.g. National Board of Medical Examiners subject examinations);
d. The independent variables were health profession education program admission criteria, applicant screening measures or entry measures;

e. Clinical performance was not measured in either a clinical workplace setting or in a clinical examination conducted externally to the education program utilizing authentic or standardised patients; or

f. The paper was an abstract, review, dissertation or discussion.

The exclusion criteria listed above were applied to ensure reasonable consistency between papers in the interpretation of 'summative assessment' and 'clinical performance' between papers to allow for a cohesive synthesis of the information. Review papers were used to provide background and supporting information. To ensure maximal search saturation a secondary search of the reference lists of papers retained for review, and papers providing background or supporting information were scanned for potentially relevant articles. These articles were then gathered and subjected to the same inclusion and exclusion criteria described above.

2.4.3 Critical Appraisal of Methodological Quality

Studies included in this review were critically appraised using a modified Downs and Black checklist (Downs & Black, 1998). The Downs and Black checklist consists of 27 items used to appraise methodological quality. The checklist was originally devised to assess the methodological quality of health care interventions, however it was appropriate to use in this review as it provided a structured format for critically appraising the papers selected for review. The protocol contains five major categories for appraisal: reporting quality, external validity, internal validity – bias, confounding, and statistical power.

The original Downs and Black checklist is scored out of 32. All items excepting Items 5 and 27 are scored on a two-point scale. A classification of 'yes' is scored as '1' point and a classification of 'no' or 'unable to determine' is scored as '0' points. Item 5, which appraises the description of confounders is scored out of '2' points, with 'yes' scoring '2' points, a 'partial description' scoring '1' point and 'no' scoring '0' points. Item 27 concerning the statistical power of the sample size was originally scored out of '5' points. For the purposes of this review Item 27 was adjusted to be scored out of either '1' point where power is reported and '0' points where power was not reported. As a result of these

adjustments, the modified total possible score was 28. This modification has been previously applied and reported in the literature (Cocke & Orr, 2015).

To allow for a quality grading of the studies, the total critical appraisal score for each study was converted into a percentage by dividing the study's raw score by 28 and multiplying by 100 to give a critical appraisal percentage. The total critical appraisal percentage was then categorised as either of 'good', 'fair' or 'poor' quality using the ranking described by Kennelly (2011). When applied to the modified Downs and Black scoring Kennelly's model categorises papers with critcal appraisal scores of 71% or greater as good quality, 54 - 70% as fair quality and 53% or less as poor quality.

Each paper was individually rated by two assessors (RT and NM) with the level of agreement determined by a Kappa analysis conducted by a third person (RO). Following the Kappa analysis any discrepancies in scores between the two scoring authors (RT and NM) were settled by consensus. Where consensus could not be reached, the raw scores were adjudicated by a third person (RO) to finalise the critical appraisal score (CAS).

2.4.4 Data Extraction and Synthesis

Data from each paper included in the review were extracted by a single author (RT) and confirmed by the fellow authors. Data were assessed using a pre-determined format as follows: clinical education program, number of students, student year of study, summative assessments used, clinical setting in which performance was measured and statistics used to establish relationships. Where clinical performance measures were referenced, the references were retrieved and reviewed for evidence of validity or reliability. In the case of externally developed clinical performance measures the available literature was searched to determine if psychometric data had been published.

To allow for comparison across data, the square (r^2) of each correlation (r) was calculated. Squaring the correlation gives the variance which measures the proportion of variability in one variable that is explained by the relationship with the other variable (Gravetter & Wallnau, 1996). In this review, the variance describes the proportion of variability in students' clinical performance explained by summative assessment scores.

2.5 Results

2.5.1 Literature Search and Selection

The results of the search are reported in Figure 8. After the application of inclusion and exclusion criteria 18 papers were retained for final analysis. Excluded papers and the reasons for their exclusion are listed in Section 2.8 (Supplementary Material).

2.5.2 Study Participants

The papers retained for the final review reported on summative coursework assessments and student performance in the clinical setting and are summarised in Table 3. Across these papers seven different clinical professions were represented: medicine or osteopathic medicine (12), pharmacy (1), physiotherapy (1), dietetics (1), speech pathology (1), dentistry (1) and dental hygiene (1). Student populations studied were from the United States of America (11), Australia (2), Canada (1), the United Kingdom (1), New Zealand (1), South Korea (1) and Hong Kong (1).

Table 3. Summary of critical review papers

Author and Population	n	Summative Assessments	Clinical Setting	Clinical Performance Measure (CPM)	CPM Evidence of Validity or Reliability
Baker et al. (2006) **Osteopathic Medicine, USA**	70	• Year 3 OSCE	COMLEX-USA Level 2-PE	COMLEX-USA Level 2-PE (Pass or fail)	Y (Gimpel, Boulet, & Errichetti, 2003; Sandella, Smith, & Dowling, 2014)
Berg et al. (2008) **Medicine, USA**	217	• Year 3 OSCE - Data Gathering, Documentation and Communication/Interpersonal Skills subsets	USMLE Step 2 CS	USMLE Step 2 CS - Data Gathering, Documentation & Communication/Interpersonal skills subsets	Y (De Champlain et al., 2006)
Campos-Outcalt, Watkins, Fulginiti, Kutob, & Gordon (1999) **Medicine, USA**	308	• 4th Year OSCE	First year of residency training – environment not specified	Residency director ratings	N
Carr et al. (2014) **Medicine, Australia**	200	• Year 4 and 5 OSCE • Year 5 Written Examination – 5 Essay and 5 SAQ • Year 6 Written Examination – 100 EMQ and 10 SAQ	Tertiary hospitals	Junior Doctor Assessment Tool (JDAT)	Y (Carr et al., 2013)
Cope, Baker, Foster, & Boisvert (2007) **Osteopathic Medicine, USA**	70	• Year 3 OSCE	Clinical rotations – environment not specified	Clinical Education Grade Form	N
Dong et al. (2014) **Medicine, USA**	806	• Year 2 & 3 OSCE	USMLE Step 2 CS	USMLE Step 2 CS – Integrated Clinical Encounter (ICE) and Communication and Interpersonal Skills (CIS) component	Y (De Champlain et al., 2006)

Ferguson & Kreiter (2004)
 	
Medicine, USA
	157	• Year 2 MCQ Examination
		• Second semester case-based learning evaluations
		  - Group participation
		  - Written reports
			Third year clerkships	Clinical evaluation form	Y (Kreiter, Ferguson, Lee, Brennan, & Densen, 1998)

Author and Population	n	Summative Assessments	Clinical Setting	Clinical Performance Measure (CPM)	CPM Evidence of Validity or Reliability
Gadbury-Amyot et al. (2005) **Dental Hygiene, USA**	74	• Portfolio	Central Region Dental Testing Service (CRTDS)	CRTDS clinical licensure examination	Y (Haladyna, 2011)
Graham, Zubiaurre Bitzer, & Anderson (2013) **Dentistry, USA**	145	• OSCE	Third year clinical training – environment not specified	Clinical productivity value – combined score for successfully completed clinical procedures	N
Han & Chung (2016) **Medicine, South Korea**	63	• OSCE • Clinical Performance Examination	Internship – Chonnam National University Hospital	Intern performance scores – 5-pt Likert scale	N
Hawker, Walker, Barrington, & Andrianopoulos (2010) **Dietetics, Australia**	193	• OSCE	7-week clinical placements – environment not specified	Clinical teacher assessments with standardised rubric	N

Ho et al. (2014)
 	
Speech Pathology, Hong Kong
	71	• PBL Evaluations –
		  ○ Reading form
		  ○ Reflective journal
		  ○ Tutorial process assessments
			Year 2 – In-house clinic
			
			Year 3 – Hospitals and community rehabilitation centers
				COMPASS®
				
				Hong Kong University Clinical Forms
					Y
					(M. McAllister, 2005; McAllister et al., 2010)
					
					N

Kahn, Merrill, Anderson, & Szerlip. (2001) **Medicine, USA**	50	• OSCE	First year of residency training – environment not specified	Residency program director overall evaluations – 5-pt Likert Scale	N
LaRochelle, Dong, & Durning (2015) **Medicine, USA**	514	• OSCE • Written examinations – x 2 MCQ and x 1 cumulative essay	First year of residency training – environment not specified	Program director evaluation form (PGY-1 PD)	Y (Dong, Durning, Gilliland, Swygert, & Artino, 2015)

Author and Population	n	Summative Assessments	Clinical Setting	Clinical Performance Measure (CPM)	CPM Evidence of Validity or Reliability
McLaughlin, Khanova, Scolaro, Rodgers, & Cox (2015) **Pharmacy, USA**	289	• Year 2 (Spring and Fall) OSCE and Year 3 (Fall) OSCE	Year 4 Advanced Pharmacy Practice Experiences	Online evaluation form	N
Probert, Cahill, McCann, & Ben-Shlomo (2003) **Medicine, UK**	30	• OSCE	Hospitals 1 year after graduation	Senior doctor assessments – 5-pt Likert scale	N
Wessel et al. (2003) **Physiotherapy, Canada**	48	• OSCE	6-week clinical practicum – environment not specified	Physiotherapy Clinical Performance Instrument (PT CPI)	Y (Roach et al., 2002; Straube & Campbell, 2003)
Wilkinson & Frampton (2004) **Medicine, New Zealand**	117	• OSCE • Written 1 – 3 x 3hr short and long essay • Written 2 – 1 x 3hr short essay and 2x3hr EMQ	Trainee internship year	Global rating instrument	Y (Wilkinson & Frampton, 2003)

COMLEX-USA Level 2-PE – Comprehensive Osteopathic Medical Licensing Examination of the United States Level 2-Performance Evaluation; CPM – Clinical performance measure;

CRDTS – Central Region Dental Testing Service; n - number of students; OSCE – Objective Structured Clinical Examination; SAQ – Short Answer Question; MCQ – Multiple Choice

Question; EMQ – Extended Matching Question; PBL – Problem-Based Learning; USMLE Step 2 CS - United States Medical Licensing Examination Step 2 Clinical Skills

The mode of coursework summative assessment investigated most commonly was the OSCE, with only three papers not featuring an OSCE as a summative assessment (Ferguson & Kreiter, 2004; Gadbury-Amyot et al., 2005; Ho et al., 2014). Written examinations featured in four papers (Carr et al., 2014; Ferguson & Kreiter, 2004; LaRochelle et al., 2015; Wilkinson & Frampton, 2004) and problem-based learning (PBL) evaluations (Ho et al., 2014), case-based learning evaluations (Ferguson & Kreiter, 2004) and student portfolios (Gadbury-Amyot et al., 2005) each featured in one paper.

Measures of clinical performance used in the medical programs were: the United States Medical Licensing Examination Step 2 Clinical Skills (USMLE Step 2 CS) (Berg et al., 2008; Dong et al., 2014); the Comprehensive Osteopathic Medical Licensing Examination of the United States Level 2-Performance Evaluation (COMLEX-USA Level 2-PE) (Baker et al., 2006); a Clinical Education Grade Form (Cope et al., 2007); a standardised Clinical Evaluation Form (Ferguson & Kreiter, 2004); intern performance scores (Han & Chung, 2016); senior doctor assessments (Probert et al., 2003); the Junior Doctor Assessment Tool (JDAT) (Carr et al., 2014); a global rating instrument (Wilkinson & Frampton, 2004); program director evaluations (LaRochelle et al., 2015) and residency program director assessments (Campos-Outcalt et al., 1999; Kahn et al., 2001). A variety of clinical performance measures were used amongst the allied health programs: the Physiotherapy Clinical Performance Instrument (PT CPI) (Wessel et al., 2003); the National Dental Hygiene Examination (NDHE) (Gadbury-Amyot et al., 2005); the Hong Kong University (HKU) speech pathology clinical evaluation form and COMPASS®: Competency Based Assessment in Speech Pathology (Ho et al., 2014); a standardised dietetics clinical teacher evaluation rubric (Hawker et al., 2010); an online evaluation form of pharmacy student performance (McLaughlin et al., 2015) and a dental clinical productivity value (Graham et al., 2013).

2.5.3 Critical Appraisal of Methodological Quality

Percentage scores based on the modified Downs and Black (Downs & Black, 1998) checklist ranged from 29% (Ferguson & Kreiter, 2004) - 68% (Wilkinson & Frampton, 2004) with a mean percentage of 56.15% (± 8.29%). The level of agreement between raters was considered as 'almost perfect' (Viera & Garrett, 2005) (k = 0.852). When graded against the criteria established by Kennelly (2011), two papers were categorised

as 'poor' quality with a critical percentage scores of 29% (Ferguson & Kreiter, 2004) and 50% (Kahn et al., 2001), the remainder were categorized as 'fair' quality (54-68%). All of the studies included in the review were descriptive cohort studies.

Analysis of the mean and standard deviations of the categories of the modified Downs and Black checklist were conducted and showed the mean score achieved in the *'reporting'* category to be 5.94 points (± 1.35 points) out of a possible 11 points. Most of the studies appraised had good *'external validity'* with a mean score in this category of 2.5/3 points. The mean score in the *'internal validity – bias'* category was 4.33 points (± 0.69 points) out of a possible 7 points. Similarly, the mean score for the *'internal validity – confounding'* category was 2.94 points (± 0.85 points) out of a possible 6 points.

The critical review findings are displayed in Table 4. All but four papers (Graham et al., 2013; Hawker et al., 2010; LaRochelle et al., 2015; Probert et al., 2003) used either Pearson's correlation, Spearman's rho or point-biserial correlations to identify the relationship between summative assessment scores and clinical performance ratings. One paper reported correlations but did not specify the type (Cope et al., 2007). Variances are listed in Table 5 and ranged from 1.4% - 39.7%.

Table 4. Critical Review Findings

Author and Population	Statistic	Findings		CAP
Baker et al. (2006) **Osteopathic Medicine**	• Point biseral correlations	Significant (p < 0.01) correlation between pass/failure of COMLEX-USA Level 2-PE and OSCE: Total OSCE score Physical examination subscore	r = 0.33 r = 0.40	64% Fair
Berg et al. (2008) **Medicine**	• Pearson's correlation	Significant (p < 0.05) correlations between the same subsets across tests. Data gathering Documentation Communication/personal	r = 0.18 r = 0.35 r = 0.32	54% Fair
Campos-Outcalt et al. (1999) **Medicine**	• Pearson's correlation	Significant (p < 0.01) correlations between residency director ratings and OSCE: Total OSCE score	r = 0.305	57% Fair
Carr et al. (2014) **Medicine**	• Pearson's correlation • Linear regression with Bonferroni adjustment	Significant correlations between the overall JDAT Score and the: Year 6 Written Year 4 OSCE Year 5 OSCE Linear regression model found individual summative assessments did not demonstrate a significant influence on overall JDAT score (p-values of 0.141-0.859).	r = 0.178, p = 0.014 r = 0.137, p = 0.027 r = 0.161, p = 0.022	64% Fair
Cope et al. (2007) **Osteopathic medicine**	• Correlations	Significant (p < 0.05*; p < 0.01**) correlations between subscores of the Clinical Evaluation Grade Form and OSCE measures: OSCE Total and Subscores 1**, 2*, 3 - 5** History taking and Subscores 1, 3 - 5** Physical Examination and Subscores 1, 3, 5* SOAP Note Form and Subscores 1**, 2*, 3**, 5*	r = 0.25 - 0.43 r = 0.31 - 0.40 r = 0.24 - 0.29 r = 0.28 - 0.34	54% Fair
Dong et al. (2014) **Medicine**	• Pearson's correlation	Significant correlations between USMLE Step 2 CS components and OSCEs. Year 2 OSCE and Integrated Clinical Encounter Component r = 0.25 Year 2 OSCE and Communication and Interpersonal Skills Component r = 0.26 Year 3 OSCE and Integrated Clinical Encounter Component r = 0.16 Year 3 OSCE and Communication and Interpersonal Skills Component r = 0.27		57% Fair

Author and Population	Statistic	Findings		CAP

Ferguson & Kreiter (2004) **Medicine**	• Pearson's correlation	Significant correlations between clinical evaluation form and:
		    MCQ Written examination
		    Case based learning reports – Group participation
		                                         - Written reports
			r = 0.27, p = 0.0009
			r = 0.28, p = 0.0004
			r = 0.21, p = 0.009
				29%
				Poor

Gadbury-Amyot et al. (2005) **Dental Hygiene**	• Pearson's correlation • Linear regression	Significant (p < 0.05) correlation between Portfolio total score and CRDTS score A prediction model using two factors predicted 13.9% of the variance in Central Region Dental Service Testing scores	r = 0.27	54% Fair

Graham et al. (2013) **Dentistry**	• Polynomial regression	Significant (p < 0.001) correlation between OSCE and clinical productivity value
		    2010 Cohort
		    2011 Cohort
			r = 0.614
			r = 0.54
				61%
				Fair

Han & Chung (2016) **Medicine**	• Pearson's correlation	Significant correlation between mean intern performance scores and OSCE
		
		Significant correlation between mean intern performance and CPX subsets
		    Patient-physician interaction
		    Clinical skills
			r = 0.278, p < 0.028
			
			r = 0.503, p < 0.001
			r = 0.278, p < 0.027
				57%
				Fair

Hawker et al. (2010) **Dietetics**	• Linear regression	Identified a β coefficient of 0.66 (p < 0.0001) between individual OSCE scores and placement scores		61% Fair

Ho et al. (2014) **Speech Pathology**	• Spearman's rho	Significant correlations (p < 0.01**; p < 0.05*) between:
		treatment skills and interpersonal skill subsets of the HKU clinical form and
		    Reflective journal
		    Tutorial process
		
		COMPASS® generic competencies and tutorial process r = 0.315 - 0.407**
		COMPASS® overall occupational competency scores and
		    Reflective journal
		    Tutorial process
			r = 0.331**, 0.272*
			r = 0.242*, 0.280*
			
			r = 0.271*
			r = 0.367**
				54%
				Fair

Kahn et al. (2001) **Medicine**	• Pearson's correlations • Spearman's rho	No significant correlations between OSCE and program director overall evaluations.	r = 0.22, p = 0.15	50% Poor

Author and Population	Statistic	Findings	CAP
LaRochelle et al. (2015) **Medicine**	• Multiple linear regression	The OSCE was a significant predictor of PGY1-PD Medical Expertise scores in a model containing multiple independent variables (β = 0.134, p = 0.013). The written examination were not significant predictors of PGY1-PD scores, although approached statistical significance (β = 0.266, p = 0.07). The OSCE was the only significant predictor of PGYI-PD Professionalism scores in a model containing multiple independent variables (β = 0.124, p < 0.026)	54% Fair

McLaughlin et al. (2015) **Pharmacy**	• Pearson's correlations	Significant (p < 0.05*; p < 0.01**) correlations between OSCEs and specific APPEs: acute care, ambulatory care, clinical specialty, and community
		Year 2 Fall OSCE and all four APPEs r = 0.13* - 0.14*
		Year 2 Spring OSCE and acute care APPE r = 0.12*
		Year 3 Fall OSCE and:
		    acute care APPE r = 0.12*
		    ambulatory care APPE r = 0.25**
		    clinical specialty APPE r = 0.13*
			57% Fair

Probert et al. (2003) **Medicine**	• Logistic regression	No statistically significant results. OR 1.64, 95% CI 0.50-5.41 OSCE showed trend of positive association with senior doctor assessments	57% Fair
Wessel et al. (2003) **Physiotherapy**	• Spearman's rank correlations	No significant correlations between OSCE average score and Physiotherapy Clinical Performance Instrument average score.	61% Fair

Wilkinson & Frampton (2004) **Medicine**	- Pearson's correlation	Significant (p < 0.01*, p < 0.001**) correlations between global rating instrument:
		Total score and: OSCE r = 0.59**
		    Written 2 r = 0.54**
		Clinical skills subset and: OSCE r = 0.63**
		    Written 2 r = 0.57**
		Humanistic subset and: OSCE r = 0.44**
		    Written 2 r = 0.41*
			68% Fair

APPE – Advanced Pharmacy Practice Experiences; CAP – Critical appraisal percentage; COMLEX-USA Level 2-PE – Comprehensive Osteopathic Medical Licensing Examination of the United States Level 2-Performance Evaluation; CPX - Clinical performance examination; HKU – Hong Kong University; JDAT – Junior Doctor Assessment Tool; OSCE – Objective Structured Clinical Examination; PGY-1 PD - Program director evaluation form; USMLE Step 2 CS – United States Medical Licensing Examination Step 2 Clinical Skills

Table 5. Proportion of variability accounted for by the relationship between summative assessment and clinical performance

Study	Relationship	Correlation (r)	p- value	Variance (r^2)	%
Baker et al. (2006) Osteopathic Medicine	OSCE measures and COMLEX-USA Level 2-PE Pass or Failure 　　OSCE Total score 　　OSCE Physical Examination subscore	0.33 0.40	< 0.01 < 0.01	0.109 0.16	10.9% 16%
Berg et al. (2008) Medicine	OSCE and USMLE Step 2 CS Data Gathering	0.18	< 0.05	0.032	3.2%
	OSCE and USMLE Step 2 CS Documentation	0.35	< 0.05	0.123	12.3%
	OSCE and USMLE Step 2 CS and Communication/Personal	0.32	< 0.05	0.102	10.2%
Campos-Outcalt et al. (1999) Medicine	OSCE total score and residency director ratings	0.305	< 0.01	0.093	9.3%
Carr et al. (2014) Medicine	Year 5 Written exam (5 modified essay questions + 5 SAQ) and JDAT overall score	0.076	0.148	0.022	2.2%
	Year 6 Written exam (100 EMQ + 10 SAQ) and JDAT overall score	0.178	0.014	0.032	3.2%
	Year 4 OSCE and JDAT overall score	0.137	0.027	0.019	1.9%
	Year 5 OSCE and JDAT overall score	0.161	0.022	0.026	2.6%
Cope et al. (2007) Osteopathic medicine	OSCE measures and Clinical Evaluation Grade Form subscores 　　OSCE Total and Subscores 1, 3 - 5 　　OSCE Total and Subscore 2 　　OSCE - History score and Subscores 1, 3 - 5 　　OSCE - Physical Examination score and Subscores 1, 3, 5 　　OSCE - SOAP Note Form score and Subscore 1, 3 　　OSCE - SOAP Note Form score and Subscores 2,5	0.31 - 0.43 0.25 0.31 - 0.40 0.24 - 0.29 0.34 - 0.38 0.28 - 0.30	< 0.01 < 0.05 < 0.01 < 0.05 < 0.01 < 0.05	0.096 - 0.185 0.063 0.096 - 0.16 0.058 - 0.084 0.116 - 0.144 0.078 - 0.090	9.6% - 18.5% 6.3% 9.6% - 16% 5.8% - 8.4% 11.6% - 14.4% 7.8% - 9%
Dong et al. (2014) Medicine	Year 2 OSCE and USMLE Step 2 CS ICE Component	0.25	< 0.01	0.063	6.3%
	Year 2 OSCE and USMLE Step 2 CS CIS Component	0.26	< 0.01	0.068	6.8%
	Year 3 OSCE and USMLE Step 2 CS ICE Component	0.16	< 0.01	0.026	2.6%

Study	Relationship	Correlation (r)	p- value	Variance (r^2)	%
	Year 3 OSCE and USMLE Step 2 CS CIS Component	0.27	< 0.01	0.073	7.3%
Ferguson & Kreiter (2004)	MCQ Written Examination and Clinical Evaluation Form	0.27	0.0009	0.073	7.3%
	Case based learning measures and Clinical Evaluation Form	0.28	0.0004	0.078	7.8%
Medicine	Case based learning group participation	0.21	0.009	0.044	4.4%
	Case based learning written reports				
Gadbury-Amyot et al. (2005) Dental Hygiene	Portfolio and CRDTS clinical licensure examination	0.27	< 0.05	0.073	7.3%
Graham et al. (2013) Dentistry	OSCE and Clinical Productivity Value	2010 Cohort = 0.614	< 0.0001	0.377	37.7%
		2011 Cohort = 0.54	< 0.0001	0.292	29.2%
Han & Chung (2016)	OSCE and Intern performance score	0.278	0.028	0.077	7.7%
	CPX clinical skills subset and intern performance score	0.278	0.027	0.077	7.7%
Medicine	CPX patient-physician interaction subset and intern performance score	0.503	< 0.001	0.253	25.3%
Ho et al. (2014) Speech Pathology	PBL Evaluation Reflective Journal and HKU clinical form				
	- Treatment skill subset	0.331	< 0.01	0.110	11.0%
	- Interpersonal skill subset	0.272	< 0.05	0.074	7.4%
	PBL Evaluation Tutorial Process and HKU clinical form				
	- Treatment skill subset	0.242	< 0.05	0.059	5.9%
	- Interpersonal skill subset	0.280	< 0.05	0.078	7.8%
	PBL Evaluation Tutorial Process and COMPASS® generic competencies	0.315 - 0.407	< 0.01	0.099 - 0.166	9.9% - 16.6%
	PBL Evaluation Reflective Journal and COMPASS® Occupational competencies overall score	0.271	< 0.05	0.073	7.3%
	PBL Evaluation Tutorial Process and COMPASS® Occupational competencies overall score	0.367	< 0.01	0.135	13.5%
Kahn et al. (2001) Medicine	OSCE and Program director evaluations	0.22	0.15	0.048	4.8%

Study	Relationship	Correlation (r)	p- value	Variance (r^2)	%
McLaughlin et al. (2015) Pharmacy	Year 2 Fall OSCE and APPE online evaluations	0.13 - 0.14	< 0.05	0.017 - 0.020	1.7% - 2%
	Year 2 Spring OSCE and Acute Care APPE	0.12	< 0.05	0.014	1.4%
	Year 3 OSCE and:				
	- acute care APPE	0.12	< 0.05	0.014	1.4%
	- ambulatory care APPE	0.25	< 0.01	0.063	6.3%
	- clinical specialty APPE	0.13	< 0.05	0.017	1.7%
Wessel et al. (2003) Physiotherapy	OSCE and Physiotherapy Clinical Performance Instrument	-0.13	Reported as not significant	0.017	1.7%
Wilkinson & Frampton (2004) Medicine	OSCE and Global rating instrument				
	- total score	0.59	< 0.001	0.348	34.8%
	- clinical skills subscale	0.63	< 0.001	0.397	39.7%
	- Humanistic subscale	0.44	< 0.001	0.194	19.4%
	Written 1 (3 x 3hr short and long essay questions) and Global rating instrument				
	- total score	0.17	0.201	0.029	2.9%
	- clinical skills subscale	0.24	0.071	0.058	5.8%
	- humanistic subscale	0.05	0.738	0.003	0.03%
	Written 2 (1x3hr short essay and 2 x 3hr EMQ) and Global rating instrument				
	- total score	0.54	< 0.001	0.292	29.2%
	- clinical skills subscale	0.57	< 0.001	0.325	32.5%
	- humanistic subscale	0.41	< 0.001	0.168	16.8%

APPE – Advanced Pharmacy Practice Experiences; CIS - Communication and Interpersonal Skills; COMLEX-USA Level 2-PE – Comprehensive Osteopathic Medical Licensing Examination of the United States Level 2-Performance Evaluation; CPX – Clinical Performance examination; CRDTS – Central Region Dental Testing Service; EMQ – Extended Matching Question; HKU – Hong Kong University; ICE - Integrated Clinical Encounter; JDAT – Junior Doctor Assessment Tool; MCQ – Multiple Choice Question; OSCE – Objective Structured Clinical Examination; PBL – Problem-Based Learning; PT CPI – Physiotherapy Clinical Performance Instrument; SAQ – Short Answer Question; USMLE Step 2 CS – United States Medical Licensing Examination Step 2 Clinical Skills

2.5.3.1 Objective Structured Clinical Examination

Three of the studies (20%) investigating the predictive ability of the OSCE found no significant relationship (Kahn et al., 2001; Probert et al., 2003; Wessel et al., 2003). OSCE did not predict physiotherapy student clinical performance on the PT CPI (Wessel et al., 2003), or medical student performance measured by either program director evaluations (Kahn et al., 2001) or senior doctor evaluations (Probert et al., 2003). Nine of 12 studies (75%) in the medical profession identified a significant positive relationship between medical student OSCE scores and clinical performance (Baker et al., 2006; Campos-Outcalt et al., 1999; Carr et al., 2014; Cope et al., 2007; Dong et al., 2014; Ferguson & Kreiter, 2004; Han & Chung, 2016; LaRochelle et al., 2015; Wilkinson & Frampton, 2004), with OSCE scores explaining between 1.9% and 39.7% of the variability in medical student clinical performance. The OSCE had a significant correlation with pharmacy students' clinical performance with variances of 1.4 - 6.3% (McLaughlin et al., 2015). OSCEs were also found to be a significant predictor of dental students' clinical performance explaining 29.2% - 37.7% of the variability in clinical productivity values (Graham et al., 2013). A significant relationship was reported between pre-clinical OSCE scores and the clinical performance of dietetic students (β = 0.66; 95% CI 0.46 - 0.86; P < 0.0001) (Hawker et al., 2010).

2.5.3.2 Written Examinations

Four of the studies evaluating medical student performance reported on the predictive ability of written examinations (Carr et al., 2014; Ferguson & Kreiter, 2004; LaRochelle et al., 2015; Wilkinson & Frampton, 2004). Two papers reported on written examinations containing long essay questions and, in both cases, they did not predict student clinical performance (LaRochelle et al., 2015; Wilkinson & Frampton, 2004). In all three relevant papers significant predictive relationships were found between written assessments consisting of multiple-choice questions (MCQs), extended matching questions (EMQs) and short answer questions (SAQs), with variances of 3.2%, 7.3% and 29.2% (Carr et al., 2014; Ferguson & Kreiter, 2004; Wilkinson & Frampton, 2004).

2.5.3.3 Other Assessments

One paper (Gadbury-Amyot et al., 2005) reported on the use of a portfolio assessment and found it predicted 7.3% of the variability in dental hygiene student clinical

performance. A PBL evaluation consisting of three assessment items predicted 5.9 - 16.6% of speech pathology student clinical performance on treatment skill and interpersonal skill subsets (Ho et al., 2014). Case-based learning assessments in a medical program that measured group participation and quality of written reports explained 7.3% and 4.8% of the variance students clinical performance respectively (Ferguson & Kreiter, 2004).

2.5.3.4 Prediction Models

A prediction model for medical student clinical performance incorporating Year 4 and 5 OSCEs, Year 5 and 6 written examinations, scores from Year 6 clinical attachments and overall GPA identified that no individual summative assessment significantly influenced the clinical performance score; the best overall predictor of clinical performance measured by the JDAT was overall GPA (Carr et al., 2014). A second paper (Wilkinson & Frampton, 2004) combined the OSCE and written examination results of medical students in a multiple regression model and found that the OSCE added significantly to the correlation with clinical performance scores. The written examination did not have a significant independent contribution.

2.6 Discussion

The aim of this review was to critically appraise and discuss the findings of existing research investigating the ability of summative assessments used within the non-clinical components of an academic curriculum to predict clinical performance across the breadth of health profession education. Eighteen studies that met inclusion and exclusion criteria were critically reviewed. The overall methodological quality of the literature that was investigated to inform this review was considered to be 'fair'. None of the studies included in the review were found to report on: (i) the principle confounders, (ii) the power of the research and (iii) attempts to blind either participants or those measuring clinical performance. The studies that scored more highly clearly described the summative assessment being investigated and the main findings, as well as reported actual probability values and the characteristics of students lost to follow up.

The OSCE is well established in health education programs worldwide. It is a mode of assessment specifically designed to provide a valid and reliable measure of students' clinical competence in a simulated environment (Harden & Gleeson, 1979). Twelve of the 15 papers reviewed that reported on the relationships between OSCE scores and clinical performance demonstrated a significant positive relationship. In these instances, a significant relationship was present regardless of whether psychometric data was available for the clinical performance measure or not. Of note, the three studies (Kahn et al., 2001; Probert et al., 2003; Wessel et al., 2003) that did not identify a significant relationship had the smallest sample sizes of all the papers in the review. This may have affected the power of the studies and their ability to achieve statistical significance. This is supported by two (Kahn et al., 2001; Probert et al., 2003) of the three papers which identified that there was a positive trend towards the OSCE predicting student performance and that statistical significance may have been reached with a larger sample size. The clinical performance measures used by studies included in this review assessed similar domains of competency to OSCEs, although in more complex and often less structured environments. OSCEs assess student performance at the 'shows how' level of Miller's pyramid (Khan, Ramachandran, et al., 2013); it is likely that the clinical performance measures also evaluate students at the 'shows how' level as there is a strong argument that 'does' can only be measured when the candidate is unaware of being observed or assessed (Khan & Ramachandran, 2012). The similarities between both the domains of competence and the levels of performance measured provides some explanation for the consistent positive relationship reported between students OSCE scores and their future clinical performance.

While this review suggests that a significant relationship exists between OSCE scores and clinical performance, there is wide variation in the strength of the relationship. With the OSCE explaining between 1.9% (Carr et al., 2014) and 39.7% (Wilkinson & Frampton, 2004) of the variation in student clinical performance, the strength of the relationships may have been influenced by other factors that in turn may vary between programs. One such factor is the structure of the OSCE itself. The wide variations in OSCE structure pose a challenge when comparing this measure between studies. For example, the dietetic OSCE had only three stations (Hawker et al., 2010) whereas the dentistry OSCE had 35 stations (Graham et al., 2013). The OSCEs described in studies on medical students ranged from

5 (Dong et al., 2014) to 18 (Wilkinson & Frampton, 2004) stations. The papers with the two strongest predictive relationships between OSCE and student clinical performance described OSCEs with 18 x 5min stations (Wilkinson & Frampton, 2004) and 35 x 2min stations (Graham et al., 2013) which suggests that longer OSCE assessments may be better predictors of performance. This finding is supported by a systematic review (Brannick et al., 2011) of the reliability of the OSCE in medical education programs which identified that while scores on OSCEs are not always very reliable, better reliability was associated with a greater number of stations. This is attributed to a wider sampling of cases across the increased number of stations. Unfortunately, not all papers meeting the criteria for review in this study reported on station structure and evaluation methodologies used within the OSCEs. This limited the ability to further discuss the impact of OSCE structure on the predictive ability of the assessment but may explain the large differences in variance.

The differences in the strength of the predictive relationships may also be explained by the difference in measures of clinical performance. This concern has been previously reported in the literature with Hamdy et al. (2006) noting that a limitation of their systematic review was the lack of a widely-used measure of clinical performance. The findings of the present review also need to be considered in light of the limitations imposed by the variety of clinical performance measures used.

A variance of 1.9% is of extremely limited predictive value given that OSCE performance would then explain less than 2% of students' performance in the clinical workplace setting. However, a variance of 37.7% indicates a strong predictive relationship. A predictive relationship of this strength would be valuable for assisting to identify students at risk of poor performance in the clinical setting. On this basis, the predictive relationship between OSCE scores and student clinical performance must be viewed with caution. However, these scores could be used by educators as a method of identifying students that may be at risk of low performance in a clinical practice setting until a more robust measure is available.

As only one paper was identified for each of the portfolio, case-based and problem-based learning assessments there is inadequate data to draw conclusions about these modes of assessment. Four papers in the review did investigate written assessments. Both papers

that investigated written assessment batteries containing long essay questions (LaRochelle et al., 2015; Wilkinson & Frampton, 2004) found no significant correlation with clinical performance scores, however all four papers that investigated written assessments consisting of EMQ, MCQ and SAQs did identify a significant positive relationship. This supports literature advocating the use of EMQs or MCQs in written examinations rather than essay questions (Lukhele et al., 1994). Like the findings for the OSCE, there was a large difference in the strength of the relationship between papers reviewed. An EMQ/MCQ written assessment explained 29.2% (Wilkinson & Frampton, 2004) of the variation in students overall clinical performance measured by a global rating instrument, but only 3.2% (Carr et al., 2014) when clinical performance was measured by the JDAT. While other program factors other than the choice of clinical performance measure may also influence these relationships, there is a large difference in the ability of the MCQ/EMQ written assessments to predict clinical performance. This highlights the need for research to occur where a standard measure of clinical performance is used to allow for comparison between studies. The findings of this review suggest that there is limited evidence to support the use of SMQ, MCQ and EMQ written assessments to predict students' clinical performance and that the written examinations should be used as a predictive measure with caution.

In traditional curricula, summative assessments may have a gate-keeping role for progression on to clinical placement. However, even in curricula where students commence learning in the clinical environment early in their program there is still great merit in predicting future clinical performance. The early identification of students at risk of poor performance allows for targeted remediation prior to clinical experiences, as well as the implementation of focused support whilst the student is embedded in the clinical environment. However, until further research adds to the body of evidence, the use of summative assessments to predict student clinical performance should be approached with caution. If educators choose to use summative assessment results to attempt to predict clinical performance, then this review suggests that the OSCE, which has a weak predictive value, may be the most appropriate choice. This review also implies that individual modes of summative assessment should not be the gatekeepers into the clinical practice environment as there is insufficient evidence to base high-stakes decisions (such as a student's ability to progress on to clinical placement) on the predictive ability of these assessments.

In addition to the differences in the structure of summative assessments investigated and clinical performance measures used that this review has already discussed, a potential limitation of the research reviewed is that only students who completed their program of study were included. Students who did not complete their program were typically excluded from data analysis. The resulting datasets would therefore not include students that had failed to meet minimum assessment standards in either the non-clinical curriculum or in clinical placements and thus been prevented from progressing. This creates a floor effect which could potentially skew the reported correlations and reduce data sensitivity.

Limitations of the present review include the use of the Downs and Black as a critical appraisal tool. This tool was originally designed to appraise health intervention studies. While it has enabled a standardised critique of the studies in this review, it may be that the papers have been appraised more harshly when applied to the same critique as an interventional study. Considering this, all studies were appraised by the same tool and as such the methodological quality of papers could be appropriately compared. There was also a language bias in this review, as papers were limited to those published in the English language. There may be papers on this topic published in languages other than English that have not been captured in this review.

