Evaluating the quality of clinical teaching in osteopathy

Thesis submitted in fulfilment of the requirements for the degree of
Doctor of Philosophy

Institute of Health and Sport
Victoria University

Brett Robert Vaughan

October 2020

ABSTRACT

Clinical education is a core component of all pre-professional health professions courses. This experiential learning allows the student to utilise the skills and knowledge developed in the classroom, workshop and simulated learning environments, in patient care. Supervision of patient care is undertaken by qualified health professionals with their role being the facilitation of the students' learning to prepare them for clinical practice. This approach to teaching and learning contrasts with academic teaching whereby clinical teaching is embedded in the patient care environment rather than a classroom or simulated care environment. In osteopathy, this patient care and clinical education is undertaken in student-led clinics. Student-led clinics are broadly defined as an education and clinical environment where the student leads patient care under supervision of a qualified health professional.

Clinical teaching and supervision in the student-led context relates to supervision of patient care by the student, and the demonstration of patient care skills, knowledge and attitudes by the clinical educator. As part of the quality assurance for this education, students are asked to provide feedback about the quality of clinical teaching and supervision they receive. Most often, this feedback is in the form a questionnaire. Questionnaires in the literature have been developed for the hospital and tertiary care contexts. However, there is no such questionnaire for the student-led clinical environment. This Thesis by Publication presents the development and testing of a questionnaire to allow students to evaluate the quality of clinical teaching they receive. The conceptual framework was adopted for this thesis was Kane's approach to validity – developing a validity argument for the score derived from the questionnaire. To develop evidence for the validity argument, four studies were undertaken.

Study one reviewed the clinical teaching evaluation and quality literature and developed an initial questionnaire item bank. An initial bank of 83 items was reviewed by stakeholders and refined to 56 items. The 56 items were then tested with osteopathy students. The resultant exploratory factor analysis identified a 5-factor, 30

item measure. Study two utilised item response theory (Rasch analysis) to refine the 30-item questionnaire and provide additional evidence for the measurement properties. This study resulted in the development of a 12-item measure of clinical teaching quality that demonstrated fir to the Rasch measurement model and the properties of measurement invariance. Study three explored the reliability of the questionnaire, including the application of generalisability analysis to determine the number of questionnaires needed for a single educator to draw reliable inferences from the data. This study identified that eight ratings of a clinical educator by individual students would be reliable. Study four evaluated the relationship between student perceptions of clinical teaching and clinical educator self-efficacy and self-evaluation of their clinical teaching quality. This final study demonstrated that those clinical educators with high self-efficacy were rated lower by students than those clinical educators who reported low self-efficacy with respect to clinical teaching.

This research developed a 12-item questionnaire that is feasible, reliable and provides evidence for the validity of the score derived from the questionnaire. The questionnaire is designed to be incorporated into a suite of quality assurance evaluation tools to ensure that patients receive high-quality care and students receive the best possible educational experience to prepare them for their life as a health professional.

DECLARATION OF AUTHENTICITY

I, Brett Robert Vaughan, declare that the PhD Thesis by Publication entitled **Evaluating the quality of clinical teaching in osteopathy** is no more than 100,000 words in length including quotes and exclusive of tables, figures, appendices, bibliography, references and footnotes. This thesis contains no material that has been submitted previously, in whole or in part, for the award of any other academic degree or diploma. Except where otherwise indicated, this thesis is my own work.

Signature 

Date October 5th, 2020

DETAILS OF INCLUDED PAPERS

Chapter	Title	Publication status	Publication title and details
2	Developing a clinical teaching quality questionnaire for use in a university osteopathic pre-registration teaching program	Published	BMC Medical Education, 2015, 15(70): doi: 10.1186/s12909-015-0358-6 Scimago Journal Rank: Q1 H-Index: 54 2-year/5-year Impact Factor: 1.83/2.23 Source Normalized Impact per Paper: 1.52
3	Exploring the measurement properties of the osteopathy clinical teaching questionnaire using Rasch analysis	Published	Chiropractic & Manual Therapies, 2018, 26(13): doi: 10.1186/s12998-018-0182-2 Scimago Journal Rank: Q1 H-Index: 30 2-year/5-year Impact Factor: 1.51/Not available Source Normalized Impact per Paper: 1.02
4	Evaluation of teaching in a student-led clinic environment: Assessing the reliability of a questionnaire	Published	International Journal of Osteopathic Medicine, 2019, 31;28-35: doi: 10.1016/j.ijosm.2018.11.001 Scimago Journal Rank: Q2 H-Index: 17 2-year/5-year Impact Factor: 1.20/0.95 Source Normalized Impact per Paper: 0.45
5	Clinical educator self-efficacy, self-evaluation and its relationship with student evaluations of clinical teaching	Published	BMC Medical Education, 2020, 20(347): doi: 10.1186/s12909-020-02278-z Scimago Journal Rank: Q1 H-Index: 54 2-year/5-year Impact Factor: 1.83/2.23 Source Normalized Impact per Paper: 1.52

OTHER WORKS DERIVED FROM THE THESIS

Conference Presentations

Vaughan, B. (2015, July). Developing a clinical teaching quality questionnaire for use in a university osteopathic pre-registration teaching program. Poster session presented at the Australian and New Zealand Association of Health Professional Educators Conference 2015, Newcastle, Australia.

Vaughan, B. (2015, October). Exploring the construct validity of the Osteopathy Clinical Teaching Questionnaire using Rasch analysis. Poster session presented at the Chiropractic and Osteopathic College of Australasia Conference, Sydney, Australia. (Best postgraduate student poster prize)

Vaughan, B. (2017, December). Exploring the construct validity of the Osteopathy Clinical Teaching Questionnaire using Rasch analysis. Podium session presented at the British School of Osteopathy Education Conference, London, United Kingdom.

Vaughan, B. (2018, July). Mokken scale analysis of a questionnaire to evaluate teaching in a student-led clinic. Poster session presented at the Australian and New Zealand Association of Health Professional Educators Conference 2018, Hobart, Australia.

Vaughan, B. (2018, July). Clinical educator self-efficacy and its relationship with student evaluations of clinical teaching. Poster session presented at the Australian and New Zealand Association of Health Professional Educators Conference 2018, Hobart, Australia.

Scholarship of Teaching and Learning

Cochrane, A., Cham, K., Guest, D., Vaughan B. (2019). Evaluating the effectiveness of clinical teaching through the student and the clinical educator lens. Melbourne School of Health Sciences Education Scholarship Grant ($4,983.48)

Noyer, A., Baker, T., Draper-Rodi, J., Thomson, O., MacMillan, A., Vaughan, B. (2020). The relationship between student perception of clinical teaching quality and clinical educator characteristics in a UK-based osteopathic educational institution.

Professional Presentations

Vaughan, B. (2018, November). Evaluating the quality of clinical teaching in osteopathy. Department of Medical Education, The University of Melbourne, Melbourne, Australia.

ACKNOWLEDGEMENTS

Over my doctoral journey I have been fortunate enough to work with many people, all of whom have played a role in one way or another. Firstly, to my doctoral supervisors Patrick McLaughlin, Keri Moore and Jane Mulcahy, thank you for allowing me to pursue a project of my own choosing, for the sage and pointed advice from the very beginning to the conclusion, your spirit of generosity in allowing me to sole author the publications in this thesis, and for answering the inevitable "but why?" question. I hope that this thesis is a strong reflection of the guidance and support offered throughout the journey.

To my colleagues from Victoria University, and particularly those in the osteopathy program, thank you for supporting the work I undertook as part of this thesis. A special thanks to Tracy Denning, Kylie Fitzgerald, Douglas Wong and Michael Fleischmann for their support and for listening to me talk about the publications in this thesis, what I had found and how things may have been done differently. It was cathartic.

I would also like to extend my thanks to my colleagues in the Department of Medical Education at The University of Melbourne. Everyone has been wonderfully supportive of me finishing my thesis, offering encouragement and guidance in the process. I would particularly like to thank Clare Delany, Robyn Woodward-Kron, Anna Ryan and Stephen Trumble for their support and constant questioning of when I would finish. I did get there.

To my osteopathy colleagues who kindly participated in the studies, and the students who also did, thank you. This work would never have been possible without you all completing questionnaires and allowing me to talk endlessly about the project. I hope that this thesis provides a small step forward in our knowledge of how we do clinical education in osteopathy. Thanks are also extended to Georgina Fargher for her generosity and time editing this thesis.

Finally, to my family and friends who put up with me saying no to events and gatherings, and long periods stuck at a desk reading or writing. Thank you all for your support in giving me the latitude to pursue this work.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

LIST OF ABBREVIATIONS

EMRs – Electronic medical records

OCTQ – Osteopathy Clinical Teaching Questionnaire

SECT – Self-efficacy of Clinical Teachers tool

SET – Student evaluation of teaching

CHAPTER 1

Introduction

"Teaching is a complex, dynamic process subject to myriad contextual factors, although it may be no less amenable to valid measurement than other educational constructs such as empathy and professionalism" (Conigliaro & Stratton, 2010, p. 385).

Learning in the clinical context is vital for the education of future health professionals. But how do we ensure that the quality of clinical teaching is suitable for this learning environment? Evaluation of the quality of clinical teaching from the perspective of the learner can provide an insight. This thesis seeks to complement the quality assurance framework for osteopathy clinical education. The objective is to develop a measure that affords learners an opportunity to evaluate the quality of clinical teaching.

Measures used to evaluate clinical teaching effectiveness or quality should be suited to the specific requirements of the clinical learning environment (Brown et al., 2013; Fluit et al., 2014) and treat teaching effectiveness as a latent construct (Spooren et al., 2013). Learners also appreciate the role of different measures for the classroom and clinical learning environments (Pettit et al., 2015). With these considerations as guiding principles, the thesis describes the development of a questionnaire for the student-led clinical learning environment – where the learner leads the care of a patient under the supervision of a qualified health professional – which is the dominant environment for osteopathy clinical education.

Researcher context

My clinical background in the health professions is as an osteopath. I am in private osteopathy practice and combine this with academic, research and governance roles in osteopathy and the broader health professions. The basis of this thesis was informed by my time as an educator, a clinical educator and clinical coordinator for the osteopathy program at Victoria University. The latter role I was able to job share with another colleague. Our ambition was to broadly improve the clinical learning environment through the implementation of contemporary clinical (workplace-based) assessments and provide the clinical educators with the support and tools needed to

enhance the learner experience though improved teaching. I had identified the latter as a significant way in which the learning environment could be improved. However, the only evaluation tool available within the university was the institution-wide student evaluation of teaching. As will be described in the section on 'The evaluation of clinical teaching' in this chapter, these tools were not designed for use in a clinical learning environment, and therefore the data derived from them was of little value.

This thesis evolved from initial thinking about how to evaluate the quality of clinical teaching in the osteopathy clinical learning environment to identification of potential tools to measure this quality. In this thesis, I present four studies that describe the development, testing and application of a questionnaire that allows learners to evaluate the quality of the clinical teaching they receive. I am now fortunate in my role as a lecturer in clinical education within the Department of Medical Education at The University of Melbourne to be able to contribute to the professional development of clinical educators from a range of health professions on a daily basis, whilst still contributing to the development of the osteopathy profession and education of future osteopaths.

Thesis structure

The thesis is structured as six chapters with peer-review publications comprising chapters 2 through to 5. Chapter 1 provides a context for the development and testing of a questionnaire to evaluate the quality of clinical teaching in osteopathy.

The purpose of chapter 2 is to describe the item selection, content validation and refinement of the items for the questionnaire to evaluate the quality of clinical teaching in osteopathy student-led clinics. This chapter comprises the first peer-review publication of the thesis. To begin the development of the measure, tools to allow learners to evaluate the clinical educator tools were identified in the literature. Items were extracted from these measures, and then stakeholder input was sought as to the applicability and suitability of items for a student-led clinical learning environment. Exploratory factor analysis was used to refine the measure and identify the underlying factor structure. The outcomes of this work were then described in the

context of an argument-based approach to validity (Kane, 1992) – the conceptual framework for the thesis.

Chapter 3 builds on the work presented in chapter 2 by testing and refining the measure using a modern test theory approach (Rasch analysis). To inform this analysis, data were collected from osteopathy teaching institutions in three countries (Australia, New Zealand and the United Kingdom). The Rasch analysis, along with complementary analyses of the unidimensionality and internal structure of the questionnaire, refined the measure to 12 items capturing the construct of *clinical teaching quality in osteopathy student-led clinics*. The peer review publication in chapter 3 also describes how the results of the study build into the argument-based approach to validity.

Chapter 4 utilises the 12-item questionnaire derived from analyses in chapter 3 to explore its reliability (inter- and intra-learner reliability). The peer review publication in chapter 4 particularly focuses on the number of questionnaires to be completed by learners to obtain a reliable indication about clinical educator performance. Analysis of the internal structure is also presented to reinforce the results presented in chapter 3. The peer review publication in chapter 4 also presents the study results in the context of the argument-based approach to validity.

Chapter 5 is the final peer-review publication of the thesis. This chapter describes a study exploring the relationship between learner and clinical educator self-evaluation using the 12-item questionnaire described in chapter 3. The concept of clinical educator self-efficacy is also described in this chapter and its relationship to learner evaluation of clinical teaching quality is presented. The study outcomes are also described with an emphasis on how they may be able to be used to inform professional development of clinical educators.

As the final chapter, chapter 6 brings together the preceding four peer-review publications. Taken together, the four publications set out in chapters 2 to 5 demonstrate how the questionnaire was developed, provide evidence to argue for the validity of the scores derived from the questionnaire, and demonstrate the capacity to

effectively measure clinical teaching quality in osteopathy student-led clinical learning environments. How the questionnaire may be utilised within a quality assurance framework to evaluate a learners' clinical education as a part of their osteopathy program of study is also described. To conclude the chapter, suggestions for other uses of the questionnaire are provided, along with a discussion of potential future research opportunities that will continue to develop the validity argument for the interpretation of the scores derived from the questionnaire.

Introduction to the chapter

This chapter provides the context for the development and testing of a questionnaire to allow learners to evaluate the quality of clinical teaching in osteopathy. The chapter begins by describing the argument-based approach to validity that underpins this thesis. Validity as a measurement property is then described along with the rationale for the 'interpretation and use' approach to validity articulated by Kane (1992). An overview of osteopathy as a health profession and how clinical education is undertaken in osteopathy pre-professional training programs is subsequently described. Next is a discussion of clinical education in the health professions and of who the clinical educator is. The chapter concludes with an exploration of the student evaluation of teaching (SET) literature, and how this literature pertains to learner evaluations of clinical teaching.

Theoretical perspective

"Validity is not a property of the test. Rather, it is a property of the proposed interpretations and uses of the test scores" (Kane, 2013, p. 3).

The theoretical perspective of this thesis centres on validity – a concept at the core of any assessment or evaluation. The concept and definition of validity has constantly evolved. However, it is classically defined as the tool/assessment measuring what it was designed to measure (Borsboom & Markus, 2013; Hathcoat, 2013). Initial conceptions of validity encompassed content, criterion and construct validity. Later, validity evolved to only construct validity with subcategories describing face validity,

content validity, predictive validity, and discriminative validity. These terms have provided researchers with discrete types of validity, or potentially a tick-box approach to demonstrating validity. Further, authors routinely claim that their assessment or evaluation is 'valid'. These discrete approaches to validity suggest there is a finite endpoint to achieving validity, when it is likely that changes in context, participants and so, will affect this interpretation.

Much of the literature that evaluates an assessment/evaluation tool or approach typically examines the reliability and aspects of the validity of the tool, and in some cases the efficiency and feasibility of administration. Study of these elements contributes to the defensibility of an assessment/evaluation decision but is not the whole picture. The discussion below highlights key literature with respect to the traditional conceptions of validity and then introduces the concept of an argument-based approach. This thesis advances the use of the validity argument framework proposed by Kane (1992) to address gaps that exist in our understanding of the validity of these tools and the scores derived from them, particularly in the evaluation of clinical teaching.

Traditional conceptions of validity and sources of validity in clinical teaching evaluations

Hewson and Jensen (1990) conclude that "The inventory was shown to have validity, and to be reliable with internal consistency correlations" (p. 518). Historically, studies developing assessments (Cook et al., 2014) and clinical teaching measures drew similar conclusions whereby traditional conceptions of measurement properties were evaluated in the one work. Authors of clinical teaching evaluations have typically described face validity, content validity, concurrent validity, and structural validity. Additionally, authors have often described the internal consistency of the clinical teaching evaluation, and reliability in terms of test-retest and inter-rater. This approach to validity and reliability is still encouraged in the development and testing of patient-reported outcome measures, and it is advocated that each measurement property be investigated in its own study (Consensus-based Standards for the

Selection of Health Measurement Instruments, 2010). However, this conception of validity is not as widely utilised in the educational measurement literature.

The concept of validity has evolved in the educational measurement literature. This evolution has resulted in a shift in how the concept is described in relation to individual evaluations of clinical teaching. For educational and psychological measures, the American Psychological and Education Research Associations have followed the work of Messick (1989) and published standards ('the Standards') identifying five sources of validity evidence by: (1) Content; (2) Response process; (3) Internal structure; (4) Relations to other variables; and, (5) Consequences (American Educational Research Association et al., 1999). It is evident there is overlap between some of the aforementioned traditional conceptions, and an evolution to include the 'consequences' of the measure outcome. This evolving understanding of validity provides a frame and common language by which educational and psychological measures could be developed and tested. Systematic reviews of clinical teaching measures by Beckman et al. (2005) and Fluit (2010) both utilised the Standards to summarise measures of clinical teaching quality and effectiveness, and this framing has also been used to explore validity evidence of student evaluations of teaching (Ory & Ryan, 2001).

Argument-based approaches to validity

The underlying principles of contemporary or 'unified' approaches to validity is the understanding that an assessment or evaluation tool or process, and the score from that tool, is not in itself valid, and that testing in different contexts or populations necessitates collection of additional evidence. This is particularly important as "values and societal norms change" (Hubley & Zumbo, 2011, p. 221) creating the need for additional evidence. In unified validity theory, it is the interpretation of the score derived from these tools or processes that is validated. This interpretation is an explicit statement of intended use of the score, and the fairness, equity, and systemic effects of score use - termed *consequences* (Zumbo & Hubley, 2016). Although the role of *consequences* in validity is not universally agreed (Kane, 2016a), it is accepted

that they can be positive and negative, and the process "of measuring or testing engenders consequences" (Zumbo & Hubley, 2016, p. 299).

By evaluating the consequences of the score interpretation, researchers begin to provide evidence for the interpretation of the score derived from the assessment or evaluation tool. Researchers build an 'argument' for the validity of the scores derived from an assessment or evaluation tool by combining evidence from multiple sources. Where a researcher sources this evidence is guided by the intended use of the derived score from the tool. Frameworks to guide the collection and presentation of this evidence can be utilised, as well as guiding the researcher as to the nature and volume of evidence to support the argument. Kane (2016b) states that "the argument-based approach allows for a wide range of interpretations and uses, but it requires that the claims being made be critically evaluated" (p. 309). Proponents of argument-based approaches to validity also suggest that the collection of evidence is an ongoing process – changes in context, population, knowledge and so on, will influence the validity argument and necessitate collection of further evidence (Hubley & Zumbo, 2011; Royal, 2017). LeBaron Wallace (2011) also contends that argument-based approaches are suited to evaluation studies as a single dataset can have multiple uses and explore multiple perspectives.

Kane's approach to the validity argument

Kane (2016a) describes three frameworks to gather evidence for validity: *interpretation only*; *consequences-as-indicators*; and the *interpretation and use* model. Collectively these are *interpretative arguments* that provide "an explicit statement of the inferences and assumptions inherent in the proposed interpretations and uses" (Colliver et al., 2012, p. 368). All three frameworks start with the same premise described previously – the assessment or evaluation tool itself is not valid, it is the interpretation of the score that one argues and provides evidence to support validity. How each of these frameworks is implemented determines the volume and type of evidence that could be provided to argue for the validity of the score interpretation (Royal, 2017). Further, "the overall validity argument is only as strong as the weakest link in the chain of inference" (Clauser et al., 2012, p. 166) - collecting

more evidence for a part of the argument that is already strong does little to advance the validity argument.

Although the *interpretation only* and *consequences-as-indicators* frameworks are briefly described, it is *interpretation and use* that is the conceptual framework underpinning this thesis given argument-based approaches are suitable for this type of investigation (LeBaron Wallace, 2011).

Interpretation only validity framework

The interpretation only framework is the most limited of those frameworks described in this chapter but also the simplest to evaluate. This framework focuses on the interpretation of the score derived from an assessment or evaluation tool, and deals with consequences as a separate issue (Kane, 2016a). Kane (2016a) argues this separation of interpretation from consequences narrows the scope of the argument that can be made for the validity of score interpretation as the reader is not presented with any evidence for the score use, or the impact of this use. Arguably it is difficult to separate the interpretation from the score use (Kane, 2013). For example, in the workplace-based assessment context, efforts are directed toward reducing the variance attributable to the examiner to improve the fairness for the assessee. These issues suggest there are limited scenarios where an interpretation only framework would be suitable.

Consequences-as-indicators validity framework

This second framework explores the consequences (both positive and negative) of the assessment or evaluation. This framework is consistent with that described by Messick (1989) (and adopted through the Standards) where there is a stronger emphasis on the use of psychometrics to identify positive and negative consequences. Consequences can be positive insofar as the score is representative of an intended outcome or related to another conceptually similar variable, and negative in that there is differential group impact of the score interpretation or unintended consequences for

particular aspects of the test population or potentially society (Hubley & Zumbo, 2011). The negative consequences, in particular, can be challenging to evaluate and are likely to be the element missing in many validity studies.

Messick (1989) further argues that construct under-representation and irrelevant sources of score variance are negative consequences. When present, they highlight there may be an issue within the assessment/evaluation program that requires additional consideration. In this framework, only those negative consequences that can be attributed to the test/evaluation are considered to count against the validity argument (Kane, 2016a). Construct under-representation and irrelevant sources of score variance are readily evaluated during the item development and testing phases using quantitative structural equation modelling and item response theory approaches (Cook & Beckman, 2006; Handley et al., 2008), along with qualitative approaches such as think-aloud protocols (Ercikan et al., 2010; Padilla & Benítez, 2014). The aforementioned quantitative approaches are now more accessible to researchers through improvements in technology reducing the computationally intensive nature of these statistical techniques. These improvements ensure that researchers are able to address construct issues early in the development phases.

Interpretation and use validity framework

The basis for an argument-based approach is twofold: one is to detail the proposed *interpretation* of the score derived from the assessment or evaluation; and two is to describe the scores *use* through development of a *validity argument* (Cook et al., 2015; Kane, 2016a). In the context of evaluation, LeBaron Wallace (2011) suggests that the *interpretation* could be described as specification and the *validity argument* as evaluation. Consideration of consequences is built into the validity argument (Knorr & Klusmann, 2015), and in contrast to the *consequences-as-indicators* argument, negative consequences, regardless of their source, are considered to count as evidence against validity (Kane, 2016a).

Schuwirth and van der Vleuten (2012) and Clauser et al. (2012) posit that Kane's approach provides somewhat directed methodology to develop the validity argument.

However, Kane himself states that "I do not want to suggest that all IUAs [interpretation and use arguments] have to follow any particular pattern" (Kane, 2013, p. 10). From an evaluation standpoint Kane's approach provides a degree of flexibility above and beyond that described by Messick (1989) and the Standards. Further, the *interpretation and use* framework is applicable to a range of quantitative and qualitative tools (Cook et al., 2015) – both common approaches to data collection in evaluation studies and programs. Although the Standards and argument-based approaches are both described in health professions education, it appears that the latter approach proposed by Kane is finding favour in the assessment context (Clauser et al., 2012; Cook et al., 2015; Hatala et al., 2015; Schuwirth & van der Vleuten, 2012; Vaughan & Moore, 2015). For the aforementioned reasons, and the flexibility provided by Kane's *interpretation and use* approach to a validity argument, it is used as the conceptual framework for this thesis.

'Interpretation and use' as a conceptual framework

Kane (1992, 2016a) contends four inferences can be used to develop the validity argument. The four inferences are: (1) scoring; (2) generalisation; (3) extrapolation; and (4) interpretation or implications. Put simply, the objective is to evaluate "the evidence…that progressively transforms an isolated observation into a defensible action" (Hatala et al., 2015, p. 1152) or a defensible decision. The proposed *use* of the score is to be defined before making an inference, and then the interpretation and use are evaluated by a validity argument (Kane, 2016a). It is important to appreciate that all four inferences do not need to be addressed or explored with the same stringency depending on the proposed *use*. For example, where the score use is formative, generalisation evidence may be of less importance given the focus is on the individual. The four sections below provide an overview of each of these inferences.

Scoring

The scoring inference is the most commonly presented evidence for assessments and educational measures, with the score resulting from some form of measurement (Kane, 1992). For assessments such as multiple-choice tests, scoring rules leave no

room for interpretation of the outcome, that is, the item is either answered correctly or incorrectly. By contrast, observations such as those in evaluations or workplace-based assessments require interpretation by the examiner, or person completing an evaluation form. For these assessments and evaluations, there needs to be sufficient opportunity to observe the behaviour/task to make a credible observation and judgement (Clauser et al., 2012) and the properties of the assessor/evaluator described (Kane, 1992). Accuracy of the scoring is also an important consideration and can be demonstrated through appropriate use of the measurement scale (i.e. each option on the Likert-type scale is used by respondents) and ensuring that factors beyond those being measured (i.e. gender, age, experience) do not influence the scoring (Clauser et al., 2012). Construction of the items on the assessment or evaluation can also influence the scoring inferences. For example, dichotomous versus Likert-type responses or scale anchors will guide the scoring of the item.

Generalisation

The generalisation inference takes the test or evaluation score and makes a broader interpretation across the breadth or 'universe' of similar observations (Kane, 1992) or an "estimate of some more general attribute" (Kane, 2013, p. 10). Underpinning the generalisation inference is the assumption that the items are taken from the "universe of possibilities" (Cook et al., 2015, p. 567) to ensure adequate representation of the construct to be measured. Further, test/measurement administrators are rarely interested in a single score. Decisions are often based on multiple scores. However, this requires consideration of issues such as stability of responses, and - across assessors/evaluators – the reliability of the test or measure (Kane, 1992). Clauser et al. (2012, p. 171) contend that "generalizability or reliability is central to the validity argument regardless of the purpose" and this may be measured through evaluation of score stability across administrations and across raters (Birenbaum, 2007). Kane (1992) contends that whilst generalisation evidence is a key element of the validity argument, it cannot be the only evidence presented.

Extrapolation

Linking the measurement or observation to the other behaviours or performances comprising the construct being measured is the basis of the extrapolation argument (Kane, 1992). This is typically demonstrated through the relationship between the score and performance on a conceptually related task (Cook et al., 2015). In cases where it is difficult to explore this relationship, other evidence may be considered but it is likely to make a weaker contribution to the validity argument. Content validity, 'expert' input into the measurement tool etc. is considered to be weak evidence in this respect, however evidence of the structure of the tool (i.e. exploratory/confirmatory factor analysis), and the relationship of the scores to other related measures (i.e. convergent validity, divergent validity and concurrent validity) may provide evidence (Clauser et al., 2012). Consideration also needs to be given to the negative consequences that have been articulated previously: construct-irrelevant variance; and construct under-representation (Kane, 2013). These consequences may influence the ability to make a plausible link to real-world performance.

Interpretation or implications

Evidence for this aspect of the validity argument is based on the defensibility of the decision or interpretation of the score, typically based on some level of theory or at a minimum "loose collections of general assumptions" (Kane, 1992, p. 530). Kane (2006) argues that it does not necessarily follow that evidence supporting the interpretation of the score justifies the use of the score. That is, the score may be justifiable based on its accuracy but may not be able to be implemented due to its cost, for example. This inference also requires consideration of the impact or consequences for the use of the score on relevant stakeholders, and additional consideration of the unintended consequences from score use. Cook et al. (2015) contend this inference is the most important of the four but may also be the most challenging to evaluate. In order to understand the stakeholder context, the next section of the chapter will provide an overview of the profession of osteopathy and how osteopaths are educated.

Osteopathy as a health profession

"Osteopathic practitioners use their understanding of the relationship between structure and function to optimize the body's self-regulating, self-healing capabilities" (World Health Organization, 2010).

Osteopathy is a health profession that focuses on the management of musculoskeletal complaints using a variety of manual therapy 'hands-on' techniques. Osteopaths are also trained to recognise when osteopathy or manual therapy interventions may not be appropriate. The predominant musculoskeletal complaints managed by osteopaths include low back pain, neck pain and shoulder issues (Adams et al., 2018; Burke et al., 2013; Fawkes et al., 2013; Morin & Aubin, 2014; Vaughan et al., 2014b) and headaches (Cerritelli et al., 2017). There is also some evidence for the use of osteopathy in the management of chronic obstructive pulmonary disease (COPD), irritable bowel syndrome and visceral complaints, asthma and peripheral arterial disease (Attali et al., 2013; Cicchitti et al., 2015; Florance et al., 2012; Guillaud et al., 2018; King, 2013), and in the management of pregnancy-related musculoskeletal complaints (Frawley et al., 2016).

Osteopaths utilise a range of manual therapy techniques including articulation, mobilisation, muscle energy technique, soft tissue massage, high velocity low amplitude manipulation, strain-counterstrain, and Osteopathy in the Cranial Field (Adams et al., 2018; Burke et al., 2013; Orrock, 2009; Van Dun et al., 2016; Wilkinson et al., 2015) in the management of musculoskeletal complaints, although practice approaches can vary. For example, in parts of Europe there is more emphasis on the use of techniques directed towards the viscera in patient care (Cerritelli et al., 2019; Van Dun et al., 2016). Osteopaths may also use adjunct therapies including dry needling, exercise prescription and patient education with respect to diet and nutrition, ergonomics, and pain management (Adams et al., 2018).

Osteopathy education

Those undertaking their education in the United States undertake a Doctor of Osteopathy (DO) program - equivalent to a Doctor of Medicine (MD) program. This provides equivalent access to speciality education, in addition to rights to prescribe medicines and undertake surgical procedures. As there is a substantial difference in practice between the US and non-US countries, this thesis will focus on the practice of the latter.

Osteopathy education in non-United States countries

Osteopaths in non-United States countries undertake their education in either a university or private college setting. Osteopathy is largely an unregulated health profession around the world. Programs of study in countries where the profession is unregulated can structure their curriculum in any way they see fit. However, there is some guidance as to the suggested content of a program of study through the World Health Organisation's *Benchmarks for Training in Osteopathy* (World Health Organization, 2010). It should be noted that programs of study are not required to use these *Benchmarks* to achieve a particular qualification standing, rather they can be used to inform curricula design and program duration.

In Australia, New Zealand and the United Kingdom a program of study leads to eligibility to apply for registration as an osteopath under government legislation. Programs of study in these countries are accredited against a set of standards approved by a profession-specific regulatory authority. In the United Kingdom, the General Osteopathic Council (GOsC) is the regulatory authority. The same role undertaken by the Osteopathic Council of New Zealand (OCNZ) under the Health Practitioners Competence Assurance Act 2003. Both the GOsC and OCNZ are established by law in their respective countries and have a role in both regulating the conduct of those in the profession, in addition to accrediting programs of study leading to eligibility to apply for registration as an osteopath.

In Australia, the regulatory authority is the Osteopathy Board of Australia which is established under the Health Practitioner Regulation National Law (Parliament of Queensland, 2009). The Osteopathy Board of Australia delegates the authority to accredit programs of study leading to eligibility to apply for registration as an osteopath in Australia to the Australian Osteopathic Accreditation Council ('the Council'). At the time of writing this thesis, there were three accredited programs of study in Australia. At the time of this thesis, the standards are undergoing revision to ensure they are consistent with the revised Capabilities for Osteopathic Practice (Osteopathy Board of Australia, 2019). The latter document describes the standards of practice for the Australian profession and have been developed utilising a framework consistent with the Canadian Medical Education Directives for Specialists (CanMEDS) roles (Frank & Danoff, 2007).

It is also important to highlight that the osteopathy profession has yet to explore its signature pedagogy, described by Shulman (2005) as instruction in "critical aspects of three fundamental dimensions of professional work – to *think*, to *perform* and to *act with integrity*" (p. 52). Other health professions that have either begun to explore this aspect of their educational practice include social work (Earls Larrison & Korr, 2013; Wayne et al., 2010), psychology (Goodyear et al., 2006) occupational therapy (Schaber et al., 2012), nursing (Long et al., 2012) and physical therapy (Anderson & Tunney, 2014). In the latter profession, the signature pedagogy is thought to be clinical education (Anderson & Tunney, 2014) and this may well be the same for osteopathy.

Clinical education in osteopathy

Clinical education forms part of the requirements for an osteopathy program of study in Australia, New Zealand and the United Kingdom. The United Kingdom *Guidance for Osteopathic Pre-registration Training* states that a graduate should have undertaken "a minimum of 1,000 hours of clinical practice" (General Osteopathic Council, 2015, p. 14) with a similar requirement suggested in the *Benchmarks for Training in Osteopathy.* The *Benchmarks* suggest that a program of study should contain 1,000 hours of "practical supervised clinical experience" (World Health

Organization, 2010, p. 10) as it forms "an essential component of the training of osteopathic practitioners and should take place in an appropriate osteopathic clinical environment so that high-quality clinical support and teaching can be provided" (World Health Organization, 2010, p. 10). However, what defines "an appropriate osteopathic clinical environment" is open to debate given the limited literature on osteopathy clinical education. The accreditation standards for osteopathy programs of study in Australia and New Zealand are silent on the volume of clinical education. In Australia, accreditation of osteopathy courses having moved from an input-based to an outcomes-focused framework. This framework allows the education provider to supply evidence to support sufficiency of clinical learning. The use of an outcomes-focused framework for accreditation is consistent with many other regulated health professions in Australasia.

As previously highlighted, minimal literature exists on clinical education in osteopathy. Where literature does exist, its focus is on programs of study in the United States. Programs of study in United States involve students in general or family practice, and hospital in-patient and out-patient environments. In contrast, clinical education in non-United States programs of study commonly takes place in student-led, on-campus clinic environments (Quality Assurance Agency for Higher Education, 2015; Vaughan et al., 2014a). Community placements (Quality Assurance Agency for Higher Education, 2015) and those in private practice (Moore & Field, 2017) may also form part of the clinical education program. The discussion of clinical education in osteopathy programs in this thesis is largely an anecdotal description, based on the experiences of the author, and informed by a commentary from Vaughan et al. (2014a). The discussion presented here is also limited to the Australian, New Zealand and United Kingdom context given these are the major jurisdictions with regulation underpinning clinical practice and education standards in the profession.