Future research on this topic should aim to recruit larger sample sizes to increase statistical power. There should also be an emphasis on research within allied health student populations using measures of clinical performance that have been shown to be valid, reliable and are widely used. This approach would allow for a more rigorous comparison between programs and even professions to be conducted, aiding in the generalisation of findings across the allied health professions.

2.7 Conclusion

The findings of this review suggest that assessments used within an academic curriculum do have significant positive relationships with the clinical performance of health professional students. To use these assessments as predictive measures caution is required due to a small body of evidence and large variations in the predictive strength of

the relationships identified. The OSCE may be the most appropriate choice at this time for educators planning to use summative assessment scores to identify students that may be at risk of poor performance in a clinical workplace environment. Further research, with larger sample sizes, is required to determine the ability of summative assessments to predict the future clinical performance of health profession students particularly in allied health student populations.

2.8 Supplementary Material

2.8.1 Studies Excluded from Critical Review and Reasons for Exclusion

Exclusion Criteria

a. The independent variable is a formative assessment
b. Individual coursework summative assessments methods were not specified (e.g. used overall grade-point average)
c. The independent variable was a standardised assessment limited to use by a single health profession (e.g. National Board of Medical Examiners subject examinations)
d. The independent variables were health profession education program admission criteria, applicant screening measures or entry measures
e. Studies did not measure the dependent variable clinical performance in either a clinical workplace setting or in a clinical examination conducted externally to the education program utilising real or standardised patients
f. Paper was an abstract, review, dissertation or discussion

Table 6. Studies excluded from critical review and reasons for exclusion

Reference	Reason for exclusion
Alexander GL, Davis WK, Yan AC, Fantone III JC. Following medical school graduates into practice: residency directors' assessments after the first year of residency. Acad Med. 2000;75(10):S15-S17.	B
Andonian L. Emotional intelligence, self-efficacy, and occupational therapy students' fieldwork performance. Occup Ther Health Care. 2013;27(3):201-215.	A

Reference	Reason for exclusion
Andriole DA, Jeffe DB, Whelan AJ. What predicts surgical internship performance? Am J Surg. 2004;188(2):161-4.	B
Amos DE, Massagli TL. Medical school achievements as predictors of performance in a physical medicine and rehabilitation residency. Acad Med. 1996;71(6):678.	B
Artino AR, Gillilan WR, Waechter DM, Cruess D, Calloway M, Durning SJ. Does self-reported clinical experience predict performance in medical school and internship? Med Educ. 2012;46(2):172-8.	D
Baggs T, Barnett D, McCullough K. The value of traditional cognitive variables for predicting performance in graduate Speech-Language Pathology programs. J Allied Health. 2015;44(1):10-6.	B, D
Basco WT, Jr Gilbert GE, Chessman AW, Blue AV. The ability of a medical school admission process to predict clinical performance and patients' satisfaction. Acad Med. 2000;75(7):743-7.	D
Beauvais AM, Brady N, O'Shea ER, Griffin MT. Emotional intelligence and nursing performance among nursing students. Nurse Educ Today. 2011;31(4):396-401.	A
Bolender JS. Predictors of certification scores in family nurse practitioners: personality, academic, and demographic factors. Cardinal Stritch University. 2001.	F
Boyse TD, Patterson SK, Cohan RH, Korobkin M, Fitzgerald JT, Oh MS, Quint DJ. Does medical school performance predict radiology resident performance? Acad Radiol. 2002;9(4):437-45.	B, C
Brailovsky C, Charlin B, Beausoleil S, Coté S, Van der Vleuten C. Measurement of clinical reflective capacity early in training as a predictor of clinical reasoning performance at the end of residency: an experimental study on the script concordance test. Med Educ. 2001;35(5):430-6.	E
Brown G, Imel B, Nelson A, Hale LS, Jansen N. Correlations between PANCE performance, physician assistant program grade point average, and selection criteria. J Physician Assist Educ. 2013;24(1):42-4.	B
Burish MJ, Fredericks CA, Engstrom JW, Tateo VL, Josephson SA. Predicting success: What medical student measures predict resident performance in neurology? Clin Neurol Neurosurg. 2015;135:69-72.	B, C, E

Reference	Reason for exclusion
Callahan CA, Hojat M, Veloski J, Erdmann JB, Gonnella JS. The predictive validity of three versions of the MCAT in relation to performance in medical school, residency, and licensing examinations: a longitudinal study of 36 classes of Jefferson Medical College. Acad Med. 2010:85(6):980-7.	D
Carpio B, O'Mara L, Hezekiah J. Predictors of success on the Canadian Nurses Association testing service (CNATS) examination. Can J Nurs Res. 1996;28(4):115-23.	B
Corcoran J, Halverson A L, Schindler N. A formative midterm test increases accuracy of identifying students at risk of failing a third year surgery clerkship. Am J Surg. 2014;207(2):260-2.	A
Curtis DA, Lind SL, Brear S, Finzen FC. The correlation of student performance in preclinical and clinical prosthodontic assessments. J Dent Educ. 2007;71(3):365-72.	E
Danielson JA, Wu TF, Molgaard LK, Preast VA. Relationships among common measures of student performance and scores on the North American Veterinary Licensing Examination. J Am Vet Med Assoc. 2011;238(4):454-61.	E
Denton GD, Durning SJ, Wimmer AP, Pangaro LN, Hemmer PA. Is a faculty developed pretest equivalent to pre-third year GPA or USMLE step 1 as a predictor of third-year internal medicine clerkship outcomes? Teach Learn Med. 2004;16(4):329-32.	A
Dixon D. Prediction of Osteopathic Medical School Performance on the basis of MCAT score, GPA, sex, undergraduate major, and undergraduate institution. J Am Osteopath Assoc. 2012: 112(4):175-81.	B, D
Dong T, Saguil A, Artino AR, Jr., Gilliland WR, Waechter DM, Lopreaito J, Flanagan A, Durning SJ. Relationship between OSCE scores and other typical medical school performance indicators: a 5-year cohort study. Mil Med. 2012;177(9 Suppl):44-6.	E
Edgar S, Mercer A, Hamer P. Admission interview scores are associated with clinical performance in an undergraduate physiotherapy course: an observational study. Physiotherapy. 2014;12(14).	D

Reference	Reason for exclusion
Efurd MG. Predictors of Academic Success for the National Board Dental Hygiene Examination and the Southern Regional Testing Agency Clinical Exam. ProQuest LLC, Michigan. 2012.	F
Elliott MJ. Academic Predictors of National Council Licensure Examination for Registered Nurses Pass Rates. Walden University. 2011.	F
Ennulat CW, Garrubba C, DeLong D. Evaluation of multiple variables predicting the likelihood of passage and failure of PANCE. J Physician Assist Educ. 2011;22(1):7-18.	E
Evans P, Goodson LB, Schoffman SI. Relationship between academic achievement and student performance on the Comprehensive Osteopathic Medical Licensing Examination-USA level 2. J Am Osteopath Assoc. 2003;103(7):331-6.	B
Evans P, Goodson LB, Schoffman SI, Baker HH. Relations between academic performance by medical students and COMLEX-USA Level 2: a multisite analysis. J Am Osteopath Assoc. 2003;103(11):551-6.	B
George AB, Schuster A, Helmer SD, Drake RM, Silkey B, Cusick TE, Osland JS, Ammar AD. Do medical student's surgical examination scores correlate with performance markers? Am J Surg. 2014;208(6):1040-6.	E
Gonnella JS, Erdmann JB, Hojat M. An empirical study of the predictive validity of number grades in medical school using 3 decades of longitudinal data: implications for a grading system. Med Educ. 2004;38(4):425-34	B
Greenburg DL, Durning SJ, Cohen DL, Cruess D, Jackson JL. Identifying medical students likely to exhibit poor professionalism and knowledge during internship. J Gen Intern Med. 2007;22(12):1711-7.	B
Hamdy H, Prasad K, Anderson MB, Scherpbier A, Williams R, Zwierstra R, Cuddihy H. BEME systematic review: predictive values of measurements obtained in medical schools and future performance in medical practice. Med Teach. 2006;28(2):103-16.	F
Harfmann KL, Zirwas MJ. Can performance in medical school predict performance in residency? A compilation and review of correlative studies. J Am Acad Dermatol. 2011;65(5):1010-22.	F

Reference	Reason for exclusion
Hojat M, Gonnella JS, Mangione S, Nasca TJ, Veloski JJ, Erdmann JB, Callahan CA, Magee M. Empathy in medical students as related to academic performance, clinical competence and gender. Med Educ. 2002;36(6):522-7.	A, E
Howard L, Jerosch-Herold C. Can entry qualifications be used to predict fieldwork and academic outcomes in occupational therapy and physiotherapy students? The British Journal of Occupational Therapy. 2000;63(7):329-34.	D
Hu Y, Martindale JR, LeGallo RD, White CB, McGahren ED, Schroen AT. Relationships between preclinical course grades and standardized exam performance. Adv Health Sci Educ Theory Pract. 2016;21(2):389-99.	B
Inda KC. Relationship between clinical reasoning skills and certification exam performance in occupational therapy candidates. Nova Southeastern University. 2007.	F
Kirchner GL & Holm MB. Prediction of academic and clinical performance of occupational therapy students in an entry-level master's program. Am J Occup Ther. 1997;51(9):775-9.	D
Kirchner GL, Stone RG, Holm MB. Use of Admission Criteria to Predict Performance of Students in an Entry-Level Master's Program on Fieldwork Placements and in Academic Courses. Occup Ther Health Care. 2001;13(1):1-10.	D
Kosmahl EM. Factors related to physical therapist license examination scores. Journal of Physical Therapy Education. 2005;19(2):52-6.	E
Kulatunga-Moruzi C, & Norman GR. Validity of admissions measures in predicting performance outcomes: the contribution of cognitive and non-cognitive dimensions. Teach Learn Med. 2002;14(1):34-42.	D
Lawson DM, Till H. Predictors of performance of students from the Canadian Memorial Chiropractic College on the licensure examinations of the Canadian Chiropractic Examining Board. J Manipulative Physiol Ther. 2006;29(7):566-69.	E
Lewis ES. A study of emotional intelligence, cognitive intelligence and clinical performance of physical therapy students. University of Massachusetts Lowell. 2004.	F

Reference	Reason for exclusion
Lewis E. Emotional intelligence as a predictor for clinical performance in professional physical therapy students. Internet Journal of Allied Health Sciences & Practice. 2010;8(4):1-8.	A
Lievens F, Sackett PR. The validity of interpersonal skills assessment via situational judgment tests for predicting academic success and job performance. J Appl Psychol. 2012;97(2):460-8.	D
Luedtke-Hoffmann K, Dillon L, Utsey C, Tomaka J. Is there a relationship between performance during physical therapist clinical education and scores on the National Physical Therapy Examination (NPTE)? Journal of Physical Therapy Education. 2012;26(2):41-9.	B, C, E
Lurie SJ, Lambert DR, Nofziger AC, Epstein RM, Grady-Weliky TA. Relationship between peer assessment during medical school, dean's letter rankings, and ratings by internship directors. J Gen Intern Med. 2007;22(1):13-6.	A
Martin IG & Jolly B. Predictive validity and estimated cut score of an objective structured clinical examination (OSCE) used as an assessment of clinical skills at the end of the first clinical year. Med Educ. 2002;36(5):418-25.	E
Massey SL, Lee L, Young S, Holmerud D. The relationship between formative and summative examination and PANCE results: a multi-program study. J Physician Assist Educ. 2013;24(1):24-34.	B, E
McCauley DT. Predictors of Canadian registered nurse examination performance: Nursing coursework, clinical courses, and length to program completion. D'Youville College. 2014.	F
Merrick HW, Nowacek G, Boyer J, Robertson J. Comparison of the objective structured clinical examination with the performance of third-year medical students in surgery. Am J Surg 2000;179(4):286-8.	A
Middlemas DA, Manning JM, Gazzillo LM, Young J. Predicting Performance on the National Athletic Trainers' Association Board of Certification Examination From Grade Point Average and Number of Clinical Hours. J Athl Train. 2001;36(2):136-40.	B, E

Reference	Reason for exclusion
Morris J & Farmer A. The predictive strength of entry grades and biographical factors on the academic and clinical performance of physiotherapy students. Physiother Theory Pract. 1999;15(3):165-73.	D
Morrison CA, Ross LP, Sample L, Butler A. Relationship between performance on the NBME(R) Comprehensive Clinical Science Self-Assessment and USMLE(R) Step 2 Clinical Knowledge for USMGs and IMGs. Teach Learn Med. 2014;26(4):373-8.	A, C, E
Muller ES, Harik P, Margolis M, Clauser B, Mckinley D, Boulet JR. An examination of the relationship between clinical skills examination performance and performance on USMLE Step 2. Acad Med. 2003:78(10):S27-9.	B
Nadasan T & Puckree T. Do the selection criteria for admittance to the physiotherapy program predict students' performance? South African Journal of Physiotherapy. 2003;59(3):20.	D
Nelson LP, Maramaldi P, Kinnunen TH, Kalenderian E. Early performance in a humanistic medicine course as a predictor of dental students' later clinical performance. J Dent Educ. 2013;77(8):1006-12.	B
Nunez DW, Taleghani M, Wathen WF, Abdellatif HM. Typodont versus live patient: predicting dental students' clinical performance. J Dent Educ. 2012;76(4):407-13.	E
Park SE, Susarla SM, Massey W. Do admissions data and NBDE Part I scores predict clinical performance among dental students? J Dent Educ. 2006;70(5):518-24.	C, D
Pearson SA, Rolfe IE, Henry RL. The relationship between assessment measures at Newcastle Medical School (Australia) and performance ratings during internship. Med Educ. 1998;32(1):40-5.	B
Pepple DJ, Young LE, Gordon-Strachan GM, Carroll RG. Pre-clinical grades predict clinical performance in the MBBS stage II examination at the University of the West Indies, Mona Campus. Niger J Physiol Sci. 2013;28(2)201-4.	B, E
Peskun C, Detsky A, Shandling M. Effectiveness of medical school admissions criteria in predicting residency ranking four years later. Med Educ. 2007;41(1):57-64.	D

Reference	Reason for exclusion
Ranney RR, Gunsolley JC, Miller LS, Wood M. The relationship between performance in a dental school and performance on a clinical examination for licensure: a nine-year study. J Am Dent Assoc. 2004;135(8):1146-53.	B
Rice EW. The relationship between emotional intelligence, self-efficacy, and clinical performance in associate degree nursing students. Capella University. 2013.	A
Ripkey DR, Case SM, Swanson DB. Identifying students at risk for poor performance on the USMLE Step 2. Acad Med. 1999;74(10 Suppl):S45-8.	C, E
Roberts CM. Relationships among admission variables, professional education outcome measures, and job performance of University of Missouri physical therapy graduates. University of Missouri, Columbia. 1996.	F
Roberts WL, Pugliano G, Langenau E, Boulet JR. Modeling relationships between traditional preadmission measures and clinical skills performance on a medical licensure examination. Adv Health Sci Educ Theory Pract. 2012;17(3):403-17.	D
Roth KS, Riley WT, Brandt RB, Seibel HR. Prediction of students' USMLE step 2 performances based on premedical credentials related to verbal skills. Acad Med. 1996;71(2);176-80.	B, E
Saguil A, Dong T, Gingerich RJ, Swygert K, LaRochelle JS, Artino Jr AR, Cruess DF, Durning SJ. Does the MCAT Predict Medical School and PGY-1 Performance? Mil Med. 2015;180(4):4-11.	D
Sandow PL, Jones AC, Peek CW, Courts FJ, Watson RE. Correlation of admission criteria with dental school performance and attrition. J Dent Educ. 2002;66(3):385-92.	D,E
Sawhill A, Butler A, Ripkey D, Swanson DB, Subhiyah R, Thelman J, Angelucci K. Using the NBME self-assessments to project performance on USMLE Step 1 and Step 2: impact of test administration conditions. Acad Med. 2004;79 Suppl 10:55-57.	A, C, E
Scior K, Bradley CE, Potts HW, Woolf K, & de CWAC. What predicts performance during clinical psychology training? Br J Clin Psych. 2014;53(2):194-212.	D

Reference	Reason for exclusion
Scott JN, Markert RJ, Dunn MM. Critical thinking: change during medical school and relationship to performance in clinical clerkships. Med Educ. 1998;32(1):14-8.	A
Shelledy DC, Gardner DD, Carpenter ME, Murphy DL. The relationship between general critical thinking ability and student performance. Respiratory Care Education Annual. 2004;13:23-8.	A, B
Silver B & Hodgson CS. Evaluating GPAs and MCAT scores as predictors of NBME I and clerkship performances based on students' data from one undergraduate institution. Acad Med. 1997;72(5):394-6.	B
Simon SR, Bui A, Day S, Berti D, Volkan K. The relationship between second-year medical students' OSCE scores and USMLE Step 2 scores. J Eval Clin Pract. 2007;13(6):901-5.	E
Sisola SW. Moral reasoning as a predictor of clinical practice: the development of physical therapy students across the professional curriculum. Journal of Physical Therapy Education. 2000;14(3):26-34.	A
Stacey DG & Whittaker JM Predicting academic performance and clinical competency for international dental students: seeking the most efficient and effective measures. J Dent Educ. 2005;69(2):270-80.	D
Stewart CM, Bates RE, Jr., Smith GE. Relationship between performance in dental school and performance on a dental licensure examination: an eight-year study. J Dent Educ. 2005;69(8):864-9.	B
Tan K, Meredith P, McKenna K. Predictors of occupational therapy students' clinical performance: an exploratory study. Australian Occupational Therapy Journal. 2004;51(1):25-33.	A,D
Thieman TJ, Weddle ML, Moore MA. Predicting academic, clinical, and licensure examination performance in a professional (entry-level) master's degree program in physical therapy. Journal of Physical Therapy Education. 2003;17(2):32-7.	D
Tomlin G. The use of interactive video client simulation scores to predict clinical performance of occupational therapy students. Am J Occup Ther. 2005;59(1):50-6.	B

Reference	Reason for exclusion
Townsend AH, McLIvenny S, Miller CJ, Dunn EV. The use of an objective structured clinical examination (OSCE) for formative and summative assessment in a general practice clinical attachment and its relationship to final medical school examination performance. Med Educ. 2001;35(9):841-6.	E
Velayo BC, Stark PC, Eisen SE, Kugel G. Using dental students' preclinical performance as an indicator of clinical success. J Dent Educ. 2014;78(6):823-8.	E
Vendrely AM. An investigation of the relationships between academic performance, clinical performance, critical thinking, and success on the physical therapy education program. Loyola University of Chicago. 2002.	A, B, F
Victoroff KZ & Boyatzis RE. What is the relationship between emotional intelligence and dental student clinical performance? J Dent Educ. 2013;77(4):416-26.	A
Wass V, Van Der Vleuten C, Shatzer J, Jones R. Assessment of clinical competence. Lancet. 2001;357(9260):945-9.	F
Watson CJ, Barnes CA, Williamson JW. Determinants of clinical performance in a physical therapy program. J Allied Health. 2000;29(3):150-6.	D
Wettstein RB, Wilkins RL, Gardner DD, Restrepo RD. Critical-thinking ability in respiratory care students and its correlation with age, educational background, and performance on national board examinations. Respir Care. 2011;56(3):284-9.	A
White CB, Dey EL, Fantone JC. Analysis of factors that predict clinical performance in medical school. Adv Health Sci Educ Theory Pract. 2009;14(4):455-64.	D
Wiggers TB, Holton RH. Predictive value of a senior comprehensive examination as to performance on a national certification examination. Clin Lab Sci. 2001;14(1):21-6.	E
Williams KB, Glasnapp DR, Tilliss TS, Osborn J, Wilkins K, Mitchell S, Kershbaum W, Schmidt C. Predictive validity of critical thinking skills for initial clinical dental hygiene performance. J Dent Educ. 2003;67(11):1180-92.	A, E

Reference	Reason for exclusion
Williams RG, Klamen DL. Twenty Questions game performance on medical school entrance predicts clinical performance. Med Educ. 2015;49(9):920-7.	A
Williams KB, Schmidt C, Tilliss TS, Wilkins K, Glasnapp DR. Predictive validity of critical thinking skills and disposition for the national board dental hygiene examination: a preliminary investigation. J Dent Educ. 2006;70(5):536-44.	A
Woloschuk W, McLaughlin K, Wright B. Is undergraduate performance predictive of postgraduate performance? Teach Learn Med. 2010;22(3):202-4.	B
Yoho RM, Tallerico V, Vardaxis V. Relationship between student academic and clinical performance in podiatric medical education at Des Moines University. J Am Podiatr Med Assoc. 2012;102(4):314-8.	B

Chapter 3. Undergraduate grade-point average as a selection criterion for a postgraduate entry-level physiotherapy program

This chapter has been published in *Australian Journal of Clinical Education:* Terry, R., Hing, W. A., Orr, R. M., & Milne, N. (2018). Undergraduate Grade-Point Average as a Selection Criterion for a Postgraduate Entry-Level Physiotherapy Program. *Australian Journal of Clinical Education*, *3*(1), [3]. Reproduced under a CC BY-NC-ND license

The formatting of this chapter is in accordance with the guidelines of *Australian Journal of Clinical Education.*

3.1 Prelude

The systematic review conducted in Chapter 2 (Terry, Hing, Orr, & Milne, 2017) aimed to identify relationships between students' performance in coursework summative assessments. Evidence of weak predictive relationships between the performance of medical students on written assessments consisting of MCQ/EMQs and their future clinical performance was identified. Evidence of predictive relationships between medical students OSCE scores and their future clinical performance was also identified, although with significant variability in the strength of the relationships (variances ranged from 1.9% - 39.7%). However, no such relationships were identified for physiotherapy students. Indeed, only one paper from the physiotherapy profession was included in the study (Wessel et al., 2003), and this showed no relationship between student performance on OSCE and future clinical practice. This is a clear indication that further study should be conducted within the physiotherapy profession, to support the ongoing use of a medical-type assessment framework in physiotherapy education programs.

This chapter describes a study conducted to identify the relationships between undergraduate grade-point average and the performance of students in an entry-level postgraduate physiotherapy program. An initial search of the literature identified that

research conducted within the physiotherapy profession tended to focus entry measures to undergraduate programs. The global trend in physiotherapy is moving toward postgraduate entry-level programs. Postgraduate programs use varying methods to determine students' eligibility to enter a program, and these may be competitive. This study investigates a commonly used selection method to determine students' entry into postgraduate programs, undergraduate grade-point average (UGPA).

3.2 Abstract

Aim: Undergraduate GPA (UGPA) is commonly used to select students into postgraduate physiotherapy programs. The aim of this study was to explore the relationships between UGPA and the academic and clinical performance of postgraduate entry-level physiotherapy students.

Method: A retrospective cohort study of students from four cohorts (2010 - 2013) of a postgraduate entry-level physiotherapy program. UGPA, average pre-clinical coursework marks and clinical performance scores were investigated. Clinical performance was measured by the Assessment of Physiotherapy Practice. Normality tests, descriptive analysis and correlations between variables were calculated. Participants were then grouped according to UGPA and a one-way ANOVA was performed to determine differences in clinical or coursework performance between groups.

Results: Data from 121 students were analysed. There were no significant relationships identified between UGPA and pre-clinical coursework or clinical performance scores. There were no significant differences in academic or clinical performance between groups when students were classified by UGPA.

Conclusion: These findings indicate a need to reconsider the use of UGPA as a sole selection criterion and supports the inclusion of other criteria to select students into competitive programs. Minimum UGPA entry requirements for postgraduate physiotherapy programs should be reviewed to ensure all suitable applicants are eligible for admission.

3.3 Introduction

In the last decade, the global trend in physiotherapy education has shifted towards postgraduate entry-level programs (Commission on Accreditation In Physical Therapy Education, 2018; Council of Canadian Physiotherapy University Programs, 2009). In Australia, prospective physiotherapy students can still choose between an undergraduate and postgraduate entry-level education. Of the 28 accredited Australian entry-level programs there are 17 Bachelor programs, eight Master and three Doctor of Physiotherapy programs (Australian Health Practitioner Regulation Agency, 2013). Appropriate selection of candidates into these postgraduate programs represents high stakes decisions for both applicants and institutions. Applicants are required to have completed set prerequisites to be eligible for admission. This represents a significant time and financial investment from the applicant, so selection processes must reflect the significance of this commitment with the use of valid and reliable assessments. The rising costs of delivering, and receiving, a professional tertiary education has resulted in community demands that institutions have a moral obligation to admit students who are equipped academically and personally for success within their chosen program of study. This is necessary to avoid failing students accruing fruitless debt, and institutions and clinical education partners pouring resources into students that will not graduate to be employed within the health sector.

3.4 Literature Review

Previous literature from within the physiotherapy profession investigating program admission requirements has tended to focus on entry measures into undergraduate degrees (Edgar, Mercer, & Hamer, 2014; Howard & Jerosch-Herold, 2000; Morris & Farmer, 1999; Payton, 1997; Watson, Barnes, & Williamson, 2000). Entry requirements into undergraduate physiotherapy programs differ from those of postgraduate programs. Undergraduate physiotherapy programs typically use academic school-leaving scores (or those equivalent) to determine entry. Contemporary literature suggests that while significant relationships do exist between academic admission scores and a student's future academic performance in an undergraduate physiotherapy program, the relationships are weak and of limited predictive value (Edgar et al., 2014; Howard & Jerosch-Herold, 2000; Morris & Farmer, 1999; Watson et al., 2000). Significant relationships between prior academic performance and subsequent academic performance in undergraduate medical programs have also been

identified (Howard & Jerosch-Herold, 2000; Mercer & Puddey, 2011; Morris & Farmer, 1999; Watson et al., 2000). However, contemporary studies have found no relationship between previous academic performance and the future clinical performance of undergraduate physiotherapy students (Edgar et al., 2014; Watson et al., 2000).

Postgraduate physiotherapy programs tend to place greater emphasis on candidate selection. Candidate selection is used by medical and allied health schools in an attempt to ensure that students entering the program possess the traits desired by their profession and have the potential to successfully complete the program. This is especially true where there is competition for limited places. Tools used in postgraduate entry-level programs may include: pre-admission academic grades, aptitude tests, interviews, written submissions and letters of reference (Salvatori, 2001).

Previous academic performance is a common criterion assessed on application to postgraduate study. Predictive relationships between undergraduate grade point average (UGPA) and both academic and clinical performance in postgraduate medical education programs have been identified (Kulatunga-Moruzi & Norman, 2002). Utzman, Riddle & Jewell (2007) analysed data from 20 postgraduate entry-level physiotherapy programs in the United States of America and concluded that UGPA, in conjunction with other academic admissions measures, could be used to estimate students' academic risk. However, their study did not investigate relationships between undergraduate academic performance and specific physiotherapy program performance outcomes, such as performance on clinical placement. An earlier study by Thieman, Weddle & Moore (2003) did investigate the relationship between overall undergraduate academic performance and students' performance in a Masters of Physiotherapy program (MPT). Thieman and colleagues identified a moderate relationship between students' overall UGPA and their overall MPT GPA, but found no significant relationship between UGPA and clinical performance. The conflicting findings between research conducted in undergraduate and postgraduate physiotherapy education programs, as well as those within the medical profession, is a clear indication of the need for further research on this topic.

A search of entry requirements listed on official university websites for the 11 postgraduate entry-level physiotherapy programs currently on offer in Australia (see pages 99-100 for programs and websites referenced) revealed that all programs listed prerequisite subjects

necessary to be eligible for admission. Ten of the 11 programs used prior academic performance as an admission criterion: seven programs listed a minimum desired UGPA or equivalent, a further three stated that a competitive UGPA was required. Of these ten programs, seven did not use any other measures to determine applicants' appropriateness or rank.

Undergraduate academic performance, as measured by students UGPA, is the most common method used in Australia to determine admission into a postgraduate entry-level physiotherapy program, with most programs using it in isolation to select students. The use of UGPA is a more cost-effective method of selection for institutions compared to the time and financial cost of administering interviews and additional testing. The postgraduate program investigated in this study uses UGPA as evidence for likely academic success, alongside a semi-structured interview and personal statements as part of its selection process. The administrative and academic time spent interviewing and reviewing personal statements represents a significant financial investment by the university into the selection process, given that the number of applications well exceeds the available places. This raises the question as to whether these additional selection measures represent a good investment for the university, and the wider physiotherapy profession that will be serviced by the graduate, if UGPA is in fact a valid predictor of students future clinical practice performance.

Entry into the host program is competitive, with applicants required to meet set academic pre-requisites, submit supporting documentation, and undertake a semi-structured interview process. Applicants are first vetted to ensure they have met academic prerequisites. All eligible applicants are then interviewed by faculty staff. There is some evidence of small to moderate relationships between applicants' performance on an admission interview and their future clinical performance in an Australian undergraduate physiotherapy program (Edgar et al., 2014). The interviewers review applicants' personal statements and academic record prior to the interview. A semi-structured interview is then conducted where applicants are required to respond to set questions that are scored, and also respond to questions raised by the review of submitted documents. The interview process is aimed at ensuring applicants are knowledgeable about both the physiotherapy profession and the program itself, can demonstrate the personal qualities desired as a health-professional program graduate, and are equipped to meet the demands of the program. The interviewers consider factors including previous industry experience, applicants' personal reflections and

motivation for physiotherapy as a career, their reasoning for entry into a postgraduate physiotherapy program, and their academic history. A minimum UGPA requirement is advertised, however applicants who strongly display the personal attributes desired by the physiotherapy profession but have a UGPA lower than the published threshold may be offered a place if the selection panel note an upward trajectory in their grades over the final year of their undergraduate program. The applicant is questioned to seek clarification or further detail on any information given in personal statements and during the interview itself. The applicant is then ranked into bands based on their responses during the interview. Offers are made based on these rankings, and UGPA is used as the factor to differentiate between appropriate candidates within the same band.

The selection process is intended to admit students who are likely to succeed on clinical placement and in the future as competent health professionals. There is a clear bias to this process in that only applicants who display the personal attributes, communication skills and commitment to the profession are admitted into the program. However, the host program still uses UGPA as a factor to differentiate between students of equal rank after the interview process. This assumes that UGPA is a valid predictor of applicant's future performance. Given that applicants are competitively selected based on their ability to demonstrate the attributes, behaviours and communication skills required of a health professional during the interview process, it is an implicit assumption by faculty that future variations in academic and clinical performance as a student can be partially explained by the student's previous academic performance. This is clearly a widely held assumption, given the reliance on UGPA as an admission criterion to select students into Australian postgraduate physiotherapy programs. However, there is currently no clear evidence supporting this assumption given the limited amount of published literature investigating relationships between admission measures and student performance in contemporary physiotherapy education, particularly in the Australian context. Rather, there are conflicting findings between research conducted in undergraduate and postgraduate physiotherapy education, as well as that conducted in the medical profession.

3.5 Aims

The dominant method of determining entry into postgraduate physiotherapy programs in Australia is through applicant UGPA with this approach used by the host institution to discriminate between equally ranked applicants. The aim of this study was to determine if a relationship existed between UGPA and the academic and clinical performance of postgraduate entry-level physiotherapy students in the host institution. This study also aimed to determine if differences in performance existed between students when grouped according to UGPA. These findings will then be related to current practice and implications for student selection into postgraduate physiotherapy discussed. This study focused on students' performance in core areas of physiotherapy that are likely to be consistently represented in curriculums across Australia and internationally.

3.6 Methods

This was a retrospective cohort study. Participants in this study were students from four consecutive cohorts of an Australian postgraduate, entry-level physiotherapy program. The student numbers in each cohort are shown in Table 7. All students who enrolled in their first year of study between 2010 and 2013 were included within the study. Students were excluded if they did not complete any clinical placements. Ethics approval was received from the host institution.

Table 7. Student numbers by cohort

Cohort Entry	Number of students
A	27
B	30
C	30
D	36

Preclinical coursework data and clinical performance data were retrieved for four core areas of physiotherapy: Cardiorespiratory (CR), Orthopaedics (Ortho), Neurological (Neuro) and Musculoskeletal (MSK) Physiotherapy. Data were retrieved from electronic records and

archived hardcopies stored at the host university. Pre-clinical coursework data retrieved consisted of total subject marks for the coursework subjects undertaken in the four core areas. Marks were retrieved as a percentage out of 100. These six pre-clinical coursework subjects totalled 60 credit points, out of a required 240 credit points. Clinical performance was measured by the Assessment of Physiotherapy Practice (APP) instrument. The APP is well described in the literature (Dalton et al., 2009) and is a valid (Dalton et al., 2011) and reliable (Dalton et al., 2012) tool used to measure physiotherapy student clinical performance. A Level 2 recommendation for its use was given in a recent systematic review on the edumetric and psychometric properties of clinical performance tools (O'Connor et al., 2018). Clinical placements were embedded within the program and were completed directly after the relevant unit of study. Students completed five-week clinical placements in each of the core clinical areas: CR, Ortho, Neuro and MSK. Students were immersed in authentic clinical environments and practiced their profession under the supervision of a clinical educator, a practicing clinician from that workplace. The clinical educator completed the APP formatively at mid-unit and summatively at end-unit. Summative APP scores were retrieved for the core placements as a percentage out of a possible 100%. The CR and Ortho placements were undertaken in the second semester of study for the program, and the Neuro and MSK placements took place in the third semester of study.

3.6.1 Data Management and Statistical Analysis

Undergraduate GPA was retrieved by the University admissions office and provided to the research team in a non-identifiable form. UGPA scores were retrieved in four different formats: a score out of four (n = 60), a score out of seven (n = 52), a score out of 100 (n = 6), or a score out of another numeral (n = 3). These scores were converted to a score out of seven using methodology adopted by the Queensland Tertiary Admissions Centre (QTAC) to allow for a consistent comparison across all applicants.

Students' mean APP score across all placements completed was calculated to produce the Mean APP. Pre-clinical coursework performance was determined by calculating the mean subject mark (i.e., Mean Pre-Clinical Coursework) for the six pre-clinical coursework subjects completed across the four core areas of physiotherapy investigated. All data was made non-identifiable. Data was entered in the Statistical Package for the Social Sciences

(SPSS) software (Chicago, IL) Version 24 with significance set at $p < 0.05$. Student data were profiled using descriptive statistics.

Tests for normality were performed to identify the appropriate analyses to undertake. Pearson's correlations were undertaken unless the data was not normally distributed in which case relationships were explored using Spearman's rho correlations. Correlations were calculated between UGPA and the dependent variables: Mean Pre-Clinical Coursework, Mean APP, and APP scores from individual clinical placements. Linear regressions were then performed to explore the influence of UGPA on students' performance on each of the dependent variables. To explore if there were differences in performance between students who did or did not meet the various minimum GPA requirements for entry into a postgraduate physiotherapy program, students were grouped based on undergraduate GPA. Groups were chosen based on common minimum GPA requirements across the 11 Australian postgraduate programs. These groups were: < 4.5, 4.5 - 4.99, 5.00 - 5.49, 5.50 - 5.99, ≥ 6.0. A one-way ANOVA and Tukey post-hoc analysis (with testing of assumptions) was performed to determine if mean pre-clinical coursework and clinical performances were different between students according to their undergraduate GPA. Homogeneity of the variances was assessed using Levene's test. Alpha levels were set at 0.05 a priori.

3.7 Results

There were 123 students enrolled in their first year of study between 2010-2013 (male n = 65, female n = 58), the numbers in each cohort are displayed in Table 8. Two students withdrew in the early stages of the program for personal reasons and as they did not progress to clinical placement their data were excluded from analysis (male n = 1, female n = 1). Data from 121 students were analysed. Of these 118 (97.5%) graduated from the program.

Descriptive statistics are displayed in Tables 8 and 9. The frequency and distribution of UGPAs are displayed in Figure 9. The dependent variable of Mean Pre-Clinical Coursework was normally distributed. The distributions of all other dependant variables were skewed to the left with a dominance of scores at the higher end.
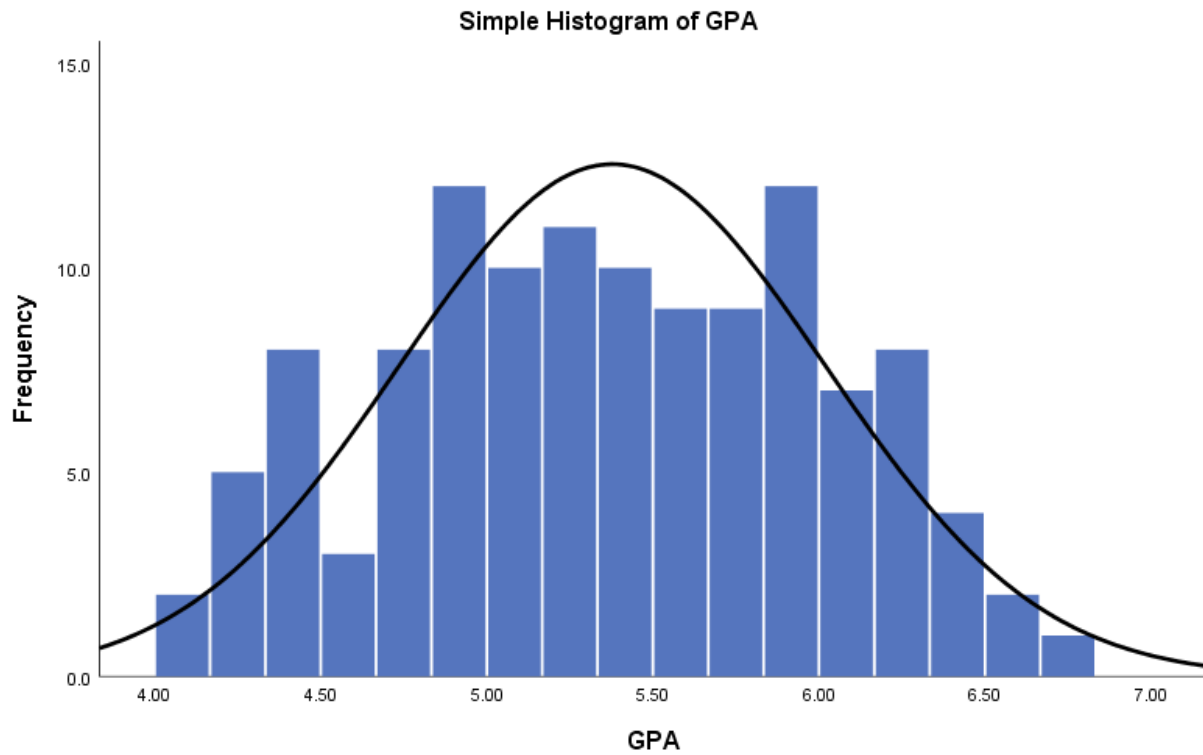
Figure 9. Frequency and distribution of undergraduate GPA scores

Table 8. Descriptive statistics of normally distributed data

Variable	N	Minimum	Maximum	Mean	Std Dev
Undergraduate GPA	121	4.12	6.81	5.38	0.64
Mean Pre-clinical Coursework	121	62.95	89.46	75.14	5.30

GPA – grade-point average
a Undergraduate grade-point average (GPA) is displayed as a score out of 7.
b All other variables displayed as percentages out of 100

Table 9. Descriptive statistics of data with a left-skewed distribution

Variable	N	Minimum	Maximum	Median	Interquartile Range
Mean APP	121	36.25	97.50	77.92	72.50 - 85.07
Orthopaedic APP	118	36.25	100.00	77.50	66.25 - 90.23
Neurological APP	116	48.75	100.00	85.00	70.31 - 93.75
Musculoskeletal APP	116	42.50	100.00	81.88	67.81 - 92.50
Cardiorespiratory APP	119	36.25	100.00	75.00	67.50 - 85.00

APP – Assessment of Physiotherapy Practice

[b] Variables displayed as percentages out of 100

There were no significant relationships identified between undergraduate GPA and measures of coursework or clinical performance taken in the first half of the participants program of study. These findings are displayed in Table 10.