In osteopathy courses, students participate in clinical education predominantly in the mid to later years of a program following a period of classroom study. Over the last 30 years, clinical education in osteopathy has been undertaken in the university-based, student-led clinical environment due to a lack of access to private practice placements

and placements in the public health system. Broadly, this student-led environment allows senior learners to consult with a member of the general public under the supervision of qualified, registered osteopaths. In this environment, these senior learners have responsibility for patient care across the clinical history taking, patient assessment, clinical reasoning and management. These learners are also involved in some of the administrative functions of the clinic including making patient appointments and invoicing (Vaughan et al., 2014a). The goal of this education is to ensure that graduates are ready to competently practice in the profession given there is no requirement to undertake graduate-year training. In this learning environment, learners are expected to be exposed to a range of clinical presentations and age groups, reflecting the populations they are likely to see in practice (Vaughan et al., 2020a).

Consistent with other areas of osteopathy education, we know little about those osteopaths who choose to participate in clinical education beyond work by Vaughan et al. (2020b). Work in other health professions suggest availability and clinical experience play a role in recruitment to a clinical education role (Altmann, 2006; Rodger et al., 2008). The latter finding was supported by work from the Australian osteopathy practice-based research network (Vaughan et al., 2020b). These authors identified that Australian osteopathy clinical educators were more likely to be older and have a greater number of years of clinical experience than their non-clinical education peers. Vaughan et al. (2014a) describe the minimum duration of clinical experience to be three years before being able to take on a clinical educator role although it is not known if this is consistent across education programs in the United Kingdom, New Zealand or Australia. Approximately 15% of Australian osteopaths participate in clinical supervision (Vaughan et al., 2020b) but how osteopaths come to this educator role is not known. Numerous authors suggest health professionals choose to become clinical educators as a strategy to improve work satisfaction, a desire to interact with learners, improve their own clinical skills and ‘give back’ (Bennett, 2003; Bing-You & Harvey, 1991; Currens & Bithell, 2000; Krueger et al., 2009; Sevenhuysen & Haines, 2011). Further, how osteopathy clinical educators develop professionally in their role is also unknown. The preceding discussion

highlights significant gaps in our knowledge about the osteopathy clinical educator and this thesis seeks to begin addressing some of these aspects of practice.

Clinical educator to learner ratios for osteopathy clinical education are not prescribed in the Australian accreditation standards. However, they are typically in the range of 1:5 to 1:10 (Vaughan et al., 2014a). This is significantly different to many allied health professions where the range is often 1:1 to 1:3 (McAllister & Nagarajan, 2015). The influence of these ratios on the clinical education experience is not known but is historically consistent in the osteopathy profession. Arguably, these ratios are likely to limit the time an educator can spend with an individual learner both during patient care and in other clinical learning opportunities.

The typical arrangement with respect to patient care in osteopathy clinical education is what Figueiró-Filho et al. (2014) refer to as "minimum faculty member supervision" (p. 366) . In the osteopathy clinical learning environment, the learner will initially undertake the clinical history with the patient without supervision (Regan-Smith et al., 2002). The learner is then expected to present this information to the clinical educator in another area in the clinic where the differential diagnoses, assessment and management plan are discussed. This approach is analogous to the traditional ambulatory care clinical learning environment (Beach et al., 1991) and is repeated for each learner the educator is working with. Following the discussion, the learner returns to the consultation to undertake the patient assessment and treatment. The clinical educator will then come into the consultation room and observe aspects of the assessment and treatment taking place. They may also provide advice to the learner as to how to proceed, or potentially demonstrate aspects of patient care. It is important to note that the educator will not be present in the room for the entirety of the consultation. Following the consultation, the clinical educator will debrief with the learner and review and sign-off on the clinical notes for the consultation.

Clinical education

"The practice of clinical education is a complicated task, undertaken within complex and dynamic environments characterised by sometimes competing demands" (Abey et al., 2013, p. 226).

Clinical learning

Clinical education forms a substantial portion of a program of study leading to a health professions qualification and can be defined as:

"the clinical placement providing opportunities to blend [conceptual and procedural pre-occupational] kinds of knowledge, and transfer them in ways that enhance the development of competent practitioners" (Newton et al., 2009, p. 315).

The conceptual and procedural knowledge referred to by Newton et al. (2009) is typically taught and assessed in the classroom, and potentially simulation laboratories, prior to a learner entering the clinical education phase of their training. During their clinical education, learners will typically begin at the periphery of a professions' practice and observe practitioners in their patient care role (Eberle et al., 2014; Lave & Wenger, 1991). As they proceed towards graduation, learners are progressively engaged in more of the patient care role. Towards the end of training, it is expected learners have had multiple opportunities to participate in direct patient care. It is through these opportunities that students have engaged in clinical decision making and practices that they would be expected to undertake as professionals in the discipline.

Student-led clinical learning environment

The settings where learners participate in clinical education are variable across the professions but typically include general or family practice, tertiary or teaching hospital in-patient and out-patient environments, and less often, community-based or primary-care practices. The substantive literature in clinical education has focused on

how learning occurs in these environments. However, with limitations in the volume of workplace learning sites and associated costs (Allan et al., 2011; Moore et al., 2018), an appreciation of the need for a breadth of placement environments, and in some cases profession-specific health system factors (i.e. not having practice rights in tertiary care settings), has seen the development of student-run clinics (Schutte et al., 2015), student-led clinics, and university clinics (Allan et al., 2011; Moore et al., 2018).

Schutte et al. (2015) posit that student-led clinical environments may "provide the most realistic setting for context-based learning and legitimate early clinical experiences with responsibility for patient care" (p. 249). The common feature is the student leading the consultation with a patient under supervision of a qualified professional (Schutte et al., 2015). Student-run clinics have been predominantly described in United States and Canadian health professions education as a clinical learning environment (Haggarty & Dalcin, 2014; Schutte et al., 2015; Smith et al., 2014). In these countries, student-run clinics are a strategy by which care can be provided to underserviced communities where services are often provided at no charge (Smith et al., 2014). At a health system level, these clinics may also reduce the burden on emergency and primary care environments (Gertz et al., 2011), although this has not been definitively established (Haines et al., 2014). The benefits of student-led clinics include exposure to systems-based care/practice (Colbert et al., 2010; Meah et al., 2009; Sheu et al., 2013), contextual inter-professional care (Fiddes et al., 2013; Holmqvist et al., 2012; Kent et al., 2016; Moskowitz et al., 2006; Seif et al., 2014), development of clinical reasoning (Seif et al., 2014), clinical leadership (Black et al., 2013), and fostering humanistic attitudes towards patients (Black et al., 2013; Modi et al., 2017).

Vaughan et al. (2014a) provided the first description of the osteopathy clinical learning environment. These authors' description is consistent with that of a university clinic (Allan et al., 2011). There is little in the literature that explores the university clinic beyond work by Allan et al. (2011) and Moore et al. (2018) although the emergence of a special interest group through the Australian and New Zealand Association of Health Professional Educators (2017) suggests an appreciation of the

role these clinics play in educating future health professionals. It is also important to note that this research is specific to the Australian context.

Allan et al. (2011) explored the role of context in the establishment and ongoing operation of university clinics. These authors identified that most clinics were established with student learning as the primary focus, were often uni-professional (although some institutions had multiple clinics on-site), and had challenges with recruiting the required volume and breadth of patients. Moore et al. (2018) identified additional challenges for university clinics including cost and financial sustainability, competition with local healthcare providers, student timetabling (e.g. semester/exam breaks, other education demands) and educator demands (e.g. availability, training).

The aforementioned studies also identify the benefits of establishing a university clinic as a learning environment. These benefits include facilitation of clinical learning opportunities for early year learners, exposure to the running of healthcare practices, interprofessional learning, and contributing to healthcare delivery (Moore et al., 2018). With respect to clinical educators, benefits of these university clinics include the opportunity for institutions to recruit high-quality clinicians and educators, develop early-year clinicians as educators, and allow academics in the discipline an opportunity to maintain currency of their skills and knowledge (Moore et al., 2018).

The clinical teacher

"Excellence in clinical teaching is at the top of the pyramid of complexity and expertise" (Irby, 2014, p. 777).

Cooke et al. (2010) and Vanek et al. (1996) advocate clinical teaching as one of medicine's signature pedagogies, and clinical teaching is typically described in terms of the characteristics of *good* clinical teachers (Conigliaro & Stratton, 2010). However, a significant challenge in describing clinical supervision/teaching is the lack of an agreed definition. How supervision is defined is often related to historical, professional, cultural and contextual factors (Martin et al., 2014). By way of example, two definitions of clinical supervision are provided here:

"The provision of guidance and feedback on matters of personal, professional and educational development in the context of a trainee's experience of providing safe and appropriate patient care" (Kilminster et al., 2007, p. 3).

"A formal process of professional support and learning which enables individual practitioners to develop knowledge and competence, and is acknowledged to be a life-long process" (Martin et al., 2014, p. 1).

Effective clinical teaching balances patient safety with learning (Chen et al., 2015), has been demonstrated to influence learner assessment outcomes (Blue et al., 1999; Grant et al., 2003; Stern et al., 2000), and may also positively impact patient outcomes (Farnan et al., 2012; Prideaux et al., 2000; Snowdon et al., 2017). Numerous authors have summarised the characteristics of effective or excellent clinical educators (Irby, 1995; Sutkin et al., 2008; Ullian et al., 1994) and articulated the teaching and learning strategies these educators may employ (Heidenreich et al., 2000). Health profession educators also appear to separate the teaching and clinical supervision roles (Moore et al., 2018), although confidence in one of these roles may not necessarily translate to the other.

The two aforementioned aspects of clinical teaching (personal and professional characteristics of effective clinical teachers and effective clinical teaching strategies) map to the *interpersonal* and *clinical teaching* domains of clinical education described by Beckman et al. (2004a). These authors' review of clinical teaching effectiveness measures argued that "future studies should consider developing assessment tools comprised solely of interpersonal and clinical teaching domains" (p. 976) and that these domains "may adequately assess the proficiency of clinical teachers" (p. 975). A brief overview of the most commonly identified characteristics is described in the next section, drawing on the aforementioned works and other literature from across the medical and allied health professions. These characteristics are those identified by learners and clinical educators. In describing these characteristics, consideration should be given to the fact that interpersonal

characteristics are likely inherent to the individual clinical educator, and that clinical teaching capability can be developed (Branch et al., 1997; Koh, 2008).

Interpersonal domain

The interpersonal domain captures those personal characteristics of clinical educators that appear to be associated with their effectiveness in this role. The most commonly described interpersonal characteristic irrespective of the profession is role-modelling (Al Kadri et al., 2011; Arah et al., 2012; Benbassat, 2014; Blue et al., 1999; Boerebach et al., 2012b; Branch et al., 1997; Buchel & Edwards, 2005; Cross, 1995; Gerzina et al., 2005; Huff et al., 2014; Jochemsen-van der Leeuw et al., 2013; Kilminster et al., 2007; Kumar & Greenhill, 2016; Lee et al., 2002; Okoronkwo et al., 2013; Pinard et al., 201; Wright et al., 1998). Others have segmented role modelling further into:

- clinical, teaching and personal domains (Cruess et al., 2008);
- patient care, teaching and personal qualities (Jochemsen-van der Leeuw et al., 2013); and,
- professional, personal and educational domains (Matthews, 2000).

These 'categories' suggest significant overlap of role-modelling with the interpersonal aspect of what is considered to be effective clinical teaching (Wright et al., 1998). Other characteristics that are consistently identified in the literature with respect to the interpersonal domain are listed in Table 1.

Table 1. Characteristics of effective clinical educators in the interpersonal domain.

Characteristic	**Reference**
Respect for the learner, patient and colleagues	Alweshahi et al. (2007); Boerebach et al. (2012b); Branch et al. (1997); Cole and Wessel (2008); Cuesta-Briand et al. (2014); Goldie et al. (2015); Irby and Bowen (2004); Kilminster et al. (2007); Rutz et al. (2019)
Enthusiasm	Beckman et al. (2003); Bennett (2003); Buchel and Edwards (2005); Cole and Wessel (2008); Goldie et al. (2015); Huff et al. (2014); Kilminster et al. (2007); Schönwetter et al. (2006); Schultz et al. (2004)
Rapport	Cottrell et al. (2002); Kisiel et al. (2010); Ramani and Leinster (2008); Schönwetter et al. (2006)
Encouraging	Masunaga and Hitchcock (2010)
Approachable	Alweshahi et al. (2007); Beckman and Lee (2009); Bennett (2003); Fernandez (1998); Goldie et al. (2015); Huff et al. (2014)
Adaptable and flexible	Al Kadri et al. (2011); Irby (2014); Okoronkwo et al. (2013)
Interest in teaching	Jahan et al. (2008); Ramani and Leinster (2008); Schultz et al. (2004)
Extraverted personality type	Scheepers et al. (2014)

The aforementioned description and volume of literature describing additional characteristics (Table 1) support the contention of Beckman et al. (2004a) that the interpersonal domain is a key element of clinical teaching.

Clinical teaching domain

The clinical teaching domain refers to the teaching and learning strategies that are associated with *good* or *effective* clinical educators. Cuesta-Briand et al. (2014) identified that learners perceive a 'good doctor' [medical practitioner] to also be a 'good teacher'. Consistent with this description is the predominant theme through the literature of the clinical knowledge and skills of the educator (Al Kadri et al., 2011; Branch et al., 1997; Buchel & Edwards, 2005; Fernandez, 1998; Henzi et al., 2006; Jahan et al., 2008; Kelly, 2007; Kilminster et al., 2007; Lauber et al., 2003; Okoronkwo et al., 2013; Ramani & Leinster, 2008; Singh et al., 2013). Additional clinical educator characteristics associated with the clinical teaching domain identified in the literature are described in Table 2.

Table 2. Characteristics of effective clinical educators in the clinical teaching domain.

Characteristic	**Reference**
Demonstration of clinical skills	Bing-You and Harvey (1991); Cole and Wessel (2008); Cottrell et al. (2002); Curtis et al. (1998); Gerzina et al. (2005); Kernan et al. (2008); Lauber et al. (2003)
Assessing the learner	Al-Kadri et al. (2013); Edgar and Connaughton (2014); Gerzina et al. (2005); Hays (2008)
Facilitator of learning	Bennett (2003); Bing-You and Harvey (1991); Gerzina et al. (2005); Healey (2008)
Establish a positive learning environment	Blue et al. (1999); Irby and Bowen (2004); Kernan et al. (2008); Kilminster and Jolly (2001)

Organisation and planning	Burns et al. (2006); Irby (2014); Ramani and Leinster (2008); Schönwetter et al. (2006)
Scaffolding learning	Chen et al. (2015)
Resource developer	Alweshahi et al. (2007)
Feedback provider	Alweshahi et al. (2007); Boerebach et al. (2012b); Buchel and Edwards (2005); Cole and Wessel (2008); Healey (2008); Kelly (2007); Kernan et al. (2008); Kilminster et al. (2007); Schultz et al. (2004); Torre et al. (2003)
Availability and time spent with learners	Buchel and Edwards (2005); Huff et al. (2014); Kelly (2007); Kumar and Greenhill (2016); Ramani and Leinster (2008); Robinson (2015); Schultz et al. (2004); Wright et al. (1998)
Links educational outcomes to patient care activities	Al-Kadri et al. (2013); Chen et al. (2016)
Undertaken formal clinical education studies	Arah et al. (2012)

Two additional characteristics associated with effective clinical teaching in the literature appear to be related to employment status. Having a full time clinical academic appointment (Allison-Jones & Hirt, 2004) or formal clinical education component in their professional role (Kumar & Greenhill, 2016) have both been associated with effective clinical teaching. These outcomes suggest that employment status may influence how learners perceive a clinician as an educator.

The evaluation of teaching

"We believe that the most productive evaluation processes for enhancing teaching quality are those which have the intention of developing the teacher professionally" (Nygaard & Belluigi, 2011, p. 663).

Evaluation in the educational context is designed to explore the impact of an intervention, strategy or process on student learning (Nygaard & Belluigi, 2011) or as Cox and Swanson (2002, p. 251) described "evaluation is the determination of worth". Data derived from such evaluations help to inform changes to improve student learning at the formative end through to summative decisions about curricula and program structure (Bassett et al., 2017; Nygaard & Belluigi, 2011). There are multiple sources of evaluation data including that from learners, faculty, administrators and other institutional stakeholders – each bringing their own perspective to the evaluation (Jahangiri et al., 2008; Nygaard & Belluigi, 2011).

Student evaluations of teaching (SETs)

In the higher education context, student evaluations of teaching (SET), learning design and institutional experience are ubiquitous (Blair & Valdez Noel, 2014; Bush et al., 2018; Linse, 2017; Peterson et al., 2019; Richardson, 2005). Learners are routinely asked to express their opinion of the teaching and learning experience for both formative and summative purposes (Bassett et al., 2017). Blair and Valdez Noel (2014, p. 881) summarise that SETs capture "student perceptions of quality, ability and clarity."

For educators, learner evaluations can act as a feedback mechanism to guide changes to curricula and teaching (Darwin, 2017; Emerson & Records, 2007; Golding & Adam, 2016; Kember et al., 2002; McAuley et al., 2017; Richardson, 2005; Spooren et al., 2013) and be used as part-evidence in promotion or tenure applications (Appling et al., 2001; Linse, 2017). For regulators and accreditors, evaluations are required to be completed as part of quality assurance processes (Kember et al., 2002). For administrators, evaluations from learners are efficient (Uttl et al., 2017), provide

evidence to assist with tenure and promotion decisions (Fraile & Bosch-Morell, 2015; Kember et al., 2002; Linse, 2017; McAuley et al., 2017; Oermann et al., 2018), merit-based pay, guide professional development, and inform changes at subject, curriculum and institutional level (Darwin, 2017; McAuley et al., 2017). Darwin (2017) also suggests that evaluations may have a more normative function – comparing the performance of academics. However, these administrative outcomes have led to a call for SETs to not be used as a summative decision making tool (Hornstein, 2017).