Table 10. Correlations between undergraduate GPA and measures of clinical and academic performance

Independent Variable	N	Correlation	p-value
Mean Pre-clinical Coursework	121	0.14	0.13
Mean APP	121	0.15	0.09
Cardiorespiratory APP	119	0.02	0.86
Orthopaedic APP	118	0.17[‡]	0.07
Neurological APP	116	0.11[‡]	0.25
Musculoskeletal APP	116	0.09 [‡]	0.36

APP – Assessment of Physiotherapy Practice

[‡] Spearman's rho used to calculate correlation

Linear regressions established that undergraduate GPA did not significantly predict students' mean pre-clinical coursework performance, $F(1, 119) = 2.34$, $R^2adj = 0.011$, $p = 0.13$, or mean clinical performance, $F(1, 119) = 1.62$, $R^2adj = 0.005$, $p = 0.21$. Undergraduate GPA also did not predict student performance on clinical placement in the

core areas of: cardiorespiratory, $F_{(1, 117)} = 0.001$, R2adj = -0.009, p = 0.98; orthopaedics, $F_{(1, 116)} = 3.56$, R2adj = 0.021, p = 0.06; neurological, $F_{(1, 114)} = 2.38$, R2adj = 0.012, p = 0.13; or musculoskeletal physiotherapy, $F_{(1, 114)} = 1.05$, R2adj = < 0.001, p = 0.31.

When students were grouped by GPA the number of students in each group were as follows: < 4.5, n = 12 (9.9%); 4.50 - 4.99, n = 20 (16.5%); 5.00 - 5.49, n = 34 (28.1%); 5.50 - 5.99, n = 30 (24.8%); and ≥ 6.0, n = 25 (20.7%). A one-way ANOVA determined there was no significant differences between the mean pre-clinical coursework marks or clinical performance scores of students when classified into five groups based on their undergraduate GPA. These findings are displayed in Figures 10 - 15. Testing of assumptions identified that there was homogeneity of variances and outliers were retained. The data were normally distributed in all groups except for the Neuro APP and MSK APP variables. For these two variables it was determined that although the ANOVA was an acceptable test to utilise, an Independent Samples Kruskal-Wallis Test may also be appropriate. As such the analysis was repeated using this Independent Samples Kruskal-Wallis Test which confirmed that there were no significant differences between the distribution of students' performance on Neuro or MSK placements when classified into groups based on their undergraduate GPA.
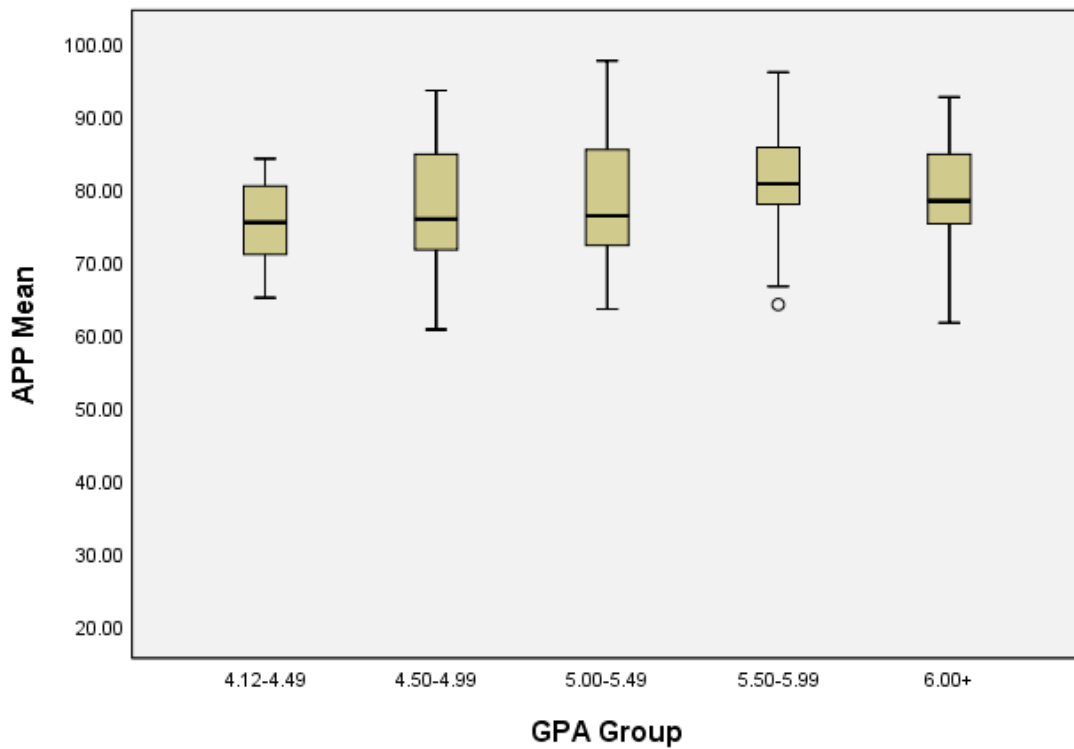
o represents outliers

Figure 10. Boxplot of comparison of students' Mean APP scores when classified by undergraduate GPA
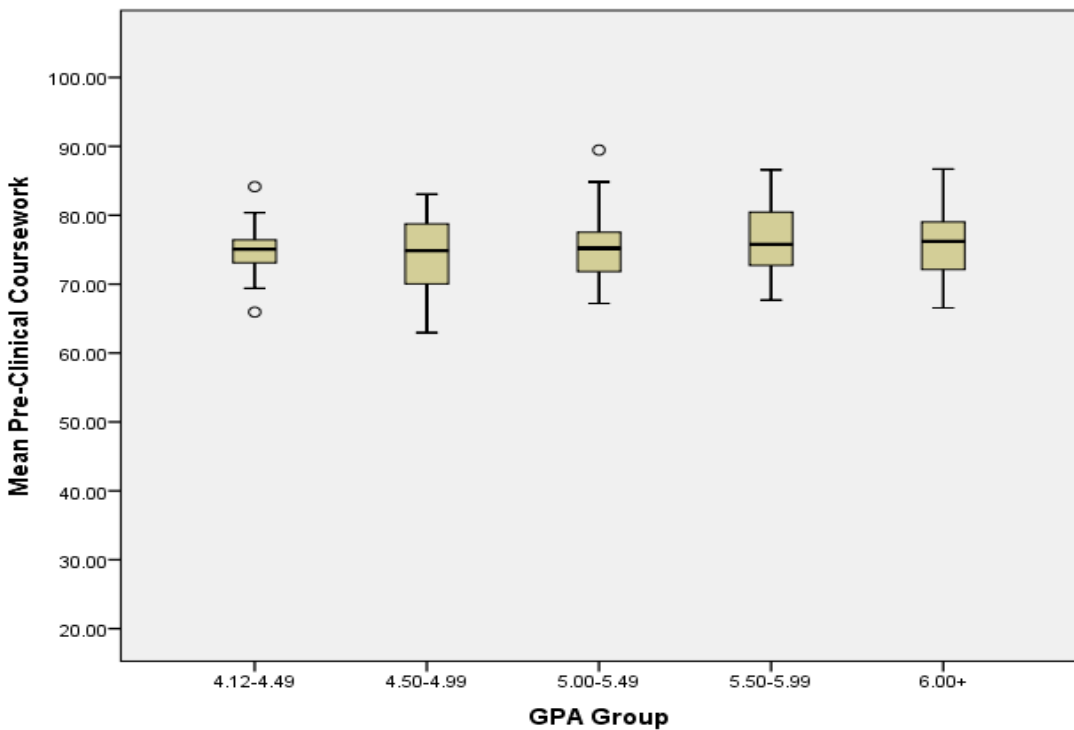
o represents outliers

Figure 11. Boxplot of comparison of students' Mean Pre-Clinical Coursework scores when classified by undergraduate GPA
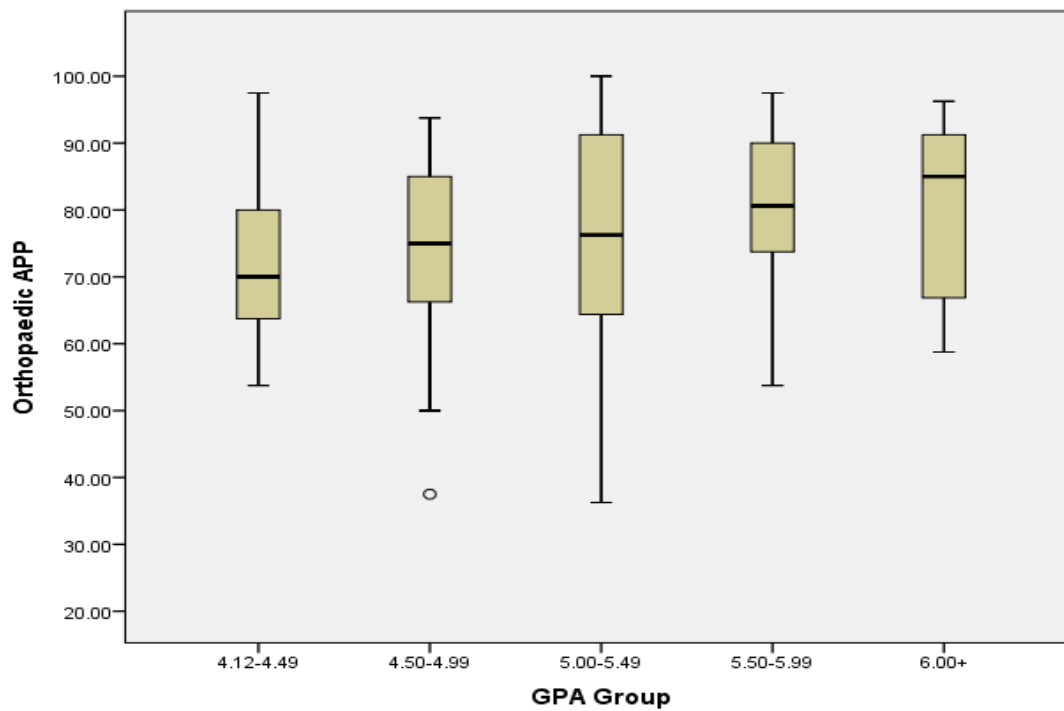
^O represents outliers

Figure 12. Boxplot of comparison of students' Orthopaedic APP scores when classified by undergraduate GPA (first or second placement)
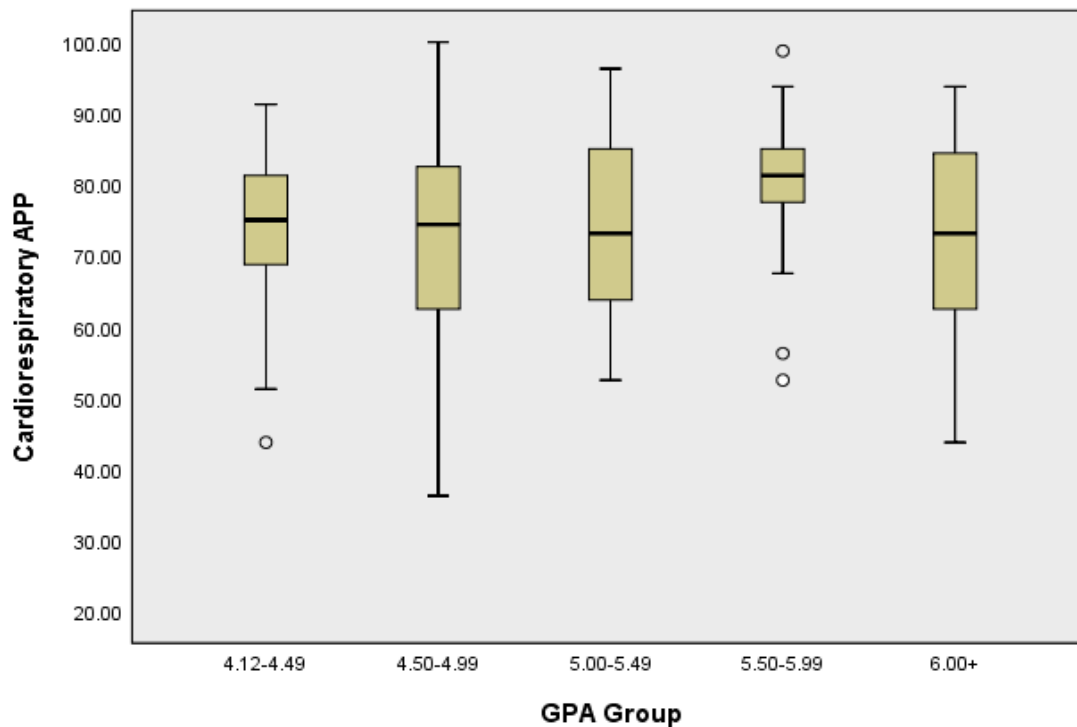
^O represents outliers

Figure 13. Boxplot of comparison of students' Cardiorespiratory APP scores when classified by undergraduate GPA (first or second placement)
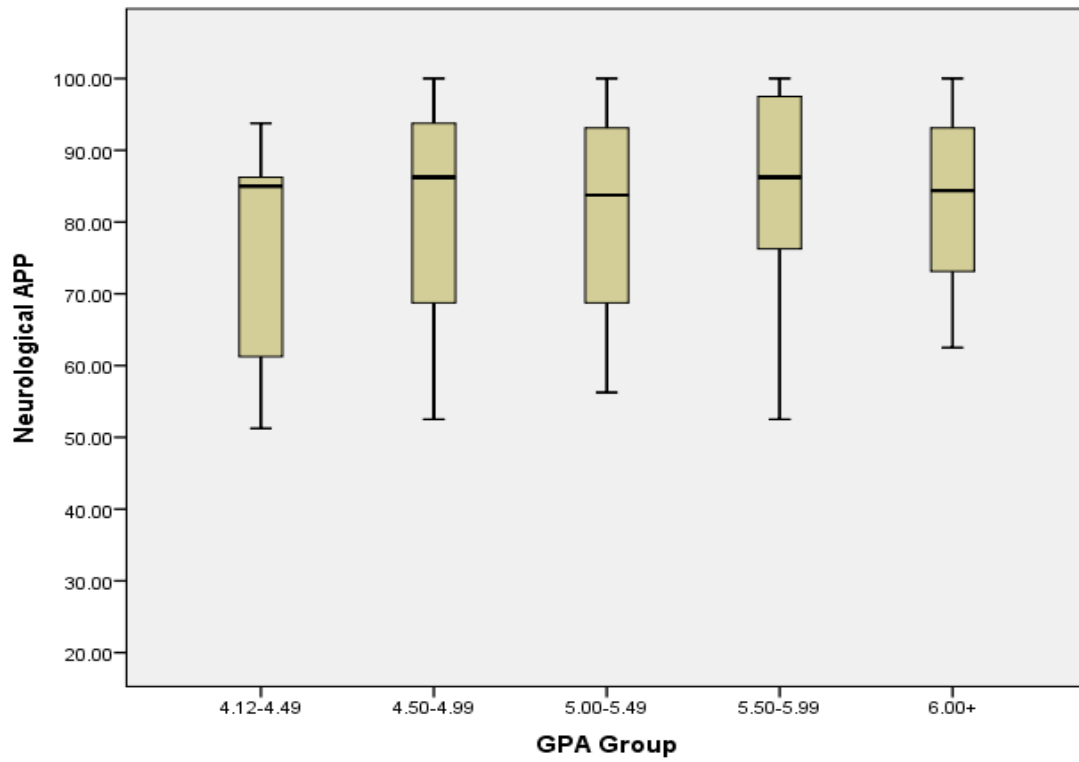
Figure 14. Boxplot of comparison of students' Neurological APP scores when classified by undergraduate GPA (third or fourth placement)
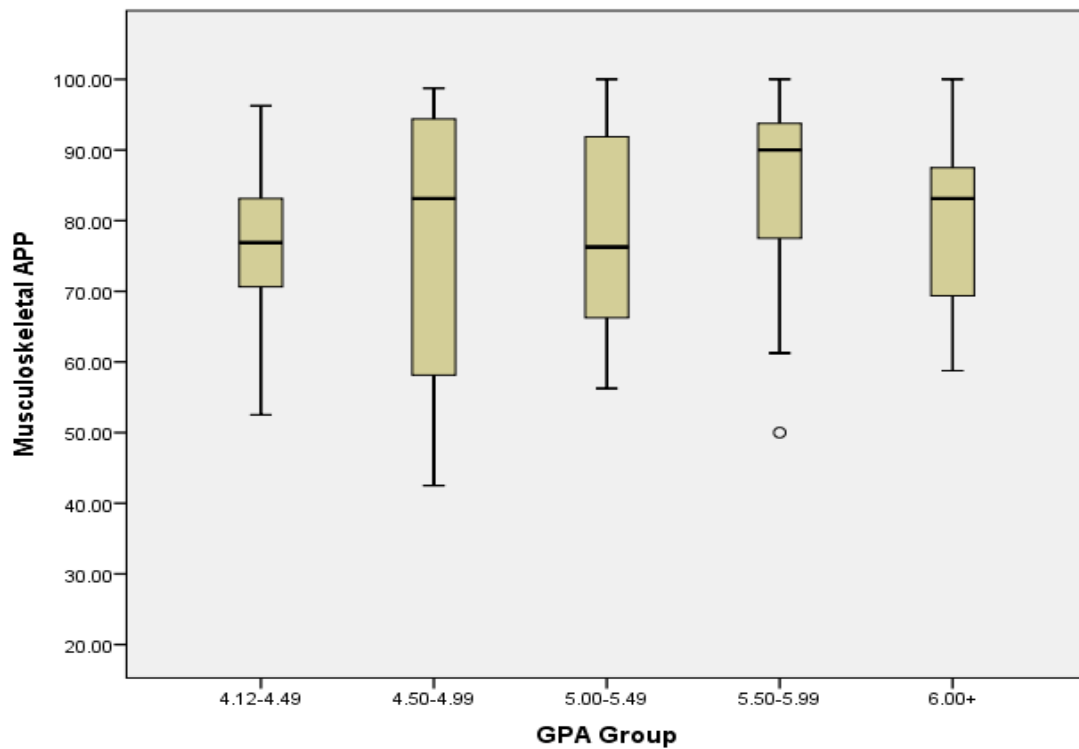
O represents outliers

Figure 15. Boxplot of comparison of students' Musculoskeletal APP scores when classified by undergraduate GPA (third or fourth placement)

3.8 Discussion

The aims of this study were to explore the relationship between undergraduate academic performance as measured by UGPA and the clinical and academic performance of physiotherapy students, and to determine if there were differences in performance between groups of students whose UGPA may be higher or lower than common GPA cut-offs for entry into postgraduate physiotherapy programs. The authors did not investigate relationships between other selection measures utilised by the host institution (personal statements and semi-structured interview). Due the range of UGPA scores within the study cohort (4.12 - 6.81 out of 7) and the normal distribution of these scores, relationships existing between UGPA and the dependent variables should be evident even in the presence of other selection measures used to admit students into the host program.

This study found no significant relationships between students' UGPA and their early academic or clinical performance in a postgraduate physiotherapy program. These findings are surprising given that the ability of entry GPA to predict the future performance of medical students is well accepted within the medical profession. Much of the literature supporting that acceptance is conducted within undergraduate medical programs and the GPA utilised are school-leaving scores. However, contemporary literature also reports significant relationships between UGPA and postgraduate academic and/or clinical performance in entry-level programs not only in the medical (Dixon, 2012; Sladek, Bond, Frost, & Prior, 2016) but also the speech and language (Baggs, Barnett, & McCullough, 2015) professions. In these studies, (Baggs et al., 2015; Sladek et al., 2016) clinical performance was not measured by a tool with demonstrated validity or reliability, or that is widely used amongst the medical or speech and language professions. The present study used the APP to measure clinical performance, a valid and reliable instrument (Dalton et al., 2011; Dalton et al., 2012) which is consistently used across Australian physiotherapy education programs. The APP is administered by clinical educators who are usually not formally affiliated with the education provider. Clinical educators are assessing student performance against a consistent benchmark standard – that of a new graduate on their first day employment (Dalton et al., 2009). This standard is set by the Australian physiotherapy profession, and so the finding that UGPA is not related to students' performance in clinical practice may be generalisable to programs beyond the host institution. However, to confirm this supposition

further research using larger sample sizes and across multiple institutions would be of benefit.

At the time of the present study, seven Australian postgraduate entry-level programs specified minimum UGPA requirements to be eligible for admission, which ranged from 4.5 - 5.5. This study found there was no significant difference in mean pre-clinical coursework or clinical performance scores between students when grouped by their UGPA. Given that there was no difference in average clinical performance even between the lowest and highest UGPA groups, this suggests that using UGPA may not be an appropriate method to rank applicants to determine entry into competitive programs. It also suggests that minimum GPA requirements into postgraduate programs could be reviewed and potentially revised. It is worthy of note that 97.5% of students within this study successfully graduated from the program, despite 54.5% having an UGPA lower than 5.5. A GPA of 5.5 has been highlighted as this is the highest minimum GPA entry requirement for the postgraduate physiotherapy programs currently on offer within Australia. However, three programs require a 'competitive GPA' suggesting that students with lower UGPAs may not be successful at gaining entry into these programs. Competitive or higher minimum UGPA requirements may prevent applicants with the personal and academic attributes to succeed within postgraduate physiotherapy education from being eligible for admission. There is some evidence that traditional measures of achievement discriminate against defined groups (James, Yates, & Nicholson, 2010), so a competitive or an unnecessarily high GPA requirement may bias against suitable applicants from minority or low socio-economic backgrounds from accessing postgraduate physiotherapy education.

Once academic pre-requisites have been met, UGPA may not be appropriate to be used as the sole admission criterion to select students into postgraduate physiotherapy programs. Other measurement tools should therefore be considered for use alongside UGPA. As UGPA is a measure of academic achievement, additional measures utilised should aim to measure other attributes that will be required by the student along their physiotherapy education and training pathway (Prideaux et al., 2011).

Admission measures currently used by postgraduate physiotherapy programs in Australia other than UGPA include semi-structured interviews, Multiple Mini Interviews (MMIs) and personal statements. The evidence supporting the validity and reliability of semi-structured

admission interviews is not robust, however there is some evidence to suggest that admission interview scores are related to students future performance in undergraduate physiotherapy (Edgar et al., 2014; Watson et al., 2000) and medical schools (Sladek et al., 2016). A 2013 systematic review on the use of MMIs for student selection in health profession training concluded that the MMI is reliable, acceptable and feasible (Pau et al., 2013), suggesting the MMI could be considered by programs to determine its appropriateness for implementation in view of their applicant demographic. There is limited evidence for the predictive validity of letters of reference or personal statements for the future performance of health professional students (Patterson et al., 2016; Siu & Reiter, 2009). The Ottawa 2010 Conference consensus statement on assessment and selection for health care students identified emotional intelligence and personality testing as interesting areas for further development (Prideaux et al., 2011). A 2016 systematic review (Patterson et al., 2016) into the effectiveness of selection methods within medical education identified positive evidence for the predictive validity of personality and emotional intelligence tests, however cautioned against their use without considering the impact on the diversity of the medical profession. The same review also identified that there was a good level consensus for the use of situational judgment tests as an acceptable criterion for selection into medical school. The physiotherapy profession would benefit from consideration and investigation into the use of additional selection criterion such as MMI's, situational judgement tests, personality and emotional intelligence testing.

The present study is limited to the investigation of a single measure utilising a sample from a single institution. There are always challenges in generalising the findings of research conducted on a study population from a single institution. Certainly, there is likely to be variation in methods and timing of the teaching and assessment of core curricula between postgraduate physiotherapy programs, which may limit the generalisability of the finding that UGPA was not related to students' early coursework performance. There may also be variations in the amount of support available to students while they are in the clinical environment. The host program has in place a comprehensive clinical support structure. Early support and intervention is offered to students identified as being at risk either before they enter into the clinical environment, or as the placement progresses. This additional support offered by university academics is designed to enhance student performance in clinical practice and therefore may contribute to an increase in APP scores at the end of the placement.

3.9 Implications for Practice

The findings of this study suggest that UGPA, when used as one component of selection, is not related to students' future performance. Therefore, UGPA may not be the most appropriate method of selecting students into entry-level postgraduate physiotherapy education programs. Further research is needed in this area utilising larger sample sizes across multiple institutions to confirm these findings and ensure they are applicable to the wider physiotherapy program outside the host institution.

Students who performed strongly at an undergraduate level did not perform better, either academically or clinically, when compared to their peers with lower UGPAs. This may have implications for planned staffing and support structures within postgraduate physiotherapy programs. A cohort with strong prior academic performance on admission may still contain students needing additional support and enhancement throughout the program. This study suggests that support staffing should not be planned based on a cohorts' perceived academic strength. It also suggests that students should not be categorised either into academic streams, or judgements made about their expected clinical performance based on their UGPA.

The finding that there was no difference in either clinical or academic performance between students with UGPAs ranging from 4.12 - 6.81 indicates that there is a need to reflect on the appropriateness of minimum UGPA requirements. Individual institutions should consider the purpose served by a minimum UGPA requirement. UGPA did not predict success in a postgraduate physiotherapy program, suggesting that minimum GPA requirements could be reviewed and potentially lowered while still retaining the quality of graduate desired by the physiotherapy profession. Limitations on cohort size may require methods of controlling admissions into a program. Minimum UGPA is one method of limiting those who are eligible to enter into a program, however this may result in students who are equipped to successfully complete the program being ineligible to enter. This could feasibly impact on the diversity of the student cohort and the future physiotherapy profession.

The host institution will continue to utilise additional selection measures to admit students into their program. It is clear that there is a strong need for further research into admission and selection measures for postgraduate physiotherapy programs, investigating not only the

measures currently used, but also emerging areas of interest such as personality and emotional intelligence testing. Entry into postgraduate physiotherapy programs is competitive and high-stakes for all involved, so it is essential that the criteria used to select-in and rank applicants for admission are valid measures of the attributes that are both required and desired to produce the highest quality health professionals.

3.10 Conclusion

In Australian postgraduate entry-level physiotherapy education, there is a trend toward the use of UGPA as the sole selection criterion for program admission. In this study population UGPA, when used as one component of selection, was not related to students' early academic or clinical performance in core areas of practice. These findings indicate a need to reconsider the use of UGPA as the sole selection criterion and support the inclusion of other criteria to select students into competitive programs. There were no significant differences in the clinical or coursework performance of students when grouped according to their UGPA. This suggests that if minimum GPA requirements for entry into postgraduate physiotherapy programs are to be retained, they should be reviewed to ensure all suitable applicants with the personal and academic attributes for success within the physiotherapy profession meet program eligibility criteria.

3.11 Supplementary Material

3.11.1 Postgraduate Entry-level Physiotherapy Programs Offered in April 2018

Listed below are the official university websites accessed on 10th April 2018 to determine the entry requirements for postgraduate entry-level physiotherapy programs currently offered within Australia:

1. Bond University Doctor of Physiotherapy. https://bond.edu.au/program/doctor-physiotherapy#entry_requirements
2. Curtin University Masters of Physiotherapy. http://courses.curtin.edu.au/course_overview/postgraduate/Master-Physiotherapy

3. Flinders University Masters of Physiotherapy. http://www.flinders.edu.au/courses/rules/postgrad/mpt.cfm

4. Griffith University Masters of Physiotherapy. https://degrees.griffith.edu.au/Program/5320/HowToApply/Domestic#can-i-apply

5. La Trobe University Masters of Physiotherapy. https://www.latrobe.edu.au/courses/master-of-physiotherapy-practice

6. Macquarie University Doctor of Physiotherapy. https://courses.mq.edu.au/2018/domestic/postgraduate/doctor-of-physiotherapy/entry-requirements#content

7. The University of Melbourne Doctor of Physiotherapy. http://mdhs-study.unimelb.edu.au/degrees/doctor-of-physiotherapy/entry-requirements

8. The University of Queensland Masters of Physiotherapy. https://future-students.uq.edu.au/study/program/Master-of-Physiotherapy-Studies-5267

9. The University of Sydney Masters of Physiotherapy. https://sydney.edu.au/courses/courses/pc/master-of-physiotherapy.html

10. University of Canberra Masters of Physiotherapy. http://www.canberra.edu.au/coursesandunits/course?course_cd = 768AA

11. UTS Masters of Physiotherapy. https://www.uts.edu.au/future-students/find-a-course/master-physiotherapy

Chapter 4. Relationships between early pre-clinical summative assessment scores and the clinical performance of physiotherapy students

The following manuscript has been published in *Journal of Allied Health*.

Terry, R., Hing, W., Orr, R., & Milne, N. (2020). Relationships between pre-clinical summative assessment scores and the clinical performance of physiotherapy students. *Journal of Allied Health*, *49*(1), 13E - 19E. Reproduced with permission from Association of Schools of Allied Health Professionals.

The formatting of this chapter is in accordance with the guidelines of *Journal of Allied Health*.

4.1 Prelude

Chapter 1 discussed systematic reviews by Hamdy et al. (2006) and Harfmann and Zirwas (2011) which reported that OSCE scores and pre-clinical GPA were predictor variables for students' future clinical performance in medical school and as junior doctors. The systematic review reported in Chapter 2 identified that while there was some evidence of a significant positive relationship between pre-clinical summative assessments and the clinical performance of health profession students, there was a lack of evidence to support such relationships within the physiotherapy profession. Only the one paper investigated a physiotherapy student cohort, and this found no relationship between their performance on an OSCE and their future clinical practice (Wessel et al., 2003).

This chapter describes a study on the relationships between the pre-clinical summative assessments completed by postgraduate entry-level physiotherapy students and their future clinical performance. It investigates summative assessments commonly used within physiotherapy education programs, OSCEs, written examinations and oral presentations.

4.2 Abstract

Aim: Education providers need to ensure that students allocated to a clinical placement are optimised for success. The aim of this study was to determine the relationships between physiotherapy students' summative assessment scores in pre-clinical coursework and their future performance in clinical practice.

Methods: Selected as potential subjects were 123 students from four consecutive intakes (2010 - 2013) of an Australian entry-level Doctor of Physiotherapy program. Retrospective cohort summative assessment data for pre-clinical (Objective Structured Clinical Examinations (OSCEs), written examinations, and seminar presentations) and clinical practice (clinical practice scores) subjects in core areas of physiotherapy were retrieved. Clinical practice performance was assessed using the reliable and validated Assessment of Physiotherapy Practice instrument. A descriptive analysis, Pearson's correlations and multiple regressions were performed between mean pre-clinical and clinical performance scores.

Results: Assessment data from 118 students were analysed. Pre-clinical assessment scores were positively related to clinical performance (OSCE $r = 0.57$, $p < 0.001$; written examination $r = 0.39$, $p < 0.001$; seminar presentations $r = 0.29$, $p = 0.012$). A multiple regression model identified OSCE as an independent contributor to clinical performance scores (adjusted $R^2 = 0.33$, $p < 0.001$).

Discussion and Conclusion: OSCE scores were strongly related to clinical performance and explained 32% of physiotherapy students' future clinical performance. Pre-clinical OSCE scores could provide opportunity to implement proactive support and enhancement strategies to increase stakeholder satisfaction and maintain quality placement experiences.

4.3 Introduction

The value of clinical education in the allied health professions is well-recognised. Accreditation standards for registrable professions like physiotherapy require entry-level education programs to include clinical education components undertaken in authentic clinical environments (Australian Physiotherapy Council, 2020; Health & Care Professions Council, 2017). There are increasing challenges for education providers to source and maintain good quality clinical placements including fiscal restraints, reductions in health sector staffing and changing staff profile, increasing student cohort numbers and a proliferation of new programs (Rodger et al., 2008). While collaboration and innovation must occur to widen the placement opportunities available, the need to nurture and enhance existing quality clinical placements should not be overlooked. Finding opportunities to optimise student preparedness for clinical placements could provide a means of maximising students' ability to engage with the learning experience and provide quality client care - minimising stress to clinical educators and the need for repeat placements. The ability to identify students likely to benefit from additional preparation prior to entering clinical placement, or require additional support during placement, is an opportunity to proactively implement strategies to promote a successful and satisfying clinical practice experience for all stakeholders.

There are several points during a student's education that can be explored as potential predictors of clinical performance. Studies have reported on the ability of admission criteria (Edgar et al., 2014; Howard & Jerosch-Herold, 2000; Morris & Farmer, 1999; Nadasan & Puckree, 2003; Thieman et al., 2003) and overall academic performance (Luedtke-Hoffmann, Dillon, Utsey, & Tomaka, 2012; Vendrely, 2007) to predict the clinical performance of allied health students. However, these measures take place either before or after the education program has been completed. This limits their utility as they do not reflect the factors that may influence a student's performance at varying points throughout the program. Summative assessments embedded within discreet coursework subjects are direct measures of the knowledge, skills and behaviours that must be implemented in a subsequent clinical practice experience. However, there may be students that meet academic progression requirements overall but would benefit from further upskilling and enhancement prior to commencing clinical placement.

Commonly used assessment methods in health profession programs include written examinations and Objective Structured Clinical Examinations (OSCEs). There is evidence within the medical profession that OSCEs (Baker et al., 2006; Berg et al., 2008; Campos-Outcalt et al., 1999; Carr et al., 2014; Cope et al., 2007; Dong et al., 2014; Han & Chung, 2016; LaRochelle et al., 2015; Wilkinson & Frampton, 2004) and written examinations of multiple choice or extended matching questions (Carr et al., 2014; Wilkinson & Frampton, 2004) are significantly correlated with clinical performance. OSCE scores were reported to explain 1.9% - 39.7% of the variability in medical students' clinical performance (Baker et al., 2006; Berg et al., 2008; Campos-Outcalt et al., 1999; Carr et al., 2014; Cope et al., 2007; Dong et al., 2014; Han & Chung, 2016; LaRochelle et al., 2015; Wilkinson & Frampton, 2004). The variety of clinical performance measures used is one factor that may explain the large differences in strengths of relationships identified as nine studies (Baker et al., 2006; Berg et al., 2008; Campos-Outcalt et al., 1999; Carr et al., 2014; Cope et al., 2007; Dong et al., 2014; Han & Chung, 2016; LaRochelle et al., 2015; Wilkinson & Frampton, 2004) used eight different methods of assessing clinical performance.

There are many similarities between education in the medical and allied health professions. However, allied health students in Australia are expected to have the ability to be independent, autonomous practitioners at the point of graduation. This is very different to the support structure in place for new graduate medical doctors. It is likely then that the expectation of students' clinical performance between the professions will also be different, which limits the generalisability of relationships identified in medical students.

Literature investigating relationships between specific coursework summative assessments and students' future clinical performance in physiotherapy is limited, as it is within the wider allied health professions (Terry et al., 2017). One study of a Canadian physiotherapy program found no significant correlation between OSCE scores and clinical performance as measured by the Clinical Performance Instrument (Wessel et al., 2003). In contrast to this finding, single studies among dentistry (Graham et al., 2013), dietetic (Hawker et al., 2010) and pharmacy (McLaughlin et al., 2015) students have identified significant positive relationships between OSCE scores and clinical performance. The limited and conflicting research conducted within the physiotherapy and wider allied health professions is indicative of the need for further research in this area. The aim of the present study was to determine if relationships existed between specific coursework

summative assessments used in pre-clinical coursework and the future clinical performance of physiotherapy students, employing a well-described measure of clinical performance that is valid, reliable and widely-used.

4.4 Material and Methods

This was a retrospective cohort study. Participants were students from four consecutive cohorts (2010 - 2013) of an entry-level Australian Doctor of Physiotherapy program. Participants were excluded if they had not completed the first year of the program, including all required core coursework subjects and corresponding clinical placements. Assessment data were retrieved from electronic records and archived hardcopies. Participants' data were analysed if there was complete assessment data for a core area of physiotherapy, consisting of all summative assessment scores in the pre-clinical coursework subject immediately preceding the corresponding clinical placement, and the placement's clinical performance score.

Participants completed discreet pre-clinical coursework subjects and a corresponding clinical placement in four core clinical areas of physiotherapy: Cardiorespiratory (CR), Orthopaedics (Ortho), Musculoskeletal Outpatients (MSK), and Neurological (Neuro) physiotherapy. Coursework covered in preparation for clinical placements included theory and practical skills related to the assessment and treatment of the following: in CR, conditions involving the respiratory system, cardiac system, the acutely unwell patient and the post-surgical patient; in Ortho, acute musculoskeletal injuries, orthopaedic surgery, and rheumatoid and osteoarthritis; in MSK, chronic musculoskeletal conditions, the spinal column and deeper exploration of the management of acute musculoskeletal conditions; and in Neuro acute and chronic neurological, and age-related conditions.

Pre-clinical coursework summative assessments investigated were the OSCE, written examinations and seminar presentations (Table 11). The position of the assessment items within the program are described in Table 12. Total subject marks were also investigated.

Table 11. Pre-clinical summative assessments completed

Core Area	Written Examinations	OSCEs	Seminar Presentations
Cardiorespiratory	1 x 90min SAQ	2 x 15min stations	1 x 15min presentation
Orthopaedics	1 x 120min MCQ/SAQ	2 x 15min stations	None
Neurological	1 x 120min MCQ/SAQ	3 x 15min stations and 1 x 30min station	None
Musculoskeletal	1 x 120min SAQ	3 x 15min stations	None

MCQ – Multiple Choice Questions, OSCE – Objective Structured Clinical Examination, SAQ – Short Answer Question

The written assessments consisted either entirely of short answer questions (SAQs), or a combination of SAQs and multiple-choice questions (MCQs). The written examinations aimed to assess students' theoretical knowledge of relevant anatomy, physiology, pathology, clinical reasoning, and ability to identify appropriate physiotherapy assessments and interventions. Written assessments accounted for 40 - 50% of students' total subject mark.

Table 12. Representation of the first three semesters of the host Doctor of Physiotherapy program by semester

Semester 1		Semester 2		Semester 3	
Principles of Physiotherapy	CR PT I	CR PT II OSCE + Written + Seminar	CR Clinical Placement APP		
	MSK PT I	MSK PT II (Ortho) OSCE + Written	Ortho Clinical Placement APP	MSK PT III (Outpatients) OSCE + Written	MSK Clinical Placement APP
				Neuro PT OCSE + Written	Neuro Clinical Placement APP

APP – Assessment of Physiotherapy Practice; CR – Cardiorespiratory; MSK – Musculoskeletal; Neuro – Neurological; OSCE – Objective Structured Clinical Examination; Ortho – Orthopaedic; PT – Physiotherapy; Seminar – Seminar Presentation; Written – Written Examination

OSCEs were conducted either one or two days after the written assessments and all were delivered in the same conditions. Examiners were licensed physiotherapists, used detailed mark sheets, and were briefed on the applications of these prior to the examination. Students completed 2 - 3 x 15min stations per OSCE. One OSCE also had an additional 30min station at the commencement of the examination where students were required to watch and analyse video material. Each station was structured around a clinical case study and may have had marks allocated for professional conduct, communication, clinical reasoning, technique performance and safety. OSCEs accounted for 45 - 50% of students' total subject mark.

Seminar presentations were utilised in one core coursework subject (CR) and required students to deliver a 10 - 15-minute seminar on a relevant disease or pathology. Seminars were independently marked by two assessors against a standardised rubric with criteria including quality of communication, organisation of material, content and knowledge of subject matter, use of evidence-based information and effectiveness of delivery. The mean score of the two assessors constituted the students' allocated mark. The seminar presentations constituted 15% of the students' overall CR subject mark.

Students who met the passing standard in pre-clinical coursework progressed into the corresponding clinical placement. Participants completed a maximum of four clinical placements, one in each of CR, Ortho, MSK and Neuro. Each clinical placement was embedded in the program and occurred within the same semester of study in which the relevant pre-clinical coursework was completed. Students were allocated to clinical placements where they spent five weeks, full-time, immersed in that clinical environment practicing under the supervision of a clinical educator. Clinical educator details were not known at the time of allocation and clinical educators were not informed of students results in pre-clinical coursework, other than they had met the required minimum standard.

Clinical performance was measured by the Assessment of Physiotherapy Practice (APP) instrument. The APP is commonly used by physiotherapy education programs to assess the clinical performance of students across Australia and New Zealand. The APP is a valid (Dalton et al., 2011) and reliable (Dalton et al., 2012) measure of clinical performance and is well described in the literature (Dalton et al., 2009). A systematic review by O'Connor et al. (2018) synthesised evidence relating to the edumetric and psychometric properties of

clinical performance tools used by the physiotherapy profession and awarded the APP level A evidence for validity, and level B evidence for reliability and edumetric evidence. The APP measures student performance across seven domains of practice and totals 20 items. The domains of practice examined by the APP are professionalism, communication, assessment, analysis and planning, intervention, evidence-based practice, and risk management. Items are accompanied by performance indicators, a list of observable behaviours assessors can refer to, to assist in determining the standard to which an item is being demonstrated. The APP was completed by the clinical educator formatively at the mid-point of the clinical placement and summatively at the end of the five weeks of placement. Final APP total scores account for 70% of students total core clinical practice subject marks. For this study, final APP total scores, and scores for each of the seven domains of practice were retrieved.

4.4.1 Data Management and Statistical Analysis

Retrieved raw assessment data were converted to a percentage out of 100, as is done routinely by the university to calculate weighted marks. Data were made non-identifiable. Students' mean score for each assessment type as well as total subject score were calculated and entered in the Statistical Package for the Social Sciences (SPSS) software (Chicago, IL) version 24 with significance set at $p < 0.05$. Student data were profiled using descriptive statistics. Tests for normality were performed to identify the appropriate analyses to undertake. Pearson's correlations were calculated between independent variable mean scores (OSCE, written examination, seminar presentation and total subject mark) and APP mean scores utilising a Bonferroni adjusted p-value of 0.0125 (0.05/4) to determine significance. Correlations were calculated between the independent variables and specific domains of the APP also utilising a Bonferroni adjustment. Strengths of correlations were based on the rating scale of Cohen (1988) and are displayed in Table 13. A multiple regression analysis was performed to determine the effect of each independent assessment item on clinical performance scores.

Table 13. Strength of associations by Cohen (1988)

Coefficient value	Strength of association
0.1 < r < 0.3	weak correlation
0.3 < r < 0.5	moderate correlation
r > 0.5	strong correlation

4.5 Results

The 2010 - 2013 cohorts had a total of 123 students enrolled (male n = 65, female n = 58). Five students did not complete the first year of the program. Data from 118 students met the inclusion criteria and their retrieved assessment data were analysed (male n = 62, female n = 56).

The mean scores for each assessment item are provided in Table 14. Students performed best on the seminar presentation with a mean score of 90.54 (± 6.24)%. Written examinations had the lowest mean score of 71.88 (± 6.48)%. The mean APP scores of raw assessment data before conversion to a score out of 100 are shown in Figure 16.
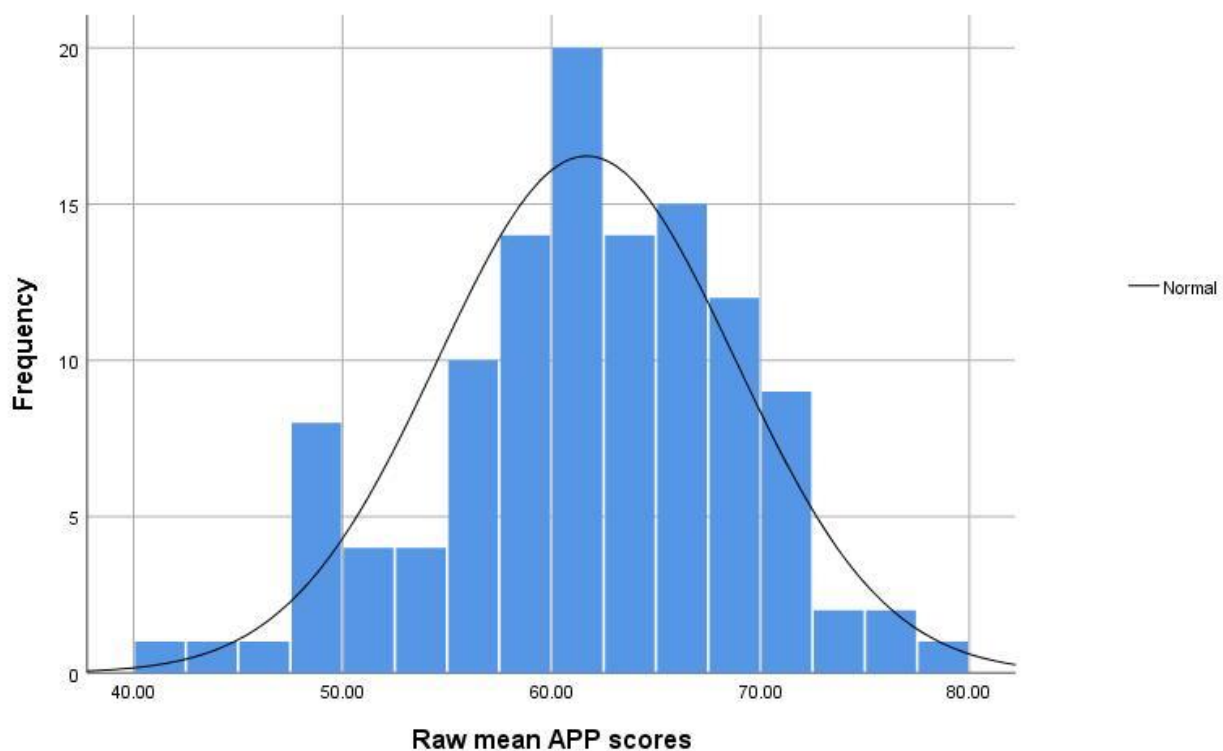
Figure 16. Distribution of raw mean APP scores

The mean scores of each coursework assessment item were significantly correlated with mean clinical performance scores as reported in Table 14. The OSCE (r = 0.57, p < 0.001) and total subject mark (r = 0.55, p < 0.001) had the strongest significant correlations. Written assessment and the seminar presentation demonstrated significant correlations of moderate and weak strength respectively.

Table 14. Descriptive statistics of mean assessment scores across core areas

Assessment	n	Minimum %	Max %	Mean %	Std Dev
APP	118	60.63	97.50	78.67	8.41
OSCE	118	63.12	93.01	76.91	5.81
Written Examination	118	52.09	86.04	71.88	6.48
Seminar Presentation	118	64.00	100.00	90.54	6.24
Total Subject Mark	118	60.35	89.11	74.93	5.36

APP – Assessment of Physiotherapy Practice, OSCE – Objective Structured Clinical Examination

Table 15. Relationships between mean pre-clinical summative assessment scores and mean APP scores across core areas

Coursework Assessment	r	r^2	p
OSCE	0.57	0.32	< .001*
Written Examination	0.39	0.15	< .001*
Seminar Presentation	0.29	0.08	.012*
Total Subject Mark	0.55	0.30	< .001*

Significance is set at p < 0.0125

APP – Assessment of Physiotherapy Practice, r - Pearson's correlation coefficient, r^2 – variance

Correlations between coursework summative assessments and specific domains of the APP are reported in Table 16. The OSCE demonstrated significant moderate to strong correlations with all domains of the APP. Correlations ranged from r = 0.36 (Professionalism) to r = 0.62 (Intervention). The written examination showed significant weak or moderate correlations with all domains of the APP (r = 0.29 - 0.42), except

110

professionalism and risk management. The seminar assessment demonstrated significant weak correlations with the communication and intervention domains (both r = 0.23). Variances are displayed in decimal form in Table 17.

Table 16. Correlations between mean pre-clinical summative assessment scores and mean performance on APP domains across core areas

APP Domains	OSCE			Written Examination			Seminar Presentation			Total Subject Mark		
	r	r^2	p	r	r^2	p	r	r^2	p	r	r^2	p
Professionalism	0.36	0.13	<0.001*	0.16	0.03	0.09	0.11	0.01	0.25	0.31	0.10	0.001*
Communication	0.45	0.20	<0.001*	0.34	0.12	<0.001*	0.23	0.05	0.01*	0.45	0.21	<0.001*
Assessment	0.56	0.31	0.001*	0.40	0.16	0.001*	0.19	0.04	0.04	0.54	0.29	0.001*
Analysis & Planning	0.57	0.32	<0.001*	0.42	0.18	0.001*	0.20	0.04	0.03	0.56	0.31	0.001*
Intervention	0.62	0.38	0.001*	0.42	0.18	<0.001*	0.23	0.05	0.01*	0.59	0.35	0.001*
EBP	0.42	0.18	0.001*	0.29	0.08	0.001*	0.15	0.02	0.10	0.41	0.17	<0.001*
Risk Management	0.45	0.20	0.001*	0.20	0.04	0.03	0.20	0.04	0.03	0.37	0.14	<0.001*

*Significant finding of p <0.0125

APP – Assessment of Physiotherapy Practice; EBP - Evidenced based practice; r - Pearson's correlation coefficient

Table 17. Variances of significant relationships between mean pre-clinical summative assessment scores and mean performance on APP domains across core areas

APP Domains	OSCE		Written		Seminar		Total Subject Mark	
	r^2	p	r^2	p	r^2	p	r^2	p
Professionalism	0.13	<0.001	0.03	0.09	0.01	0.25	0.10	<0.001
Communication	0.20	<0.001	0.12	<0.001	0.05	0.01	0.21	<0.001
Assessment	0.31	<0.001	0.16	<0.001	0.04	0.04	0.29	<0.001
Analysis & Planning	0.32	<0.001	0.18	<0.001	0.04	0.03	0.31	<0.001
Intervention	0.38	<0.001	0.18	<0.001	0.05	0.01	0.35	<0.001
EBP	0.18	<0.001	0.08	<0.001	0.02	0.10	0.17	<0.001
Risk Management	0.20	<0.001	0.04	0.03	0.04	0.03	0.14	<0.001

Significance is set at p <0.0125

APP – Assessment of Physiotherapy Practice; EBP - Evidenced based practice; r^2 – variance

Analysis of the regression model determined that all assumptions of a multiple regression were met. A multiple regression model including mean OSCE, written and seminar scores was determined to be the best fit (F[3,114] = 20.26, p < 0.001, R^2_{adj} = 0.33, S_{EE} = 6.88) and was a moderate predictor of APP scores. The OSCE (β = 0.49, p < 0.001, 95%CI 0.46-0.98) was a significant independent contributor to the relationship with clinical performance. The written examination and seminar presentation were not significant contributors to the model (β = 0.09, p = 0.31, 95%CI -0.11 - 0.36 and β = 0.14, p = 0.08, 95%CI -0.02 - 0.39, respectively).

4.6 Discussion

The aim of this study was to determine the relationships between specific coursework summative assessments and the future clinical performance of physiotherapy students. The OSCE showed the strongest correlation with students' clinical performance although all three summative assessment items were significantly correlated with overall clinical performance. Given that an appropriately designed OSCE aims to measure students' clinical performance against a set standard of competence (Khan, Ramachandran, et al., 2013), this is not surprising.

The OSCE investigated in the present study demonstrated a strong relationship (r = 0.57, p < 0.001) with students' future clinical performance, and explained 32% of the variation (r^2 = 0.32) in students' overall performance on the APP. These findings support the assumption that performance in a pre-clinical OSCE is related to students' future performance in authentic clinical environments. They also suggest that pre-clinical OSCE scores could be effectively used as an early indicator not only of students' future overall clinical performance, but also their performance in specific domains of practice. OSCE scores were related to performance in all domains of practice. Of note, OSCE scores were strongly related to performance in the 'intervention', 'assessment' and 'analysis and planning' domains where they explained 31 - 38% of the variation (r^2 = 0.31 - 0.38) in students' clinical performance. This suggests students with low-passing OSCE scores may benefit from enhancement in these domains prior to, or early in, placement. These findings may be influenced by the structure of the stations included in the OCSEs informing this study and the allocation of marks within individual OSCE stations. They may

also be influenced by the timing of the OSCEs; OSCEs were consistently held after the written examinations.

The seminar presentation demonstrated a weak relationship (r = 0.29, p = 0.012) with students' future clinical performance, but explained only 8% of the variation (r^2 = 0.08) in students' overall performance on the APP. The only domains of practice significantly related to students' seminar presentation scores were the 'communication' and 'intervention' domains, both of which explained only 5% of the variation (r^2 = 0.05) in student performance on the APP. These correlations may have been skewed by the fact that the mean score for the seminar presentations was markedly higher than any other pre-clinical assessment. The findings are also limited by the size of the dataset for the seminar presentations, as it is utilised in only one of the four coursework subjects investigated. It should be noted that the seminar presentations investigated differed from traditional oral examinations, which have been described as "a face-to-face interaction between an examinee and one or more examiners" (Tekian & Yudkowsky, 2009). The seminars investigated were pre-prepared presentations with only a short question-time of 2 - 3 minutes allotted, so comparisons with existing literature reporting on traditional oral presentations will be extremely limited.

The written examinations demonstrated a moderate correlation (r = 0.39, p < 0.001) with students' future clinical performance, and explained 15% of the variation (r^2 = 0.15) in students' performance on the APP. These findings support the assumption that students' demonstration of knowledge in a written assessment is related to the students' application of that knowledge in a clinical setting. In clinical placement students are routinely required to apply skills assessed in written format: clear and succinct communication of knowledge and ideas, analysis of case studies, identification of appropriate assessments and interventions, and identification of potential risks. There was no significant relationship between the written examination and the 'professional behaviour' or 'risk management' domains. The written assessments did not include specific ethical scenarios, which may have contributed to the lack of a significant relationship. It is also possible that the behaviours required to demonstrate adequate performance in the professionalism and risk management domains within a fluid and reactive authentic environment are difficult to draw out in a controlled written examination.

Clinical placements are increasingly becoming a precious commodity (Rodger et al., 2008), therefore education providers should make every effort to ensure that students allocated to a clinical placement are optimised for success. The relationships identified in this study support the use of written assessments and OSCEs in physiotherapy education and confirm the assumption that the standards achieved in these assessments are an indicator of a student's expected performance in clinical practice. The strong relationship between OSCE scores and students' future clinical performance could be used as an opportunity to identify students who may benefit from positive enhancement strategies prior to the student entering the clinical environment, particularly in the assessment, analysis and planning, and intervention domains of practice. The identification of students likely to require increased clinical educator input provides opportunity for these students to be more closely monitored whilst on placement, and additional support offered to both student and clinical educator in a timely and responsive way. This is important as university support for clinical educators has been reported as a facilitator for the provision of clinical placements (McMahon, Cusack, & O'Donoghue, 2014). Proactive learning experiences and assistance could be utilised pre-clinically as well as early in the clinical placement to maximise the chance of a successful and satisfying experience for all stakeholders. The findings of this study may be used as a foundation to justify future research into the design, implementation, and evaluation of early intervention programs to enhance clinical placement preparedness.

Strengths of this study include the use of consecutive cohorts of students and a consistent pre-clinical and clinical program of assessment. Furthermore, the APP clinical performance instrument, which is used by most physiotherapy education programs in Australia and New Zealand, has demonstrated to be a valid and reliable assessment tool. The thorough description of the APP instrument available in the literature may enable programs to compare their clinical performance instruments to the APP and interpret the findings accordingly. A potential weakness of this study is that it is limited to a single institution. Future research on this topic should aim to recruit larger participant numbers drawn from multiple institutions and should also investigate differences in pre-clinical summative assessment format.

4.7 Conclusion

The pre-clinical OSCEs investigated in this study were found to have a strong relationship with clinical performance, explaining up to 32% of the variation in students' future clinical performance. OSCE scores were significantly related to students' clinical performance in all domains of practice and could potentially be used as measures to identify students who could benefit from proactive strategies for enhancement prior to entry into their associated clinical placements. OSCE scores may also offer education providers the opportunity to selectively establish early support for both student and clinical educator during clinical practice to increase satisfaction in the clinical placement experience for all stakeholders.

Chapter 5. Not all physiotherapy Objective Structured Clinical Examinations are created equal

5.1 Prelude

The study described in Chapter 4 identified that, of the commonly used summative assessments within a physiotherapy education program, mean Objective Structured Clinical Examination (OSCE) scores had the strongest relationship with students' future supervised clinical practice performance. While OSCEs are a commonly used assessment in physiotherapy education programs, the evidence supporting the use of OSCE in the assessment of physiotherapy students is currently limited and conflicting. Where significant relationships have been identified in other studies, the strength of those relationships are highly variable.

This chapter aims to further explore the appropriateness of OSCEs as a summative assessment in physiotherapy education, by investigating the construct validity of individual OSCEs used within an entry-level extended master's physiotherapy program. Sources of evidence for the construct validity of an assessment include predictive validity, through demonstrated relationships with other variables such as external assessments. Student scores on multiple OSCEs will be compared to students' future clinical performance in the corresponding area of practice, as assessed by a widely used and validated external measure of clinical competence, the Assessment of Physiotherapy Practice.

5.2 Abstract

Introduction: The use of Objective Structured Clinical Examinations (OSCEs) is well researched in the medical profession, but evidence supporting their use in the physiotherapy profession is limited and conflicting. The construct validity of OSCEs used within an extended master's physiotherapy program was tested by exploring the relationships between students' OSCE scores and subsequent outcomes in an external assessment of clinical competence.

Method: Retrospective data from students in the 2011 - 2013 intakes of an Australian extended master's entry-level physiotherapy program were analysed. Spearman's correlations between OSCE scores from six discreet coursework subjects and students' corresponding supervised clinical practice outcomes, as measured by the Assessment of Physiotherapy Practice (APP), were analysed.

Results: OSCE scores from five of the six coursework subjects investigated were significantly related with students' clinical practice performance in the associated placement. Correlations ranged from rs = 0.21, p = 0.049 (OSCE F) to rs = 0.42, p = <0.001 (OSCE C). OSCE A did not demonstrate construct validity (rs = -0.03, p = 0.78).

Discussion: While not all OSCEs investigated demonstrated construct validity, the volume of evidence suggests that well-constructed OSCEs may be a valid way to assess physiotherapy students for the competencies required for supervised clinical practice. Differences in the intrinsic qualities of an OSCE may impact their validity. This may explain differences in findings in the wider literature relating to the psychometric value of OSCEs in health profession education.

5.3 Introduction

The Objective Structured Clinical Examination (OSCE) is a commonly used measure of assessment in physiotherapy education programs. OSCEs aim to measure students' ability to apply knowledge, their clinical skills, attitudes and problem-solving abilities (Harden, Stevenson, Downie, & Wilson, 1975). Student performance on OSCEs may be used for both formative and summative assessment purposes and, in some programs, OSCE scores may be utilised as a hurdle assessment for progression to clinical placement.

The use of OSCEs is well researched within the medical profession and the body of work synthesised into guidelines for educators (Khan, Gaunt, Ramachandran, & Pushkar, 2013; Khan, Ramachandran, et al., 2013; Pell, Fuller, Homer, & Roberts, 2010). However, there is limited research exploring the validity of OSCEs within the physiotherapy profession (Bobos et al., 2021; Gorman, Lazaro, Fairchild, & Kennedy, 2010; Silva, Lunardi, Mendes, Souza, & Carvalho, 2011; Swift, Spake, & Gajewski, 2013; Wessel et al., 2003). A systematic review and meta-analysis of six studies by Bobos et al. (2021) explored the psychometric measurements of OSCEs used in physiotherapy programs. Four of the studies included in this review reported on the construct validity of OSCEs (Gorman et al., 2010; Ladyshewsky, Baker, Jones, & Nelson, 2000; Silva et al., 2011; Wessel et al., 2003). Three studies reported correlations between other forms of assessment or academic performance, these being: supervised clinical practice scores (Wessel et al., 2003), total course score and total GPA (Gorman et al., 2010), and scores in a non-OSCE examination (Silva et al., 2011). Correlation values varied from very weak (r = -0.13) relationships between OSCE scores and clinical performance, to strong (r = 0.78) relationships with individual course grades. The use of OSCEs as a hurdle assessment to progress within a program assumes that there is a relationship between students' performance on OSCE and their future performance, but the variable relationships identified between supervised clinical practice performance (r = -0.13, weak) and overall GPA (r = 0.44, moderate) in the review by Bobos et al. (2021) raise questions as to the appropriateness of using OSCEs to predict students' future performance. However, the paucity of research currently available is insufficient to draw conclusions.

In Australia and New Zealand, the clinical competence of all physiotherapy students undertaking five-week, supervised, clinical practice placements is assessed using the Assessment of Physiotherapy Practice (APP). OSCEs are used in physiotherapy programs to assess a student's ability to demonstrate practical competencies required in clinical practice, while the APP measures clinical competence in an authentic clinical setting. As such, one may expect that there would be commonality between the domains of practice assessed within the OSCE and the APP. A systematic review by Ribeiro, Ferla, and Amorim (2019) that reported on seven studies which investigated the use of OSCEs to evaluate physiotherapy knowledge and skills, suggested that OSCEs could be used effectively to evaluate competence in physiotherapy education programs. Four of the studies in the review described the evaluation of student competence in skills within the cognitive, psychomotor, and affective domains of clinical practice (Cacho et al., 2016; Edgar et al., 2014; Maloney, Storr, Morgan, & Ilic, 2013; Silva et al., 2011). However, none of the studies in the review explored the relationship between OSCE scores (i.e., practical competencies) and clinical skill competence in an authentic clinical setting.

Two studies have reported on the relationship between students' performance on OSCE during coursework education and their performance in supervised clinical practice (Terry, Hing, Orr, & Milne, 2020; Wessel et al., 2003). Wessel et al. (2003) found no significant relationship between student performance on one OSCE and clinical practice in a Canadian two-year post-baccalaureate program. In contrast, in a study of an Australian entry-level postgraduate program, Terry et al. (2020) identified a strong significant relationship between students' mean performance over four summative coursework OSCEs and their mean performance in supervised clinical practice. This limited and conflicting research investigating the validity of OSCEs in physiotherapy education against performance in clinical practice indicates a need for deeper exploration into this topic.

This study aimed to examine the construct validity of OSCEs used within an extended master's physiotherapy program by investigating the relationships between student scores on individual OSCEs across multiple subjects and subsequent student outcomes in clinical practice. The findings of this study will contribute to the evidence investigating the construct validity of OSCEs in the assessment of physiotherapy student clinical competence.

5.4 Methods

Retrospective data from students in the 2011 - 2013 intakes of an Australian extended master's entry-level physiotherapy program were utilised. The study protocol was approved by the host institution's human research ethics committee (Protocol No. RO1733). Participants' data were included in the study if they had completed at least one clinical placement. Students' OSCE scores from the cardiorespiratory (I and II), neurological, and musculoskeletal (I and II) subjects within the physiotherapy curriculum, totalling six OSCEs, were retrieved from archived hardcopies and electronic records. Student APP scores from the four corresponding clinical placements were also retrieved.

The six OSCEs investigated were coded alphabetically from A – F. The six OSCEs were administrated and delivered using the same organisational approach. All OSCEs utilised a mark sheet and examiners were experienced physiotherapists in the relevant clinical area. Examiners were briefed prior to the OSCEs. OSCEs A-D consisted of 2 x 15 min stations, OSCE's E and F both consisted of 3 x 15min stations, with one of these OSCEs having an additional 30min station. Each station contained a clinical case with marks allocated to some, or all of the following categories: clinical reasoning, communication with the client, performance of assessment and therapy techniques, safety, and professional conduct. Stations were marked, scores combined and then converted to a percentage out of 100.

Students completed five-week clinical placements in the corresponding clinical areas within authentic physiotherapy workplaces: cardiorespiratory, neurological, and musculoskeletal (I and II). Students were supervised by a clinical educator who was a registered physiotherapist employed by the host clinical facility at which the student was placed and who had no oversight of the students' final subject grades. Student clinical practice performance was measured using the APP which was completed by the clinical educator formatively at mid-placement and summatively at end-placement. The APP instrument is well-described in the literature with demonstrated validity and reliability (Dalton et al., 2011; Dalton et al., 2012; Dalton et al., 2009). The instrument consists of twenty items across seven domains of practice, these being professionalism, communication, assessment, analysis and planning, intervention, evidence-based practice, and risk management. The four end-placement APPs were coded numerically

from I - IV. Raw end-placement APP scores were converted into a percentage out of 100. Data were deidentified and entered in SPSS version 24 (SPSS Inc., Delaware, USA).

Data were tested for normality using the Shapiro-Wilks test and descriptive statistics calculated. Spearman's rank order correlations were calculated between total OSCE scores and the total APP score on the corresponding clinical placements. The strength of correlation between variables was determined as: r_s 0.1 to 3 considered weak, 0.3 to 0.5 considered moderate and > 0.5 considered strong (J. A. Cohen, 1988).

5.5 Results

There were 93 students who entered the extended master's physiotherapy program between 2011-2013. Two students did not complete any clinical placements and were excluded from the study, leaving data from 91 (male = 48, female = 43) students for analysis.

Descriptive statistics for all OSCE and clinical placement scores are displayed in Table 18. One student did not complete OSCEs E or F, nor the corresponding clinical placements. Complete APP data from four student clinical placements were unobtainable.

Correlations between pre-clinical OSCE scores and students' performance in clinical practice, as measured by the APP, are displayed in Table 19. The dependent variables, APP score, were not normally distributed for all clinical placements (see Table 18) therefore Spearman's rank order correlations were used to calculate correlations. The majority of OSCE results (n = 5; 83%) correlated with the corresponding clinical placement. OSCE C demonstrated the strongest relationship, a moderate correlation strength, with students' clinical performance score (r_s = 0.42, p < 0.001) while OSCE A did not correlate with the APP score in the corresponding clinical placement.

Table 18. Descriptive statistics for OSCE and clinical placement scores

Variable	N	Min	Max	Mean (± Std Dev)	Median	IQ Range	Shapiro-WilksSig.*
OSCE A	91	59.38	96.25	83.09 (± 8.25)	84.38	78.75 - 89.38	< 0.0001
OSCE B	91	64.38	93.13	80.87 (± 6.71)	81.25	76.88 - 85.00	0.24
OSCE C	91	56.00	96.00	78.75 (± 9.26)	79.00	71.25 - 86.00	0.09
OSCE D	91	53.00	96.00	75.06 (± 10.80)	75.63	67.00 - 84.38	0.30
OSCE E	90	57.00	93.92	72.84 (± 8.32)	72.54	67.13 - 79.46	0.38
OSCE F	90	56.00	96.67	78.00 (± 8.55)	78.17	73.17 - 84.44	0.20
APP I	90	31.73	97.50	70.92 (± 13.55)	64.76	62.46 - 80.31	0.48
APP II	89	36.25	100.00	77.99 (± 14.65)	78.75	65.63 - 91.08	0.001
APP III	90	51.25	100.00	81.10 (± 13.71)	84.38	68.44 - 93.75	< 0.0001
APP IV	89	46.25	100.00	79.89 (± 14.66)	81.25	67.50 - 93.75	< 0.0001

* Shapiro-Wilks test of normality

Table 19. Correlations between OSCE scores and the corresponding clinical placement scores as measured by the APP

OSCE	n	rs	p	Correlation strength
A	90	-0.03	0.78	Weak
B	90	**0.28**	**0.008***	Weak
C	89	**0.42**	**< 0.001***	Moderate
D	89	**0.39**	**< 0.001***	Moderate
E	89	**0.30**	**0.004***	Weak
F	90	**0.21**	**0.049***	Weak

5.6 Discussion

Construct validity can be supported by five sources of evidence (Downing, 2003); one of these is an assessment's relationships to other variables such as independent external measures. Except for OSCE A (p = 0.78), all the OSCEs investigated in this study demonstrated significant relationships with students' future clinical performance. There was, however, variability in the strength of these relationships, with correlations ranging

from weak (rs = 0.21) to strong (rs = 0.42). While the use of OSCEs has been well researched in medical education (Khan, Gaunt, et al., 2013; Khan, Ramachandran, et al., 2013; Pell et al., 2010), there has been considerably less published literature from within nursing and allied health professions. The volume of evidence from this current study adds to the emerging body of work supporting the use of OSCEs as an appropriate assessment method for physiotherapy students.

It is evident from the findings of this study, that not all physiotherapy OSCEs demonstrate construct validity. The relationship between OSCE A and students' future clinical performance was not significant. Yet, for the same students with the same outcome measure, OSCEs B - F were significantly related to future clinical performance using the APP. Furthermore, the strength of the significant correlations of OSCEs B - F ranged from weak to moderate. In the only other comparable physiotherapy focused study by Wessel et al. (2003), with a cohort of 48 physical therapy students in their first year of a two-year physical therapy degree, no significant relationship between OSCE scores and clinical performance was found. It should be noted that the OSCE investigated by Wessel and colleagues (2003) was the first time those students had experienced this type of assessment. Consequently, it is possible that student inexperience with the assessment format could have impacted student performance. 'Fear of the unknown' was one of the triggers for student anxiety during OSCE identified by Zhang and Walton (2018). In this current study, all students had experienced one OSCE prior to completing OSCEs A - F. Furthermore, OSCEs A and C took place in the same semester of study, on the same day. Stations from each examination were integrated into the same session, so students were under the same conditions and entered the two OSCEs with the same level of experience.

A further potential reason for the differences in findings between the study by Wessel et al. (2003) and the findings of this study can be found in the number of OSCEs investigated. In the study by Wessel et al. (2003) one OSCE, centred around chronic musculoskeletal conditions, informed the study. That is unlike the multiple OSCEs which informed this current study. The hypothesis that the difference in findings can be explained by the number of OSCEs investigated is supported by this study where OSCE A was not significantly correlated with future clinical competence. If only one subject was used to inform this study, the subject assessed using OSCE A, then no significant correlations between OSCE and student performance in clinical practice would have been identified.

The findings of this study investigating multiple OSCEs across a range of coursework subjects may explain why there are inconsistent findings regarding OSCE correlations with clinical performance between studies.

As mentioned previously, two of the OSCEs in the present study were conducted on the same day under the same conditions, yet the correlations between their OSCE scores and clinical practice scores were notably different: OSCE A (r_s = -0.03, p = 0.78) and OSCE C (r_s = 0.42, p = <0.001). A supposition for these differences in findings may lie in the intrinsic qualities of the OSCEs themselves. While each OSCE followed a similar blueprint regarding the constructs to be assessed, the proportion to which each construct was represented within the OSCEs may have differed. The structure and content of the mark sheets utilised in the OSCEs may also have differed.

Daniels and Pugh (2018) highlighted that OSCE scoring rubrics should be unambiguous and should clearly outline the operational definition of success for each of the checklist items. Assessor training is an important component of conducting an OSCE (Daniels & Pugh, 2018) and may go some way to mitigating the effect of an ambiguous item if the OSCE developer explains the intention of the item and how scores are to be applied (Schwartzman, Hsu, Law, & Chung, 2011). Both OSCEs provided assessors with their material prior to the OSCE and held a verbal briefing on the day of the assessment. However, this assumes the assessor will encode the training material into a mental scoring rubric that matches that of the presenter (Bejar, 2012). The influence of an assessor's mental rubric would be particularly relevant where a station requires the presentation of a spoken constructed response; personal attributes of an assessor may lead to a mental rubric that contains components not explicitly listed, or expected to be demonstrated (Bejar, 2012), as a result of an ambiguous checklist item.

Differences in the intrinsic qualities of the OSCEs B – F may also explain why, even though correlations were significant, the strength of correlations between OSCEs and their associated APP scores varied. Identifying differences in structure and content between OSCEs with lowest and highest relationships with external assessments could contribute to the current body of literature about OSCE design.

Some limitations to this study require consideration. In physiotherapy education, the structure of OSCE stations vary from program to program and in some cases, subject to subject with specific details of OSCE station construct not always provided. This creates a challenge when comparing between OSCEs and makes benchmarking between institutions, or applying the findings of the research, difficult to achieve. Explicit details of the OSCEs investigated are not given here as this would make the OSCEs themselves and the educators involved potentially identifiable, further contributing to this systemic limitation. Finally, all coursework subjects investigated in this study were convened by different educators. As such, the teaching styles and personalities of the educators may have influenced student engagement in their subjects, which may in turn have affected student outcomes.

In conclusion, this study investigated OSCEs featured in an entry-level extended master's physiotherapy education program. Evidence of construct validity was identified for five of the six OSCEs investigated, demonstrating significant weak to moderate relationships to the external measure of clinical competence, the APP. One OSCE investigated did not have a significant relationship with clinical performance. This suggests the construct validity of individual OSCEs should not be assumed and further explains why there are discrepancies in the broader research in relation to findings of OSCE relationships with clinical competence. Based on the findings of the research reported in this study, future research would benefit from investigating the intrinsic qualities of OSCEs, both with and without construct validity. A detailed exploration of OSCE content and design would be of benefit to provide guidance to physiotherapy educators looking to evaluate OSCEs used within their programs.

Chapter 6. A content analysis of two physiotherapy OSCEs with differing predictive validity of future clinical performance

6.1 Prelude

Based on the study reported in Chapter 5, it was identified that while most Objective Structured Clinical Examinations (OSCEs) from an extended master's entry-level physiotherapy program demonstrated construct validity with a valid and reliable externally applied assessment, it was not the case for all OSCEs.

The OSCEs investigated in Chapter 5 were implemented and administrated by the same faculty at the host institution. There was consistency in the student cohorts, the pre-test conditions, and the organisation of the OSCEs on test day. The content of the OSCEs however, their stations, mark sheets and instructions to examiners, were designed by the conveners of the individual coursework subject.

Content validity is a domain of construct validity (Taylor et al., 2021). It is important to establish that the content of any assessment is valid, to ensure that the content of the assessment is representative of all aspects of the construct it is aiming to test. The findings described in Chapter 5 suggest that the intrinsic properties of an OSCE may influence its validity. The study reported in this chapter aims to further explore the qualities of an OSCE with and without demonstrated construct validity.

6.2 Abstract

Aim: To apply a conventional content analysis approach to critically describe the content of two Objective Structured Clinical Examinations (OSCEs) featured within an extended master's entry-level physiotherapy program: one that has previously been identified to have a significant, moderate, relationship with an externally applied measure of students' clinical performance, the Assessment of Physiotherapy Practice (APP), and one that was identified to be unrelated to student performance in clinical practice.