The educator perspective of student evaluations of teaching

Evaluations of an individual teacher provided by learners are no longer considered to be controversial (Arah et al., 2011), particularly for personal use (i.e. improving or changing individual teaching practices) (Uttl et al., 2017). Academics report the student voice is an important one to hear with respect to their teaching practice (Blair & Valdez Noel, 2014; Darwin, 2017; Golding & Adam, 2016). However, it is widely reported that educators have a tendency to discount or ignore the results of SETs where they perceive the evaluation to:

- be a personality or popularity outcome (Linse, 2017);
- be influenced by biases such as educator gender (particularly female educators), class size, nationality etc. (Bassett et al., 2017; Boring, 2017; Fan et al., 2019; MacNell et al., 2015; Spooren & Christiaens, 2017; Uttl et al., 2017; Zabaleta, 2007);
- be an unreliable measure of teaching quality or performance (Spooren & Christiaens, 2017);
- not be required to change teaching practice (i.e. lack of incentive) (Kember et al., 2002);
- not take into account all aspects of their teaching (Emery et al., 2003);
- be negatively affected by the subject content being studied rather than the teaching itself (Looi & Anderson, 2018);
- take into account system factors beyond the control of the individual educator (Emery et al., 2003);
- have low response rates (Linse, 2017);
- be derived from anonymous responses (Spooren & Christiaens, 2017);

- be a result of acquiescence bias (Valencia, 2020);
- be completed by students who are not equipped or 'expert' enough to evaluate content accuracy, contemporaneity or 'effective' teaching (Spooren & Christiaens, 2017); and/or,
- have a variable or biased relationship to other measures of teaching and student learning (Bassett et al., 2017; Linse, 2017; Uttl et al., 2017).

Moreover, some say teaching quality evaluation scores are thought to be "the extent to which student expectations are met" (Spooren et al., 2013, p. 599) associated with the 'student as consumer' perspective (Tran & Do, 2020) and not related to teaching effectiveness (i.e. student outcomes) (Boring et al., 2016; Uttl et al., 2017).

Multiple tensions have also been identified with widespread, institutional level use of SETs. On the one hand, the use of SETs is rationalised by institutions as a mechanism to foster high quality learning outcomes, encourage innovations in teaching practice, and also improve teaching practice. Conversely, SETs are also used in the marketing of courses, as an accountability measure for various stakeholders, and for employment and tenure decisions (Boring, 2017; Darwin, 2017). How these tensions affect the individual educator appears to vary based on their individual context (i.e. tenure, teaching experience) (Darwin, 2017) with tenured academics potentially placing little weight on SETs over other measures of their performance (i.e. research output). The use of SETs may also encourage educators to chase high SET scores, i.e., by using entertainment rather than sound pedagogical approaches to achieve these (Adams & Umbach, 2012) and competing with other academics in the same department or teaching area (Emery et al., 2003). Further, non-tenured academics tend to be more wary of the evaluation knowing that poor evaluations may result in them not being employed in subsequent teaching periods or may affect their ability to apply for a tenured position in the future (Darwin, 2017). An additional tension identified by part-time academics with workplace commitments (e.g. health professionals) was a need to maintain a balance between student experience and ensuring profession standards were met (Darwin, 2017). The emerging consensus in the literature is that SETs should not be used for summative decisions, that is, promotion, tenure or merit-based pay (e.g. Hornstein, 2017;

Spooren et al., 2013). Where such decisions are desired, institutions should ensure additional data are sought to develop a more holistic view of educator performance, potentially through development of a teaching portfolio (Hornstein, 2017) or peer review teaching (Dawson & Hocker, 2019).

When administering SETs, consideration should also be given to learner attitudes towards the evaluation. To facilitate completion of SETs, learners need to have a motivation for completing the SET, that is, the feedback is going to be acted upon and improve teaching practices (Bassett et al., 2017; Blair & Valdez Noel, 2014; Chen & Hoshower, 2003; McAuley et al., 2017). Supporting this notion of ensuring learners are aware of the use of the SET results, McClain et al. (2018) identified learners will complete evaluations less honestly where results are used for summative purposes or they believe there is no purpose to the evaluation. Orienting the learner to the purpose of the SET is also thought to increase reliability (McClain et al., 2018) and highlighting potential biases to learners when completing the evaluation can reduce their subsequent influence on SET outcomes (Peterson et al., 2019). Multiple authors also describe that consideration should be directed towards data collection times to ensure learners are not being 'over-evaluated' (Bush et al., 2018; McAuley et al., 2017; Spooren & Christiaens, 2017; Spooren et al., 2007), resulting in learners not responding accurately or appropriately (Dunegan & Hrivnak, 2003; Uijtdehaage & O'Neal, 2015). Further, those learners who value the SET are likely to provide higher SET scores (Spooren & Christiaens, 2017; Worthington, 2002). Higher SET scores also appear to be associated with student personality traits - higher levels of agreeableness and lower levels of neuroticism (McCann & Gardner, 2014; Patrick, 2011).

Developing student evaluations of teaching

In the development of the items comprising SETs, relevant stakeholders including academics and learners should be involved as they may have differing conceptions of teaching effectiveness or quality (Spooren et al., 2013). Structurally, SETs typically comprise a range of items and a scale on which the learner is asked to respond to the item. However, there is no agreement as to whether SETs should be uni- or multi-

dimensional due to a lack of conceptual frameworks and no clear definition of 'effective teaching' (Spooren et al., 2013). Effective SETs comprise no more than 20 items (Bush et al., 2018) with Likert-type scales being the most common method of responding to the item (Bush et al., 2018; Spooren et al., 2007). Five point Likert-type scales appear to be the most accurate, particularly when compared to larger numbers of scale points (Spooren et al., 2007). Using positively worded items only and having the items reviewed to identify poor item wording; items that are difficult to comprehend; and, irrelevant items, is also advocated as these issues may encourage acquiescence (Spooren et al., 2007). It is also suggested that a global rating item be included to capture the learner's overall impression of the teaching (Spooren et al., 2013) and this score may assist in detecting a 'halo' effect (McClain et al., 2018). Both closed and open ended items are advocated on SETs (Spooren et al., 2007) to further capture learner impressions. With respect to delivery method (i.e. paper versus online), the literature is ambivalent (Spooren et al., 2007), and response rates can be improved by allowing learners time to complete SETs in class time (Young et al., 2019).

The evaluation of clinical teaching

"We suspect that evaluations of clinical teachers with whom students usually establish personal, one-to-one relationships do not suffer from the same pitfalls as SETs in large multi-instructor courses" (Uijtdehaage & O'Neal, 2015, p. 932).

Like other learning environments in higher education, evaluations of clinical teaching completed by learners are frequently utilised by institutions, program leaders and faculty administrators (Beckman et al., 2004b; Snell et al., 2000). Broadly speaking, clinical educators display positive attitudes towards evaluations as a way of improving their teaching (Berk et al., 2005; Skeff et al., 1992; Yuan et al., 2014) and view learners as a credible source of feedback on their performance suggesting an acceptance of these measures as part of their teaching practice (Dudek et al., 2016; McOwen et al., 2007). There is also evidence that learner ratings appear to be somewhat consistent with educator self-ratings (Allison-Jones & Hirt, 2004).

There are a range of questionnaires published in the literature to evaluate the quality of clinical teaching. One of the first questionnaires to evaluate quality of teaching in the clinical environment was published by Reichsman et al. (1964). Numerous health profession educators have since developed questionnaires to evaluate quality in their own teaching environment with only a small number being used across multiple contexts. Beckman et al. (2005), Beckman et al. (2004a) and Fluit (2010) have undertaken systematic reviews of clinical teaching quality questionnaires to identify evidence for their validity, reliability and other outcomes. Subsequent studies have also been undertaken to explore the quality of clinical teaching across a range of professions and specialties. The volume of questionnaires, and varying evidence for their measurement properties, presents a challenge for educators and administrators to decide on which questionnaire to use or whether to develop a measure for their own context.

Challenges in the evaluation of clinical teaching

Numerous challenges have been identified in the evaluation of clinical teaching with the most significant being that there is no consistent or accepted definition of clinical teaching effectiveness (Berk, 2013). Without this definition, measurement of the construct could be challenging. However, effective teaching may be the consistent display of the interpersonal and clinical teaching characteristics described previously in this chapter.

An additional measurement challenge for clinical teaching evaluations is posed where studies have not achieved sufficient power to demonstrate change in teaching effectiveness (Baker, 2010), and the presence of a ceiling effect (Baker, 2010; Copeland & Hewson, 2000; Fluit et al., 2013). The latter may limit the ability to identify any further improvement in performance. Steiner et al. (2003) also highlight that minimum clinically [educationally] important difference values have not been described for clinical teaching effectiveness measures, and this appears to persist today. Such a value would be a valuable addition to interpretation as it could indicate when meaningful changes in teaching and supervision practice have occurred.

In clinical education, learners may not be willing to complete lengthy measures (Bierer & Hull, 2007; Schiekirka et al., 2012) and appear to see it as a time consuming task (Myers et al., 2012; Zibrowski et al., 2016). There is a risk with both of these factors, along with evaluation 'fatigue' that learners may complete the measure 'mindlessly' thereby limiting the utility of the data (Uijtdehaage & O'Neal, 2015). Learners often receive little instruction on how to complete the measures and what the ratings represent (Pettit et al., 2015). In some cases, learners may be reluctant to complete the measure as they are unsure as to the outcome of the response they provide (Myers et al., 2012; Pettit et al., 2015) or only complete the evaluation if they feel that the program or institution values their response (Zibrowski et al., 2016). Myers et al. (2012) also highlight that learners may only engage with the measure where they wish to evaluate educators at the extremes (i.e. very high and very low performers).

The literature describes variable opinions with respect to anonymity of clinical teaching evaluations (Daberkow et al., 2005; Dudek et al., 2016). With small numbers of learners completing clinical teaching evaluations on a single educator, there may still be a feeling of lack of anonymity on the part of the learner (Albanese, 1999; Myers et al., 2012) and the learners may moderate the feedback they provide to ensure they are not negatively impacted (Dudek et al., 2016). That said, Afonso et al. (2005) suggest anonymity can increase the reliability of the evaluations.

Like anonymity, the influence of demographic variables on clinical teaching effectiveness measures appears to be equivocal in the literature. More senior learners may be more accurate judges (Baker, 2010; Fluit et al., 2014) and learners at different training levels may rate the same educators differently (Mazor et al., 2002; Torre et al., 2003). Gender interactions are consistently highlighted as influencing clinical teaching evaluations. Literature suggests that female teachers are more likely to receive lower ratings compared to their male counterparts across a variety of disciplines (Leone-Perkins et al., 1999; Morgan et al., 2016; Steiner et al., 2003), particularly those with low female representation (Fassiotto et al., 2018). However, this is not a consistent finding (Fluit et al., 2014; McOwen et al., 2007). Clinical experience may also influence teaching evaluations. For example, more senior

educators may be rated lower in some professions (e.g. surgery) (Callcut et al., 2004) potentially due to a lack of social congruence (Steiner et al., 2003; Yew & Yong, 2014). To summarise, although the literature suggests a number of challenges with the use of evaluations of clinical teaching, many are not insurmountable.

Opportunities in the evaluation of clinical teaching

There are numerous opportunities within the evaluation of clinical teaching space that deserve consideration. A significant consideration for the institution is the cost-effective nature of learners providing evaluations of their clinical educators (Beckman et al., 2004b), particularly in comparison to peer-review (Coverdale et al., 2010). Ease of administration in the busy clinical environment is also an advantage of this strategy where learners are able to provide ratings of teaching quality for multiple educators in a short time-frame (Coverdale et al., 2010).

All of the questionnaires identified in systematic reviews by Beckman et al. (2005) and Fluit (2010) are quantitative in nature, contributing to ease of administration and efficiency in data collection. To provide additional data and opportunities for the learner to provide feedback, questionnaires often provide a section for the learner to provide qualitative comments on educator performance. Van der Leeuw et al. (2013a) have identified that quantitative outcomes appear to be consistent with qualitative comments (i.e. high performing educators receive qualitative comments reflecting this performance, and vice-versa), and the opportunity to provide qualitative comments is often taken up by learners. Multiple authors have also demonstrated that improved clinical teaching can result from evaluations, particularly where quantitative outcomes are combined with qualitative comments (Baker, 2010; Fluit et al., 2013; Jahangiri et al., 2008) and where the data are used to inform faculty development (Bardella et al., 2005). Dudek et al. (2016) also suggest that learners may use these evaluations to develop their ability to provide and receive feedback in the future. The aforementioned works suggest learner participation in clinical teaching evaluations is valuable for both the educator and learner alike.

Importantly, even high performing clinical educators can improve their teaching effectiveness scores (Van der Leeuw et al., 2016), suggesting a benefit in continued evaluation of teaching practice (Maker et al., 2006). Positive change in clinical teaching evaluations may also be amplified if the educator has previously undertaken faculty development (Coverdale et al., 2010). Chandrasekhar et al. (2013) suggest this effect make take time to become visible in evaluation data, consistent with behavioural change theories (Baker, 2010; Breckwoldt et al., 2014) and supports the longitudinal collection of evaluation data.

From an evaluation design perspective, combining various data sources should lead to greater trustworthiness and acceptability of the evaluation (Arah et al., 2011; Berk, 2009; Gusic et al., 2013; Snell et al., 2000; Zabar et al., 2004). Student and self-evaluations are accepted and widely reported data sources in the context of clinical teaching evaluations (Arah et al., 2011; Boerebach et al., 2012a; Van der Leeuw et al., 2013b). These sources of data can be efficiently surveyed and are key stakeholders in clinical teaching evaluations.

Significance of the studies

Clinical education is an essential component of the curriculum in the training of future allied health professionals, including osteopaths. Teaching institutions place substantial financial, physical, and human resources into clinical education in order to ensure students are able to translate theory and simulated classroom learning into real world practice. One aspect of ensuring these resources are used efficiently and that translation of knowledge occurs is to evaluate the quality of teaching provided by clinical educators. There are numerous questionnaires that can be used for this purpose (Beckman et al., 2004a; Fluit, 2010). However, these questionnaires have typically been developed for use in an in-patient or hospital-based ambulatory learning environment. As Snell et al. (2000) describe, “research should include the development and validation of new measures that acknowledge a variety of teaching roles and facilitate comparisons between learning contexts such as institutions or disciplines” (p. 869). In order to develop an understanding of the quality of clinical education in any setting, it is important to evaluate different perspectives of the

clinical teaching provided. This thesis aims to develop a psychometrically sound questionnaire to allow learners to evaluate the quality of clinical teaching in osteopathy student-led clinical learning environments.

To present evidence for a psychometrically sound questionnaire, this thesis will describe the use of a Classical Test Theory approach in exploratory factor analysis (Gaskin & Happell, 2014), modern test theory/item response theory (De Champlain, 2010), and generalisability theory (Bloch & Norman, 2012; Briesch et al., 2014) in the subsequent chapters. Modern test theory, in particular Rasch analysis (Andrich, 1988; Pallant & Tennant, 2007), has not previously been used in the developmental stages for a questionnaire to evaluate clinical teaching quality, thereby presenting a unique contribution to this literature.

Evaluation of teaching quality, including clinical teaching, is a substantial contributor to program or institutional quality assurance processes. These evaluations assist an institution to establish aspects of the teaching that are performing well and identify those areas in need of improvement. Further, some institutions use the results from such evaluations to make employment and promotion decisions. As such, the scores derived from these evaluation tools should be defensible. This thesis uses the *interpretation and use* argument approach to validity described by Kane (1992) to present evidence supporting score validity and defensibility of the interpretation of these scores.

Evaluation of clinical teaching quality also falls within the purview of the institution's quality assurance processes. However, the tools needed to undertake this evaluation should be different to those used in the classroom setting (Brown et al., 2013; Fluit et al., 2014). To that end, the outcome of the thesis will be the development of a questionnaire that allows learners to evaluate the quality of clinical teaching for use in osteopathy student-led teaching clinic environments.