Method: Items on OSCE mark sheets were analysed by two independent coders. First cycle codes were applied from a combination of pre-set and self-determined codes. Second cycle coding was conducted by a single researcher to determine the naturally occurring themes within the OSCE. Variations within first cycle coding were resolved by consensus. Items on OSCE mark sheets were then mapped and aligned to domains of practice measured by the APP.

Results: There was a 78% agreement in first cycle coding between the two coders for the OSCE related to students' future clinical performance (OSCE R), compared to 67% agreement for the OSCE unrelated to students' future performance (OSCE UnR). Second cycle coding identified six naturally occurring themes: communication, assessment, treatment, safety, knowledge, and education. When mapped to the domains of the APP the majority of the content fell within the assessment (47% OSCE R, 21% OSCE UnR) and treatment (29% OSCE R, 20% OSCE UnR) domains. OSCE UnR included a greater proportion of content within the communication and education domains compared to OSCE R. The evidence-based practice and risk management domains of the APP were not explicitly assessed. On OSCE UnR, a single key construct could not be determined for 12% of the items on the OSCE mark sheets

Conclusion: Assessment, treatment, and safe practice skills as measured by the APP were the main content areas covered by the OSCE with construct validity, OSCE R. In comparison, OSCE UnR contained less content examining specific assessment and treatment skills and more ambiguous items on the mark sheet. Any domains of practice included on the APP that are not covered within an OSCE should be explicitly assessed elsewhere within the program of assessment.

6.3 Introduction

Entry-level physiotherapy education programs in Australia require students to complete clinical placements in authentic workplaces as a compulsory part of their training (Australian Physiotherapy Council, 2017). To provide and support a clinical placement requires significant investment from both the education provider and the clinical partner. It is essential that students are adequately prepared for clinical placement to ensure that the investment of all stakeholders, including the student and the clients who consent to interact with them, is maximised. Students who are not successful in achieving the minimum standard of competence within the allotted clinical placement period require further investment in the form of remediation, reassessment, and additional clinical placement time.

Coursework educators can have an inkling of the students whose underlying knowledge and skills may not be adequate in a clinical environment, and may choose to go on to offer informal remediation or support prior to clinical placement. However, without objective evidence that students have not achieved the minimum standard of competence needed for clinical placement within the classroom, there is no justification for the imposition of a formal remediation process or to halt the student's progression to clinical placement.

The OSCE is a commonly used form of assessment within the physiotherapy profession, and has been shown to be a valid method of assessing students' problem-solving, clinical skills, and professional attributes (Ribeiro et al., 2019). A study by Terry et al. (2020) found that the OSCE had the strongest relationship with students' future clinical performance when compared to written assessments or seminar presentations. These studies contribute to the psychometric evidence supporting the use of OSCEs in physiotherapy education. While psychometric evidence is important, Hodges (2003) suggested that qualitative research into the OSCE is required to determine its contextual validity. For an OSCE to demonstrate context validity, it must accurately assess a student's ability to apply their learning to the practice of physiotherapy. Students then progress in their training to apply these skills in authentic clinical environments. It is, therefore, important for physiotherapy educators to know whether the OSCEs utilised within physiotherapy programs accurately assess the various domains of physiotherapy practice which students are required to demonstrate when they are placed in authentic clinical environments.

In Australia and New Zealand, the clinical competence of physiotherapy student performance in five-week supervised clinical practice placements is assessed using the Assessment of Physiotherapy Practice (APP), an instrument which is well-described in the literature with demonstrated validity and reliability (Dalton et al., 2011; Dalton et al., 2012; Dalton et al., 2009). The instrument consists of twenty items under the umbrella of seven domains of practice: professionalism, communication, assessment, analysis and planning, intervention, evidence-based practice, and risk management.

Recommendations exist to guide health profession educators on the design, delivery, and analysis of OSCEs (Khan, Gaunt, et al., 2013; Pell et al., 2010); however, these guidelines are targeted towards medical education. There is much more limited research in the field of assessment that is targeted specifically to the allied health professions such as physiotherapy. Swift et al. (2013) published one example of a marking rubric utilised in a musculoskeletal physiotherapy OSCE. Swift's rubric was investigated as a study tool for students but an example rubric of one OSCE station was provided. As OSCEs are used in physiotherapy programs to assess a student's ability to demonstrate the competencies required in clinical practice, while the APP measures clinical competence in an authentic clinical setting, it can be expected that there would be commonality between the domains of practice assessed within both the OSCE and the APP. A systematic review by Ribeiro et al. (2019) reviewed seven studies investigating the use of OSCEs to evaluate physiotherapy knowledge and skills. The review found that OSCEs could be used to evaluate competence in physiotherapy teaching. However, none of the studies reported in the review by Ribeiro et al. (2019), investigated the use of the OSCE against clinical skill competency in an authentic clinical setting. Furthermore, there is currently no known published information regarding how domains of practice commonly assessed within physiotherapy OSCEs compare to those included in measures of clinical practice performance such as the nationally and internationally utilised APP.

This study aimed to explore content themes utilised within an OSCE demonstrated to have construct validity through a significant moderate relationship with students' subsequent performance in clinical practice as measured by the APP, and to compare this to an OSCE with no significant performance to student's future clinical performance. This information could assist educators to plan future assessments that will more likely identify which

students will be unsuccessful on placement to allow for early remediation. It was hypothesised that there would be significant alignment between the content assessed within the OSCE with predictive validity and the seven domains of the APP.

6.4 Methods

This study was an exploratory study involving three phases (Figure 17). Phase one was selecting two OSCEs for content analysis, based on their established relationships with students' future clinical performance, and is described in Chapter Five of this thesis. The individual OSCEs that demonstrated the strongest and weakest relationships to student performance in the corresponding clinical placements were identified.

Phase One

Identify coursework OSCEs with the strongest and weakest relationship to student performance in clinical practice.

(Chapter 5)

Phase Two

A conventional content analysis of identified OSCE utilising first and second cycle coding to identify constructs assessed by checklist items. Naturally-occurring themes within constructs identified.

Phase Three

OSCE content mapped to APP performance indicators and aligned to a domain of the APP.

Figure 17. Flowchart of study phases

Phase two of this study was the conduct of a content analysis within the most contemporary versions of the identified OSCEs (the 2013 intake). Content analysis is a research method that has been applied in social sciences and educational research and is used to make inferences from text (Weber, 1990). Codes, in the form of a word or short phrases, are applied to sections of text to capture the essence of the text so that valid inferences can be made (Saldaña, 2021). In this present study, codes were used to capture the key construct assessed by each item on the OSCE mark sheets. A

conventional content analysis approach (Weber, 1990) with first and second cycle coding was used in this study.

All questions within the identified OSCEs were analysed for their content. OSCE mark sheet items were entered into a Microsoft Excel spreadsheet and the mark allocation for each separate item within mark sheets were recorded. The text items on OSCE mark sheets were analysed for the primary construct being assessed in each item. First cycle coding was performed by two independent coders (RT and RO) using descriptive codes to identify the key constructs assessed in each item of the OSCE. Coders were provided with the students' questions and examiner mark sheets. Neither coder was involved in the design or development of the OSCEs investigated. Codes were applied based on the coders' impression of the content being assessed in that OSCE item. A pre-determined list of codes was established by RT and approved by RO, with coders also invited to use additional self-generated codes as they found appropriate. Second cycle coding was then performed by a single coder (RT), and the naturally occurring themes within the constructs assessed in the OSCE identified. During second cycle coding, variances between the first cycle coding from the independent coders were identified. OSCE items that could be interpreted as different constructs were settled by consensus. Where consensus could not be reached, the second-cycle code was designated by a third person (WH). The marks allocated to each naturally occurring theme were determined as a percentage of the total OSCE marks.

Phase three of the study mapped and aligned the constructs assessed on the OSCEs to those assessed on the APP. Second cycle codes were mapped to the seven APP domains of practice by two investigators (RT and RO): professionalism, communication, assessment, analysis and planning, intervention, evidence-based practice, and risk management and their accompanying performance indicators. Alignments between OSCE constructs and APP performance indicators were identified. When identifying the performance indicators that aligned with second orders codes, consideration was given to the scope of clinical performance that could be demonstrated within the OSCE. For example, APP performance indicators requiring reporting to senior clinicians were not included as this behaviour could not be authentically demonstrated within the structure of the OSCE. All performance indicators relevant to the scope of the OSCE were identified. Where second cycle codes could be placed in multiple domains within the APP, the

original student question on the OSCE was referenced to aid alignment. The APP Performance Indicators are displayed in Appendix 1.

6.5 Results

Phase one results are reported in the previous chapter of this thesis (Chapter Five, pp. 121-22). The first OSCE selected had the largest relationship between a coursework OSCE and clinical practice performance measured by the APP. This relationship was of moderate strength ($r_s = 0.42$, $p < 0.001$). The second OSCE selected did not have a significant relationship with students' future clinical performance ($r_s = -0.03$, $p = 0.78$). Both OSCEs consisted of two x 15 min stations and each station contained a clinical case study. Examiners were licensed physiotherapists with experience in assessing OSCE. Detailed mark sheets were used for each station and examiners were briefed on the utilisation of the mark sheets prior to the commencement of the OSCE.

6.5.1 Phase Two Results

Content Analysis of the OSCE Related to student performance on the APP (OSCE R)

Content of OSCE R first cycle coding revealed 27 unique codes. Approximately half (49%) of the items were coded using the same codes by the two coders. 29% of items were attributed different codes by the two coders, however the codes utilised had the same meaning (for example: knowledge of outcome measures vs theory). Overall, there was 78% agreement in first cycle coding between the two reviewers, indicating a substantial level of agreement on the constructs assessed by the items coded (Viera & Garrett, 2005). Of all items coded, 22% required moderation between the two coders to determine the key construct assessed by the OSCE item. On all occasions consensus was achieved, and a third coder was not required.

Second cycle coding identified 20 codes and six naturally occurring themes within the constructs assessed in the OSCE: communication, assessment, treatment, safety, knowledge, and education. These themes, and the second cycle codes that were attributed to each theme, are displayed in Table 20.

Table 20. Naturally occurring themes and the second cycle codes attributed to each theme within an OSCE that is significantly related to student performance in clinical practice (OSCE R)

	Naturally occurring themes within constructs assessed (% of OSCE R Content)					
	Communication (13%)	**Assessment (41%)**	**Treatment (19%)**	**Safety (16%)**	**Knowledge (8%)**	**Education (4%)**
Second cycle codes	Communicates clearly	Demonstrates assessment	Demonstrates intervention	Is safe	States correct theoretical knowledge	Educates client
	Communicates role	Demonstrates client interview	Identifies appropriate intervention	Ensures client safety	Interprets results	
	Gains consent	Monitors client during assessment	Prescribes exercises	Practices infection control	Answers extension question	
	Extension marks - communication	Identifies appropriate assessment	Monitors client during intervention	Identifies risks		

Content Analysis of the OSCE unrelated to student performance on the APP (OSCE UnR)

Content of OSCE UnR first cycle coding revealed 26 codes, nine of which were unique. Of these, 51% of the items on the OSCE mark sheet were coded by both coders using the same code, while 16% of the items were coded using different codes that had the same semantic meaning. This meant that overall, there was a 67% agreement in first cycle codes between the two reviewers. In total, 33% of the items on the mark sheet required moderation between the two coders. In all cases consensus was achieved. However, the two reviewers agreed that a single key construct could not be determined for 12% of the items on the OSCE mark sheets. The items are indicated in Table 21 using a number of asterixis ('*') to describe the number of constructs determined to be assessed within the one item (e.g., ** noting two constructs assessed within one item). In each instance where there was more than one construct assessed within a single mark sheet item, there was no key or standard of performance included on the mark sheet to guide assessors on how marks should be allocated. Second cycle coding identified 21 codes.

Table 21. Naturally occurring themes and the second cycle codes attributed to each theme within an OSCE that is not related to student performance in clinical practice (OSCE UnR)

	Naturally occurring themes (% of OSCE UnR Content)					
	Communication **(19%)**	**Assessment** **(21%)**	**Treatment** **(20%)**	**Safety** **(12%)**	**Knowledge** **(17%)**	**Education** **(12%)**
Second cycle codes	Communicates clearly	Demonstrates assessment	Demonstrates intervention **	Ensures client safety	Applies knowledge***	Client education
	Communicates role	Demonstrates client interview	Prescribes exercises	Practices infection control	Goal setting	Coaching**,***
	Gains consent	Monitors client during assessment	Monitors client during intervention	Safety****	States correct theoretical knowledge	
	Gives clear instructions**	Concludes assessment***	Coordinates intervention***		Patient positioning	

Communication

Communication refers to the student's verbal communication with the client during the OSCE station. Of the total content within the OSCE, 13% of the constructs assessed in OSCE R were determined to be within the communication theme, compared to 19% of OSCE UnR. Mark sheet items coded as *communicates clearly* (representing 32% of the marks within the *communication* theme of OSCE R, and 4% of OSCE UnR) included the requirement of students to introduce themselves, while the item coded as *communicates role* (32% OSCE R, 26% OSCE UnR) required the student to explain their role and the planned interaction. The code *extension marks – communication* (5% OSCE R, 0% OSCE UnR) were given to items which required students to demonstrate finesse in their communication with the client. The code *gains consent* (32% OSCE R, 13% OSCE UnR) required the student to explain the procedure to be performed and request explicit consent be provided by the client. Items coded as *gives clear instructions* (0% OSCE R, 57% OSCE UnR) required students to give succinct, explicit instructions to clients regarding their participation in an assessment or intervention. One checklist item attributed this code was identified as assessing two separate constructs within the single item (see Table 21). This item required students to *give clear instructions*, but also specified that students needed to clarify the client's understanding.

Assessment

Assessment refers to all mark sheet items associated with the assessment of the client's condition during the OSCE station. There were five codes identified that pertained to assessment, comprising 41% of the content within OSCE R compared to 21% of OSCE UnR. Items coded as *demonstrates assessment* (57% of the marks in the *assessment* theme in OSCE R, 78% in OSCE UnR) required students to correctly perform a physical test, while items coded as *identifies appropriate assessment* (7% OSCE R, 0% OSCE UnR) required students to state their assessment of choice based on information contained in the case scenario. Items coded as *demonstrates client interview* (28% OSCE R, 12% OSCE UnR*)* required students to effectively conduct specific components of a client interview relevant to the stations' client. Items coded as *monitors client during assessment* (8% both OSCEs) required students to demonstrate that they were actively monitoring the clients' response while conducting specific tests. The final code within the assessment theme, *concludes assessment* (0% OSCE R, 2% OSCE UnR) encompassed a diverse range of actions to be taken by the students before leaving the client at the conclusion of the assessment within the one item, for example, encouraging the client to redress.

Treatment

Treatment refers to all mark sheet items associated with the management of a clients' condition. There were five codes identified that pertained to the management of a client: *demonstrates intervention, identifies appropriate intervention, prescribes exercises,* and *monitors client during intervention.* These treatment items comprised 19% and 20% of the content within OSCE R and UnR respectively. Items coded as *demonstrates intervention* (61% of the marks within *intervention* theme in OSCE R, 70% in OSCE UnR) required students to effectively perform a physical treatment technique. One item attributed this code in OSCE UnR was determined to be assessing two separate constructs within the single item. This item required the demonstration of a physical treatment technique (*demonstrates intervention),* but also required students to provide education to the client about the therapeutic effect of the technique (*client education*). Items coded as *identifies appropriate intervention* (18% OSCE R, 0% OSCE UnR) required students to identify the most appropriate intervention for the clients' presentation. Items coded as *prescribes*

exercises (18% OSCE R, 4% OSCE UnR) required students to prescribe the most appropriate exercises for the clients' presentation. Items coded as *monitors client during intervention* (4% OSCE R, 9% OSCE UnR) required students to demonstrate that they were actively monitoring the clients' response while performing specific treatment techniques. All items coded as *coordinates intervention* (0% OSCE R, 17% OSCE UnR) were determined to assess multiple constructs. On these items, students were required to select the appropriate treatment technique, instruct both the patient and a physiotherapy assistant on their roles during the intervention, and correctly execute the maneuver.

Safety

Safety refers to all mark sheet items associated with maintaining the safety, of either client or physiotherapist, throughout the interaction. Of the content within the OSCE, 16% of OSCE R fell within the *safety* theme compared to 12% of OSCE UnR. Items coded as *is safe* (50% of the marks within *safety* theme in OSCE R, 0% UnR) required students to demonstrate appropriate ergonomics when completing manual tasks, to identify and avoid recognized contraindications to assessment and treatment, and to identify and implement safe practices before, during, and after client interactions. The other codes in this theme were *identifies risks* (25% OSCE R, 0% OSCE UnR), *ensures client safety* (25% OSCE R, 50% OSCE UnR), and *practices infection control* (0% OSCE R, 29% OSCE UnR). In OSCE R, mark sheet items coded as *practices infection* control did not have marks allocated and were marked as 'performed' or 'not performed'. The final code in this theme, *safety* (0% OSCE R, 21% OSCE UnR), was attributed to an item that was determined to be assessing multiple constructs within the one item: client position, therapist position, environmental set-up, and use of equipment as needed.

Knowledge

Knowledge refers to all mark sheet items associated with requiring students to demonstrate theoretical knowledge. The items within the *knowledge* theme comprised 8% of the content of OSCE R, compared to 17% of OSCE UnR. Items coded as *states theoretical knowledge* (55% and 54% of marks within the *knowledge* theme of OSCEs R and UnR respectively) required students to identify differential diagnoses based on a clients' current presentation, while items coded as *interprets results* (18% OSCE R, 0% OSCE UnR) required students to correctly interpret test results relating to the clients'

presenting condition. Items coded as *extension marks* (27% OSCE R, 0% OSCE UnR) required students to demonstrate excellent theoretical knowledge and confidence pertaining to the clients' presentation during their interaction with the client. OSCE UnR uniquely contained three codes: *applies knowledge****, *goal setting*, and *patient positioning.* Items coded as *applies knowledge**** (7% OSCE UnR) were identified as assessing three separate constructs as they required students to state correct theoretical knowledge to the examiner, prescribe an exercise, and then instruct the patient in the performance of the exercise. Items coded as *goal setting* (20% OSCE UnR) required students to list goals of a specific treatment technique while *patient positioning* (20% OSCE UnR) required students to state the appropriate start position for specific assessments and interventions.

Education

The provision of education requires students to communicate effectively, whilst providing correct information and instruction to clients. *Education* was determined to be a separate theme as it incorporated aspects of the *knowledge*, *intervention*, and *communication* themes (Candela, Piacentine, Bobay, & Weiss, 2018). Overall, 4% of the content of OSCE R fell within this theme, compared to 12% of OSCE UnR. Mark sheet items coded as *educates client* (100% OSCE R, 57% OSCE UnR) required students to prescribe the correct self-management for a condition including exercises, dosage, and use of aids and equipment. Students may have also been required to teach the exercises and explain their reasoning to the client. Items in OSCE UnR coded as *coaching* (43% OSCE UnR) required students to provide education to a client and, in addition, to provide feedback to the client about their performance. Two items coded as coaching were determined to be assessing multiple constructs. These items required students to provide feedback, correct their technique, and provide education on dosage.

A comparison of the proportion of content with OSCE R and OSCE UnR is displayed in Table 22.

Table 22. Proportion of OSCE content by theme

Theme	Percentage of OSCE Content	
	OSCE R (<u>related</u> to future clinical performance)	**OSCE UnR (<u>unrelated</u> to future clinical performance)**
Communication	13%	19%
Assessment	41%	21%
Treatment	19%	20%
Safety	16%	12%
Knowledge	8%	17%
Education	4%	12%
TOTAL	101%*	101%*

*Total does not equal 100% due to rounding to nearest whole percentages

6.5.2 Phase Three Results

Phase three of this study investigated the proportion of OSCE content within each domain of physiotherapy practice measured by the APP. The APP items and performance indicators that align with the second cycle codes are listed in Table 23. Where codes align with more than one domain of the APP, all domains have been listed with the domain that aligns best highlighted in bold, capitalised text.

Table 23. Alignment between OSCE second cycle codes and Assessment of Physiotherapy Practice (APP) domains

Code and Naturally Occurring Theme	Relevant APP Items and/or Performance Indicators	APP Domain	Code applied to OSCE Related (R)	Unrelated (UnR)
COMMUNICATION THEME				
Communicates clearly	Item 5: Communicates effectively and appropriately - Verbal/non-verbal (all performance indicators)	**COMMUNICATION**	Y	Y
Communicates role	Item 3: Demonstrates ethical, legal & culturally responsive practice - wears an identification badge & identifies self Item 5: Communicates effectively and appropriately – Verbal/non-verbal - greets others appropriately - questions effectively to gain appropriate information - uses suitable language & avoids jargon - demonstrates an appropriate range of communication styles (with e.g., clients, carers, administrative & support staff, health professionals, care team) - recognises barriers to optimal communication - integrates communication technology into practice as required - uses a range of communication strategies to optimise client rapport & understanding (e.g., hearing impairment, non-English speaking, cognitive impairment, consideration of non-verbal communication) - actively explains to clients & relevant others their role in care, decision-making & preventing adverse events	**COMMUNICATION** or Professional Behaviour	Y	Y
Extension marks – communication	Item 5: Communicates effectively and appropriately - Verbal/non-verbal (all performance indicators)	**COMMUNICATION**	Y	N
Gains consent	Item 1: Demonstrates an understanding of client rights and consent - obtains & records informed consent according to protocol - allows sufficient time to discuss the risks & benefits of the proposed treatment with clients & carers	**PROFESSIONAL BEHAVIOUR**	Y	Y

Code and Naturally Occurring Theme	Relevant APP Items and/or Performance Indicators	APP Domain	Code applied to OSCE Related (R)	Unrelated (UnR)
Gives clear instructions	Item 5: Communicates effectively and appropriately – Verbal/non-verbal - provides clear instructions - uses suitable language & avoids jargon - uses a range of communication strategies to optimise client rapport & understanding (e.g. hearing impairment, non-English speaking, cognitive impairment, consideration of non-verbal communication)	**COMMUNICATION**	N	Y
ASSESSMENT THEME				
Demonstrates assessment	Item 7: Conducts an appropriate client interview - conducts appropriate assessment with consideration of the social, personal, environmental & biopsychosocial factors that influence function, health & disability - seeks appropriate supplementary information, accessing other information, records, test results as appropriate & with client's consent Item 8: Selects and measures relevant health indicators and outcomes - communicates the treatment evaluation process & outcomes to the client & relevant others Item 9: Performs appropriate physical assessment procedures - considers client comfort & safety - respects client's need for privacy & modesty (e.g., provides draping or gown) - structures systematic, safe & goal-oriented assessment processes accommodating limitations imposed by client's health status - demonstrates sensitive & appropriate handling during the assessment process - applies tests & measurements safely, accurately & consistently - sensibly modifies assessment in response to client profile, feedback & relevant findings - performs appropriate tests to refine diagnosis - completes assessment in acceptable time	**ASSESSMENT**	Y	Y
Demonstrates client interview	Item 7: Conducts an appropriate client interview (all performance indicators)	**ASSESSMENT**	Y	Y

Code and Naturally Occurring Theme	Relevant APP Items and/or Performance Indicators	APP Domain	Code applied to OSCE Related (R)	Unrelated (UnR)
Identifies appropriate assessment	Item 7: Conducts an appropriate client interview - asks relevant & comprehensive questions - responds appropriately to important client cues - seeks appropriate supplementary information, accessing other information, records, test results as appropriate & with client's consent - generates diagnostic hypotheses, identifying priorities & urgency of further assessment & intervention Item 8: Selects and measures relevant health indicators and outcomes - selects appropriate variable/s to be measured at baseline from WHO ICF domains of impairment, activity limitation & participation restriction - identifies & justifies variables to be measured to monitor treatment response & outcome - selects appropriate tests/outcome measures for each variable for the purpose of diagnosis, monitoring & outcome evaluation Item 9: Performs appropriate physical assessment procedures - structures systematic, safe & goal-oriented assessment processes accommodating limitations imposed by client's health status - plans assessment structure & reasoning process using information from client history & supportive information - sensibly modifies assessment in response to client profile, feedback & relevant findings Item 20: Identifies adverse events and near misses and minimises risk associated with assessment and interventions - describes relevant contraindications & precautions associated with assessment & treatment	**ASSESSMENT** or Risk Management	Y	N
Monitors client during assessment	Item 7: Conducts an appropriate client interview - responds appropriately to important client cues	**ASSESSMENT** or Risk Management	Y	Y

Code and Naturally Occurring Theme	Relevant APP Items and/or Performance Indicators	APP Domain	Code applied to OSCE Related (R)	Unrelated (UnR)
	Item 8: Selects and measures relevant health indicators and outcomes - identifies & justifies variables to be measured to monitor treatment response & outcome - selects appropriate tests/outcome measures for each variable for the purpose of diagnosis, monitoring & outcome evaluation Item 9: Performs appropriate physical assessment procedures - demonstrates sensitive & appropriate handling during the assessment process - sensibly modifies assessment in response in response to client profile, feedback & relevant findings Item 20: Identifies adverse events and near misses and minimises risk associated with assessment and interventions - monitors client safety during assessment & treatment			
Concludes assessment	Item 9: Performs appropriate physical assessment procedures - considers client comfort & safety - respects client's need for privacy & modesty (e.g. provides draping or gown) - completes relevant documentation to the required standard (e.g. client record, statistical information, referral letters)	**ASSESSMENT**	N	Y
TREATMENT THEME				
Demonstrates intervention	Item 14: Performs interventions appropriately - demonstrates appropriate client handling skills in performance of interventions - performs techniques at appropriate standard - minimizes risk of adverse events to client & self in performance of intervention (including observance of infection control procedures &manual handling standards) - prepares environment for client including necessary equipment for treatment - completes intervention in acceptable time Item 16: Monitors the effect of intervention - makes modifications to intervention based on therapist evaluation & client feedback	**INTERVENTION**	Y	Y

Code and Naturally Occurring Theme	Relevant APP Items and/or Performance Indicators	APP Domain	Code applied to OSCE Related (R)	Unrelated (UnR)
	Item 17: Progresses intervention appropriately - demonstrates or describes safe & sensible treatment progressions - makes decisions regarding modifications, continuation or cessation of intervention in consultation with the client, based on best available evidence			
Identifies appropriate intervention	Item 4: Demonstrates collaborative practice - acknowledges expertise & role of other healthcare professionals & refers/liaises as appropriate to access relevant services Item 13: Selects appropriate intervention in collaboration with the client - all performance indicators relevant Item 14: Performs intervention appropriately - identifies when group activity might be an appropriate intervention - refers client to other professional/s when physiotherapy intervention is not appropriate, or requires a multi-disciplinary approach Item 17: Progresses intervention appropriately - demonstrates or describes safe & sensible treatment progressions - makes decisions regarding modifications, continuation or cessation of intervention in consultation with the client, based on best available evidence. Item 19: Applies evidenced based practice in client-centred care - considers the research evidence, client preferences, clinical expertise & available resources in making treatment decisions & advising clients Item 20: Identifies adverse events and near misses and minimises risk associated with assessment and interventions - describes relevant contraindications & precautions associated with assessment & treatment	**ANALYSIS & PLANNING** or Professional Behaviour or Intervention or Evidence-Based Practice or Risk Management	Y	N

Code and Naturally Occurring Theme	Relevant APP Items and/or Performance Indicators	APP Domain	Code applied to OSCE Related (R)	Unrelated (UnR)
Prescribes exercises	Item 13: Selects appropriate intervention in collaboration with the client - identifies & justifies options for interventions based on client needs, clinical guidelines, best evidence & available resources - demonstrates understanding of contraindications & precautions in selection of intervention strategies Item 14: Performs interventions appropriately - performs techniques at appropriate standard - minimizes risk of adverse events to client & self in performance of intervention (including observance of infection control procedures & manual handling standards) Item 15: Is an effective educator - provides information using a range of strategies that demonstrate consideration of client needs - confirms client's/relevant others' understanding of given information - uses appropriate strategies to motivate the client & relevant others to participate & to take responsibility for achieving defined goals - discusses expectations of physiotherapy intervention & its outcomes - educates the client in self evaluation	**ANALYSIS & PLANNING** or Intervention	Y	Y
Monitors client during intervention	Item 8: Selects and measures relevant health indicators and outcomes - identifies & justifies variables to be measured to monitor treatment response & outcome - selects appropriate tests/outcome measures for each variable for the purpose of diagnosis, monitoring & outcome evaluation Item 16: Monitors the effect of intervention (all performance indicators) Item 20: Identifies adverse events and near misses and minimises risk associated with assessment and interventions - monitors client safety during assessment & treatment	**INTERVENTION** or Assessment or Risk Management	Y	Y

Code and Naturally Occurring Theme	Relevant APP Items and/or Performance Indicators	APP Domain	Code applied to OSCE Related (R)	Unrelated (UnR)
Coordinates intervention	Items 4: Demonstrates collaborative practice - demonstrates understanding of team processes - acknowledges expertise & role of other healthcare professionals & refers/liaises as appropriate to access relevant services - cooperates with other people who are treating & caring for clients - guides & motivates support staff (where appropriate) - works collaboratively & respectfully with support staff Item 5: Communicates effectively and appropriately – Verbal/non-verbal - gives appropriate, positive reinforcement - provides clear instructions - uses suitable language & avoids jargon - actively explains to clients & relevant others their role in care, decision-making & preventing adverse events Item 13: Selects appropriate intervention in collaboration with the client - identifies & justifies options for interventions based on client needs, clinical guidelines, best evidence & available resources - demonstrates a suitable range of skills & approaches to intervention - balances needs of clients & relevant others with the need for efficient & effective intervention - demonstrates understanding of contraindications & precautions in selection of intervention strategies Item 14: Performs interventions appropriately - considers the scheduling of treatment in relation to other procedures e.g. medication for pain, wound care. - demonstrates appropriate client handling skills in performance of interventions - performs techniques at appropriate standard - minimizes risk of adverse events to client & self in performance of intervention (including observance of infection control procedures & manual handling standards) - prepares environment for client including necessary equipment for treatment	**PROFESSIONAL BEAVIOUR** or **COMMUNICATION** or **INTERVENTION** or Analysis and Planning	N	Y

Code and Naturally Occurring Theme	Relevant APP Items and/or Performance Indicators	APP Domain	Code applied to OSCE Related (R)	Unrelated (UnR)
	Item 15: Is an effective educator - applies adult learning principles in education of clients & relevant others - educates assistants & relevant others to implement safe & effective therapy - confirms client's/relevant others' understanding of given information			
SAFETY THEME				
Ensures client safety	Item 3: Demonstrates ethical, legal & culturally responsive practice - observes infection control, & workplace health & safety policies - demonstrates skills in culturally safe & responsive client-centred practice - acts within bounds of personal competence, recognizing personal & professional strengths & limitations Item 6: Demonstrates clear and accurate documentation - seeks appropriate supplementary information, accessing other information, records, test results as appropriate &with client's consent Item 7: Conducts and appropriate client-centred interview 7 - positions person safely & comfortably for interview - provides a culturally safe environment for the client Item 8: Selects and measures relevant health indicators and outcomes - identifies & justifies variables to be measured to monitor treatment response & outcome. Item 9: Performs appropriate physical assessment procedures - considers client comfort & safety - structures systematic, safe & goal-oriented assessment processes accommodating limitations imposed by client's health status - applies tests & measurements safely, accurately & consistently Item 11: Identifies and prioritises client's problems - justifies prioritisation of problem list based on knowledge & clinical reasoning	**ASSESSMENT** or Professional Behaviour or Communication or Analysis & Planning or Intervention or Risk Management	Y	Y

Code and Naturally Occurring Theme	Relevant APP Items and/or Performance Indicators	APP Domain	Code applied to OSCE Related (R)	Unrelated (UnR)
	Item 13: Selects appropriate intervention in collaboration with the client - demonstrates understanding of contraindications & precautions in selection of intervention strategies Item 14: Performs interventions appropriately - minimizes risk of adverse events to client & self in performance of intervention (including observance of infection control procedures & manual handling standards) Item 15: Is an effective educator - educates assistants & relevant others to implement safe & effective therapy Item 17: Progresses the intervention appropriately - demonstrates or describes safe & sensible treatment progressions Item 20: Identifies adverse events/near misses and minimises risk associated with assessment and interventions (all performance indicators)			
Identifies risks	Item 1: Demonstrates an understanding of client rights and consent - advises supervisor or other appropriate person if a client might be at risk Item 9: Performs appropriate physical assessment procedures - structures systematic, safe & goal-oriented assessment processes accommodating limitations imposed by client's health status - sensibly modifies assessment in response to client profile, feedback & relevant findings Item 11: Identifies and prioritises client's problems - generates a list of problems from the assessment - justifies prioritisation of problem list based on knowledge & clinical reasoning - collaborates with client to prioritise problems	**INTERVENTION** or Professional Behaviour or Assessment or Analysis & Planning or Risk Management	Y	N

Code and Naturally Occurring Theme	Relevant APP Items and/or Performance Indicators	APP Domain	Code applied to OSCE Related (R)	Unrelated (UnR)
	Item 13: Selects appropriate intervention in collaboration with the client - demonstrates understanding of contraindications & precautions in selection of intervention strategies Item 14: Performs interventions appropriately - minimizes risk of adverse events to client & self in performance of intervention (including observance of infection control procedures & manual handling standards) Item 16: Monitors the effects of the intervention - makes modifications to intervention based on therapist evaluation & client feedback Item 17: Progresses intervention appropriately - makes decisions regarding modifications, continuation or cessation of intervention in consultation with the client, based on best available evidence Item 20: Identifies adverse events and near misses and minimises risk associated with assessment and interventions - describes relevant contraindications & precautions associated with assessment & treatment - recognises & reports adverse events & near misses to appropriate members of the team - reports inappropriate or unsafe behaviour of a co-worker or situations that are unsafe - prior to client contact, reports any personal issues(physical/mental) that may impact on client care			
Is safe	Item 3: Demonstrates ethical, legal & culturally responsive practice - follows policies & procedures of the facility - advises appropriate staff of circumstances that may affect adequate work performance - observes infection control, & workplace health & safety policies - arrives fit to work	**ASSESSMENT** or **INTERVENTION** or Professional Behaviour	Y	N

Code and Naturally Occurring Theme	Relevant APP Items and/or Performance Indicators	APP Domain	Code applied to OSCE Related (R)	Unrelated (UnR)
	- demonstrates skills in culturally safe & responsive client-centred practice acts within bounds of personal competence, recognizing personal & professional strengths& limitations Item 7: Conducts an appropriate client-centred interview - positions person safely & comfortably for interview - provides a culturally safe environment for the client Item 9: Performs appropriate physical assessment procedures - considers client comfort & safety - structures systematic, safe & goal-oriented assessment processes accommodating limitations imposed by client's health status - applies tests & measurements safely, accurately & consistently Item 13: Selects appropriate intervention in collaboration with the client - demonstrates understanding of contraindications & precautions in selection of intervention strategies Item 14: Performs interventions appropriately - minimizes risk of adverse events to client & self in performance of intervention (including observance of infection control procedures &manual handling standards) - demonstrates appropriate client handling skills in performance of interventions - recognises when to enlist assistance of others to complete workload Item 15: Is an effective educator - educates assistants & relevant others to implement safe & effective therapy Item 17: Progresses intervention appropriately - demonstrates or describes safe & sensible treatment progressions Item 20: Identifies adverse events/near misses and minimises risk associated with assessment and interventions - all performance indicators relevant	or Analysis & Planning or Risk Management		

Code and Naturally Occurring Theme	Relevant APP Items and/or Performance Indicators	APP Domain	Code applied to OSCE Related (R)	Unrelated (UnR)
Practices infection control	Item 3: Demonstrates ethical, legal & culturally responsive practice - Follows policies and procedures of the facility - observes infections control, & workplace health & safety policies Item 14: Performs interventions appropriately - minimizes risk of adverse events to client & self in performance of intervention (including observance of infection control procedures & manual handling standards) Item 20: Identifies adverse events and near misses and minimises risk associated with assessment and interventions - complies with organizational health & safety requirements	**PROFESSIONAL BEHAVIOUR** or Intervention or Risk Management	Y	Y
Safety	Item 3: Demonstrates ethical, legal & culturally responsive practice - follows policies & procedures of the facility - advises appropriate staff of circumstances that may affect adequate work performance - observes infection control, & workplace health & safety policies - arrives fit to work - recognises inappropriate or unethical health practice - demonstrates skills in culturally safe & responsive client-centred practice - acts within bounds of personal competence, recognizing personal & professional strengths & limitations Item 6: Demonstrates clear and accurate documentation - seeks appropriate supplementary information, accessing other information, records, test results as appropriate & with client's consent Item 7: Conducts and appropriate client-centred interview - positions person safely & comfortably for interview - provides a culturally safe environment for the client Item 8: Selects and measures relevant health indicators and outcomes - identifies & justifies variables to be measured to monitor treatment response & outcome.	**INTERVENTION** or **RISK MANAGEMENT** Professional Behaviour or Assessment or Analysis & Planning	N	Y

Code and Naturally Occurring Theme	Relevant APP Items and/or Performance Indicators	APP Domain	Code applied to OSCE	
			Related (R)	Unrelated (UnR)
	Item 9: Performs appropriate physical assessment procedures - considers client comfort & safety - structures systematic, safe & goal-oriented assessment processes accommodating limitations imposed by client's health status - applies tests & measurements safely, accurately & consistently Item 11: Identifies and prioritises client's problems - justifies prioritisation of problem list based on knowledge & clinical reasoning Item 13: Selects appropriate intervention in collaboration with the client - demonstrates understanding of contraindications & precautions in selection of intervention strategies Item 14: Performs interventions appropriately - minimizes risk of adverse events to client & self in performance of intervention (including observance of infection control procedures & manual handling standards) Item 15: Is an effective educator - educates assistants & relevant others to implement safe & effective therapy Item 17: Progresses the intervention appropriately - demonstrates or describes safe & sensible treatment progressions - makes decisions regarding modifications, continuation or cessation of intervention in consultation with the client, based on best available evidence Item 20: Identifies adverse events and near misses and minimises risk associated with assessment and interventions - complies with workplace guidelines on manual handling - complies with organizational health & safety requirements			

Code and Naturally Occurring Theme	Relevant APP Items and/or Performance Indicators	APP Domain	Code applied to OSCE Related (R)	Unrelated (UnR)
KNOWLEDGE THEME				
Applies knowledge	Item 7: Conducts and appropriate client-centred interview - positions person safely & comfortably for interview - generates diagnostic hypotheses, identifying priorities & urgency of further assessment & intervention Item 8: Selects and measures relevant health indicators and outcomes - selects appropriate variable/s to be measured at baseline from WHO ICF domains of impairment, activity limitation & participation restriction - identifies & justifies variables to be measured to monitor treatment response & outcome - selects appropriate tests/outcome measures for each variable for the purpose of diagnosis, monitoring & outcome evaluation - links outcome variables with treatment goals Item 9: Performs appropriate physical assessment procedures - structures systematic, safe & goal-oriented assessment processes accommodating limitations imposed by client's health status - plans assessment structure & reasoning process using information from client history & supportive information - applies tests & measurements safely, accurately & consistently Item 10: Appropriately interprets assessment findings - describes the implications of test results - describes the presentation & expected course of common clinical conditions - relates signs & symptoms to pathology - relates signs, symptoms & pathology to environmental tasks & demands - interprets findings at each stage of assessment to progressively negate or reinforce hypothesis/es - makes justifiable decisions regarding diagnoses based on knowledge & clinical reasoning	**ANALYSIS AND PLANNING** or **INTERVENTION** or Assessment or Risk Management	N	Y

Code and Naturally Occurring Theme	Relevant APP Items and/or Performance Indicators	APP Domain	Code applied to OSCE Related (R)	Unrelated (UnR)
	Item 11: Identifies and prioritises client's problems - generates a list of problems from the assessment - justifies prioritisation of problem list based on knowledge & clinical reasoning; Item 13: Selects appropriate intervention in collaboration with the client - identifies & justifies options for interventions based on client needs, clinical guidelines, best evidence & available resources - describes acceptable rationale (e.g. likely effectiveness) for treatment choices - demonstrates understanding of contraindications & precautions in selection of intervention strategies Item 14: Performs interventions appropriately - identifies when group activity might be an appropriate intervention Item 15: Is an effective educator - applies adult learning principles in education of clients & relevant others - develops a realistic self-management program for prevention & management in collaboration with the client Item 16: Monitors the effect of intervention - incorporates relevant evaluation procedures/outcome measures in the physiotherapy plan Item 17: Progresses the intervention appropriately - demonstrates or describes safe & sensible treatment progressions - makes decisions regarding modifications, continuation or cessation of intervention in consultation with the client, based on best available evidence Item 18: Undertakes discharge planning - describes strategies that may be useful for maintaining or improving health status following discharge			

Code and Naturally Occurring Theme	Relevant APP Items and/or Performance Indicators	APP Domain	Code applied to OSCE Related (R)	Unrelated (UnR)
	Item 20: Identifies adverse events and near misses and minimises risk associated with assessment and interventions - describes relevant contraindications & precautions associated with assessment & treatment			
Goal Setting	Item 7: Conducts and appropriate client-centred interview - identifies client's goals & expectations Item 12: Sets realistic short and long term goals with the client - formulates goals that are specific, measurable, achievable & relevant, with specified timeframe - considers physical, emotional & financial costs, & relates them to likely gains of intervention Item 15: Is an effective educator - encourages & acknowledges achievement of short & long term goals	**ANALYSIS AND PLANNING** or Assessment or Intervention	N	Y
Interprets results	Item 7: Conducts an appropriate client-centred interview - generates diagnostic hypotheses, identifying priorities & urgency of further assessment & intervention Item 8: Selects and measures relevant health indicators and outcomes - identifies, documents & acts on factors that may compromise treatment outcomes - links outcome variables with treatment goals Item 9: Performs appropriate physical assessment procedures - plans assessment structure & reasoning process using information from client history & supportive information Item 10: Appropriately interprets assessment findings - describes the implications of test results - relates signs & symptoms to pathology - relates signs, symptoms & pathology to environmental tasks & demands	**ANALYSIS & PLANNING** or Assessment	Y	N

Code and Naturally Occurring Theme	Relevant APP Items and/or Performance Indicators	APP Domain	Code applied to OSCE Related (R)	Unrelated (UnR)
	- interprets findings at each stage of assessment to progressively negate or reinforce hypothesis/es - prioritises important assessment findings Item 11: Identifies and prioritises client's problems - generates a list of problems from the assessment - justifies prioritisation of problem list based on knowledge & clinical reasoning			
States correct theoretical knowledge of assessment	Item 8: Selects and measures relevant health indicators and outcomes - selects appropriate variable/s to be measured at baseline from WHO ICF domains of impairment, activity limitation & participation restriction - identifies & justifies variables to be measured to monitor treatment response & outcome - selects appropriate tests/outcome measures for each variable for the purpose of diagnosis, monitoring & outcome evaluation Item 9: Performs appropriate physical assessment procedures - structures systematic, safe & goal-oriented assessment processes accommodating limitations imposed by client's health status - sensibly modifies assessment in response to client profile, feedback & relevant findings Item 10: Appropriately interprets assessment findings - describes the implications of test results - relates signs & symptoms to pathology - relates signs, symptoms & pathology to environmental tasks & demands - interprets findings at each stage of assessment to progressively negate or reinforce hypothesis/es Item 16: Monitors the effects of intervention - incorporates relevant evaluation procedures/outcome measures in the physiotherapy plan	**ASSESSMENT** or **ANALYSIS & PLANNING** or Intervention	Y	N

Code and Naturally Occurring Theme	Relevant APP Items and/or Performance Indicators	APP Domain	Code applied to OSCE Related (R)	Unrelated (UnR)
	Item 19: Applies evidence-based practice in client-centred care - considers the research evidence, client preferences, clinical expertise & available resources in making treatment decisions & advising clients Item 20: Identifies adverse events and near misses and minimises risk associated with assessment and interventions - describes relevant contraindications & precautions associated with assessment & treatment			
Patient positioning	Item 7: Conducts and appropriate client-centred interview - positions person safely & comfortably for interview Item 9: Performs appropriate physical assessment procedures - considers client comfort & safety; - respects client's need for privacy & modesty (e.g., provides draping or gown)	**ASSESSMENT**	N	Y
States correct Theoretical knowledge	Item 6: Demonstrates clear and accurate documentation - incorporates relevant evaluation procedures/outcome measures in the physiotherapy plan Item 8: Selects and measures relevant health indicators and outcomes - selects appropriate variable/s to be measured at baseline from WHO ICF domains of impairment, activity limitation & participation restriction - selects appropriate tests/outcome measures for each variable for the purpose of diagnosis, monitoring & outcome evaluation Item 9: Performs appropriate physical assessment procedures - structures systematic, safe & goal-oriented assessment processes accommodating limitations imposed by client's health status - sensibly modifies assessment in response to client profile, feedback & relevant findings	**ASSESSMENT** or **ANALYSIS AND PLANNING** or Intervention or Risk Management	N	Y

Code and Naturally Occurring Theme	Relevant APP Items and/or Performance Indicators	APP Domain	Code applied to OSCE Related (R) Unrelated (UnR)
	Item 10: Appropriately interprets assessment findings - describes the implications of test results - interprets findings at each stage of assessment to progressively negate or reinforce hypothesis/es; - relates signs & symptoms to pathology - relates signs, symptoms & pathology to environmental tasks & demands Item 11: Identifies and prioritises client's problems - generates a list of problems from the assessment Item 13: Selects appropriate intervention in collaboration with the client - demonstrates understanding of contraindications & precautions in selection of intervention strategies - selects appropriate intervention in collaboration with the client Item 14: Performs interventions appropriately - identifies when group activity might be an appropriate intervention Item 15: Is an effective educator - discusses expectations of physiotherapy intervention & its outcomes Item 16: Monitors the effect of intervention - incorporates relevant evaluation procedures/outcome measures in the physiotherapy plan Item 17: Progresses intervention appropriately - makes decisions regarding modifications, continuation or cessation of intervention in consultation with the client, based on best available evidence Item 19: Applies evidence-based practice in client-centred care - considers the research evidence, client preferences, clinical expertise & available resources in making treatment decisions & advising clients - assists clients & carers to identify reliable & accurate health information		

Code and Naturally Occurring Theme	Relevant APP Items and/or Performance Indicators	APP Domain	Code applied to OSCE Related (R)	Unrelated (UnR)
	Item 20: Identifies adverse events and near misses and minimises risk associated with assessment and interventions. - describes relevant contraindications & precautions associated with assessment & treatment			
EDUCATION THEME				
Educates client	Item 5: Communicates effectively and appropriately – Verbal/non-verbal - provides clear instructions - uses suitable language & avoids jargon - integrates communication technology into practice as required - uses a range of communication strategies to optimise client rapport & understanding (e.g. hearing impairment, non-English speaking, cognitive impairment, consideration of non-verbal communication) - actively explains to clients & relevant others their role in care, decision-making & preventing adverse events Item 6: Demonstrates clear and accurate documentation - adapts written material for a range of audiences (e.g., provides translated material for non-English speaking people, considers reading ability, age of client) Item 13: Selects appropriate intervention in collaboration with the client - engages with client to explain assessment findings, discuss intervention strategies &develop an acceptable plan - advises client about the effects of treatment or no treatment Item 15: Is an effective educator - demonstrates skill in client education & health promotion e.g., modifies approach to suit client age group &/or cultural needs - applies adult learning principles in education of clients & relevant others	**INTERVENTION** or Communication or Analysis & Planning or Risk Management	Y	Y

Code and Naturally Occurring Theme	Relevant APP Items and/or Performance Indicators	APP Domain	Code applied to OSCE Related (R)	Unrelated (UnR)
	- develops a realistic self-management program for prevention & management in collaboration with the client - provides information using a range of strategies that demonstrate consideration of client needs - confirms client's/relevant others' understanding of given information - uses appropriate strategies to motivate the client & relevant others to participate & to take responsibility for achieving defined goals - discusses expectations of physiotherapy intervention & its outcomes - provides feedback to client regarding health status - educates the client in self evaluation - encourages & acknowledges achievement of short- & long-term goals Item 19: Applies evidence-based practice in client-centred care - considers the research evidence, client preferences, clinical expertise & available resources in making treatment decisions & advising clients - assists clients & carers to identify reliable & accurate health information			
Coaching	Item 5: Communicates effectively and appropriately – Verbal/non-verbal - gives appropriate, positive reinforcement - provides clear instructions - uses suitable language & avoids jargon Item 11: Identifies and prioritises client's problems - collaborates with client to prioritise problems - negotiates realistic short term treatment goals in partnership with client - negotiates realistic long term treatment goals in partnership with client Item 13: Selects appropriate intervention in collaboration with the client - engages with client to explain assessment findings, discuss intervention strategies & develop an acceptable plan - advises client about the effects of treatment or no treatment	**INTERVENTION** Or Communication or Analysis and Planning	N	Y

Code and Naturally Occurring Theme	Relevant APP Items and/or Performance Indicators	APP Domain	Code applied to OSCE Related (R)	Unrelated (UnR)
	Item 15: Is an effective educator			
	- demonstrates skill in client education & health promotion e.g., modifies approach to suit client age group &/or cultural needs			
	- applies adult learning principles in education of clients & relevant others			
	- develops a realistic self-management program for prevention & management in collaboration with the client			
	- confirms client's/relevant others' understanding of given information			
	- uses appropriate strategies to motivate the client & relevant others to participate & to take responsibility for achieving defined goals			
	- discusses expectations of physiotherapy intervention & its outcomes			
	- provides feedback to client regarding health status			
	- encourages & acknowledges achievement of short & long term goals			
	Item 17: Progresses intervention appropriately			
	- makes decisions regarding modifications, continuation or cessation of intervention in consultation with the client, based on best available evidence			

OSCE content within the *communication* theme aligned with performance indicators in the *communication* and *professional behaviour* domains of the APP. Mark sheet items coded as 'communicates role' and 'communicates clearly' were determined to best align with the *communication* domain of the APP, while items coded as 'gains consent' best aligned to the *professional behaviour* domain. Content within the *assessment* theme was described by performance indicators in both the *assessment* and *risk management* domains of the APP, but best aligned with those in the *assessment* domain.

OSCE content within the *treatment* theme was described by performance indicators across the *professional behaviour, analysis and planning*, *intervention* and *risk management* domains of the APP. Items coded as 'identifies appropriate intervention' and 'prescribes exercises' were determined to best align with the *analysis and planning* domain of the APP, while items coded as 'demonstrates intervention' best aligned to the *intervention* domain of the APP.

No codes aligned specifically to the *evidenced-based practice* domain of the APP. Codes in the *safety* theme aligned to performance indicators across all domains of the APP except *communication* and *evidence-based practice* themes.

The proportions of each APP domain examined within the OSCEs are displayed in Table 24. Three codes, totaling 7% of OSCE UnR could not be aligned to a single domain of the APP and have been noted as being 'Indeterminate'. These codes were 'safety', 'coordinates intervention' and 'applied knowledge'.

Table 24. A comparison of the proportion of OSCE content described by APP domain

APP Domain	Percentage of OSCE Content	
	OSCE R (<u>related</u> to future clinical performance)	OSCE UnR (<u>unrelated</u> to future clinical performance)
Assessment	47%	28%
Intervention	29%	30%
Analysis & Planning	11%	13%
Communication	9%	17%
Professional Behaviour	4%	6%
Evidence-Based Practice	0%	0%
Risk Management	0%	0%
Indeterminate	0%	7%
TOTAL	100%	101%*

*Total does not equal 100% due to rounding to nearest whole percentages

6.6 Discussion

The two OSCEs investigated had previously been demonstrated (see Chapter 5) to have: a moderately strong relationship with students' future clinical performance (OSCE R: $r_s = 0.42$, $p < 0.001$); and no significant relationship with students' future clinical performance as measured by the APP (OSCE UnR: $r_s = -0.03$, $p = 0.78$). Following a content analysis approach, the OSCEs were coded and mapped to the domains of practice of the APP. Neither of the coders were involved with the design, development, or implementation of the OSCEs investigated. Nor did the coders teach into the subjects assessed by the OSCEs, with the coders each from different clinical backgrounds. Despite this, there was substantial agreement between first cycle codes assigned by the two coders for 78% of the items in OSCE R. This suggests that there was a clearly discernible construct being assessed by those items. Discussion was needed between the two investigators for 22% of the items in OSCE R to determine the second cycle coding. Consensus was achieved for all items without the need for a moderator. This suggests that the disagreements were predominately semantic, rather than about the construct being assessed.

When coding OSCE UnR, there was a 67% agreement on first order codes between the two coders on the constructs assessed by the items on the OSCE mark sheet. Consensus was achieved for the remaining 33% without need for a moderator. However, for 12% of the mark sheet items, while the two reviewers agreed on a second cycle code, they also agreed that there was more than one clear construct being assessed, even when broader more complex constructs were considered.

While consensus was achieved on the semantic meaning of all OSCE items, discussions were required. Daniels and Pugh (2018) highlight that OSCE scoring rubrics should be unambiguous and clearly outline the operational definition of success for checklist items. When alternative interpretations of a successful performance on checklist items are possible, this impacts on the validity of ratings made by examiners. Training OSCE assessors is an important component of implementing an OSCE (Daniels & Pugh, 2018) and may help mitigate the effect of ambiguous items if the OSCE developer explains the intention of the items and how scores are to be applied (Schwartzman et al., 2011). However, this assumes the assessor will encode the training material into a mental scoring rubric that matches that of the trainer (Bejar, 2012). As discussed by Bejar (2012), the influence of an assessor's mental rubric would be particularly relevant where a station requires the presentation of a spoken constructed response. In this instance, if there is an ambiguous checklist item, the personal attributes of an assessor may lead to a mental rubric that contains components that are not explicitly listed, or even expected to be demonstrated.

The 12% of the items on the mark sheets of OSCE UnR that did not assess a single clear construct of the OSCE were ambiguous. OSCE R items that were coded either as *answers extension questions* or *extension marks - communication* (see Table 20) during second cycle coding were also a source of ambiguity and required considerable discussion between the two coders. The generic second cycle codes applied were eventually decided upon because these items did not have a single clear construct that was being assessed. Chahine, Holmes, and Kowalewski (2016, p. 611) stated in their paper investigating the hidden assumptions of OSCE examiners that, "in order to provide precise estimates of candidate performance, two elements at the very least are needed: congruent definitions

of a construct... and consistent measurement". OSCE developers should refer to the OSCE blueprint to clarify the intention behind such items (Khan, Gaunt, et al., 2013), explicitly train assessors as to the aim of the OSCE (van der Want, Bloemendaal, & van der Hage, 2021), and ensure the scoring rubric accurately defines both the construct and the definition of a successful performance.

Ambiguous items on the OSCEs were not accompanied by instructions to outline the distribution of marks allocated to the item, nor what a successful performance should look like. Additionally, 31% of the total OSCE UnR content fell within the combined *communication* and *education* themes, compared to 17% of OSCE R. Mark sheet items assessing constructs within these two themes were not accompanied by a standard of what successful communication and education should look like. This impacts the validity of the ratings made by assessors for these items (Daniels & Pugh, 2018) and may contribute to the reason OSCE UnR was unrelated to student performance in subsequent clinical practice.

There were six themes within the constructs assessed by the OSCE that naturally emerged from the second cycle coding: *communication, assessment, treatment, safety, education,* and *knowledge.* Education as a construct is complex, as it incorporates the elements of *knowledge*, *intervention*, and *communication*, which are constructs in their own right. There was not one single station that assessed students' ability to provide patient education, rather this was incorporated as one or more items within a station that also assessed other skills. Effective education requires the correct theoretical knowledge of what to advise the client, combined with the ability to educate effectively using appropriate communication and teaching methods (Candela et al., 2018). Therefore, when asked to provide patient education within an OSCE station, students are required to demonstrate the effective application of multiple cognitive and psychomotor skills. Checklist style marking may reduce inter-rater variability (Regehr et al., 1998), however the determination of whether the delivery of patient education was effective or not may lend itself more to the holistic appraisal that can be achieved using a global rating scale (Swanson & van der Vleuten, 2013). Swanson and van der Vleuten (2013) highlighted the risk of checklists rewarding rote-learning and mechanical performance. In OSCEs utilising checklists for marking, items assessing education skills may be better separated into

component parts to achieve a more accurate and reliable assessment of students' ability rather than a single checklist item rewarding a trivial, or mechanical, attempt to provide education.

Of the sample OSCE content, 17% of the content within OSCE UnR fell within the knowledge theme compared to 8% of OSCE R. The assessment of cognitive skills could feasibly be assessed in a written format (Khan, Gaunt, et al., 2013). Removing theoretical knowledge questions from the OSCE and incorporating them instead into a written assessment provides an opportunity to increase the sampling of psychomotor skills and non-clinical skills. Appropriate sampling is required to ensure a reliable generalisation of the student's overall competence (Boursicot, 2010). While knowledge is a critical component of the clinical reasoning process, the OSCE aims to assess that a student can appropriately apply that knowledge at the 'shows how' level of Miller's pyramid (Wass et al, 2001). Removing items that award marks for theoretical knowledge without requiring the student to demonstrate their application of that knowledge (such as through the synthesisation of assessment findings or their responses to client cues/behaviours), allows increased sampling of clinical reasoning and practical skills within the OSCE. Factual knowledge is essential, but the recognition of when that knowledge is applicable and the appropriate implementation of that knowledge via a clinical reasoning process is representative of what students must demonstrate within clinical practice; having knowledge does not equate to having the ability to meet minimum standards of practice if students cannot apply that knowledge safely and effectively to the client or case in front of them (Kiesewetter et al., 2016).

The performance indicators that accompany the APP are a list of observable behaviours that can be considered when scoring students' performance on each item of the APP. This list is not exhaustive, but it is comprehensive. Of the 20 second cycle codes utilised in OSCE R, 13 of these were described by performance indicators across more than one domain of the APP. In each of these cases second cycle codes could be clearly aligned with one domain that fit best. The exception was the code *states correct theoretical knowledge of assessment,* which could not be clearly aligned to one domain of practice without referring to the original student question. The difficulty in determining the construct as it related to the practical application of the knowledge examined by these items is

further indication that these items warrant review. As discussed by Chahine et al. (2016), a congruent understanding of the construct to be assessed is an essential element for accurate measurement of student performance.

When analysed according to naturally occurring themes, 41% of OSCE R fell within the *assessment* theme. When the OSCE R content was aligned to domains of the APP, the proportion of content in the APP domain *'assessment'* increased to 47%. This is compared to 28% of the content of OSCE UnR. As OSCEs should be blueprinted to the curriculum covered within a course (Boursicot, 2010), this should be interpreted as being representative of the emphasis on assessment skills within the course in question *if* accurate blueprinting has occurred. Physiotherapists in Australia and New Zealand are first-contact practitioners, so well-developed assessment skills are imperative. While it is expected that new physiotherapists may require support to manage complex clients, it is essential that they are able to conduct a safe and effective assessment at a basic level and be able to communicate their findings and hypotheses to a senior therapist (Australian Physiotherapy Council, 2017; Physiotherapy Board of Australia & Physiotherapy Board of New Zealand, 2015). While there are only three items in the *assessment* domain of the APP (15% of the APP), an effective assessment is the foundation of being able to manage a client appropriately. A poor initial assessment may negatively impact on the physiotherapists ability to provide an effective intervention (Monie, Fazey, & Singer, 2016). Conversely, strong assessment skills can lead to enhanced patient outcomes (Bird, Thompson, & Williams, 2016; Sutton, Govier, Prince, & Morphett, 2015). Compared to OSCE R which had 60% of its content focused on assessment and treatment skills, OSCE UnR dedicated 41% of its content to the assessment of these skills. The coursework subject being assessed by OSCE UnR was a pre-clinical subject for a specific area of practice. It was not an introductory subject, nor a subject with a specific focus on communication and therapeutic relationships. Educators should ensure that blueprinting occurs so that the content of the OSCE accurately represents the constructs contained within the unit of study.

The content analysis identified that 43% of the OSCE UnR marks was dedicated to constructs in the themes of *communication, education, and safety.* It has been suggested that when assessing a student's performance on OSCE in these areas, the use of global rating scales may be a more appropriate method of assessing student performance (Khan,

Gaunt, et al., 2013; Swanson & van der Vleuten, 2013). In a global rating scale, it can be discriminated whether a student is 'safe' or 'unsafe', or 'effective or 'ineffective'. The risk of mark sheet checklists rewarding rote-learning have been highlighted (Swanson & van der Vleuten, 2013). Checklists may encourage students to memorise the steps of a task, rather than achieving a deeper understanding of the reasoning or application of a skill that could be best measured by a rating scale. Given that no marking rubric detailing the expected standard of performance accompanied the mark sheets in OSCE UnR, it can be surmised that up to 43% of the OSCE may have compromised validity. This is consistent with the finding that scores achieved in OSCE UnR were an unreliable assessment of students' actual ability to apply these skills in a clinical setting, given there was no relationship between students' scores on OSCE UnR and their actual performance in clinical practice (Chapter 5, pp. 121-22).

The APP describes performance indicators related to safe clinical practice across five domains: *professional behaviour, assessment, analysis and planning, intervention* and *risk management.* Items coded with the *safety* theme therefore could potentially align with any, or all, of these domains. This is reflective of the essential requirement that student clinicians achieve safe practice across all aspects of their interactions with clients and collaborators (Physiotherapy Board of Australia & Physiotherapy Board of New Zealand, 2015), and that patient safety is viewed as an implicit program outcome in health professions curricula (Cresswell et al., 2013). It also reinforces the need for university programs to appropriately assess students' ability to identify risks and act accordingly, prior to students entering a clinical practice environment. For each of the codes within the s*afety* theme of OSCE R there was a clear primary alignment with performance indicators in either the *assessment* (*ensures client safety*) or *intervention* domains (*identifies risks, is safe, and practices infection control*). This increased the proportion of content within OSCE R that aligned to the *assessment* and *intervention* domains of the APP, as the 16% that fell within the naturally occurring theme of *safety* aligned to one of these two domains.

Integrating the assessment of safe practice into multiple OSCE stations emphasises its importance. However, if safety is a key construct to be assessed, the OSCE design must discriminate between students who are 'safe' and 'unsafe' (Khan, Gaunt, et al., 2013). Developers should consider if their OSCE scoring rubrics adequately discriminate between safe and unsafe performance, or whether safety should be assessed at a dedicated

station. The inclusion of a global rating scale allows assesses to make an overall judgement of whether safe practice was demonstrated during an OSCE station. Global rating scales scored by experts have been demonstrated to have high inter-station reliability, better construct validity, and better concurrent validity than checklists alone (Regehr et al., 1998). This would also mirror the APP, which has a global rating scale to allow the rater to make an overall judgement of the students' performance in clinical practice as 'inadequate', 'adequate', 'good' or 'excellent'.

The APP is accompanied by a list of performance indicators which are a sample of observable behaviours that assist the assessor to score a students' performance. Of the 21 second-cycle codes in OSCE UnR, 13 of these were described by performance indicators across more than one domain of the APP. In ten of these cases, there was a clear alignment to one domain of the APP that best described the construct assessed by the mark sheet item. For three codes ('safety', 'coordinates intervention' and 'applied knowledge'), there was no clear alignment to one domain of the APP. These codes were classified as the domain being 'indeterminate'. Given that student performance cannot be accurately assessed if the construct being assessed is unclear (Chahine et al., 2016), this may explain why student performance in OSCE UnR did not relate to their future performance in clinical practice. Each of the three codes with an indeterminate domain of the APP were also identified as assessing multiple constructs within the one item. These items were intended to assess students' ability to demonstrate safe practice, the performance of a two-person intervention, and the application of theoretical knowledge to a case study. The APP has a clearly defined bench-mark of what constitutes a passing standard of performance: independent completion of tasks expected of a new-graduate physiotherapist on their first day of practice (Dalton et al., 2009). More specifically, this equates to being able to independently manage an uncomplicated client by performing a safe assessment, identifying the main problems, and delivering a safe and effective treatment in a reasonable timeframe. Without guidance of what the expected standard of performance in the OSCE looks like for these items, it is possible students were awarded passing marks in the OSCE when in fact they did not possess the skills to complete these tasks to an adequate standard in their future clinical practice.

Neither OSCE investigated contained content assessed in the *evidence-based practice* domain, or specific performance indicators included in the *risk management* domain of the APP. An OSCE is but one test within a program of assessment, so this finding should be considered within the context of that program. Van der Vleuten et al. (2012, p. 205) describes a program of assessment as "an arrangement of assessment methods planned to optimise its fitness for purpose". A program of assessment acknowledges that no one testing method can be expected to cover all competencies required to be acquired by a learner over a course of study. The subjects containing the OSCEs investigated in this paper also featured a written examination and a seminar assessment. The program of assessment containing the OSCE investigated should be considered to ensure that students' ability to consider and apply the principles of evidence-based practice is included in an appropriate assessment format. However, the inclusion of specific evidenced-based items into the OSCE should also be considered. Daniels et al. (2014) recommended that when developing OSCE checklists, preference should be given to evidence-based items over non-evidenced based items to increase the reliability of the OSCE. As assessment is acknowledged to be a key driver of student learning (van der Vleuten et al., 2012), emphasising evidence-based practice across multiple testing methods increases its salience for the learners.

As should all forms of assessment, OSCE checklists must discriminate between levels of student achievement. Considerations for educators constructing OSCE checklists have been discussed: the need for clear, unambiguous checklist items; inclusion of a global rating scale; and the inclusion of evidence-based items. Items on the OSCEs investigated that were most unambiguous and had the best agreement between coders were those focusing on gaining consent, practicing infection control, demonstrating a physical assessment technique and ensuring client safety during the performance of the technique, and interpreting results. Items that had the poorest agreement and required moderation were those that focused on the client interview, monitoring a client, demonstrating manual handling, identifying possible risks, extension marks, and items determined to be assessing multiple constructs. The items with poor agreement did not have a clear standard of performance written on the mark sheet. They also tended to be items that would lend themselves to the use of a global rating scale rating the effectiveness of the student performance, rather than just whether the item was performed or not. For

example, a checklist item that requires students to select and perform a specific manual handling technique, does not mean the technique is performed safely. Future research should investigate if these trends are consistent across OSCEs that are not related to students' future clinical practice performance.

The findings of this study demonstrate the challenges surrounding research into the use of OSCEs in physiotherapy and other health profession education programs. Research into two OSCEs from the same program, and the same cohort of students, had two very different outcomes. If either of these OSCEs were investigated in isolation the resultant publication could either support or discountenance the use of OSCE in physiotherapy education. However, the evidence suggests that a well-constructed OSCE is a valid method of assessment in an entry-level postgraduate physiotherapy program (Chapter 5, pp. 116-25). Thus, the findings of this study, not only explain why some of the discrepancies in the broader literature surrounding the use of OSCEs in physiotherapy education may exist, but more importantly why these discrepancies may exist.

Most studies exploring the use of OSCEs in physiotherapy programs have focused on the psychometric properties of the assessment (Gorman et al., 2010; Ladyshewsky et al., 2000; Silva et al., 2011). An OSCE investigated by Wessel et al. (2003) identified no relationship between the OSCE investigated and students' future clinical performance. The OSCE consisted of eight stations and marking was conducted using a checklist. These authors hypothesised that a greater number of stations may increase the validity of the OSCE and concluded that an eight-station OSCE should not be used in isolation to make decisions about clinical competence. However, a two-station OSCE was identified as having a moderately strong relationship with clinical performance (Chapter Five, pp. 121-22). The number of stations within an OSCE therefore should be but one consideration when designing an OSCE. In the study by Wessel et al. (2003) assessors were briefed prior to the OSCE, but no written guidance on the expected standard of performance was reported. Of note, students' clinical performance was assessed using the Clinical Performance Instrument (CPI), which consists of 24 items. Wessel et al. (2003) identified that only six items of these were addressed by the content assessed within the OSCE, which may have contributed to the lack of a relationships between student performance on this OSCE and in future clinical practice. Blueprinting an examination to the learning

outcomes of the course should occur to identify the key constructs and content required to be assessed, in addition to ensuring adequate sampling through an appropriate number of stations (Boursicot, 2010). Alternatively, where logistics dictate the number of stations an institution can feasibly administer within an OSCE, blueprinting should identify the number of skills that need to be covered with a single station and therefore the length of time that should be spent at each station.

6.6.1 Limitations

In physiotherapy education, the structure of OSCE stations vary from program to program, which creates a challenge when comparing between OSCEs if there is limited information published about the assessment. This makes benchmarking between institutions or applying the findings of the research difficult to achieve. Stations and items within stations in the OSCE investigated are still utilised in the host institutions bank of OSCE questions, so specific detail beyond what is included in this paper could potentially impact the host program. However, the detail contained in this paper with respect to the themes assessed within the OSCE, and their relationship to competencies assessed within the APP, should provide information to other institutions with which to compare the blueprinting and content of their own physiotherapy OSCEs.

A further limitation of this study is that coursework subjects delivered into the host program are all convened by different educators and are discreet, stand-alone subjects focused on a specific area of clinical practice. Student engagement and academic performance in these subjects can be influenced by students' own motivational beliefs and achievement emotions (Artino, La Rochelle, & Durning, 2010). It is possible that the engagement of some students may have differed between the subject featuring OSCE UnR, and the subject that featured OSCE R. Educator characteristics in the classroom may also influence OSCE outcomes (Anderson et al., 1991). However, educators may similarly impact on the student experience in clinical practice, and this could influence student outcomes on the Assessment of Physiotherapy Practice (Anderson et al., 1991).

6.7 Conclusion

This study investigated two OSCEs featured in an entry-level extended masters' physiotherapy education program, only one of which demonstrated construct validity. Skills relating to the assessment of a clients' condition was the main content area covered within the OSCE related to students' future clinical performance, constituting 41% of the content. This was followed by skills relating to treatment (19%) and safe practice (16%). Evidence-based practice was not explicitly assessed. Coders were more likely to agree on mark sheet items involving demonstrating physical assessment techniques, interpreting results, gaining consent and infection control. They were less likely to agree on items pertaining to the client interview, monitoring a client, demonstrating manual handling tasks, and items that assessed more than one construct within a single item. Compared to the content of the OSCE demonstrated construct validity, the OSCE that was unrelated to students' future clinical performance contained less content assessing specific assessment and treatment skills, but more content within the communication, education, and knowledge domains. Future research on the use of OSCEs in health education programs should consider that the design and the content of the OSCE may influence the validity of the OSCE investigated and take steps to mitigate this. Students should be adequately prepared for all domains of practice they are required to demonstrate on clinical placement, therefore any key domains of clinical practice not assessed within an OSCE should be assessed elsewhere within a program of assessment.

Chapter 7. Summation, discussion, and recommendations

7.1 Summation

The program of research reported in this thesis provides a unique and important contribution to the current body of literature regarding the assessment of postgraduate physiotherapy students, and the relationship of these assessments to students' performance in clinical practice settings. A critical review of the literature (Chapter 2) reporting on relationships between summative assessments and the future clinical performance of health profession students identified a paucity of research focused on the physiotherapy profession. In the subsequent studies (Chapters 3 - 6), data from four cohorts of students from a single institutions' entry-level postgraduate physiotherapy program were analysed to identify relationships between entry and summative assessment measures and students' future clinical performance. The main findings from this doctoral program of research were:

- Undergraduate grade-point average (UGPA) was not related to students' future performance in authentic clinical practice settings (measured using the Assessment of physiotherapy Practice (APP));
- There was no difference between students' future academic or clinical performance when grouped according to UGPA
- Summative assessments used in pre-clinical coursework subjects: Objective Structured Clinical Examinations (OSCEs), written assessments, and oral seminar presentations, were all related to students' future performance in authentic clinical practice settings;
- The OSCE had the strongest relationship with students' future clinical performance, but when explored individually not all OSCEs demonstrated predictive validity; and
- When an OSCE with moderate predictive validity was compared to an OSCE with no predictive validity the main differences were:
  - The OSCE with predictive validity had a higher proportion of content focusing on client assessment and intervention skills; and

- The OSCE that did not have predictive validity had a higher proportion of content assessing communication, education, and safety; more mark sheet items that were ambiguous as to the construct that was being assessed; and a higher proportion of theoretical knowledge items that could be assessed elsewhere in a program of assessment.

7.2 Discussion

Allied health professions are an important part of the health care system and play an essential role in maintaining the well-being of the community (Department of Health, 2022). Based on data from the 2019 census, there are 34 844 registered physiotherapy practitioners in Australia with an annual growth rate of 5.3%, making physiotherapy one of the fastest growing allied health professions in Australia (Department of Health, 2021). The Australian Physiotherapy Association reported a need for continued annual growth of 4.5% until 2024 to meet the forecast employment demands for physiotherapists (Australian Physiotherapy Association, 2021). Given this requirement, it is of critical importance that physiotherapy education programs continue to maintain the standards of the profession whilst meeting the demands of the community by graduating an increasing number of students.

Clinical education is a critical component of the development of health professionals (Downing & Yudkowsky, 2009). Authentic workplace experiences provide students the opportunity to practice their cognitive, psychomotor, and affective skills during the care of real clients, whilst developing their clinical reasoning processes with the supervision and support of experienced clinicians. These are essential learning experiences needed to prepare students for the responsibility of being independent autonomous practitioners after they graduate and become registered health professionals (Australian Physiotherapy Council, 2020). Predictive relationships with admission criteria and/or pre-clinical coursework assessments cannot negate the need for undertaking these critical components of a student's education. However, such relationships could provide educators with valuable information about a students' likely performance in the clinical practice setting. As forewarned is forearmed, a valid predictor of students' future clinical performance provides opportunities for programs to provide upskilling and enhancement strategies for those students at risk of poor performance. Likewise, a valid predictor of

future performance could identify students whose skills are sufficiently developed to benefit from being extended in their clinical practice experiences, and allow plans to be put in place to support this.

There is a relationship between the academic admission scores and the future academic performance of physiotherapy students in undergraduate programs (Edgar et al., 2014; Howard & Jerosch-Herold, 2000; Morris & Farmer, 1999; Paynter, Iles, & Hay, 2022; Watson et al., 2000). As such, Australian entry-level postgraduate physiotherapy programs commonly utilise prior academic performance to discriminate between candidates (Terry, Hing, Orr, & Milne, 2018). This approach is supported by research suggesting that undergraduate GPA (UGPA) is a predictor of academic performance in a postgraduate physiotherapy program in the United States of America (Roman & Buman, 2019). However, the findings detailed in Chapter 3 of this thesis suggests that there is no relationship between UGPA and the academic outcomes or clinical performance of the cohorts of physiotherapy students investigated. Factors distinct to Australian postgraduate entry-level students should be examined to explore this discrepancy. Possible factors that may enhance, or impair, a students' level of achievement within a postgraduate physiotherapy program include life experience, a dedication to the physiotherapy profession, and external commitments such as financial and family responsibilities.

At the time the research was undertaken, seven postgraduate entry-level physiotherapy programs in Australia specified minimum UGPA requirements to be eligible for admission, ranging from 4.5 - 5.5. A further three required a 'competitive GPA'. The findings of the research undertaken as part of this program of work (reported in Chapter 3) suggests that UGPA may not be an appropriate criterion to determine eligibility for admission into entry-level postgraduate physiotherapy education program in Australia if used as the sole method of selection. Furthermore, the use of a fixed or competitive UGPA as a method of student selection may rule-out applicants who would otherwise be valuable members of the profession. Thus, there is value in considering other approaches to ensure successful student selection. Wider multi-program investigations are also warranted to examine the generalisability of this finding.