References

Abey, S., Lea, S., Callaghan, L., Cotton, D., & Shaw, S. (2013). The development of a scale to assess practitioner capacity to engage in clinical education. *Journal of Further and Higher Education*, 1-18. https://doi.org/10.1080/0309877X.2013.817006

Adams, J., Sibbritt, D., Steel, A., & Peng, W. (2018). A workforce survey of Australian osteopathy: analysis of a nationally-representative sample of osteopaths from the Osteopathy Research and Innovation Network (ORION) project. *BMC Health Services Research, 18*(1), 352. https://doi.org/10.1186/s12913-018-3158-y

Adams, M. J. D., & Umbach, P. D. (2012). Nonresponse and online student evaluations of teaching: Understanding the influence of salience, fatigue, and academic environments. *Research in Higher Education, 53*(5), 576-591. https://doi.org/10.1007/s11162-011-9240-5

Afonso, N. M., Cardozo, L. J., Mascarenhas, O. A., Aranha, A. N., & Shah, C. (2005). Are anonymous evaluations a better assessment of faculty teaching performance? A comparative analysis of open and anonymous evaluation processes. *Family Medicine, 37*(1), 43-47. https://www.stfm.org/familymedicine/vol37issue1/Afonso43

Al Kadri, H. M. F., Al-Moamary, M. S., Elzubair, M., Magzoub, M. E., AlMutairi, A., Roberts, C., & van der Vleuten, C. (2011). Exploring factors affecting undergraduate medical students' study strategies in the clinical years: a qualitative study. *Advances in Health Sciences Education, 16*(5), 553-567. https://doi.org/10.1007/s10459-010-9271-2

Al-Kadri, H. M., Al-Kadi, M. T., & Van Der Vleuten, C. P. M. (2013). Workplace-based assessment and students' approaches to learning: a qualitative inquiry. *Medical Teacher, 35*(Supplement 1), S31-S38. https://doi.org/10.3109/0142159X.2013.765547

Albanese, M. (1999). Rating educational quality: factors in the erosion of professional standards. *Academic Medicine, 74*(6), 652-658. https://doi.org/10.1097/00001888-199906000-00009

Allan, J., O'Meara, P., Pope, R., Higgs, J., & Kent, J. (2011). The role of context in establishing university clinics. *Health and Social Care in the Community, 19*(2), 217-224. https://doi.org/10.1111/j.1365-2524.2010.00971.x

Allison-Jones, L. L., & Hirt, J. B. (2004). Comparing the teaching effectiveness of part-time & full-time clinical nurse faculty. *Nursing Education Perspectives, 25*(5), 238-243. https://journals.lww.com/neponline/Abstract/2004/09000/Comparing_the_Teaching_Effectiveness_of_PART_TIME.12.aspx

Altmann, T. K. (2006). Preceptor selection, orientation, and evaluation in baccalaureate nursing education. *International Journal of Nursing Education Scholarship, 3*(1). https://doi.org/10.2202/1548-923X.1014

Alweshahi, Y., Harley, D., & Cook, D. A. (2007). Students' perception of the characteristics of effective bedside teachers. *Medical Teacher, 29*(2-3), 204-209. https://doi.org/10.1080/01421590701271818

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

Anderson, J. R., & Tunney, N. M. (2014). The scholarship of teaching and learning in a physical therapy program. *New Directions for Teaching and Learning, 2014*(139), 61-76. https://doi.org/10.1002/tl.20105

Andrich, D. (1988). *Rasch Models for Measurement* (Vol. 68). Sage Publications.

Appling, S. E., Naumann, P. L., & Berk, R. A. (2001). Using a faculty evaluation triad to achieve evidence-based teaching. *Nursing and Health Care Perspectives, 22*(5), 247-247. https://doi.org/10.1007/s10459-006-9030-6

Arah, O. A., Heineman, M. J., & Lombarts, K. M. (2012). Factors influencing residents' evaluations of clinical faculty member teaching qualities and role model status. *Medical Education, 46*(4), 381-389. https://doi.org/10.1111/j.1365-2923.2011.04176.x

Arah, O. A., Hoekstra, J. B. L., Bos, A. P., & Lombarts, K. M. (2011). New tools for systematic evaluation of teaching qualities of medical faculty: results of an ongoing multi-center survey. *PLoS ONE, 6*(10), e25983. https://doi.org/10.1371/journal.pone.0025983

Attali, T. V., Bouchoucha, M., & Benamouzig, R. (2013). Treatment of refractory irritable bowel syndrome with visceral osteopathy: Short-term and long-term results of a randomized trial. *Journal of Digestive Diseases, 14*(12), 654-661. https://doi.org/10.1111/1751-2980.12098

Australian and New Zealand Association of Health Professional Educators. (2017). *An Australian and New Zealand University Health Clinics Collaborative*. Retrieved August 30, 2019, from https://www.anzahpe.org/htag-1

Baker, K. (2010). Clinical teaching improves with resident evaluation and feedback. *Anesthesiology: The Journal of the American Society of Anesthesiologists, 113*(3), 693-703. https://doi.org/10.1097/ALN.0b013e3181eaacf4

Bardella, I. J., Janosky, J., Elnicki, D. M., Ploof, D., & Kolarik, R. (2005). Observed versus reported precepting skills: teaching behaviours in a community ambulatory clerkship. *Medical Education, 39*(10), 1036-1044. https://doi.org/10.1111/j.1365-2929.2005.02269.x

Bassett, J., Cleveland, A., Acorn, D., Nix, M., & Snyder, T. (2017). Are they paying attention? Students' lack of motivation and attention potentially threaten the utility of course evaluations. *Assessment & Evaluation in Higher Education, 42*(3), 431-442. https://doi.org/10.1080/02602938.2015.1119801

Beach, P. S., McCormick, D. P., & Levine, H. G. (1991). Implementation of a mentor system to enhance student education and evaluation in the ambulatory setting. *Teaching and Learning in Medicine, 3*(3), 151-155. https://doi.org/10.1080/10401339109539500

Beckman, T. J., Cook, D. A., & Mandrekar, J. N. (2005). What is the validity evidence for assessments of clinical teaching? *Journal of General Internal Medicine, 20*(12), 1159-1164. https://doi.org/10.1111/j.1525-1497.2005.0258.x

Beckman, T. J., Ghosh, A. K., Cook, D. A., Erwin, P. J., & Mandrekar, J. N. (2004a). How reliable are assessments of clinical teaching? *Journal of General Internal Medicine, 19*(9), 971-977. https://doi.org/10.1111/j.1525-1497.2004.40066.x

Beckman, T. J., & Lee, M. C. (2009). Proposal for a collaborative approach to clinical teaching. *Mayo Clinic Proceedings, 84*(4), 339-344. https://doi.org/10.1016/S0025-6196(11)60543-0

Beckman, T. J., Lee, M. C., & Mandrekar, J. N. (2004b). A comparison of clinical teaching evaluations by resident and peer physicians. *Medical Teacher, 26*(4), 321-325. https://doi.org/10.1111/j.1525-1497.2004.40066.x

Beckman, T. J., Lee, M. C., Rohren, C. H., & Pankratz, V. S. (2003). Evaluating an instrument for the peer review of inpatient teaching. *Medical Teacher, 25*(2), 131-135. https://doi.org/10.1080/0142159031000092508

Benbassat, J. (2014). Role modeling in medical education: the importance of a reflective imitation. *Academic Medicine, 89*(4), 550-554. https://doi.org/10.1097/ACM.0000000000000189

Bennett, R. (2003). Clinical education: perceived abilities/qualities of clinical educators and team supervision of students. *Physiotherapy, 89*(7), 432-440. https://doi.org/10.1016/S0031-9406(05)60077-6

Berk, R. A. (2009). Using the 360 multisource feedback model to evaluate teaching and professionalism. *Medical Teacher, 31*(12), 1073-1080. https://doi.org/10.3109/01421590802572775

Berk, R. A. (2013). Top five flashpoints in the assessment of teaching effectiveness. *Medical Teacher, 35*(1), 15-26. https://doi.org/10.3109/0142159X.2012.732247

Berk, R. A., Naumann, P. L., & Appling, S. E. (2005). Beyond student ratings: peer observation of classroom and clinical teaching. *International Journal of Nursing Education Scholarship, 1*(1). https://doi.org/10.2202/1548-923x.1024

Bierer, S. B., & Hull, A. L. (2007). Examination of a clinical teaching effectiveness instrument used for summative faculty assessment. *Evaluation and the Health Professions, 30*(4), 339-361. https://doi.org/10.1177/0163278707307906

Bing-You, R. G., & Harvey, B. J. (1991). Factors related to residents' desire and ability to teach in the clinical setting. *Teaching and Learning in Medicine, 3*(2), 95-100. https://doi.org/10.1080/10401339109539488

Birenbaum, M. (2007). Evaluating the assessment: Sources of evidence for quality assurance. *Studies in Educational Evaluation, 33*(1), 29-49. https://doi.org/10.1016/j.stueduc.2007.01.004

Black, J. D., Palombaro, K. M., & Dole, R. L. (2013). Student experiences in creating and launching a student-led physical therapy pro bono clinic: A qualitative

investigation. *Physical Therapy, 93*(5), 637-648. https://doi.org/10.2522/ptj.20110430

Blair, E., & Valdez Noel, K. (2014). Improving higher education practice through student evaluation systems: is the student voice being heard? *Assessment & Evaluation in Higher Education, 39*(7), 879-894. https://doi.org/10.1080/02602938.2013.875984

Bloch, R., & Norman, G. (2012). Generalizability theory for the perplexed: A practical introduction and guide: AMEE Guide No. 68. *Medical Teacher, 34*(11), 960-992. https://doi.org/10.3109/0142159X.2012.703791

Blue, A. V., Griffith III, C. H., Wilson, J., Sloan, D. A., & Schwartz, R. W. (1999). Surgical teaching quality makes a difference. *The American Journal of Surgery, 177*(1), 86-89. https://doi.org/10.1080/02602938.2013.875984

Boerebach, B. C., Arah, O. A., Busch, O. R., & Lombarts, K. M. (2012a). Reliable and valid tools for measuring surgeons' teaching performance: Residents' vs. self evaluation. *Journal of Surgical Education, 69*(4), 511-520. https://doi.org/10.1016/j.jsurg.2012.04.003

Boerebach, B. C., Lombarts, K. M., Keijzer, C., Heineman, M. J., & Arah, O. A. (2012b). The teacher, the physician and the person: how faculty's teaching performance influences their role modelling. *PLoS ONE, 7*(3), e32089. https://doi.org/10.1371/journal.pone.0032089

Boring, A. (2017). Gender biases in student evaluations of teaching. *Journal of Public Economics, 145*, 27-41. https://doi.org/10.1016/j.jpubeco.2016.11.006

Boring, A., Ottoboni, K., & Stark, P. (2016). Student evaluations of teaching (mostly) do not measure teaching effectiveness. *ScienceOpen Research*. https://doi.org/10.14293/S2199-1006.1.SOR-EDU.AETBZC.v1

Borsboom, D., & Markus, K. A. (2013). Truth and evidence in validity theory. *Journal of Educational Measurement, 50*(1), 110-114. https://doi.org/10.1111/jedm.12006

Branch, W. T., Kroenke, K., & Levinson, W. (1997). The clinician-educator—present and future roles. *Journal of General Internal Medicine, 12*(Supplement 2), S1-S4. https://doi.org/10.1046/j.1525-1497.12.s2.16.x

Breckwoldt, J., Svensson, J., Lingemann, C., & Gruber, H. (2014). Does clinical teacher training always improve teaching effectiveness as opposed to no

teacher training? A randomized controlled study. *BMC Medical Education, 14*(1), 6. https://doi.org/10.1186/1472-6920-14-6

Briesch, A. M., Swaminathan, H., Welsh, M., & Chafouleas, S. M. (2014). Generalizability theory: A practical guide to study design, implementation, and interpretation. *Journal of School Psychology, 52*(1), 13-35. https://doi.org/10.1016/j.jsp.2013.11.008

Brown, T., Williams, B., & Lynch, M. (2013). Relationship between clinical fieldwork educator performance and health professional students' perceptions of their practice education learning environments. *Nursing and Health Sciences, 15*(4), 510-517. https://doi.org/10.1111/nhs.12065

Buchel, T. L., & Edwards, F. D. (2005). Characteristics of effective clinical teachers. *Family Medicine, 37*(1), 30-35. https://www.stfm.org/familymedicine/vol37issue1/Buchel30

Burke, S. R., Myers, R., & Zhang, A. L. (2013). A profile of osteopathic practice in Australia 2010–2011: a cross sectional survey. *BMC Musculoskeletal Disorders, 14*(1), 1-10. https://doi.org/10.1186/1471-2474-14-227

Burns, C., Beauchesne, M., Ryan-Krause, P., & Sawin, K. (2006). Mastering the preceptor role: challenges of clinical teaching. *Journal of Pediatric Health Care, 20*(3), 172-183. https://doi.org/10.1016/j.pedhc.2005.10.012

Bush, M. A., Rushton, S., Conklin, J. L., & Oermann, M. H. (2018). Considerations for developing a student evaluation of teaching form. *Teaching and Learning in Nursing, 13*(2), 125-128. https://doi.org/10.1016/j.teln.2017.10.002

Callcut, R. A., Rikkers, L., Lewis, B., & Chen, H. (2004). Does academic advancement impact teaching performance of surgical faculty? *Surgery, 136*(2), 277-281. https://doi.org/10.1016/j.surg.2004.03.015

Cerritelli, F., Lacorte, E., Ruffini, N., & Vanacore, N. (2017). Osteopathy for primary headache patients: a systematic review. *Journal of Pain Research, 10*, 601-611. https://doi.org/10.2147/JPR.S130501

Cerritelli, F., van Dun, P. L. S., Esteves, J. E., Consorti, G., Sciomachen, P., Lacorte, E., & Vanacore, N. (2019). The Italian Osteopathic Practitioners Estimates and RAtes (OPERA) study: A cross sectional survey. *PLoS ONE, 14*(1), e0211353. https://doi.org/10.1371/journal.pone.0211353

Chandrasekhar, A. J., Durazo-Arvizu, R., Hoyt, A., & McNulty, J. A. (2013). Do student evaluations influence the teaching skills of clerkship clinical faculty? *Educational Research and Evaluation, 19*(7), 628-635. https://doi.org/10.1080/13803611.2013.834616

Chen, H. C., Fogh, S., Kobashi, B., Teherani, A., ten Cate, O., & O'Sullivan, P. (2016). An interview study of how clinical teachers develop skills to attend to different level learners. *Medical Teacher, 38*(6), 578-584. https://doi.org/10.3109/0142159X.2015.1073238

Chen, H. C., O'Sullivan, P., Teherani, A., Fogh, S., Kobashi, B., & Ten Cate, O. (2015). Sequencing learning experiences to engage different level learners in the workplace: An interview study with excellent clinical teachers. *Medical Teacher, 37*(12), 1090-1097. https://doi.org/10.3109/0142159X.2015.1009431

Chen, Y., & Hoshower, L. B. (2003). Student evaluation of teaching effectiveness: An assessment of student perception and motivation. *Assessment & Evaluation in Higher Education, 28*(1), 71-88. https://doi.org/10.1080/02602930301683

Cicchitti, L., Martelli, M., & Cerritelli, F. (2015). Chronic inflammatory disease and osteopathy: A systematic review. *PLoS ONE, 10*(3), e0121327. https://doi.org/10.1371/journal.pone.0121327

Clauser, B. E., Margolis, M. J., Holtman, M. C., Katsufrakis, P. J., & Hawkins, R. E. (2012). Validity considerations in the assessment of professionalism. *Advances in Health Sciences Education, 17*(2), 165-181. https://doi.org/10.1007/s10459-010-9219-6

Colbert, C. Y., Ogden, P. E., Lowe, D., & Moffitt, M. J. (2010). Students learn systems-based care and facilitate system change as stakeholders in a free clinic experience. *Advances in Health Sciences Education, 15*(4), 533-545. https://doi.org/10.1007/s10459-009-9216-9

Cole, B., & Wessel, J. (2008). How clinical instructors can enhance the learning experience of physical therapy students in an introductory clinical placement. *Advances in Health Sciences Education, 13*(2), 163-179. https://doi.org/10.1007/s10459-006-9030-6

Colliver, J. A., Conlee, M. J., & Verhulst, S. J. (2012). From test validity to construct validity… and back? *Medical Education, 46*(4), 366-371. https://doi.org/10.1111/j.1365-2923.2011.04194.x

Conigliaro, R. L., & Stratton, T. D. (2010). Assessing the quality of clinical teaching: a preliminary study. *Medical Education, 44*(4), 379-386. https://doi.org/10.1111/j.1365-2923.2009.03612.x

Consensus-based Standards for the Selection of Health Measurement Instruments. (2010). *COSMIN Taxonomy of Measurement Properties*. Retrieved August 20th, 2019, from https://www.cosmin.nl/tools/cosmin-taxonomy-measurement-properties/

Cook, D. A., & Beckman, T. J. (2006). Current concepts in validity and reliability for psychometric instruments: theory and application. *The American Journal of Medicine, 119*(2), 166.e167-166.e116. https://doi.org/10.1016/j.amjmed.2005.10.036

Cook, D. A., Brydges, R., Ginsburg, S., & Hatala, R. (2015). A contemporary approach to validity arguments: a practical guide to Kane's framework. *Medical Education, 49*(6), 560-575. https://doi.org/10.1111/medu.12678

Cook, D. A., Zendejas, B., Hamstra, S. J., Hatala, R., & Brydges, R. (2014). What counts as validity evidence? Examples and prevalence in a systematic review of simulation-based assessment. *Advances in Health Sciences Education, 19*(2), 233-250. https://doi.org/10.1007/s10459-013-9458-4

Cooke, M., Irby, D. M., & O'Brien, B. C. (2010). *Educating physicians: a call for reform of medical school and residency* (Vol. 16). John Wiley & Sons.

Copeland, H. L., & Hewson, M. G. (2000). Developing and testing an instrument to measure the effectiveness of clinical teaching in an academic medical centre. *Academic Medicine, 75*(2), 161-166. https://doi.org/10.1097/00001888-200002000-00015

Cottrell, D., Kilminster, S., Jolly, B., & Grant, J. (2002). What is effective supervision and how does it happen? A critical incident study. *Medical Education, 36*(11), 1042-1049. https://doi.org/10.1046/j.1365-2923.2002.01327.x

Coverdale, J. H., Ismail, N., Mian, A., & Dewey, C. (2010). Toolbox for evaluating residents as teachers. *Academic Psychiatry, 34*(4), 298-301. https://doi.org/10.1176/appi.ap.34.4.298

Cox, S. S., & Swanson, M. S. (2002). Indentification of teaching excellence in operating room and clinic settings. *The American Journal of Surgery, 183*, 251-255. https://doi.org/10.1016/s0002-9610(02)00787-0

Cross, V. (1995). Perceptions of the ideal clinical educator in physiotherapy education. *Physiotherapy, 81*(9), 506-513. https://doi.org/10.1016/S0031-9406(05)66680-1

Cruess, S. R., Cruess, R. L., & Steinert, Y. (2008). Role modelling—making the most of a powerful teaching strategy. *British Medical Journal, 336*(7646), 718-721. https://doi.org/10.1136/bmj.39503.757847.BE

Cuesta-Briand, B., Auret, K., Johnson, P., & Playford, D. (2014). A world of difference': a qualitative study of medical students' views on professionalism and the 'good doctor. *BMC Medical Education, 14*(1), 77. https://doi.org/10.1186/1472-6920-14-77

Currens, J. A. B., & Bithell, C. P. (2000). Clinical education: listening to different perspectives. *Physiotherapy, 86*(12), 645-653. https://doi.org/10.1016/S0031-9406(05)61302-8

Curtis, N., Helion, J. G., & Domsohn, M. (1998). Student athletic trainer perceptions of clinical supervisor behaviors: a critical incident study. *Journal of Athletic Training, 33*(3), 249. http://www.ncbi.nlm.nih.gov/pmc/articles/PMC1320432/pdf/jathtrain00011-0059.pdf

Daberkow, D. W., Hilton, C., Sanders, C. V., & Chauvin, S. W. (2005). Faculty evaluations by medicine residents using known versus anonymous systems. *Medical Education Online, 10*(1), 4380. https://doi.org/10.3402/meo.v10i.4380

Darwin, S. (2017). What contemporary work are student ratings actually doing in higher education? *Studies in Educational Evaluation, 54*, 13-21. https://doi.org/10.1016/j.stueduc.2016.08.002

Dawson, S. M., & Hocker, A. D. (2019). An evidence-based framework for peer review of teaching. *Advances in Physiology Education, 44*(1), 26-31. https://doi.org/10.1152/advan.00088.2019

De Champlain, A. F. (2010). A primer on classical test theory and item response theory for assessments in medical education. *Medical Education, 44*(1), 109-117. https://doi.org/10.1111/j.1365-2923.2009.03425.x

Dudek, N. L., Dojeiji, S., Day, K., & Varpio, L. (2016). Feedback to supervisors: is anonymity really so important? *Academic Medicine, 91*(9), 1305-1312. https://doi.org/10.1111/j.1365-2923.2009.03425.x

Dunegan, K. J., & Hrivnak, M. W. (2003). Characteristics of mindless teaching evaluations and the moderating effects of image compatibility. *Journal of Management Education, 27*(3), 280-303. https://doi.org/10.1177/1052562903027003002

Earls Larrison, T., & Korr, W. S. (2013). Does social work have a signature pedagogy? *Journal of Social Work Education, 49*(2), 194-206. https://doi.org/10.1080/10437797.2013.768102

Eberle, J., Stegmann, K., & Fischer, F. (2014). Legitimate peripheral participation in Communities of Practice: Participation support structures for newcomers in faculty student councils. *Journal of the Learning Sciences, 23*(2), 216-244. https://doi.org/10.1080/10508406.2014.883978

Edgar, S., & Connaughton, J. (2014). Exploring the role and skill set of physiotherapy clinical educators in work-integrated learning. *Asia Pacific Journal of Cooperative Education, 15*(1), 29-36. http://www.apjce.org/files/APJCE_15_1_29_36.pdf

Emerson, R. J., & Records, K. (2007). Design and testing of classroom and clinical teaching evaluation tools for nursing education. *International Journal of Nursing Education Scholarship, 4*(1). https://doi.org/10.2202/1548-923X.1375

Emery, C. R., Kramer, T. R., & Tian, R. G. (2003). Return to academic standards: A critique of student evaluations of teaching effectiveness. *Quality assurance in Education, 11*(1), 37-46. https://doi.org/10.1108/09684880310462074

Ercikan, K., Arim, R., Law, D., Domene, J., Gagnon, F., & Lacroix, S. (2010). Application of think aloud protocols for examining and confirming sources of differential item functioning identified by expert reviews. *Educational Measurement: Issues and Practice, 29*(2), 24-35. https://doi.org/10.1111/j.1745-3992.2010.00173.x

Fan, Y., Shepherd, L. J., Slavich, E., Waters, D., Stone, M., Abel, R., & Johnston, E. L. (2019). Gender and cultural bias in student evaluations: Why representation matters. *PLoS ONE, 14*(2), e0209749. https://doi.org/10.1371/journal.pone.0209749

Farnan, J. M., Petty, L. A., Georgitis, E., Martin, S., Chiu, E., Prochaska, M., & Arora, V. M. (2012). A systematic review: the effect of clinical supervision on patient and residency education outcomes. *Academic Medicine, 87*(4), 428-442. https://doi.org/10.1097/ACM.0b013e31824822cc

Fassiotto, M., Li, J., Maldonado, Y., & Kothary, N. (2018). Female surgeons as counter stereotype: The impact of gender perceptions on trainee evaluations of physician faculty. *Journal of Surgical Education, 75*(5), 1140-1148. https://doi.org/10.1016/j.jsurg.2018.01.011

Fawkes, C., Leach, J., Mathias, S., & Moore, A. (2013). Profiling osteopathic practice in the UK using standardised data collection. *International Journal of Osteopathic Medicine, 16*(1), e10. https://doi.org/10.1016/j.ijosm.2013.01.006

Fernandez, E. (1998). Student perceptions of satisfaction with practicum learning. *Social Work Education, 17*(2), 173-201. https://doi.org/10.1080/02615479811220181

Fiddes, P. J., Brooks, P. M., & Komesaroff, P. (2013). The patient is the teacher: ambulatory patient-centred student-based interprofessional education where the patient is the teacher who improves patient care outcomes. *Internal Medicine Journal, 43*(7), 747-750. https://doi.org/10.1111/imj.12197

Figueiró-Filho, E. A., Amaral, E., McKinley, D., Bezuidenhout, J., & Tekian, A. (2014). Minimal supervision out-patient clinical teaching. *The Clinical Teacher, 11*(5), 365-369. https://doi.org/10.1111/tct.12237

Florance, B.-M., Frin, G., Dainese, R., Nébot-Vivinus, M.-H., Barjoan, E. M., Marjoux, S., Laurens, J.-P., Payrouse, J.-L., Hébuterne, X., & Piche, T. (2012). Osteopathy improves the severity of irritable bowel syndrome: a pilot randomized sham-controlled study. *European Journal of Gastroenterology and Hepatology, 24*(8), 944-949. https://doi.org/10.1097/MEG.0b013e3283543eb7

Fluit, C. (2010). Assessing the quality of clinical teachers. *Journal of General Internal Medicine, 25*(12), 1337-1345. https://doi.org/10.1007/s11606-010-1458-y

Fluit, C. R., Feskens, R., Bolhuis, S., Grol, R., Wensing, M., & Laan, R. (2014). Understanding resident ratings of teaching in the workplace: a multi-centre

study. *Advances in Health Sciences Education, 20*(3), 691-707. https://doi.org/10.1007/s10459-014-9559-8

Fluit, C. R. M. G., Feskens, R., Bolhuis, S., Grol, R., Wensing, M., & Laan, R. (2013). Repeated evaluations of the quality of clinical teaching by residents. *Perspectives on Medical Education, 2*(2), 87-94. https://doi.org/10.1007/s40037-013-0060-5

Fraile, R., & Bosch-Morell, F. (2015). Considering teaching history and calculating confidence intervals in student evaluations of teaching quality. *Higher Education, 70*(1), 55-72. https://doi.org/10.1007/s10734-014-9823-0

Frank, J. R., & Danoff, D. (2007). The CanMEDS initiative: implementing an outcomes-based framework of physician competencies. *Medical Teacher, 29*(7), 642-647. https://doi.org/10.1080/01421590701746983

Frawley, J., Sundberg, T., Steel, A., Sibbritt, D., Broom, A., & Adams, J. (2016). Prevalence and characteristics of women who consult with osteopathic practitioners during pregnancy; a report from the Australian Longitudinal Study on Women's Health (ALSWH). *Journal of Bodywork and Movement Therapies, 20*(1), 168-172. https://doi.org/10.1016/j.jbmt.2015.03.004

Gaskin, C. J., & Happell, B. (2014). On exploratory factor analysis: A review of recent evidence, an assessment of current practice, and recommendations for future use. *International Journal of Nursing Studies, 51*, 511-521. https://doi.org/10.1016/j.ijnurstu.2013.10.005

General Osteopathic Council. (2015). *Guidance for Osteopathic Pre-registration Education*). http://www.osteopathy.org.uk/news-and-resources/document-library/training/guidance-for-osteopathic-pre-registration-education/

Gertz, A. M., Frank, S., & Blixen, C. E. (2011). A survey of patients and providers at free clinics across the United States. *Journal of Community Health, 36*(1), 83-93. https://doi.org/10.1007/s10900-010-9286-x

Gerzina, T. M., McLean, T., & Fairley, J. (2005). Dental clinical teaching: perceptions of students and teachers. *Journal of Dental Education, 69*(12), 1377-1384. https://doi.org/10.1002/j.0022-0337.2005.69.12.tb04037.x

Goldie, J., Dowie, A., Goldie, A., Cotton, P., & Morrison, J. (2015). What makes a good clinical student and teacher? An exploratory study. *BMC Medical Education, 15*(1), 40. https://doi.org/10.1186/s12909-015-0314-5

Golding, C., & Adam, L. (2016). Evaluate to improve: useful approaches to student evaluation. *Assessment & Evaluation in Higher Education, 41*(1), 1-14. https://doi.org/10.1080/02602938.2014.976810

Goodyear, R. K., Bunch, K., & Claiborn, C. D. (2006). Current supervision scholarship in psychology: A five year review. *The Clinical Supervisor, 24*(1-2), 137-147. https://doi.org/10.1300/J001v24n01_07

Grant, J., Kilminster, S., Jolly, B., & Cottrell, D. (2003). Clinical supervision of SpRs: where does it happen, when does it happen and is it effective? *Medical Education, 37*(2), 140-148. https://doi.org/10.1046/j.1365-2923.2003.01415.x

Guillaud, A., Darbois, N., Monvoisin, R., & Pinsault, N. (2018). Reliability of diagnosis and clinical efficacy of visceral osteopathy: a systematic review. *BMC Complementary and Alternative Medicine, 18*(1), 65. https://doi.org/10.1186/s12906-018-2098-8

Gusic, M., Hageman, H., & Zenni, E. (2013). Peer review: a tool to enhance clinical teaching. *The Clinical Teacher, 10*(5), 287-290. https://doi.org/10.1111/tct.12039

Haggarty, D., & Dalcin, D. (2014). Student-run clinics in Canada: an innovative method of delivering interprofessional education. *Journal of Interprofessional Care, 28*(6), 570-572. https://doi.org/10.3109/13561820.2014.916658

Haines, T. P., Kent, F., & Keating, J. L. (2014). Interprofessional student clinics: an economic evaluation of collaborative clinical placement education. *Journal of Interprofessional Care, 28*(4), 292-298. https://doi.org/10.3109/13561820.2013.874983

Handley, L. I., Warholak Jackson, T. L., & Jackson, T. R. (2008). An evaluation of the validity of inferences made from 3 diabetes assessment instruments: A Rasch analysis. *Research in Social and Administrative Pharmacy, 4*(1), 67-81. https://doi.org/https://doi.org/10.1016/j.sapharm.2007.04.002

Hatala, R., Cook, D. A., Brydges, R., & Hawkins, R. (2015). Constructing a validity argument for the Objective Structured Assessment of Technical Skills (OSATS): A systematic review of validity evidence. *Advances in Health Sciences Education, 20*(5), 1149-1175. https://doi.org/10.1007/s10459-015-9593-1

Hathcoat, J. D. (2013). Validity semantics in educational and psychological assessment. *Practical Assessment, Research & Evaluation, 18*(9). https://doi.org/10.7275/ay6p-xw09

Hays, R. (2008). Assessment in medical education: roles for clinical teachers. *The Clinical Teacher, 5*(1), 23-27. https://doi.org/10.1111/j.1743-498X.2007.00165.x

Healey, W. E. (2008). Physical therapist student approaches to learning during clinical education experiences: a qualitative study. *Journal of Physical Therapy Education, 22*(1), 49-58. https://journals.lww.com/jopte/Abstract/2008/01000/Physical_Therapist_Student_Approaches_to_Learning.8.aspx

Heidenreich, C., Lye, P., Simpson, D., & Lourich, M. (2000). The search for effective and efficient ambulatory teaching methods through the literature. *Pediatrics, 105*(Supplement 2), 231-237. https://pediatrics.aappublications.org/content/105/Supplement_2/231

Henzi, D., Davis, E., Jasinevicius, R., & Hendricson, W. (2006). North American dental students' perspectives about their clinical education. *Journal of Dental Education, 70*(4), 361-377. https://doi.org/10.1002/j.0022-0337.2006.70.4.tb04091.x

Hewson, M. B., & Jensen, N. (1990). An inventory to improve clinical teaching in the general internal medicine clinic. *Medical Education, 24*, 518-527. https://doi.org/10.1111/j.1365-2923.1990.tb02668.x

Holmqvist, M., Courtney, C., Meili, R., & Dick, A. (2012). Student-run clinics: Opportunities for interprofessional education and increasing social accountability. *Journal of Research in Interprofessional Practice and Education, 2*(3). https://doi.org/10.22230/jripe.2012v2n3a80

Hornstein, H. A. (2017). Student evaluations of teaching are an inadequate assessment tool for evaluating faculty performance. *Cogent Education, 4*(1), 1304016. https://doi.org/10.1080/2331186X.2017.1304016

Hubley, A. M., & Zumbo, B. D. (2011). Validity and the consequences of test interpretation and use. *Social Indicators Research, 103*(2), 219. https://doi.org/10.1007/s11205-011-9843-4

Huff, N. G., Roy, B., Estrada, C. A., Centor, R. M., Castiglioni, A., Willett, L. L., Shewchuk, R. M., & Cohen, S. (2014). Teaching behaviors that define highest rated attending physicians: a study of the resident perspective. *Medical Teacher, 36*(11), 991-996. https://doi.org/10.3109/0142159X.2014.920952

Irby, D. M. (1995). Teaching and learning in ambulatory care settings: a thematic review of the literature. *Academic Medicine, 70*(10), 898-931. https://doi.org/10.1097/00001888-199510000-00014

Irby, D. M. (2014). Excellence in clinical teaching: knowledge transformation and development required. *Medical Education, 48*(8), 776-784. https://doi.org/10.1111/medu.12507

Irby, D. M., & Bowen, J. L. (2004). Time-efficient strategies for learning and performance. *The Clinical Teacher, 1*(1), 23-28. https://doi.org/10.1111/j.1743-498X.2004.00013.x

Jahan, F., Sadaf, S., Kalia, S., Khan, A., & Hamza, H. B. (2008). Attributes of an effective clinical teacher: a survey on students' and teachers' perceptions. *Journal of the College of Physicians and Surgeons Pakistan, 18*(6), 357. http://ecommons.aku.edu/pakistan_fhs_mc_fam_med/2

Jahangiri, L., Mucciolo, T. W., Choi, M., & Spielman, A. I. (2008). Assessment of teaching effectiveness in US dental schools and the value of triangulation. *Journal of Dental Education, 72*(6), 707-718. https://doi.org/10.1002/j.0022-0337.2008.72.6.tb04536.x

Jochemsen-van der Leeuw, H. G. A. R., van Dijk, N., van Etten-Jamaludin, F. S., & Wieringa-de Waard, M. (2013). The attributes of the clinical trainer as a role model: a systematic review. *Academic Medicine, 88*(1), 26-34. https://doi.org/10.1097/ACM.0b013e318276d070

Kane, M. T. (1992). An argument-based approach to validity. *Psychological Bulletin, 112*(3), 527. https://doi.org/10.1037/0033-2909.112.3.527

Kane, M. T. (2006). Validation. In R. Brennan (Ed.), *Educational measurement* (4th ed., pp. 17-64). Praeger Publishers.