Selection models that combine a range of appropriate selection criteria increase the probability of selecting students who will go on to achieve successful outcomes (Crawford,

Black, Melby, & Fitzpatrick, 2021), and may increase the diversity of the profession (Felix et al., 2012; Grabowski, 2018). In the study reported in Chapter 3 (Terry et al., 2018), the benefits of further consideration and investigation into the use of additional selection criteria was discussed. Since publication of that work, a second systematic review has verified conclusions that the Multiple Mini Interview (MMI) is a reliable, acceptable, and feasible method that could be incorporated into a model of student selection for physiotherapy programs (Yusoff, 2019), although no evidence supporting the use of the MMI specifically in the entry-level postgraduate physiotherapy context has been identified. Emotional intelligence and other non-cognitive factors were highlighted as factors to consider in the selection of health profession students (Prideaux et al., 2011), but recent evidence of predictive relationships with future academic or clinical performance is inconsistent (Chisholm-Burns, Berg-Poppe, Spivey, Karges-Brown, & Pithan, 2021; Reynolds, Bazemore, Hanebuth, Hendren, & Horn, 2021). The Grit Scale Score has been positively associated with postgraduate physiotherapy students final class rank and could be considered as a component of a model of selection (Carp, Fry, Gumerman, Pressley, & Whitman, 2020). The use of UGPA as a method of selection could be augmented with other options such as those discussed above. Proposed selection models should be researched to establish their appropriateness, cost-effectiveness, and utility in an Australian context.

Progressing from the selection of students into a postgraduate physiotherapy program, there is a need to ensure that coursework assessments completed by these students are related to their actual clinical performance. The clinical performance of physiotherapy students within Australia is measured by the APP which encompasses seven domains of practice: professionalism, communication, analysis and planning, assessment, intervention, evidenced-based practice, and risk management. The findings of the study reported in Chapter 4 (Terry et al., 2020) identified a significant relationship between the future clinical performance of postgraduate physiotherapy students and the commonly used coursework assessment items OSCE, written examinations and oral presentations. The oral presentation was weakly related to student performance on the APP and of limited predictive validity ($r^2 = 0.05$). Written assessments were moderately related to students' future clinical performance, but still of limited predictive value ($r^2 = 0.15$). The OSCE was strongly related to students' future clinical performance and was of moderate predictive validity ($r^2 = 0.32$).

While the OSCE score alone was of slightly higher predictive value than overall subject mark (r^2 = 0.32 compared to r^2 = 0.30), it is still appropriate to include multiple assessment items within a program of assessment, despite overall academic achievement within a program of study having been shown to be related to physiotherapy student clinical performance (Luedtke-Hoffmann et al., 2012; Terry et al., 2020; Vendrely, 2007). When the constructs targeted within the individual assessment items are considered, insight into their relationship with clinical performance scores becomes more apparent. It has been established that one individual item cannot meet all the needs of assessment within a course of study (van der Vleuten et al., 2012).

The oral presentations investigated targeted the collection and synthesis of information and evidence on a specific topic, and the ability to communicate information to an audience. In this style of assessment, both the preparation of the content and visual aids for presentation, and the delivery of the information, can be thoroughly planned and rehearsed. A limited amount of spontaneous question time may be allotted after a presentation. This assessment item provides students opportunity to prepare educational material and communicate it to an audience, which is a valuable learning experience for a future health professional. In clinical practice, students must apply the skills assessed by the oral presentation, but with significantly less time to plan. Authentic interactions with clients and educators in clinical practice settings are dynamic, often with limited opportunity to anticipate or rehearse the breadth and depth of knowledge synthesis and communication that may be required. While it is appropriate to include oral presentations in a program of assessment, this research suggests they are not useful as a predictive tool to identify students at risk of poor performance in clinical practice.

Written examinations target the recollection and application of theoretical knowledge. As identified in Chapter 6 of this thesis, the application of knowledge is assessed across multiple domains of the APP: analysis and planning, intervention, assessment, and risk management. Performance indicators requiring the identification of correct information or application of theoretical knowledge feature in 14 of the 20 items of the APP. Indeed, knowledge underpins the clinical reasoning required in clinical practice. This may explain why the written assessment investigated in this program of research (detailed in Chapter 5) had a greater relationship with students' future clinical performance than the oral

presentation. However, while a significant moderate relationship was identified, it was still of limited value in predicting students' future clinical performance ($r^2 = 0.15$). The written examination investigated contained a combination of multiple choice questions (MCQs) and short answer questions (SAQs). Norcini et al. (1985) suggested that both the selected response (MCQs) and constructed response (SAQs) style of written examinations are able to measure students' ability to apply clinical reasoning. However, selected response questions such as MCQ's, are accepted to have a greater reliability than constructed response questions (Hift, 2014), and allow for a wider sampling of content. As such, a written examination that did not contain SAQs may have elicited a greater predictive validity. Future research should investigate if a written examination consisting solely of MCQs have a greater predictive validity than a mixed-question format.

Of the summative assessments investigated, overall, the OSCE had the greatest predictive utility. This is consistent with evidence from the medical profession (Baker et al., 2006; Campos-Outcalt et al., 1999; Carr et al., 2014; Cope et al., 2007). Thus, the research presented in this thesis suggests that well-designed OSCEs may have the most potential utility to predict students' future clinical performance in the pre-clinical phase. However, contrary to the findings of the study reported in Chapter 4, the study by Wessel et al. (2003) found no significant relationship between physiotherapy student OSCE scores and future clinical performance. Two studies from the medical profession also found no significant relationship between OSCE scores and the clinical performance of medical students (Kahn et al., 2001; Probert et al., 2003). While the majority of the literature has identified significant predictive relationships between OSCE scores and future clinical performance of health profession students, the strength of these relationships have varied widely. Variability in the design and implementation of OSCEs may explain these discrepancies. In the studies analysed in Chapter 2, there was variation in the number of OSCE stations, the time spent at each station, as well as minimal reporting on the content and scoring of the OSCEs investigated. The intrinsic properties of an OSCE may influence the construct validity of the OSCE, and therefore the strength of any predictive relationships with external assessments that may exist.

When the content of the OSCE with predictive validity that informed this program of research was analysed thematically and compared to the APP, there was a clear alignment between the content of the OSCE and the competencies required to be

demonstrated in clinical practice. This may provide some explanation as to why UGPA did not predict students' future performance in clinical practice. UGPA is a static measure of performance at the time of entry into the program. It is a measure of academic performance, and while the application of theoretical knowledge is embedded across 14 items of the APP, this is knowledge specific to the physiotherapy profession. Conversely, UGPA is a measure of students' performance across the breadth of their undergraduate program and may not be content specific. Applications must meet the specific subject pre-requisites (anatomy, physiology, biomechanics, exercise prescription and statistics) to be eligible for admission to the host institution, but the remainder of their undergraduate program could be from other courses. Thus, the findings from the study presented in Chapter 3 (Terry et al., 2018), whereby no relationships between UGPA and academic and clinical performance were found, are not unsurprising. However, there is evidence that prerequisite science GPA is related to physiotherapy students' success on licensure examinations in the USA (Wolden, Hill, & Voorhees, 2020). Future research should investigate the relationships between prerequisite UGPA from science courses specifically and physiotherapy students' future clinical performance.

Importantly, the original findings of the study in Chapter 5 demonstrates that not *all* OSCEs have predictive validity with the clinical performance measure the APP. This not only highlights the challenges surrounding research into the use of OSCEs in physiotherapy (and other health professions), but more importantly, provides an explanation as to why discrepancies between OSCE performance and clinical performance findings may exist. As noted previously, Wessel et al. (2003) found no relationship between student performance on an OSCE and clinical practice. However, in Chapter 5 six individual OSCEs were investigated and five demonstrated significant relationships with students' future clinical performance as measured by the APP, with the strength of the correlations ranging from weak to moderate. If only one of the six OSCEs had been randomly selected for investigation, the resultant findings would be influenced by chance as to whether there was a significant or non-significant finding. These findings are important as the discrepancies indicated that other factors may be present that influence the validity of the OSCEs investigated. It also provided a unique opportunity to further investigate the structure of an OSCE with construct validity compared to one without construct validity

The two OSCEs analysed thematically were administered by a single institution with consistent organisational practices. There was also consistency in the timing and structure of the OSCEs both within the program and within each unit of study. The two OSCEs that were analysed thematically took place on the same day, so the students were under the same conditions and had the same level of experience with the OSCE format. This suggests factors intrinsic to the OSCEs themselves may explain the differences in validity.

Both OSCEs consisted of 2 x 15min stations and were associated with a pre-clinical subject in a specific area of clinical practice. However, there were differences between the two OSCEs in the level of agreement between coders, and the distribution of content within the OSCE across six themes: communication, assessment, intervention, knowledge, safety, and education. Compared to the OSCE with established construct validity, the OSCE without construct validity demonstrated the following features:

- Ambiguity of the construct being assessed by some mark sheet items (12% of total OSCE);
- Less content measuring the performance of assessment and intervention skills (41% compared to 60%);
- More content measuring communication, education, and general safety skills (43% compared to 32%); and
- More content measuring theoretical knowledge (17% compared to 8%).

There were features of both OSCEs that did not align with current best practice recommendations:

- Neither OSCE utilised a global rating scale. Global rating scales are recommended, particularly for stations assessing communication (Monti et al., 2020);
- Both OSCEs contained questions allocating marks for theoretical knowledge items which could be included elsewhere within the program of assessment, such as the written examination (Khan, Ramachandran, et al., 2013); and
- Although both OSCEs briefed examiners prior, neither OSCE provided guidelines for standards of expected performance for mark sheet items in writing or via video (Daniels & Pugh, 2018).

Encapsulating the research presented in this thesis, the volume of evidence suggests that OSCEs are a valid method of assessment in an entry-level physiotherapy program and may provide opportunity to predict students' future clinical performance. The intrinsic properties of OSCEs appear to influence their construct validity and may explain the variability in predictive relationships between OSCE scores and clinical placement outcomes across studies. There are practical challenges associated with implementing OSCEs given the time and resourcing required (Khan, Ramachandran et al., 2013; (Zamanzadeh et al., 2021). The financial costs associated with an OSCE vary considerably based on the specifics of the assessment delivered, however costs are suggested to be feasible for individual institutions to incorporate OSCEs into their program of assessment Carpenter (1995). This research on feasibility of costs was conducted in the medical program and should aim to be replicated in the physiotherapy context.

The number of students in each cohort investigated in this program of research are relatively small compared to some programs. Institutions with larger cohorts are likely to face more significant challenges when implementing OSCEs due to the resources required. Rushforth (2007) recognised the differences in nursing education to medical education where the OSCE originated and discussed variations on the OSCE used in nursing education. These included OSCEs that featured a smaller number of stations with longer time at each station (up to 30min), a pragmatic model that also allows the need for holistic consideration of a client. This thesis has demonstrated that such OSCEs can demonstrate content and construct validity.

7.2.1 Limitations and Strengths

A limitation to the analysis of student data in these studies is that they are retrospective, and they are from a single institution. While still providing opportunities for benchmarking for other institutions, this may limit the generalisability of the findings. The host institution also has generous staffing for student support during clinical placement. The substantial staffing during clinical placements offers additional support for students who may be struggling, and opportunities to extend students learning who may be doing well. This may result in higher clinical performance scores compared to institutions where only students identified as 'at risk' receive additional support.

An overall strength in this body of work is that the summative assessments investigated are commonly used assessments in both Australian international physiotherapy education programs. The content analysis of the two physiotherapy OSCEs is, to the author's knowledge, the first published literature of its kind. The two OSCEs investigated were conducted under the same conditions, at the same time in the students' program. This strongly suggests that the intrinsic properties of an OSCE may impact its validity and provides supporting evidence to the best practice guidelines currently available.

More importantly, is the measure of clinical performance against which all relationships in this thesis have been compared, the APP, is a well-researched, widely utilised instrument (Dalton et al., 2011; Dalton et al., 2012; Reubenson et al., 2020). The target construct of the APP instrument is the competency of physiotherapy students (Dalton et al., 2009). The items included on the APP closely align with the Physiotherapy Practice Thresholds of Australia and New Zealand (Physiotherapy Board of Australia & Physiotherapy Board of New Zealand, 2015), which outline the key competencies required of a competent registered physiotherapist in the two countries. The APP is also currently utilised in Qatar and Iceland (Clarity Data Solutions, 2021).

The APP and its associated performance indicators align closely with the competencies outlined in the Physiotherapist Education Framework (World Physiotherapy, 2021). This framework is based on a World Physiotherapy policy statement that describes the international expectations of entry-level physiotherapy education, and the expected standard of competence to be demonstrated within the local context. The lack of a widely-used measure of clinical performance has been identified in medical literature as a limitation in the available research (Hamdy et al., 2006). The use of the APP increases the utility of the findings in this body of work, as the competencies against which students' performance has been measured are relevant to physiotherapy students not only in Australia and New Zealand, but globally.

7.3 Recommendations

The findings of this thesis can be translated into practical recommendations regarding admission and summative assessment in physiotherapy education programs.

7.3.1 Admission into Entry-level Postgraduate Physiotherapy Education Programs

A firm eligibility criteria requirement of a UGPA above 4.5 out of 7 should be applied with caution as this may discriminate against groups of applicants with the potential to succeed as physiotherapists. Equally, a competitive UGPA to fill limited admission places should also be applied with caution and is not recommended to be used as the sole selection criterion to discriminate between applicants. Additional criterion should be considered in a model of selection to increase the ability to predict applicants' future performance and enhance the diversity of the profession.

Selection criteria into entry-level postgraduate physiotherapy programs warrant further investigation given the lack of a relationship identified between UGPA and future clinical performance, although UGPA derived from science programs of study appears promising. Models of selection under investigation should consider the holistic nature of the selection criteria included to optimise diversity within the profession. Future research should aim to investigate potential admission criteria to identify models of selection that are cost-effective to implement and effectively predict student success within the program. They should also seek to investigate the relationships between UGPA from science courses specifically and physiotherapy students' future clinical performance.

7.3.2 Summative Assessment in Physiotherapy Education Programs

Oral presentations, written examinations and OSCEs may all be valid measures of assessing physiotherapy students in pre-clinical coursework and, considered in context, could be included in a program of assessment. The composition of written assessments could be investigated to determine if the type of questions used influences the predictive validity of this assessment format. Specifically, if there is a difference in the relationship between clinical performance and written assessments that are composed of MCQs vs

SAQs. As such, future research should investigate if a written examination consisting solely of MCQs has a greater predictive validity of student clinical performance than a mixed-question format.

7.3.3 Objective Structured Clinical Examinations

OSCEs should be blueprinted to the pre-clinical coursework content and the proportion of skills to be assessed should be mapped to stations with individual mark sheets. Academics should consider the number of stations that can be feasibly implemented, and the time required at each station to ensure adequate sampling. Mark sheet items should be carefully considered. If multiple marks are allocated to one item, a single clear construct should be evident and examples of the standard of expected performance provided to assessors. Stations or mark sheet items focused on the recitation of theoretical knowledge should be identified, and considered if these can be incorporated elsewhere in the program of assessment (i.e., written examination). The use of a global rating scale should be considered, especially for stations or items mapped to assess communication, education, or safety. Resourcing should be given to appropriately train assessors. Institutions implementing OSCEs should consider the stress associated with this examination format that is experienced by students and ensure that appropriate pre-test preparation is given to minimise anticipatory stress much as possible.

While a strength of this program of research is that the OSCEs were compared to the APP, a validated and widely used tool, potential differences (be they cultural or institutional) may exist – the differences between the United States of America study by Wessel et al. (2003) and reported in this Australian context serving as an example. As such, future research should aim to replicate the studies reported in this program of research at other entry-level postgraduate physiotherapy institutions, especially in countries where clinical competence is measured using different tools. It should be noted that the APP was field-tested predominately against public health settings, with only 2% of the field tests taking place in private practice. Future research investigating the use of the APP in private practice settings would be of benefit to the profession, and whether relationships between OSCE performance and students' performance as measured by the APP are replicated in this setting. There should also be investigation of the intrinsic

properties of OSCE's designed using effective blue-printing and optimised mark sheets, to establish if the predictive validity of OSCE can be positively impacted by OSCE design.

The physiotherapy profession would benefit from research into the feasibility of implementing OSCEs in large cohorts of students. Information regarding the associated costs, both financial and human, and potential strategies to minimise the resources required whilst still maintaining the validity of this assessment format would be of benefit.

References

Allison, H., & Turpin, M. J. (2004). Development of the student placement evaluation form: A tool for assessing student fieldwork performance. *Australian Occupational Therapy Journal, 51*(3), 125-132.

American Psychological Association, American Educational Research Association, & National Council on Measurement in Education. (1999). *Standards for educational and psychological testing*. Washington, DC: American Educational Research Association.

Anastakis, D. J., Cohen, R., & Reznick, R. K. (1991). The structured oral examination as a method for assessing surgical residents. *The American Journal of Surgery, 162*(1), 67-70.

Anderson, D. C., Harris, I. B., Allen, S., Satran, L., Bland, C. J., Davis-Feickert, J. A., Poland, G.A., Miller, W. J. (1991). Comparing students' feedback about clinical instruction with their performances. *Academic Medicine, 66*(1), 29-34.

APPLinkup. (2022). Retreived from https://www.applinkup.com/Default.aspx

Artino, A. R., La Rochelle, J. S., & Durning, S. J. (2010). Second-year medical students' motivational beliefs, emotions, and achievement. *Medical Education, 44*(12), 1203-1212.

Austin, Z., O'Byrne, C., Pugsley, J., & Munoz, L. Q. (2003). Development and validation processes for an objective structured clinical examination (OSCE) for entry-to-practice certification in pharmacy: The Canadian experience. *American Journal of Pharmaceutical Education, 67*(3).

Australian Health Practitioner Regulation Agency. (2013, Jul 11 2013). Approved Programs of Study. Retrieved from http://www.ahpra.gov.au/Education/Approved-Programs-of-Study.aspx?ref=Physiotherapist

Australian Medical Council Limited. (2012). *Standards for Assessment and Accreditation of Primary Medical Programs by the AMC 2012*. Retrieved from https://www.amc.org.au/wp-content/uploads/2019/10/Standards-for-Assessment-and-Accreditation-of-Primary-Medical-Programs-by-the-Australian-Medical-Council-2012.pdf
Australian Physiotherapy Association. (2022). What is physio? Retrieved from https://choose.physio/what-is-physio. Accessed on Dec 30 2022

Australian Physiotherapy Association. (2021, Oct 01 2021). Workforce report shows physiotherapy in demand. *inMotion* Retrieved from https://australian.physio/inmotion/workforce-report-shows-physiotherapy-demand

Australian Physiotherapy Council. (2017). *Accreditation Standard For Entry-Level Physiotherapy Practitioner Programs December 2016*. Retrieved from https://physiocouncil.com.au/wp-content/uploads/2019/08/ACCREDITATION-STANDARD-V1.1-13112017.pdf

Australian Physiotherapy Council. (2006). Australian standards for physiotherapy. Canberra, APC.

Australian Physiotherapy Council. (2020). *Guidelines for Accreditation Entry-Level Physiotherapy Practitioner Programs of Study V1.4*. Retrieved from https://physiocouncil.com.au/wp-content/uploads/2021/07/GUIDELINES-FOR-ACCREDITATION-V1.4-04032021.pdf

Baggs, T., Barnett, D., & McCullough, K. (2015). The value of traditional cognitive variables for predicting performance in graduate speech-language pathology programs. *Journal of Allied Health, 44*(1), 10-16.

Baker, H. H., Cope, M. K., Adelman, M. D., Schuler, S., Foster, R. W., & Gimpel, J. R. (2006). Relationships between scores on the COMLEX-USA Level 2-Performance Evaluation and selected school-based performance measures. *The Journal of the American Osteopathic Association, 106*(5), 290-295.

Beckert, L., Wilkinson, T. J., & Sainsbury, R. (2003). A needs-based study and examination skills course improves students' performance. *Medical Education, 37*(5), 424-428.

Bejar, I. I. (2012). Rater cognition: Implications for validity. *Educational Measurement: Issues and Practice, 31*(3), 2-9.

Belland, B. R., French, B. F., & Ertmer, P. A. (2009). Validity and problem-based learning research: A review of instruments used to assess intended learning outcomes. *Interdisciplinary Journal of Problem-based Learning, 3*(1), 5.

Berg, K., Winward, M., Clauser, B. E., Veloski, J. A., Berg, D., Dillon, G. F., & Veloski, J. J. (2008). The relationship between performance on a medical school's clinical skills assessment and USMLE Step 2 CS. *Academic Medicine, 83*(10 Suppl).

Bird, S., Thompson, C., & Williams, K. E. (2016). Primary contact physiotherapy services reduce waiting and treatment times for patients presenting with musculoskeletal conditions in Australian emergency departments: an observational study. *Journal of Physiotherapy, 62*(4), 209-214.

Bobos, P., Pouliopoulou, D. V., Harriss, A., Sadi, J., Rushton, A., & MacDermid, J. C. (2021). A systematic review and meta-analysis of measurement properties of objective structured clinical examinations used in physical therapy licensure and a structured review of licensure practices in countries with well-developed regulation systems. *PLoS One, 16*(8), e0255696-e0255696.

Boursicot, K. A. M., Roberts, T.E., Burdick, W.P. (2010). Structured assessments of clinical competence. In T. Swanwick (Ed.), *Understanding Medical Education: Evidence, Theory and Practice*. Hoboken, UNITED KINGDOM: John Wiley & Sons, Incorporated.

Bramley, A. L., & McKenna, L. (2021). Entrustable professional activities in entry-level health professional education: A scoping review. *Medical Education*, 55(9), 1011-1032.

Brand, H. S., & Schoonheim-Klein, M. (2009). Is the OSCE more stressful? Examination anxiety and its consequences in different assessment methods in dental education. *European Journal of Dental Education*, *13*(3), 147–153

Brannick, M. T., Erol-Korkmaz, H. T., & Prewett, M. (2011). A systematic review of the reliability of objective structured clinical examination scores. *Medical Education, 45*(12), 1181-1189.

Brewer, M., Gribble, N., Robinson, P., Lloyd, A., & White, S. (2009). *Assessment of interprofessional competencies for health professional students in fieldwork education placements.* Paper presented at the ATN Assessment Conference 2009: Assessment in Different Dimensions.

British Medical Association. (2017). *Doctors' titles: explained*. Retrieved from London: https://www.bma.org.uk/media/1937/bma-plg-doctors-titles-explained-2017.pdf

Brown, G., Manogue, M., & Martin, M. (1999). The validity and reliability of an OSCE in dentistry. *European Journal of Dental Education, 3*(3), 117-125.

Cacho, R. d. O., Baroni, M. P., Ruaro, J. A., Lopes, J. M., Britto, H. M. J. d. S., Ferreira, T. B., Guedes, M.B.O.G., Oliveira, D. C. d. (2016). Active methods in physical therapy:

intereliability study of the OSCE method. *Revista Brasileira de Educação Médica, 40*(1), 128-137.

Caine, A. M., Copley, J., Turpin, M., Fleming, J., & Herd, C. (2021). Development of the Student Practice Evaluation Form-Revised (Second Edition) (SPEF-R2): The first action research cycle. *Australian Occupational Therapy Journal, 68*(1), 21-31.

Campos-Outcalt, D., Watkins, A., Fulginiti, J., Kutob, R., & Gordon, P. (1999). Correlations of family medicine clerkship evaluations and Objective Structured Clinical Examination scores and residency directors' ratings. *Family Medicine, 31*(2), 90-94.

Candela, L., Piacentine, L. B., Bobay, K. L., & Weiss, M. E. (2018). Teaching students to teach patients: A theory-guided approach. *Journal of Nursing Education and Practice*.

Carp, S. P. T. P., Fry, K. D. P. T., Gumerman, B. D. P. T., Pressley, K. D. P. T., & Whitman, A. D. P. T. (2020). Relationship Between Grit Scale Score and Academic Performance in a Doctor of Physical Therapy Program: A Case Study. *Journal of Allied Health, 49*(1), 29-35,35A.

Carpenter, J. L. (1995). Cost analysis of objective structured clinical examinations. *Academic Medicine, 70*(9), 828–833.

Carr, S. E., Celenza, A., & Lake, F. (2013). Assessment of Junior Doctor performance: a validation study. *BMC Medical Education, 13*(129), 129.

Carr, S. E., Celenza, A., Puddey, I. B., & Lake, F. (2014). Relationships between academic performance of medical students and their workplace performance as junior doctors. *BMC Medical Education, 14*(157), 1472-6920.

Chahine, S., Holmes, B., & Kowalewski, Z. (2016). In the Minds of OSCE Examiners: Uncovering Hidden Assumptions. *Advances in Health Sciences Education, 21*(3), 609-625.

Chen, H. C., van den Broek, W. S., & ten Cate, O. (2015). The case for use of entrustable professional activities in undergraduate medical education. *Academic Medicine, 90*(4), 431-436.

Chisholm-Burns, M. A., Berg-Poppe, P., Spivey, C. A., Karges-Brown, J., & Pithan, A. (2021). Systematic review of noncognitive factors influence on health professions students' academic performance. *Advances in Health Sciences Education Theory and Practice, 26*(4), 1373-1445.

Clarity Data Solutions. (2021). APPLinkup Select Location. Retrieved from https://www.applinkup.com/SelectRegion.aspx?cWpKWjZoVjdiRzl5aWs1bHJMdUxEQT09=STU1bVkwWlpGNElpcmlYeGZUSUtzUT09

Cocke, C., & Orr, R. (2015). The impact of physical training programs on the fitness of tactical populations: A critical review. *Journal of Australian Strength and Conditioning, 23*(1), 39-46.

Cohen, J. A. (1988). *Statistical power analysis for the behavioral sciences*. Hillsdale, NJ: Lawrence Erlbaum Associates.

Cohen, R., Reznick, R. K., Taylor, B. R., Provan, J., & Rothman, A. (1990). Reliability and validity of the objective structured clinical examination in assessing surgical residents. *The American Journal of Surgery, 160*(3), 302-305.

Commission on Accreditation In Physical Therapy Education. (2018). Aggregate Program Data 2017-2018 Physical Therapist Education Programs Fact Sheets. Retrieved from www.capteonline.org/resources

Confederation of Postgraduate Medical Education Councils. (2012). *Australian Curriculum Framework for Junior Doctors version 3.1* Retrieved from https://cpmec.org.au/app/uploads/2022/02/Austn-Curriculum-Framework-27112017_watermark.pdf

Cope, M. K., Baker, H. H., Foster, R. W., & Boisvert, C. S. (2007). Relationships between clinical rotation subscores, COMLEX-USA examination results, and school-based performance measures. Journal *of Osteopathic Medicine, 107*(11), 502-510.

Council of Canadian Physiotherapy University Programs. (2009). *Entry-to-Practice Physiotherapy Curriculum: Content Guidelines for Canadian University Programs*. Council of Canadian Physiotherapy University Programs. Retrieved from http://www.physiotherapyeducation.ca/PhysiotherapyEducation.html

Crawford, C., Black, P., Melby, V., & Fitzpatrick, B. (2021). An exploration of the predictive validity of selection criteria on progress outcomes for pre-registration nursing programmes-A systematic review. *Journal of Clinical Nursing, 30*(17-18), 2489-2513.

Cresswell, K., Howe, A., Steven, A., Smith, P., Ashcroft, D., Fairhurst, K., Bradley, F., Magnussen, C., McArthur, M., Pearson, P., Sheikh, A., Patient Safety Education Research Group. (2013). Patient safety in healthcare preregistration educational curricula: multiple case study-based investigations of eight medicine, nursing, pharmacy and physiotherapy university courses. *BMJ Quality & Safety, 22*(10), 843-854.

Cross, V., Hicks, C., & Barwell, F. (2001). Exploring the gap between evidence and judgement: Using video vignettes for practice-based assessment of physiotherapy undergraduates. *Assessment & Evaluation in Higher Education, 26*(3), 189-212.

Daelmans, H. E. M., Scherpbier, A. J. J. A., van der Vleuten, C. P., & Donker, A. J. M. (2001). Reliability of clinical oral examinations re-examined. *Medical Teacher, 23*(4), 422-424.

Dalton, M., Davidson, M., & Keating, J. (2011). The Assessment of Physiotherapy Practice (APP) is a valid measure of professional competence of physiotherapy students: a cross-sectional study with Rasch analysis. *Journal of Physiotherapy, 57*(4), 239-246.

Dalton, M., Davidson, M., & Keating, J. L. (2012). The assessment of physiotherapy practice (APP) is a reliable measure of professional competence of physiotherapy students: a reliability study. *Journal of Physiotherapy, 58*(1), 49-56.

Dalton, M., Keating, J., & Davidson, M. (2009). Development of the Assessment of Physiotherapy Practice (APP): A standardised and valid approach to assessment of clinical competence in physiotherapy. *Australian Learning and Teaching Council (ALTC) Final report*, 6-28.

Daniels, V. J., Bordage, G., Gierl, M. J., & Yudkowsky, R. (2014). Effect of clinically discriminating, evidence-based checklist items on the reliability of scores from an Internal Medicine residency OSCE. *Advances in Health Sciences Education: Theory and Practice, 19*(4), 497-506.

Daniels, V. J., & Pugh, D. (2018). Twelve tips for developing an OSCE that measures what you want. *Medical Teacher, 40*(12), 1208-1213.

Davenport, R., Hewat, S., Ferguson, A., McAllister, S., & Lincoln, M. (2018). Struggle and failure on clinical placement: a critical narrative review. *International Journal of Language & Communication Disorders, 53*(2), 218-227.

De Champlain, A., Swygert, K., Swanson, D. B., & Boulet, J. R. (2006). Assessing the underlying structure of the United States Medical Licensing Examination Step 2 test of clinical skills using confirmatory factor analysis. *Academic Medicine: Journal of the Association of American Medical Colleges, 81*(10 Suppl), S17.

Dental Board of Australia. (2020). *Scope of practice registration standard*. Retrieved from https://www.dentalboard.gov.au/registration-standards/scope-of-practice-registration-standard.aspx

Department of Health. (2021, 3 September 2021). Allied health in Australia. Retrieved from https://www.health.gov.au/health-topics/allied-health/in-australia#allied-health-workforce-numbers

Department of Health. (2022, 4 January 2022). What we're doing about allied health. Retrieved from https://www.health.gov.au/health-topics/allied-health/our-support

Desjardins, I., Touchie, C., Pugh, D., Wood, T. J., & Humphrey-Murto, S. (2014). The impact of cueing on written examinations of clinical decision making: a case study. *Medical Education, 48*(3), 255-261.

Dijkstra, J., Galbraith, R., Hodges, B., McAvoy, P. A., McCrorie, P., Southgate, L. J., van der Vleuten, C., Wass, V., Schuwirth, L. W. (2012). Expert validation of fit-for-purpose guidelines for designing programmes of assessment. *BMC Medical Education, 12*(20), 20.

Dixon, D. (2012). Prediction of Osteopathic Medical School Performance on the basis of MCAT score, GPA, sex, undergraduate major, and undergraduate institution. *Journal of the American Osteopathic Association, 112*(4), 175-181.

Dong, T., Durning, S. J., Gilliland, W. R., Swygert, K. A., & Artino, A. R., Jr. (2015). Development and initial validation of a program director's evaluation form for medical school graduates. *Military Medicine, 180*(4 Suppl), 97-103.

Dong, T., Swygert, K. A., Durning, S. J., Saguil, A., Gilliland, W. R., Cruess, D., DeZee, K.J., LaRochelle, J., Artino, A. R., Jr. (2014). Validity evidence for medical school OSCEs: associations with USMLE® step assessments. *Teaching and Learning in Medicine, 26*(4), 379-386.

Downing, S. M. (2003). Validity: on the meaningful interpretation of assessment data. *Medical Education, 37*(9), 830-837.

Downing, S. M. (2005). The Effects of Violating Standard Item Writing Principles on Tests and Students: The Consequences of Using Flawed Test Items on Achievement Examinations in Medical Education. *Advances in Health Sciences Education: Theory and Practice, 10*(2), 133-143.

Downing, S. M., & Haladyna, T. M. (2009). Validity and It's Threats. In S. M. Downing & R. Yudkowsky (Eds.), *Assessment in Health Professions Education*. Hoboken: Taylor and Francis.

Downing, S. M., & Yudkowsky, R. (2009). Introduction to assessment in the health professions. In S. M. Downing & R. Yudkowsky (Eds.), *Assessment in Health Professions Education*. Hoboken: Taylor and Francis.

Downs, S. H., & Black, N. (1998). The feasibility of creating a checklist for the assessment of the methodological quality both of randomised and non-randomised studies of health care interventions. *Journal of Epidemiology and Community Health, 52*(6), 377-384.

Driessen, E. (2017). Do portfolios have a future? *Advances in Health Sciences Education, 22*(1), 221-228.

Driessen, E., Overeem, K., van Tartwijk, J., van der Vleuten, C. P., & Muijtjens, A. M. (2006). Validity of portfolio assessment: which qualities determine ratings? *Medical Education, 40*(9), 862-866.

Driessen, E., van Tartwijk, J., van der Vleuten, C. P., & Wass, V. (2007). Portfolios in medical education: why do they meet with mixed success? A systematic review. *Medical Education, 41*(12), 1224-1233.

Edgar, S., Mercer, A., & Hamer, P. (2014). Admission interview scores are associated with clinical performance in an undergraduate physiotherapy course: an observational study. *Physiotherapy, 100*(4), 331-335.

Epstein, R. M. (2007). Assessment in medical education. *The New England Journal of Medicine, 356*(4), 387-396.

Felix, H., Laird, J., Ennulat, C., Donkers, K., Garrubba, C., Hawkins, S., & Hertweck, M. (2012). Holistic admissions process: an initiative to support diversity in medical education. *Journal of Physician Assistant Education (Physician Assistant Education As*sociation), 23(3).

Ferguson, K. J., & Kreiter, C. D. (2004). Using a longitudinal database to assess the validity of preceptors' ratings of clerkship performance. *Advances in Health Sciences Education: Theory and Practice*, 9(1), 39-46.

Ferreira, É. D. M. R., Pinto, R. Z., Arantes, P. M. M., Vieira, É. L. M., Teixeira, A. L., Ferreira, F. R., & Vaz, D. V. (2020). Stress, anxiety, self-efficacy, and the meanings that physical therapy students attribute to their experience with an objective structured clinical examination. *BMC Medical Education*, 20(1), 1-9.

Gadbury-Amyot, C. C., Bray, K. K., Branson, B. S., Holt, L., Keselyak, N., Mitchell, T. V., & Williams, K. B. (2005). Predictive validity of dental hygiene competency assessment measures on one-shot clinical licensure examinations. *Journal of Dental Education*, 69(3), 363-370.

Gadbury-Amyot, C. C., Holt, L. P., Overman, P. R., & Schmidt, C. R. (2000). Implementation of portfolio assessment in a competency-based dental hygiene program. *Journal of Dental Education, 64*(5), 375-380.

Gadbury-Amyot, C. C., McCracken, M. S., Woldt, J. L., & Brennan, R. L. (2014). Validity and reliability of portfolio assessment of student competence in two dental school populations: a four-year study. *Journal of Dental Eeducation, 78*(5), 657-667.

General Medical Council. (2013). *Good medical practice.* Manchester: General Medical Council Retrieved from http://www.gmcuk.org/Good_medical_practice___English_1215.pdf_51527435.pdf

Gimpel, J. R., Boulet, D. O., & Errichetti, A. M. (2003). Evaluating the clinical skills of osteopathic medical students. *Journal of the American Osteopathic Association, 103*(6), 267-279.

Gorman, S. L., Lazaro, R., Fairchild, J., & Kennedy, B. (2010). Development and implementation of an objective structured clinical examination (OSCE) in neuromuscular physical therapy. *Journal of Physical Therapy Education, 24*(3), 62-68. Retrieved from http://search.ebscohost.com/login.aspx?direct=true&db=cin20&AN=105096475&site=ehost-live&scope=site

Grabowski, C. J. (2018). Impact of holistic review on student interview pool diversity. *Advances in Health Sciences Education, 23*(3), 487-498.

Graham, R., Zubiaurre Bitzer, L. A., & Anderson, O. R. (2013). Reliability and predictive validity of a comprehensive preclinical OSCE in dental education. *Journal of Dental Education, 77*(2), 161-167.

Gravetter, F. J., & Wallnau, L. B. (1996). *Statistics for the behavioral sciences* (Fourth ed.). Minneapolis: West Publishing Company.

Haladyna, T. M. (2011). *An Evaluation of the Central Regional Dental Testing Services National Dental Hygiene Examination.* Retrieved from https://crdts.org/uploads/CRDTS%20Dental%20Technical%20Report-2010.pdf

Halperin, J. L., Levine, G. N., Al-Khatib, S. M., Birtcher, K. K., Bozkurt, B., Brindis, R. G., Cigarroa, J.E., Curtis, L.H., Fleisher, L.A., Gentile, F. (2016). Further evolution of the ACC/AHA clinical practice guideline recommendation classification system: a report of the American College of Cardiology/American Heart Association Task Force on Clinical Practice Guidelines. *Circulation, 133*(14), 1426-1428.

Hamdy, H., Prasad, K., Anderson, M. B., Scherpbier, A., Williams, R., Zwierstra, R., & Cuddihy, H. (2006). BEME systematic review: predictive values of measurements obtained in medical schools and future performance in medical practice. *Medical Teacher, 28*(2), 103-116.

Han, E. R., & Chung, E. K. (2016). Does medical students' clinical performance affect their actual performance during medical internship? *Singapore Medical Journal, 57*(2), 87.

Harden, R. M. (1988). What is an OSCE? *Medical Teacher, 10*(1), 19.

Harden, R. M., & Gleeson, F. A. (1979). Assessment of clinical competence using an objective structured clinical examination (OSCE). *Medical Education, 13*(1), 41.

Harden, R. M., Stevenson, M., Downie, W. W., & Wilson, G. (1975). Assessment of clinical competence using objective structured examination. *British Medical Journal, 1*(5955), 447-451.

Harfmann, K. L., & Zirwas, M. J. (2011). Can performance in medical school predict performance in residency? A compilation and review of correlative studies. *Journal of the American Academy of Dermatology, 65*(5), 1010-1022.

Hawker, J. A., Walker, K. Z., Barrington, V., & Andrianopoulos, N. (2010). Measuring the success of an objective structured clinical examination for dietetic students. *Journal of Human Nutrition and Dietetics, 23*(3), 212-216.