Kane, M. T. (2013). Validating the interpretations and uses of test scores. *Journal of Educational Measurement, 50*(1), 1-73. https://doi.org/10.1111/jedm.12000

Kane, M. T. (2016a). Explicating validity. *Assessment in Education: Principles, Policy & Practice, 23*(2), 198-211. https://doi.org/10.1080/0969594X.2015.1060192

Kane, M. T. (2016b). Validity as the evaluation of the claims based on test scores. *Assessment in Education: Principles, Policy & Practice, 23*(2), 309-311. https://doi.org/10.1080/0969594X.2016.1156645

Kelly, C. (2007). Student's perceptions of effective clinical teaching revisited. *Nurse Education Today, 27*(8), 885-892. https://doi.org/10.1016/j.nedt.2006.12.005

Kember, D., Leung, D. Y. P., & Kwan, K. (2002). Does the use of student feedback questionnaires improve the overall quality of teaching? *Assessment & Evaluation in Higher Education, 27*(5), 411-425. https://doi.org/10.1080/0260293022000009294

Kent, F., Martin, N., & L. Keating, J. (2016). Interprofessional student-led clinics: An innovative approach to the support of older people in the community. *Journal of Interprofessional Care, 30*(1), 123-128. https://doi.org/10.3109/13561820.2015.1070133

Kernan, W. N., Hershman, W., Alper, E. J., Lee, M. Y., Viscoli, C. M., Perry, J. R., & O'Connor, P. G. (2008). Disagreement between students and preceptors regarding the value of teaching behaviors for ambulatory care settings. *Teaching and Learning in Medicine, 20*(2), 143-150. https://doi.org/10.1080/10401330801991667

Kilminster, S., Cottrell, D., Grant, J., & Jolly, B. (2007). AMEE Guide No. 27: Effective educational and clinical supervision. *Medical Teacher, 29*(1), 2-19. https://doi.org/10.1080/01421590701210907

Kilminster, S., & Jolly, B. (2001). Effective supervision in clinical practice settings: a literature review. *Medical Education, 34*(10), 827-840. https://doi.org/10.1046/j.1365-2923.2000.00758.x

King, H. H. (2013). Severity of Irritable Bowel Syndrome symptoms is reduced by osteopathy. *Journal of the American Osteopathic Association, 113*(4), 357-358. https://doi.org/10.1111/j.1440-1746.2006.04741.x

Kisiel, J. B., Bundrick, J. B., & Beckman, T. J. (2010). Resident physicians' perspectives on effective outpatient teaching: a qualitative study. *Advances in*

Health Sciences Education, 15(3), 357-368. https://doi.org/10.1007/s10459-009-9202-2

Knorr, M., & Klusmann, D. (2015). The trouble with validity: what is part of it and what is not? *Medical Education, 49*(6), 550-552. https://doi.org/10.1111/medu.12738

Koh, L. C. (2008). Refocusing formative feedback to enhance learning in pre-registration nurse education. *Nurse Education in Practice, 8*(4), 223-230. https://doi.org/10.1016/j.nepr.2007.08.002

Krueger, P. M., Dane, P., Slocum, P., & Kimmelman, M. (2009). Osteopathic clinical training in three universities. *Academic Medicine, 84*(6), 712-717. https://doi.org/10.1097/ACM.0b013e3181a409b1

Kumar, K., & Greenhill, J. (2016). Factors shaping how clinical educators use their educational knowledge and skills in the clinical workplace: a qualitative study. *BMC Medical Education, 16*(1), 68. https://doi.org/10.1186/s12909-016-0590-8

Lauber, C. A., Toth, P. E., Leary, P. A., Martin, R. D., & Killian, C. B. (2003). Program directors' and clinical instructors' perceptions of important clinical-instructor behavior categories in the delivery of athletic training clinical instruction. *Journal of Athletic Training, 38*(4), 336-341. https://www.ncbi.nlm.nih.gov/pmc/articles/PMC314394/

Lave, J., & Wenger, E. (1991). *Situated Learning: Legitimate Perpiheral Participation*. Cambridge University Press.

LeBaron Wallace, T. (2011). An argument-based approach to validity in evaluation. *Evaluation, 17*(3), 233-246. https://doi.org/10.1177/1356389011410522

Lee, W. S., Cholowski, K., & Williams, A. K. (2002). Nursing students' and clinical educators' perceptions of characteristics of effective clinical educators in an Australian university school of nursing. *Journal of Advanced Nursing, 39*(5), 412-420. https://doi.org/10.1046/j.1365-2648.2002.02306.x

Leone-Perkins, M., Schnuth, R., & Kantner, T. (1999). Preceptor-student interactions in an ambulatory clerkship: Gender differences in student evaluations of teaching. *Teaching and Learning in Medicine, 11*(3), 164-167. https://doi.org/10.1207/S15328015TL110307

Linse, A. R. (2017). Interpreting and using student ratings data: Guidance for faculty serving as administrators and on evaluation committees. *Studies in Educational Evaluation, 54*, 94-106. https://doi.org/10.1016/j.stueduc.2016.12.004

Long, T. L., Brietkreuz, K. R., Diaz, D. A., McNulty, J., Engler, A. J., Poliforini, C., & Telford, C. (2012). Signature pedagogies in nursing education. In N. L. Chick, A. Haynie, & R. A. R. Gurung (Eds.), *Signature pedagogies: Approaches to teaching disciplinary habits of mind* (pp. 188-202). Stylus Sterling.

Looi, J. C. L., & Anderson, K. J. (2018). Between SET and ASP: balancing the scales of student evaluation of teaching (SET) and teachers' assessments of student performance (ASP) for medical school education in psychiatry. *Australasian Psychiatry, 26*(6), 659-661. https://doi.org/10.1177/1039856218783854

MacNell, L., Driscoll, A., & Hunt, A. N. (2015). What's in a name: Exposing gender bias in student ratings of teaching. *Innovative Higher Education, 40*(4), 291-303. https://doi.org/10.1007/s10755-014-9313-4

Maker, V. K., Lewis, M. J., & Donnelly, M. B. (2006). Ongoing faculty evaluations: developmental gain or just more pain? *Current Surgery, 63*(1), 80-84. https://doi.org/10.1016/j.cursur.2005.08.010

Martin, P., Copley, J., & Tyack, Z. (2014). Twelve tips for effective clinical supervision based on a narrative literature review and expert opinion. *Medical Teacher, 36*(3), 201-207. https://doi.org/10.3109/0142159X.2013.852166

Masunaga, H., & Hitchcock, M. A. (2010). Residents' and faculty's beliefs about the ideal clinical teacher. *Family Medicine, 42*(2), 116-120. https://www.stfm.org/familymedicine/vol42issue2/Masunaga116

Matthews, C. (2000). Role modelling: how does it influence teaching in Family Medicine? *Medical Education, 34*(6), 443-448. https://doi.org/10.1046/j.1365-2923.2000.00624.x

Mazor, K. M., Stone, S. L., Carlin, M., & Alper, E. (2002). What do medicine clerkship preceptors do best? *Academic Medicine, 77*(8), 837-840. https://doi.org/10.1097/00001888-200208000-00018

McAllister, L., & Nagarajan, S. (2015). Accreditation requirements in allied health education: Strengths, weaknesses and missed opportunities. *Journal of*

Teaching and Learning for Graduate Employability, 6(1), 2-23. https://doi.org/10.21153/jtlge2015vol6no1art570

McAuley, J. W., Backo, J. L., Sobota, K. F., Metzger, A. H., & Ulbrich, T. (2017). Identifying motivators and barriers to student completion of instructor evaluations: A multi-faceted, collaborative approach from four colleges of pharmacy. *Currents in Pharmacy Teaching and Learning, 9*(1), 20-27. https://doi.org/10.1016/j.cptl.2016.08.029

McCann, S., & Gardner, C. (2014). Student personality differences are related to their responses on instructor evaluation forms. *Assessment & Evaluation in Higher Education, 39*(4), 412-426. https://doi.org/10.1080/02602938.2013.845647

McClain, L., Gulbis, A., & Hays, D. (2018). Honesty on student evaluations of teaching: effectiveness, purpose, and timing matter! *Assessment & Evaluation in Higher Education, 43*(3), 369-385. https://doi.org/10.1080/02602938.2017.1350828

McOwen, K. S., Bellini, L. M., Guerra, C. E., & Shea, J. A. (2007). Evaluation of clinical faculty: gender and minority implications. *Academic Medicine, 82*(10), S94-S96. https://doi.org/10.1097/ACM.0b013e3181405a10

Meah, Y. S., Smith, E. L., & Thomas, D. C. (2009). Student-run health clinic: Novel arena to educate medical students on systems-based practice. *Mount Sinai Journal of Medicine, 76*(4), 344-356. https://doi.org/10.1002/msj.20128

Messick, S. (1989). Validity. In R. Linn (Ed.), *Educational measurement* (pp. 13-103). American Council on Education & Macmillan.

Modi, A., Fascelli, M., Daitch, Z., & Hojat, M. (2017). Evaluating the relationship between participation in student-run free clinics and changes in empathy in medical students. *Journal of Primary Care & Community Health, 8*(3), 122-126. https://doi.org/10.1177/2150131916685199

Moore, K., Bacon, R., Bevitt, T., Bialocerkowski, A., Ciccone, N., Haworth, N., Horstmanshof, L., Milne, N., Naumann, F., & Sanderson, B. (2018). Report: The university health clinic: Definition, educational practices and outcomes. *Focus on Health Professional Education: A Multi-Disciplinary Journal, 19*(2), 1. https://doi.org/10.11157/fohpe.v19i2.213

Moore, K., & Field, B. J. (2017). Enhancing clinical education in the private practice setting: A case study in osteopathy. *International Journal of Osteopathic Medicine, 23*, 42-50. https://doi.org/10.1016/j.ijosm.2016.06.003

Morgan, H. K., Purkiss, J. A., Porter, A. C., Lypson, M. L., Santen, S. A., Christner, J. G., Grum, C. M., & Hammoud, M. M. (2016). Student evaluation of faculty physicians: gender differences in teaching evaluations. *Journal of Women's Health, 25*(5), 453-456. https://doi.org/10.1089/jwh.2015.5475

Morin, C., & Aubin, A. (2014). Primary reasons for osteopathic consultation: a prospective survey in Quebec. *PLoS ONE, 9*(9), e106259. https://doi.org/10.1371/journal.pone.0106259 (e0121180)

Moskowitz, D., Glasco, J., Johnson, B., & Wang, G. (2006). Students in the community: an interprofessional student-run free clinic. *Journal of Interprofessional Care, 20*(3), 254-259. https://doi.org/10.1080/13561820600721091

Myers, K., Zibrowski, E. M., & Lingard, L. (2012). Engaged at the extremes: Residents' perspectives on clinical teaching assessment. *Academic Medicine, 87*(10), 1397-1400. https://doi.org/ACM.0b013e3182674488

Newton, J. M., Billett, S., Jolly, B., & Ockerby, C. M. (2009). Lost in translation: Barriers to learning in health professional clinical education. *Learning in Health and Social Care, 8*(4), 315-327. https://doi.org/10.1111/j.1473-6861.2009.00229.x

Nygaard, C., & Belluigi, D. Z. (2011). A proposed methodology for contextualised evaluation in higher education. *Assessment & Evaluation in Higher Education, 36*(6), 657-671. https://doi.org/10.1080/02602931003650037

Oermann, M. H., Conklin, J. L., Rushton, S., & Bush, M. A. (2018). Student evaluations of teaching (SET): Guidelines for their use. *Nursing Forum, 53*, 280-285. https://doi.org/10.1111/nuf.12249

Okoronkwo, I. L., Onyia-Pat, J.-L., Agbo, M.-A. E., Okpala, P. U., & Ndu, A. C. (2013). Students' perception of effective clinical teaching and teacher behaviour. *Open Journal of Nursing, 3*(1), 63-70. https://doi.org/10.4236/ojn.2013.31008

Orrock, P. (2009). Profile of members of the Australian Osteopathic Association: Part 1 – The practitioners. *International Journal of Osteopathic Medicine, 12*(1), 14-24. https://doi.org/10.1016/j.ijosm.2008.04.002

Ory, J. C., & Ryan, K. (2001). How do student ratings measure up to a new validity framework? *New Directions for Institutional Research, 2001*(109), 27-44. https://doi.org/10.1002/ir.2

Osteopathy Board of Australia. (2019). *Capabilities for Osteopathic Practice*. Retrieved July 4th, 2019, from https://www.osteopathyboard.gov.au/Codes-Guidelines/Capabilities-for-osteopathic-practice.aspx

Padilla, J.-L., & Benítez, I. (2014). Validity evidence based on response processes. *Psicothema, 26*(1), 136-144. https://doi.org/10.7334/psicothema2013.259

Pallant, J. F., & Tennant, A. (2007). An introduction to the Rasch measurement model: an example using the Hospital Anxiety and Depression Scale (HADS). *British Journal of Clinical Psychology, 46*(1), 1-18. https://doi.org/10.1348/014466506x96931

Health Practitioner Regulation National Law Act 2009 (Qld), (2009). https://www.legislation.qld.gov.au/LEGISLTN/CURRENT/H/HealthPracRNatLaw.pdf

Patrick, C. L. (2011). Student evaluations of teaching: effects of the Big Five personality traits, grades and the validity hypothesis. *Assessment & Evaluation in Higher Education, 36*(2), 239-249. https://doi.org/10.1080/02602930903308258

Peterson, D. A. M., Biederman, L. A., Andersen, D., Ditonto, T. M., & Roe, K. (2019). Mitigating gender bias in student evaluations of teaching. *PLoS ONE, 14*(5), e0216241. https://doi.org/10.1371/journal.pone.0216241

Pettit, J. E., Axelson, R. D., Ferguson, K. J., & Rosenbaum, M. E. (2015). Assessing effective teaching: what medical students value when developing evaluation instruments. *Academic Medicine, 90*(1), 94-99. https://doi.org/10.1097/ACM.0000000000000447

Pinard, A.-M., Savard, I., & Côté, L. (201). Role modelling: moving from implicit to explicit. *Clinical Teacher, 15*(5), 430-432. https://doi.org/10.1111/tct.12727

Prideaux, D., Alexander, H., Bower, A., Dacre, J., Haist, S., Jolly, B., Norcini, J., Roberts, T., Rothman, A., & Rowe, R. (2000). Clinical teaching: maintaining

an educational role for doctors in the new health care environment. *Medical Education, 34*(10), 820-826. https://doi.org/10.1046/j.1365-2923.2000.00756.x

Quality Assurance Agency for Higher Education. (2015). *Subject Benchmark Statement: Osteopathy* (United Kingdom Quality Code for Higher Education, Issue). Quality Assurance Agency. http://www.qaa.ac.uk/en/Publications/Documents/SBS-Osteopathy-15.pdf

Ramani, S., & Leinster, S. (2008). AMEE Guide no. 34: teaching in the clinical environment. *Medical Teacher, 30*(4), 347-364. https://doi.org/10.1080/01421590802061613

Regan-Smith, M., Young, W. W., & Keller, A. M. (2002). An efficient and effective teaching model for ambulatory education. *Academic Medicine, 77*(7), 593-599. https://doi.org/10.1097/00001888-200207000-00003

Reichsman, F., Browning, F. E., & Hinshaw, J. R. (1964). Observations of undergraduate clinical teaching in action. *Academic Medicine, 39*(2), 147-163. https://journals.lww.com/academicmedicine/citation/1964/02000/observations_of_undergraduate_clinical_teaching_in.8.aspx

Richardson, J. T. E. (2005). Instruments for obtaining student feedback: A review of the literature. *Assessment & Evaluation in Higher Education, 30*(4), 387-415. https://doi.org/10.1080/02602930500099193

Robinson, R. (2015). Hospitalist workload and resident evaluations. *The Clinical Teacher, 12*(1), 55-58. https://doi.org/10.1111/tct.12253

Rodger, S., Webb, G., Devitt, L., Gilbert, J., Wrightson, P., & McMeeken, J. (2008). Clinical education and practice placements in the allied health professions: an international perspective. *Journal of Allied Health, 37*(1), 53-62. https://www.ingentaconnect.com/content/asahp/jah/2008/00000037/00000001/art00009

Royal, K. D. (2017). Four tenets of modern validity theory for medical education assessment and evaluation. *Advances in Medical Education and Practice, 8*, 567. https://doi.org/10.2147/AMEP.S139492

Rutz, M., Turner, J., Pettit, K., Palmer, M. M., Perkins, A., & Cooper, D. D. (2019). Factors that contribute to resident teaching effectiveness. *Cureus, 11*(3). https://doi.org/10.7759/cureus.4290

Schaber, P., Marsh, L., Wilcox, K. J., Chick, N. L., Haynie, A., & Gurung, R. A. R. (2012). Relational learning and active engagement in occupational therapy professional education. In N. L. Chick, A. Haynie, & R. A. R. Gurung (Eds.), *Exploring more signature pedagogies: Approaches to teaching disciplinary habits of mind* (pp. 188-202). Stylus Sterling.