Hawkins, R. E., Swanson, D., Dillon, G., Clauser, B., King, A., Scoles, P., Whelan, G.P., Burdick, W.P., Bou;let, J.R., Homan, A. (2005). The introduction of clinical skills assessment into the United States medical licensing examination (USMLE): a description of USMLE step 2 clinical skills (CS). *Journal of Medical Licensure and Discipline, 91*, 22-25.

Health & Care Professions Council. (2017). *Standards of education and training*. Retrieved from
http://www.hpcuk.org/publications/index.asp?id=183#publicationSearchResults

Health Workforce Australia. (2011). *National Clinical Supervision Support Framework*. Retrieved from http://www.ntrtn.com.au/sites/default/files/hwa-national-clinical-supervision-support-framework-201110.pdf

Hift, R. J. (2014). Should essays and other "open-ended"-type questions retain a place in written summative assessment in clinical medicine? *BMC Medical Education, 14*(1), 249-249.

Ho, D. W., Whitehill, T. L., & Ciocca, V. (2014). Performance of speech-language pathology students in problem-based learning tutorials and in clinical practice. *Clinical Linguistics & Phonetics, 28*(1-2), 102-116.

Hodges, B. (2003). Validity and the OSCE. *Medical Teacher, 25*(3), 250-254.

Hodges, B., Regehr, G., Hanson, M., & McNaughton, N. (1998). Validation of an objective structured clinical examination in psychiatry. *Academic Medicine, 73*(8), 910-912.

Hodges, B., Regehr, G., McNaughton, N., Tiberius, R., & Hanson, M. (1999). OSCE checklists do not capture increasing levels of expertise. *Academic Medicine, 74*(10), 1129-1134.

Howard, L., & Jerosch-Herold, C. (2000). Can entry qualifications be used to predict fieldwork and academic outcomes in occupational therapy and physiotherapy students? *The British Journal of Occupational Therapy, 63*(7), 329-334.

Hu, J., Jones, A. Y. M., Zhou, X., Zhai, H., Ngai, S. P. C., Siu, K.-C., & Dalton, M. (2020). Acceptance of the 'Assessment of Physiotherapy Practice (Chinese)' as a standardised evaluation of professional competency in Chinese physiotherapy students: an observational study. *BMC Medical Education, 20*(1), 108.

James, D., Yates, J., & Nicholson, S. (2010). Comparison of A level and UKCAT performance in students applying to UK medical and dental schools in 2006: cohort study. *BMJ, 340*(340), c478.

Jelovsek, J. E., Kow, N., & Diwadkar, G. B. (2013). Tools for the direct observation and assessment of psychomotor skills in medical trainees: a systematic review. *Medical Education, 47*(7), 650-673.

Kahn, M. J., Merrill, W. W., Anderson, D. S., & Szerlip, H. M. (2001). Residency program director evaluations do not correlate with performance on a required 4th-year objective structured clinical examination. *Teaching and Learning in Medicine, 13*(1), 9-12.

Kennelly, J. (2011). Methodological approach to assessing the evidence. In *Reducing Racial/Ethnic Disparities in Reproductive and Perinatal Outcomes* (pp. 7-19): Springer.

Khan, K. Z., Gaunt, K., Ramachandran, S., & Pushkar, P. (2013). The Objective Structured Clinical Examination (OSCE): AMEE Guide No. 81. Part II: Organisation & Administration. *Medical Teacher, 35*(9), e1447-e1463.

Khan, K. Z., & Ramachandran, S. (2012). Conceptual framework for performance assessment: competency, competence and performance in the context of assessments in healthcare--deciphering the terminology. *Medical Teacher, 34*(11), 920-928.

Khan, K. Z., Ramachandran, S., Gaunt, K., & Pushkar, P. (2013). The Objective Structured Clinical Examination (OSCE): AMEE Guide No. 81. Part I: an historical and theoretical perspective. *Medical Teacher, 35*(9), e1437.

Kiesewetter, J., Ebersbach, R., Tsalas, N., Holzer, M., Schmidmaier, R., & Fischer, M. R. (2016). Knowledge is not enough to solve the problems–The role of diagnostic knowledge in clinical reasoning activities. *BMC Medical Education, 16*(1), 1-8.

Kirwan, G. W., Kirwan, G. W., Clark, C. R., Clark, C. R., Dalton, M., & Dalton, M. (2019). Rating of physiotherapy student clinical performance: is it possible to gain assessor consistency? *BMC Medical Education, 19*(1), 32-32.

Kreiter, C. D., Ferguson, K., Lee, W. C., Brennan, R. L., & Densen, P. (1998). A generalizability study of a new standardized rating form used to evaluate students' clinical clerkship performances. *Academic Medicine : Journal of the Association of American Medical Colleges, 73*(12), 1294.

Kulatunga-Moruzi, C., & Norman, G. R. (2002). Validity of admissions measures in predicting performance outcomes: the contribution of cognitive and non-cognitive dimensions. *Teaching and Learning in Medicine, 14*(1), 34-42.

Ladyshewsky, R., Baker, R., Jones, M., & Nelson, L. (2000). Evaluating clinical performance in physical therapy with simulated patients. *Journal of Physical Therapy Education, 14*(1), 31.

LaRochelle, J. S., Dong, T., & Durning, S. J. (2015). Preclerkship assessment of clinical skills and clinical reasoning: the longitudinal impact on student performance. *Military Medicine, 180*(4 Suppl), 43-46.

Luedtke-Hoffmann, K., Dillon, L., Utsey, C., & Tomaka, J. (2012). Is there a relationship between performance during physical therapist clinical education and scores on the National Physical Therapy Examination (NPTE)? *Journal of Physical Therapy Education, 26*(2), 41-49.

Lukhele, R., Thissen, D., & Wainer, H. (1994). On the Relative Value of Multiple-Choice, Constructed Response, and Examinee-Selected Items on Two Achievement Tests. *Journal of Educational Measurement, 31*(3), 234-250.

Machado, J. L. M., Machado, V. M. P., Grec, W., Bollela, V. R., & Vieira, J. E. (2008). Self- and peer assessment may not be an accurate measure of PBL tutorial process. *BMC Medical Education, 8*(1), 55-55.

Maloney, S., Storr, M., Morgan, P., & Ilic, D. (2013). The effect of student self-video of performance on clinical skill competency: a randomised controlled trial. *Advances in Health Sciences Education: Theory and Practice, 18*(1), 81-89.

Martínez-Pascual B, Ramírez-Adrados A, Fernández-Martínez S, Gonzalez-de-Ramos C, Fernández-Elías VE, Clemente-Suárez VJ. Autonomic stress response of physiotherapy student in the different scenarios of an objective structured clinical examination. *BMC Medical Education.* 2022 Nov 24;22(1):811.

McAllister, M. (2005). Transformative teaching in nursing education: leading by example. *Collegian, 12*(2), 11-16.

McAllister, S. (2005). *Competency based assessment of speech pathology students' performance in the workplace.* (PhD Doctorate). University of Sydney, Retrieved from http://prijipati.library.usyd.edu.au/handle/2123/1130

McAllister, S., Lincoln, M., Ferguson, A., & McAllister, L. (2010). Issues in developing valid assessments of speech pathology students' performance in the workplace. *International Journal of Language & Communication Disorders, 45*(1), 1-14.

McAllister, S., Lincoln, M., Ferguson, A., & McAllister, L. (2011). A systematic program of research regarding the assessment of speech-language pathology competencies. *International Journal of Speech-Language Pathology, 13*(6), 469-479.

McCready, T. (2007). Portfolios and the assessment of competence in nursing: a literature review. *International Journal of Nursing Studies, 44*(1), 143-151.

McLaughlin, J. E., Khanova, J., Scolaro, K., Rodgers, P. T., & Cox, W. C. (2015). Limited Predictive Utility of Admissions Scores and Objective Structured Clinical Examinations for APPE Performance. *American Journal of Pharmaceutical Education, 79*(6), 84.

McMahon, S., Cusack, T., & O'Donoghue, G. (2014). Barriers and facilitators to providing undergraduate physiotherapy clinical education in the primary care setting: a three-round Delphi study. *Physiotherapy, 100*(1), 14-19.

Memon, M. A., Joughin, G. R., & Memon, B. (2010). Oral assessment and postgraduate medical examinations: establishing conditions for validity, reliability and fairness. *Advances in Health Sciences Education: Theory and Practice, 15*(2), 277-289.

Mercer, A., & Puddey, I. B. (2011). Admission selection criteria as predictors of outcomes in an undergraduate medical course: a prospective study. *Medical Teacher, 33*(12), 997-1004.

Miller, G. E. (1990). The assessment of clinical skills/competence/performance. *Academic Medicine, 65*(9 Suppl), S63-67.

Monie, A., Fazey, P., & Singer, K. (2016). Low back pain misdiagnosis or missed diagnosis: Core principles. *Manual Therapy, 22*, 68-71.

Monti, M., Klöckner-Cronauer, C., Hautz, S. C., Schnabel, K. P., Breckwoldt, J., Junod-Perron, N., Feller, S., Bonvin, R., Huwendiek, S. (2020). Improving the assessment of communication competencies in a national licensing OSCE: lessons learned from an experts' symposium. *BMC Medical Education, 20*(1), 171-171.

Morris, J., & Farmer, A. (1999). The predictive strength of entry grades and biographical factors on the academic and clinical performance of physiotherapy students. *Physiotherapy Theory & Practice, 15*(3), 165-173.

Nadasan, T., & Puckree, T. (2003). Do the selection criteria for admittance to the physiotherapy program predict students' performance? *South African Journal of Physiotherapy, 59*(3), 20.

Norcini, J. J. (2003). Work based assessment. *BMJ, 326*(7392), 753-755.

Norcini, J. J., Swanson, D. B., Grosso, L. J., & Webster, G. D. (1985). Reliability, validity and efficiency of multiple choice question and patient management problem item formats in assessment of clinical competence. *Medical Education, 19*(3), 238.

Nursing & Midwifery Council. (2010). *Standards for competence for registered nurses*. London Retrieved from https://www.nmc.org.uk/globalassets/sitedocuments/standards/nmc-standards-for-competence-for-registered-nurses.pdf

O'Connor, A., McGarr, O., Cantillon, P., McCurtin, A., & Clifford, A. (2018). Clinical performance assessment tools in physiotherapy practice education: a systematic review. *Physiotherapy, 104*(1), 46-53. doi:10.1016/j.physio.2017.01.005

Occupational Therapy Board of Australia. (2018). Australian occupational therapy competency standards 2018. Retrieved from https://www.occupationaltherapyboard.gov.au/codes-guidelines/competencies.aspx

O'Dowd, E., Lydon, S., O'Connor, P., Madden, C., & Byrne, D. (2019). A systematic review of 7 years of research on entrustable professional activities in graduate medical education, 2011–2018. *Medical Education*, 53(3), 234-249.

Oudkerk Pool, A., Govaerts, M. J. B., Jaarsma, D., & Driessen, E. (2018). From aggregation to interpretation: how assessors judge complex data in a competency-based portfolio. *Advances in Health Sciences Education: Theory and Practice, 23*(2), 275-287.

Paget, N. S., Newble, D. I., Saunders, N. A., & Du, J. (1996). Physician assessment pilot study for the Royal Australasian College of Physicians. *Journal of Continuing Education in the Health Professions, 16*(2), 103-111.

Pangaro, L., & ten Cate, O. (2013). Frameworks for learner assessment in medicine: AMEE Guide No. 78. *Medical Teacher, 35*(6), e1197-1210. doi:10.3109/0142159X.2013.788789

Patterson, F., Knight, A., Dowell, J., Nicholson, S., Cousans, F., & Cleland, J. (2016). How effective are selection methods in medical education? A systematic review. *Medical Education, 50*(1), 36-60.

Pau, A., Jeevaratnam, K., Chen, Y. S., Fall, A. A., Khoo, C., & Nadarajah, V. D. (2013). The Multiple Mini-Interview (MMI) for student selection in health professions training - a systematic review. *Medical Teacher, 35*(12), 1027-1041.

Paynter, S., Iles, R., & Hay, M. (2022). An investigation of the predictive validity of selection tools on performance in physiotherapy training in Australia. *Physiotherapy,* 114, 1-8. doi:https://doi.org/10.1016/j.physio.2021.11.001

Payton, O. D. (1997). A meta-analysis of the literature on admissions criteria as predictions of academic performance in physical therapy education in the United States and Canada: 1983 through 1994. *Physiotherapy Canada, 49*(2), 97-102.

Pell, G., Fuller, R., Homer, M., & Roberts, T. (2010). How to measure the quality of the OSCE: A review of metrics - AMEE guide no. 49. *Medical Teacher, 32*(10), 802-811.

Pham, H., Harry, C., Devitt, P., Trigg, M., Wu, S., Barnard, J., & O'Connell, A. (2018). Choosing medical assessments: Does the multiple-choice question make the grade? *Education for Health, 31*(2), 65.

Physiotherapy Board of Australia & Physiotherapy Board of New Zealand. (2015). Physiotherapy practice thresholds in Australia and Aotearoa New Zealand. Retrieved from http://www.physiotherapyboard.gov.au/Accreditation.aspx

Pitts, J. (2010). Portfolios, personal development and reflective practice. In T. Swanwick & E. B. L. E. Library (Eds.), *Understanding Medical Education: Evidence, Theory and Practice.* Hoboken: John Wiley & Sons, Ltd.

Prideaux, D., Roberts, C., Eva, K., Centeno, A., McCrorie, P., McManus, C., Patterson, F., Powis, D., Tekian, A., Wilkinson, D. (2011). Assessment for selection for t he health care professions and specialty training: consensus statement and recommendations from the Ottawa 2010 Conference. *Medical Teacher, 33*(3), 215-223.

Probert, C. S., Cahill, D. J., McCann, G. L., & Ben-Shlomo, Y. (2003). Traditional finals and OSCEs in predicting consultant and self-reported clinical skills of PRHOs: a pilot study. *Medical Education, 37*(7), 597-602.

Rees, C. E., & Sheard, C. E. (2004). The reliability of assessment criteria for undergraduate medical students' communication skills portfolios: the Nottingham experience. *Medical Education, 38*(2), 138-144.

Regehr, G., MacRae, H., Reznick, R. K., & Szalay, D. (1998). Comparing the psychometric properties of checklists and global rating scales for assessing performance on an OSCE-format examination. *Academic Medicine, 73*(9), 993-997.

Reubenson, A., Ng, L., & Gucciardi, D. F. (2020). The Assessment of Physiotherapy Practice tool provides informative assessments of clinical and professional dimensions of student performance in undergraduate placements: a longitudinal validity and reliability study. *Journal of Physiotherapy, 66*(2), 113-119.

Reynolds, K., Bazemore, C., Hanebuth, C., Hendren, S., & Horn, M. (2021). The relationship of non-cognitive factors to academic and clinical performance in graduate rehabilitation science students in the United States: a systematic review. *Journal of Educational Evaluation for Health Professions, 18*, 31.

Reznick, R. K., Blackmore, D., Dauphinee, W. D., Rothman, A. I., & Smee, S. (1996). Large-scale high-stakes testing with an OSCE: report from the Medical Council of Canada. *Academic Medicine, 71*(1), S19-21.

Ribeiro, A. M., Ferla, A. A., & Amorim, J. S. C. d. (2019). Objective structured clinical examination in physiotherapy teaching: a systematic review. *Fisioterapia em Movimento, 32*.

Roach, K. E., Frost, J. S., Francis, N. J., Giles, S., Nordrum, J. T., & Delitto, A. (2012). Validation of the Revised Physical Therapist Clinical Performance Instrument (PT CPI): Version 2006. *Physical Therapy, 92*(3), 416-428.

Roach, K. E., Gandy, J., Deusinger, S. C., Gramet, P., Gresham, B., Hagler, P., & Rainery, Y. (2002). The development and testing of APTA Clinical Performance Instruments. American Physical Therapy Association. *Physical Therapy, 82*(4), 329-353.

Rodger, S., Chien, C. W., Turpin, M., Copley, J., Coleman, A., Brown, T., & Caine, A. M. (2016). Establishing the Validity and Reliability of the Student Practice Evaluation Form-Revised (SPEF-R) in Occupational Therapy Practice Education: A Rasch Analysis. *Evaluation & the Health Professions, 39*(1), 33-48.

Rodger, S., Coleman, A., Caine, A. M., Chien, C. W., Copley, J., Turpin, M., & Brown, T. (2014). Examining the inter-rater and test-retest reliability of the Student Practice Evaluation Form-Revised (SPEF-R) for occupational therapy students. *Australian Occupational Therapy Journal, 61*(5), 353-363.

Rodger, S., Webb, G., Devitt, L., Gilbert, J., Wrightson, P., & McMeeken, J. (2008). A clinical education and practice placements in the allied health professions: an international perspective. *Journal of Allied Health, 37*(1), 53-62.

Rodriguez, M. C. (2003). Construct equivalence of multiple-choice and constructed-response items: A random effects synthesis of correlations. *Journal of Experimental Medicine, 40*(2), 163.

Roman, G., & Buman, M. P. (2019). Preadmission predictors of graduation success from a physical therapy education program in the United States. *Journal of Educational Evaluation for Health Professions, 16:5.* doi:https://doi.org/10.3352/jeehp.2019.16.5

Rushforth, H. E. (2007). Objective structured clinical examination (OSCE): review of literature and implications for nursing education. *Nurse Education Today, 27*(5), 481-490.

Saarinen-Rahiika, H., & Binkley, J. M. (1998). Problem-based learning in physical therapy: a review of the literature and overview of the McMaster University experience. *Physical Therapy, 78*(2), 195-207.

Saldaña, J. (2021). *The coding manual for qualitative researchers*: sage.

Salvatori, P. (2001). Reliability and validity of admissions tools used to select students for the health professions. *Advances in Health Sciences Education: Theory and Practice, 6*(2), 159-175.

Sandella, J. M., Smith, L. A., & Dowling, D. J. (2014). Consistency of interrater scoring of student performances of osteopathic manipulative treatment on COMLEX-USA Level 2-PE. *Journal of the American Osteopathic Association, 114*(4), 253-258.

Schmidt, H. G. (1983). Problem-based learning: rationale and description. *Medical Education, 17*(1), 11-16.

Schwartzman, E., Hsu, D. I., Law, A. V., & Chung, E. P. (2011). Assessment of patient communication skills during OSCE: Examining effectiveness of a training program in minimizing inter-grader variability. *Patient Education and Counseling, 83*(3), 472-477.

Scriven, M. (1967). The Methodology of Evaluation In R. Tyler, R. Gagne, & M. Scriven (Eds.), Perspectives of Curriculum Evaluation. Chicago: Rand McNally.

Shorey, S., Lau, T. C., Lau, S. T., & Ang, E. (2019). Entrustable professional activities in health care education: a scoping review. *Medical Education*, 53(8), 766-777.

Silva, C. C., Lunardi, A. C., Mendes, F. A., Souza, F. F., & Carvalho, C. R. (2011). Objective structured clinical evaluation as an assessment method for undergraduate chest physical therapy students: a cross-sectional study. *Brazilian Journal of Physical Therapy, 15*, 481-486.

Siu, E., & Reiter, H. I. (2009). Overview: what's worked and what hasn't as a guide towards predictive admissions tool development. *Advances in Health Sciences Education, 14*(5), 759.

Sladek, R. M., Bond, M. J., Frost, L. K., & Prior, K. N. (2016). Predicting success in medical school: a longitudinal study of common Australian student selection tools. *BMC Medical Education, 16*, 187.

Sloan, D. A., Donnelly, M. B., Schwartz, R. W., & Strodel, W. E. (1995). The Objective Structured Clinical Examination. The new gold standard for evaluating postgraduate clinical performance. *Annals of Surgery, 222*(6), 735-742.

Smedts, A. M., Campbell, N., & Sweet, L. (2013). Work-integrated learning (WIL) supervisors and nonsupervisors of allied health professional students. *Rural Remote Health, 13*(1).

Straube, D., & Campbell, S. K. (2003). Rater discrimination using the visual analog scale of the Physical Therapist Clinical Performance Instrument. *Journal of Physical Therapy Education, 17*(1), 33.

Sutton, M., Govier, A., Prince, S., & Morphett, M. (2015). Primary-contact physiotherapists manage a minor trauma caseload in the emergency department without

misdiagnoses or adverse events: an observational study. *Journal of Physiotherapy, 61*(2), 77-80.

Swanson, D. B., & van der Vleuten, C. P. (2013). Assessment of clinical skills with standardized patients: state of the art revisited. *Teaching and learning in medicine, 25*(1), 842916.

Swift, M., Spake, E., & Gajewski, B. J. (2013). The Reliability of a Musculoskeletal Objective Structured Clinical Examination in a Professional Physical Therapist Program. *Journal of Physical Therapy Education, 27*(2), 41-48.

Taras, M. (2005). Assessment – Summative and Formative – Some Theoretical Reflections. *British Journal of Educational Studies, 53*(4), 466-478.

Tarrant, M., & Ware, J. (2008). Impact of item-writing flaws in multiple-choice questions on student achievement in high-stakes nursing assessments. *Medical Education*, 42(2), 198-206.

Taylor, D., Park, Y. S., Smith, C., Cate, O. t., & Tekian, A. (2020). Constructing approaches to entrustable professional activity development that deliver valid descriptions of professional practice. *Teaching and Learning in Medicine*, 33(1), 89-97.

Tekian, A., & Yudkowsky, R. (2009). Oral Examinations. In S. M. Downing & R. Yudkowsky (Eds.), Assessment in Health Professions Education.

Ten Cate, O., Chen, H. C., Hoff, R. G., Peters, H., Bok, H., & van der Schaaf, M. (2015). Curriculum development for the workplace using entrustable professional activities (EPAs): AMEE guide no. 99. *Medical Teache*r, 37(11), 983-1002.

Terry, R., Hing, W., Orr, R., & Milne, N. (2017). Do coursework summative assessments predict clinical performance? A systematic review. *BMC Medical Education*, 17(1), 40.

Terry, R., Hing, W., Orr, R., & Milne, N. (2020). Relationships between pre-clinical summative assessment scores and the clinical performance of physiotherapy students. *Journal of Allied Health, 49*(1), 13E-19E.

Terry, R., Hing, W., Orr, R. M., & Milne, N. (2018). Undergraduate Grade-point Average as a Selection Criterion for a Postgraduate Entry-level Physiotherapy Program. *Australian Journal of Clinical Education, 3*(1), 1-17.

Thieman, T. J., Weddle, M. L., & Moore, M. A. (2003). Predicting academic, clinical, and licensure examination performance in a professional (entry-level) master's degree program in physical therapy. *Journal of Physical Therapy Education, 17*(2), 32-37.

Tochel, C., Haig, A., Hesketh, A., Cadzow, A., Beggs, K., Colthart, I., & Peacock, H. (2009). The effectiveness of portfolios for postgraduate assessment and education: BEME Guide No 12. *Medical Teacher, 31*(4), 299-318.

Traska, M. R. (1988). Perspectives. The perils of competition. *Mcgraw Hills Med Health, 42*(36), suppl 4 p.

Turpin, M., Fitzgerald, C., & Rodger, S. (2011). Development of the Student Practice Evaluation Form Revised Edition Package: SPEF-R DEVELOPMENT. *Australian Occupational Therapy Journal, 58*(2), 67-73.

United States Medical Licensing Examination ®. (2021). Work to Relaunch USMLE Step 2 CS Discontinued [Press release]. Retrieved from https://www.usmle.org/announcements/?ContentId=309

Utzman, R. R., Riddle, D. L., & Jewell, D. V. (2007). Use of demographic and quantitative admissions data to predict academic difficulty among professional physical therapist students. *Physical Therapy, 87*(9), 1164-1180.

Valle, R., Petra, L., Martínez-Gonzáez, A., Rojas-Ramirez, J. A., Morales-Lopez, S., & Piña-Garza, B. (1999). Assessment of student performance in problem-based learning tutorial sessions. *Medical Education, 33*(11), 818-822.

van der Vleuten, C. P., & Schuwirth, L. W. (2005). Assessing professional competence: from methods to programmes. *Medical Education, 39*(3), 309-317.

van der Vleuten, C. P., Schuwirth, L. W., Driessen, E., Dijkstra, J., Tigelaar, D., Baartman, L. K., & van Tartwijk, J. (2012). A model for programmatic assessment fit for purpose. *Medical Teacher, 34*(3), 205-214.

van der Vleuten, C. P., & Schuwirth, L. W. T. (2019). Assessment in the context of problem-based learning. *Advances in Health Sciences Education : theory and practice, 24*(5), 903-914.

van der Want, A. C., Bloemendaal, P. M., & van der Hage, J. A. (2021). Examiners' Perceptions in Surgical Education: The Blind Spot in the Assessment of OSCEs. *Journal of Surgical Education, 78*(2), 590-596.

Vendrely, A. M. (2007). An investigation of the relationships among academic performance, clinical performance, critical thinking, and success on the physical therapy licensure examination. *Journal of Allied Health, 36*(2), 108-123.

Viera, A. J., & Garrett, J. M. (2005). Understanding interobserver agreement: the kappa statistic. *Family Medicine, 37*(5), 360-363.

Wass, V., van der Vleuten, C. P., Shatzer, J., & Jones, R. (2001). Assessment of clinical competence. *Lancet, 357*(9260), 945-949.

Wass, V., Wakeford, R., Neighbour, R., & van der Vleuten, C. P. (2003). Achieving acceptable reliability in oral examinations: an analysis of the Royal College of General Practitioners membership examination's oral component. *Medical Education, 37*(2), 126-131.

Watson, C. J., Barnes, C. A., & Williamson, J. W. (2000). Determinants of clinical performance in a physical therapy program. *Journal of Allied Health, 29*(3), 150-156.

Weber, R. P. (1990). *Basic content analysis*: Sage.

Wessel, J., Williams, R., Finch, E., & Gemus, M. (2003). Reliability and validity of an objective structured clinical examination for physical therapy students. *Journal of Allied Health, 32*(4), 266-269.

Wilkinson, T. J., & Frampton, C. M. (2003). Assessing performance in final year medical students. Can a postgraduate measure be used in an undergraduate setting? *Medical Education, 37*(3), 233-240.

Wilkinson, T. J., & Frampton, C. M. (2004). Comprehensive undergraduate medical assessments improve prediction of clinical performance. *Medical Education, 38*(10), 1111-1116.

Wolden, M., Hill, B., & Voorhees, S. (2020). Predicting Success for Student Physical Therapists on the National Physical Therapy Examination: Systematic Review and Meta-Analysis. *Physical Therapy, 100*(1), 73-89.

Wood, D. F. (2010). Formative Assessment. In T. Swanwick (Ed.), *Understanding Medical Education* (pp. 259-270). Oxford, UK: Wiley-Blackwell.

Wood, D. F. (2013). Formative assessment. *Understanding Medical Education*, 317-323.

World Confederation for Physical Therapy. (2019). Description of physical therapy policy statement. Retrieved from https://world.physio/sites/default/files/2020-07/PS-2019-Description-of-physical-therapy.pdf. Accessed on 30 Dec 2022

World Physiotherapy. (2021). *Physiotherapist education framework* (978-1-914952-01-2). Retrieved from London, UK: https://world.physio/sites/default/files/2021-07/Physiotherapist-education-framework-FINAL.pdf

Yusoff, M. S. B. (2019). Multiple Mini Interview as an admission tool in higher education: Insights from a systematic review. *Journal of Taibah University Medical Sciences, 14*(3), 203-240.

Zamanzadeh, V., Ghaffari, R., Valizadeh, L., Karimi-Moonaghi, H., Johnston, A. N. B., & Alizadeh, S. (2021). Challenges of objective structured clinical examination in undergraduate nursing curriculum: Experiences of faculties and students. *Nurse Education Today, 103*, 104960–104960.

Zhang, N., & Walton, D. M. (2018). Why So Stressed? A Descriptive Thematic Analysis of Physical Therapy Students' Descriptions of Causes of Anxiety during Objective Structured Clinical Exams. *Physiotherapy Canada, 70*(4), 356-362.

Appendices

Appendix 1. Assessment of Physiotherapy Practice (APP) Performance Indicators

Professional Behaviour

1. Demonstrates an understanding of client rights and consent
- obtains & records informed consent according to protocol
- recognises clients' health-care rights
- prioritises clients' rights, needs & interests
- allows sufficient time to discuss the risks &benefits of the proposed treatment with clients &carers
- refers clients to a more senior staff member for consent when appropriate
- advises supervisor or other appropriate person if a client might be at risk
- respects clients' privacy & dignity
- complies with confidentiality & privacy requirements for client's health & personal information
- applies ethical principles to the collection, maintenance, use & dissemination of data &information

2. Demonstrates commitment to learning
- responds in a positive manner to questions, suggestions &/or constructive feedback
- reviews & prepares appropriate material before &during placement
- develops & implements a plan of action in response to feedback
- seeks information/assistance as required
- demonstrates self-evaluation, reflects on progress& implements appropriate changes based on reflection
- takes responsibility for learning & seeks opportunities to meet learning needs
- uses clinic time responsibly

3. Demonstrates ethical, legal &culturally responsive practice
- follows policies & procedures of the facility
- advises appropriate staff of circumstances that may affect adequate work performance
- observes infection control, & workplace health & safety policies
- arrives fit to work
- arrives punctually & leaves at agreed time
- calls appropriate personnel to report intended absence
- wears an identification badge & identifies self
- recognises inappropriate or unethical health practice
- observes dress code
- completes projects/tasks within designated time frame
- maintains appropriate professional boundaries with clients & carers
- advocates for clients & their rights (where appropriate)
- demonstrates appropriate self-care strategies(e.g., management of stress, mental & physical health issues)
- acts ethically & applies ethical reasoning in all health care activities
- demonstrates skills in culturally safe & responsive
- client-centred practice
- acts within bounds of personal competence, recognizing personal & professional strengths& limitations

4. Demonstrates collaborative practice
- demonstrates understanding of team processes
- contributes appropriately in team meetings
- acknowledges expertise & role of other healthcare professionals & refers/liaises as appropriate to access relevant services
- advocates for the client when dealing with other services
- collaborates with the health care team & client to achieve optimal outcomes
- cooperates with other people who are treating& caring for clients
- guides & motivates support staff (where appropriate)
- works collaboratively & respectfully with support staff

Communication

5. Communicates effectively and appropriately Verbal/non-verbal
- greets others appropriately
- questions effectively to gain appropriate information
- listens carefully & is sensitive & empathetic to views of client & relevant others
- respects cultural & personal differences of others
- gives appropriate, positive reinforcement
- provides clear instructions
- uses suitable language & avoids jargon
- demonstrates an appropriate range of communication styles (with e.g., clients, carers, administrative & support staff, health professionals, care team)
- recognises barriers to optimal communication
- responds appropriately to non-verbal cues
- integrates communication technology into practice as required
- uses a range of communication strategies to optimise client rapport & understanding (e.g. Hearing impairment, non-English speaking, cognitive impairment, consideration of non-verbal

communication
- uses accredited interpreters appropriately
- maintains effective communication with clinical educators
- recognises risk of conflict & takes appropriate action to mitigate &/or resolve
- actively explains to clients & relevant others their role in care, decision-making & preventing adverse events
- actively encourages clients to provide complete information without embarrassment or hesitation
- conducts communication with client in a manner& environment that demonstrates consideration of confidentiality, privacy & client's sensitivities
- negotiates appropriately with other health professionals

10. Demonstrates clear and accurate documentation
- writes legibly
- completes relevant documentation to the required standard (e.g., client record , statistical information, referral letters)
- maintains records compliant with legislative medico-legal requirements
- complies with organisational protocols &legislation for communication
- adapts written material for a range of audiences(e.g., provides translated material for non-English speaking people, considers reading ability, age of client)

Assessment

7. Conducts an appropriate client interview
- positions person safely & comfortably for interview
- structures a systematic, purposeful interview seeking qualitative & quantitative details
- provides a culturally safe environment for the client
- asks relevant & comprehensive questions
- politely controls the interview to obtain relevant information
- responds appropriately to important client cues
- identifies client's goals & expectations
- conducts appropriate assessment with consideration of the social, personal, environmental & biopsychosocial factors that influence function, health & disability.
- seeks appropriate supplementary information, accessing other information, records, test results as appropriate & with client's consent
- generates diagnostic hypotheses, identifying priorities& urgency of further assessment & intervention
- completes assessment in acceptable time

8. Selects and measures relevant health indicators and outcomes
- selects appropriate variable/s to be measured at baseline from WHO ICF domains of impairment, activity limitation & participation restriction.
- identifies & justifies variables to be measured to monitor treatment response & outcome.
- selects appropriate tests/outcome measures for each variable for the purpose of diagnosis, monitoring &outcome evaluation.

- links outcome variables with treatment goals
- communicates the treatment evaluation process & outcomes to the client & relevant others
- identifies, documents & acts on factors that may compromise treatment outcomes

9. Performs appropriate physical assessment procedures
- considers client comfort & safety
- respects client's need for privacy & modesty (e.g., provides draping or gown)
- structures systematic, safe & goal oriented assessment processes accommodating limitations imposed by client's health status
- plans assessment structure & reasoning process using information from client history & supportive information
- demonstrates sensitive & appropriate handling during the assessment process
- applies tests & measurements safely, accurately& consistently
- sensibly modifies assessment in response to client profile, feedback & relevant findings
- performs appropriate tests to refine diagnosis
- assesses/appraises work, home or other relevant environments as required
- completes assessment in acceptable time

Analysis & Planning

10. Appropriately interprets assessment findings
- describes the implications of test results
- describes the presentation & expected course of common clinical conditions
- relates signs & symptoms to pathology
- relates signs, symptoms & pathology to environmental tasks & demands
- interprets findings at each stage of assessment to progressively negate or reinforce hypothesis/es
- makes justifiable decisions regarding diagnoses based on knowledge & clinical reasoning
- prioritises important assessment findings
- compares observed findings to expected findings

11. Identifies and prioritises client's problems
- generates a list of problems from the assessment
- justifies prioritisation of problem list based on knowledge & clinical reasoning
- collaborates with client to prioritise problems
- considers client's values, priorities & needs

12. Sets realistic short and long term goals with the client
- negotiates realistic short term treatment goals in partnership with client
- negotiates realistic long term treatment goals in partnership with client
- formulates goals that are specific, measurable,
- achievable & relevant, with specified timeframe
- considers physical, emotional & financial costs, & relates them to likely gains of intervention

13. Selects appropriate intervention in collaboration with the client

- engages with client to explain assessment findings, discuss intervention strategies & develop an acceptable plan
- identifies & justifies options for interventions based on client needs, clinical guidelines, best evidence & available resources
- considers whether physiotherapy is indicated
- demonstrates a suitable range of skills & approaches to intervention
- describes acceptable rationale (e.g. likely effectiveness) for treatment choices
- balances needs of clients & relevant others with the need for efficient & effective intervention
- demonstrates understanding of contraindications & precautions in selection of intervention strategies
- advises client about the effects of treatment or no treatment

Intervention

14. Performs interventions appropriately

- considers the scheduling of treatment in relation to other procedures e.g., medication for pain, wound care.
- demonstrates appropriate client handling skills in performance of interventions
- performs techniques at appropriate standard
- minimizes risk of adverse events to client & self in performance of intervention (including observance of infection control procedures & manual handling standards)
- prepares environment for client including necessary equipment for treatment
- identifies when group activity might be inappropriate intervention
- demonstrates skill in case management
- recognises when to enlist assistance of others to complete workload
- completes intervention in acceptable time
- refers client to other professional/s when physiotherapy intervention is not appropriate, or requires a multi-disciplinary approach

15. Is an effective educator

- demonstrates skill in client education & health promotion e.g., modifies approach to suit client age group &/or cultural needs
- applies adult learning principles in education of clients & relevant others
- educates assistants & relevant others to implement safe & effective therapy
- participates in leading educational activities for peers/staff (where appropriate)
- demonstrates skills in conducting group sessions
- develops a realistic self-management program for prevention & management in collaboration with the client
- provides information using a range of strategies that demonstrate consideration of client needs
- confirms client's/relevant others' understanding of given information
- uses appropriate strategies to motivate the client &

relevant others to participate & to take responsibility for achieving defined goals
- discusses expectations of physiotherapy intervention & its outcomes
- provides feedback to client regarding health status
- educates the client in self evaluation
- encourages & acknowledges achievement of short & long term goals

16. Monitors the effects of intervention

- incorporates relevant evaluation procedures/outcome measures in the physiotherapy plan
- monitors client response to the intervention
- makes modifications to intervention based on therapist evaluation & client feedback
- records & communicates outcomes where appropriate

17. Progresses intervention appropriately

- demonstrates or describes safe & sensible treatment progressions
- makes decisions regarding modifications, continuation or cessation of intervention in consultation with the client, based on best available evidence
- discontinues treatment in the absence of measurable benefit

18. Undertakes discharge planning

- begins discharge planning in collaboration with the health care team at the time of the initial episode of care
- discusses discharge planning with the client
- describes strategies that may be useful for maintaining or improving health status following discharge
- arranges appropriate follow-up health care to meet short & long term goals
- addresses client & carer needs for ongoing care through the coordination of appropriate services

Evidence Based Practice

19. Applies evidence based practice in client-centred care

- considers the research evidence, client preferences, clinical expertise & available resources in making treatment decisions & advising clients
- practises in accordance with relevant clinical practice guidelines
- locates & applies relevant current evidence e.g. clinical practice guidelines & systematic reviews
- assists clients & carers to identify reliable & accurate health information
- shares new evidence with colleagues
- participates in & applies quality improvement procedures when possible

Risk Management

20. Identifies adverse events and near misses and minimises risk associated with assessment and interventions

- monitors client safety during assessment & treatment

- complies with workplace guidelines on manual handling
- complies with organizational health & safety requirements
- describes relevant contraindications & precautions associated with assessment & treatment
- recognises & reports adverse events & near misses to appropriate members of the team

- implements appropriate measures in case of emergency
- reports inappropriate or unsafe behaviour of a co-worker or situations that are unsafe
- prior to client contact, reports any personal issues(physical/mental) that may impact on client care