Scheepers, R. A., Lombarts, K. M., van Aken, M. A., Heineman, M. J., & Arah, O. A. (2014). Personality traits affect teaching performance of attending physicians: Results of a multi-center observational study. *PLoS ONE, 9*(5), e98107. https://doi.org/10.1371/journal.pone.0098107

Schiekirka, S., Reinhardt, D., Heim, S., Fabry, G., Pukrop, T., Anders, S., & Raupach, T. (2012). Student perceptions of evaluation in undergraduate medical education: A qualitative study from one medical school. *BMC Medical Education, 12*(1), 45. https://doi.org/10.1186/1472-6920-12-45

Schönwetter, D. J., Lavigne, S., Mazurat, R., & Nazarko, O. (2006). Students' perceptions of effective classroom and clinical teaching in dental and dental hygiene education. *Journal of Dental Education, 70*(6), 624-635. https://doi.org/10.1002/j.0022-0337.2006.70.6.tb04118.x

Schultz, K. W., Kirby, J., Delva, D., Godwin, M., Verma, S., Birtwhistle, R., Knapper, C., & Seguin, R. (2004). Medical students' and residents' preferred site characteristics and preceptor behaviours for learning in the ambulatory setting: a cross-sectional survey. *BMC Medical Education, 4*(1), 12. https://doi.org/10.1186/1472-6920-4-12

Schutte, T., Tichelaar, J., Dekker, R. S., van Agtmael, M. A., de Vries, T. P. G. M., & Richir, M. C. (2015). Learning in student-run clinics: a systematic review. *Medical Education, 49*(3), 249-263. https://doi.org/10.1111/medu.12625

Schuwirth, L. W., & van der Vleuten, C. P. (2012). Programmatic assessment and Kane's validity perspective. *Medical Education, 46*(1), 38-48. https://doi.org/10.1111/j.1365-2923.2011.04098.x

Seif, G., Coker-Bolt, P., Kraft, S., Gonsalves, W., Simpson, K., & Johnson, E. (2014). The development of clinical reasoning and interprofessional behaviors: service-learning at a student-run free clinic. *Journal of Interprofessional Care, 28*(6), 559-564. https://doi.org/10.3109/13561820.2014.921899

Sevenhuysen, S. L., & Haines, T. (2011). The slave of duty: Why clinical educators across the continuum of care provide clinical education in physiotherapy. *Hong Kong Physiotherapy Journal, 29*(2), 64-70. https://doi.org/10.1016/j.hkpj.2011.06.002

Sheu, L., O'Brien, B., O'Sullivan, P. S., Kwong, A., & Lai, C. J. (2013). Systems-based practice learning opportunities in student-run clinics: a qualitative analysis of student experiences. *Academic Medicine, 88*(6), 831-836. https://doi.org/10.1097/ACM.0b013e31828ff92c

Shulman, L. S. (2005). Signature pedagogies in the professions. *Daedalus, 134*(3), 52-59. https://doi.org/10.1162/0011526054622015

Singh, S., Pai, D. R., Sinha, N. K., Kaur, A., Soe, H. H. K., & Barua, A. (2013). Qualities of an effective teacher: what do medical teachers think? *BMC Medical Education, 13*(1), 128. https://doi.org/10.1186/1472-6920-13-128

Skeff, K. M., Stratos, G. A., Bergen, M. R., Albright, C. L., Berman, J., Farquhar, J. W., & Sox Jr, H. C. (1992). The Stanford faculty development program: a dissemination approach to faculty development for medical teachers. *Teaching and Learning in Medicine, 4*(3), 180-187. https://doi.org/10.1080/10401339209539559

Smith, S., Thomas, R., Cruz, M., Griggs, R., Moscato, B., & Ferrara, A. (2014). Presence and characteristics of student-run free clinics in medical schools. *JAMA, 312*(22), 2407-2410. https://doi.org/10.1001/jama.2014.16066

Snell, L., Tallett, S., Haist, S., Hays, R., Norcini, J., Prince, K., Rothman, A., & Rowe, R. (2000). A review of the evaluation of clinical teaching: new perspectives and challenges. *Medical Education, 34*(10), 862-870. https://doi.org/10.1046/j.1365-2923.2000.00754.x

Snowdon, D. A., Leggat, S. G., & Taylor, N. F. (2017). Does clinical supervision of healthcare professionals improve effectiveness of care and patient experience? A systematic review. *BMC Health Services Research, 17*(1), 786. https://doi.org/10.1186/s12913-017-2739-5

Spooren, P., Brockx, B., & Mortelmans, D. (2013). On the validity of student evaluation of teaching: The state of the art. *Review of Educational Research, 83*(4), 598-642. https://doi.org/10.3102/0034654313496870

Spooren, P., & Christiaens, W. (2017). I liked your course because I believe in (the power of) student evaluations of teaching (SET). Students' perceptions of a teaching evaluation process and their relationships with SET scores. *Studies in Educational Evaluation, 54*, 43-49. https://doi.org/10.1016/j.stueduc.2016.12.003

Spooren, P., Mortelmans, D., & Denekens, J. (2007). Student evaluation of teaching quality in higher education: development of an instrument based on 10 Likert-scales. *Assessment & Evaluation in Higher Education, 32*(6), 667-679. https://doi.org/10.1080/02602930601117191

Steiner, I. P., Yoon, P. W., Kelly, K. D., Diner, B. M., Donoff, M. G., Mackey, D. S., & Rowe, B. H. (2003). Resident evaluation of clinical teachers based on teachers' certification. *Academic Emergency Medicine, 10*(7), 731-737. https://doi.org/10.1111/j.1553-2712.2003.tb00067.x

Stern, D. T., Williams, B. C., Gill, A., Gruppen, L. D., Woolliscroft, J. O., & Grum, C. M. (2000). Is there a relationship between attending physicians' and residents' teaching skills and students' examination scores? *Academic Medicine, 75*(11), 1144-1146. https://doi.org/10.1097/00001888-200011000-00024

Sutkin, G., Wagner, E., Harris, I., & Schiffer, R. (2008). What makes a good clinical teacher in medicine? A review of the literature. *Academic Medicine, 83*(5), 452-466. https://doi.org/10.1097/ACM.0b013e31816bee61

Torre, D. M., Sebastian, J. L., & Simpson, D. E. (2003). Learning activities and high-quality teaching: perceptions of third-year IM clerkship students. *Academic Medicine, 78*(8), 812-814. https://doi.org/10.1097/00001888-200308000-00015

Tran, T. T. T., & Do, T. X. (2020). Student evaluation of teaching: do teacher age, seniority, gender, and qualification matter? *Educational Studies*, 1-28. https://doi.org/10.1080/03055698.2020.1771545

Uijtdehaage, S., & O'Neal, C. (2015). A curious case of the phantom professor: mindless teaching evaluations by medical students. *Medical Education, 49*(9), 928-932. https://doi.org/10.1111/medu.12647

Ullian, J. A., Bland, C. J., & Simpson, D. E. (1994). An alternative approach to defining the role of the clinical teacher. *Academic Medicine, 69*(10), 832-838. https://doi.org/10.1097/00001888-199410000-00013

Uttl, B., White, C. A., & Gonzalez, D. W. (2017). Meta-analysis of faculty's teaching effectiveness: Student evaluation of teaching ratings and student learning are not related. *Studies in Educational Evaluation, 54*, 22-42. https://doi.org/10.1016/j.stueduc.2016.08.007

Valencia, E. (2020). Acquiescence, instructor's gender bias and validity of student evaluation of teaching. *Assessment & Evaluation in Higher Education, 45*(4), 483-495. https://doi.org/10.1080/02602938.2019.1666085

Van der Leeuw, R. M., Boerebach, B. C. M., Lombarts, K. M., Heineman, M. J., & Arah, O. A. (2016). Clinical teaching performance improvement of faculty in residency training: a prospective cohort study. *Medical Teacher, 38*(5), 464-470. https://doi.org/10.3109/0142159X.2015.1060302

Van der Leeuw, R. M., Overeem, K., Arah, O. A., Heineman, M. J., & Lombarts, K. M. (2013a). Frequency and determinants of residents' narrative feedback on the teaching performance of faculty: Narratives in numbers. *Academic Medicine, 88*(9), 1324-1331. https://doi.org/10.1097/ACM.0b013e31829e3af4

Van der Leeuw, R. M., Slootweg, I. A., Heineman, M. J., & Lombarts, K. M. (2013b). Explaining how faculty members act upon residents' feedback to improve their teaching performance. *Medical Education, 47*(11), 1089-1098. https://doi.org/10.1111/medu.12257

Van Dun, P. L., Nicolaie, M. A., & Van Messem, A. (2016). State of affairs of osteopathy in the Benelux: Benelux Osteosurvey 2013. *International Journal of Osteopathic Medicine, 20*, 3-17. https://doi.org/10.1016/j.ijosm.2016.01.003

Vanek, E. P., Snyder, C. W., Hull, A. L., & Hekelman, F. P. (1996). The relationship between teachers' confidence and use of clinical teaching skills in ambulatory care settings. *Teaching and Learning in Medicine: An International Journal, 8*(3), 137-141. https://doi.org/10.1080/10401339609539783

Vaughan, B., Fitzgerald, K., Fleischmann, M., & Mulcahy, J. (2020a). Determinants of health, health behaviours and demographic profile of patients attending an

Australian university student-led osteopathy clinic. *Chiropractic & Manual Therapies, 28*(1), 2. https://doi.org/10.1186/s12998-019-0292-5

Vaughan, B., Fleischmann, M., Fitzgerald, K., Grace, S., McLaughlin, P., Jolly, B., & Trumble, S. (2020b). Profile of an allied health clinical supervision workforce: results from a nationally representative Australian practice-based research network. *Health Professions Education, 6*(3), 376-385. https://doi.org/10.1016/j.hpe.2020.04.008

Vaughan, B., Macfarlane, C., & Florentine, P. (2014a). Clinical education in the osteopathy program at Victoria University. *International Journal of Osteopathic Medicine, 17*(3), 199-205. https://doi.org/10.1016/j.ijosm.2013.10.010

Vaughan, B., & Moore, K. (2015). The mini Clinical Evaluation Exercise (mini-CEX) in a pre-registration osteopathy program: Exploring aspects of its validity. *International Journal of Osteopathic Medicine, 19*, 61-72. https://doi.org/10.1016/j.ijosm.2015.07.002

Vaughan, B., Morrison, T., Buttigieg, D., Macfarlane, C., & Fryer, G. (2014b). Approach to low back pain-osteopathy. *Australian Family Physician, 43*(4), 197-198.

Wayne, J., Bogo, M., & Raskin, M. (2010). Field education as the signature pedagogy of social work education. *Journal of Social Work Education, 46*(3), 327-339. https://doi.org/10.5175/JSWE.2010.200900043

Wilkinson, J., Thomas, K., Freeman, J., & McKenna, B. (2015). Day-to-day practice of osteopaths using osteopathy in the cranial field, who are affiliated with the Sutherland Cranial College of Osteopathy (SCCO): A national survey by means of a standardised data collection tool. *International Journal of Osteopathic Medicine, 18*(1), 13-21. https://doi.org/10.1016/j.ijosm.2014.04.008

World Health Organization. (2010). *Benchmarks for training in osteopathy*. World Health Organisation,. Retrieved July 20, 2013, from http://www.who.int/medicines/areas/traditional/BenchmarksforTraininginOsteopathy.pdf

Worthington, A. C. (2002). The impact of student perceptions and characteristics on teaching evaluations: a case study in finance education. *Assessment &*

Evaluation in Higher Education, 27(1), 49-64. https://doi.org/10.1080/02602930120105054

Wright, S. M., Kern, D. E., Kolodner, K., Howard, D. M., & Brancati, F. L. (1998). Attributes of excellent attending-physician role models. *New England Journal of Medicine, 339*(27), 1986-1993. https://doi.org/10.1056/NEJM199812313392706

Yew, E. H. J., & Yong, J. J. Y. (2014). Student perceptions of facilitators' social congruence, use of expertise and cognitive congruence in problem-based learning. *Instructional Science, 42*(5), 795-815. https://doi.org/10.1007/s11251-013-9306-1

Young, K., Joines, J., Standish, T., & Gallagher, V. (2019). Student evaluations of teaching: the impact of faculty procedures on response rates. *Assessment & Evaluation in Higher Education, 44*(1), 37-49. https://doi.org/10.1080/02602938.2018.1467878

Yuan, D., Bridges, M., D'Amico, F. J., & Wilson, S. A. (2014). The effect of medical student feedback about resident teaching on resident teaching identity: a randomized controlled trial. *Family Medicine, 46*(1), 49-54. https://www.stfm.org/familymedicine/vol46issue1/Yuan49

Zabaleta, F. (2007). The use and misuse of student evaluations of teaching. *Teaching in Higher Education, 12*(1), 55-76. https://doi.org/10.1080/13562510601102131

Zabar, S., Hanley, K., Stevens, D. L., Kalet, A., Schwartz, M. D., Pearlman, E., Brenner, J., Kachur, E. K., & Lipkin, M. (2004). Measuring the competence of residents as teachers. *Journal of General Internal Medicine, 19*(5), 530-533. https://doi.org/10.1111/j.1525-1497.2004.30219.x

Zibrowski, E., Crukley, J., Malett, J., & Myers, K. (2016). To submit or not submit: The burden of evaluation on postgraduate medical trainees. *MedEdPublish, 5*. https://doi.org/10.15694/mep.2016.000058

Zumbo, B. D., & Hubley, A. M. (2016). Bringing consequences and side effects of testing and assessment to the foreground. *Assessment in Education: Principles, Policy & Practice, 23*(2), 299-303. https://doi.org/10.1080/0969594X.2016.1141169

CHAPTER 2

Developing a clinical teaching quality questionnaire for use in a university osteopathic pre-registration teaching program

The previous chapter has provided an overview of clinical education, clinical teaching, the education of osteopaths, and the challenges of evaluating clinical teaching. Chapter 2 describes the identification of items for a measure to evaluate the quality of clinical teaching in student-led osteopathy clinical learning environments. The items are subsequently tested with a cohort of osteopathy students and analysed using exploratory factor analysis. The study resulted in the development of a measure of clinical teaching for osteopathy student-led clinical learning environments.

Vaughan *BMC Medical Education* (2015) 15:70
DOI 10.1186/s12909-015-0358-6

BMC Medical Education

RESEARCH ARTICLE Open Access

Developing a clinical teaching quality questionnaire for use in a university osteopathic pre-registration teaching program

Brett Vaughan[1,2,3]

Abstract

Background: Clinical education is an important component of many health professional training programs. There is a range of questionnaires to assess the quality of the clinical educator however none are in student-led clinic environments. The present study developed a questionnaire to assess the quality of the clinical educators in the osteopathy program at Victoria University.

Methods: A systematic search of the literature was used to identify questionnaires that evaluated the quality of clinical teaching. Eighty-three items were extracted and reviewed for their appropriateness to include in a questionnaire by students, clinical educators and academics. A fifty-six item questionnaire was then trialled with osteopathy students. A variety of statistics were used to determine the number of factors to extract. Exploratory factor analysis (EFA) was used to investigate the factor structure.

Results: The number of factors to extract was calculated to be between 3 and 6. Review of the factor structures suggested the most appropriate fit was four and five factors. The EFA of the four-factor solution collapsed into three factors. The five-factor solution demonstrated the most stable structure. Internal consistency of the five-factor solution was greater than 0.70.

Conclusions: The five factors were labelled Learning Environment (Factor 1), Reflective Practice (Factor 2), Feedback (Factor 3) and Patient Management (Factor 4) and Modelling (Factor 5). Further research is now required to continue investigating the construct validity and reliability of the questionnaire.

Keywords: Evaluation, Exploratory factor analysis, Student-led clinic, Student-run clinic, Clinical education, Osteopathy, Osteopathic medicine

Background

Clinical education is an important component of health profession education programs, as it provides an opportunity for students to apply the skills and knowledge they have learnt in the classroom in an 'authentic' learning environment [1-3]. Clinical education usually takes the form of student management of patients under the supervision of related qualified health professionals with placement type influencing the volume and type of teaching and/or supervision [4], the type of health care provided, and degree of student involvement in health care events.

Authors have described the educational theories that may be applied to clinical education and these typically focus on those that related to workplace learning [5-7]. Although there has been no explicit discussion of the theories underlying osteopathic clinical education, Vaughan et al. [8] suggest that the Cognitive Apprenticeship model could account for aspects of the learning and student-educator interaction that takes place in the on-campus, student-led clinics. Beyond the commentary by these authors, we must explore the wider health profession education literature in order to draw on other theories. The profession with the most similarities from an education

Correspondence: brett.vaughan@vu.edu.au
[1]Centre for Chronic Disease Prevention & Management, College of Health & Biomedicine, Victoria University, Melbourne, Australia
[2]Institute of Sport, Exercise & Active Living, Victoria University, Melbourne, Australia
[3] School of Health & Human Sciences, Southern Cross University, Lismore, Australia

BioMed Central

© 2015 Vaughan; licensee BioMed Central. This is an Open Access article distributed under the terms of the Creative Commons Attribution License (http://creativecommons.org/licenses/by/4.0), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly credited. The Creative Commons Public Domain Dedication waiver (http://creativecommons.org/publicdomain/zero/1.0/) applies to the data made available in this article, unless otherwise stated.

and professional practice viewpoint is physiotherapy. Patton et al. [5] highlight there has been little in the way of literature published on the theories that underpin clinical education in physiotherapy. The subsequent commentary by these authors suggests, "…that one model or specification could address the needs of every situation would be contestable." These authors describe workplace learning, learning as practice, social learning, situated learning and reflective/critical thinking as models that can be applied to clinical education in physiotherapy. It is likely that these models are also applicable to osteopathic clinical education and readers are encouraged to review the work by Patton et al. [5] for a comprehensive description of these models.

Teaching in a clinical environment is complex [3,9,10]. It includes issues related to patients such as safety and patient census - the availability and variety of patients and illnesses, clinic operational issues such as timetables and facilities, issues related to students such as time management [3], and individual characteristics issues related to the clinical educator such as personality [4,11] and education. Cross [12] reported that students perceived there was a strong relationship between being a good physiotherapist and a good clinical teacher, however this does not appear to be a consistent theme that emerges from the literature nor is there strong evidence that this relationship improves student learning.

We know that clinical educators require clinical competency [2,13]; good clinical reasoning skills [2,12]; appropriate, relevant and up-to-date knowledge [14]; good interpersonal skills; and supervision and teaching skills [12,14-18]. These attributes also extend to the provision of timely student feedback [14,18-23], regular observation of students [10], role-modelling [10,12,17,24-29] and the development of a positive, professional and supportive learning environment [15,22,30,31]. Sutkin et al. [15] provide an extensive list of characteristics of a 'good' clinical educator based on their systematic search of the literature.

Although the list of clinical educator attributes is extensive, there is no research that consistently demonstrates which attributes contribute to effective student learning [32] and further research is required [14]. Students have an opinion and expectation as to what constitutes a good clinical educator [10]. That said, arguably, one of the most effective ways of determining the impact of clinical educator attributes on students learning is to explore students perspectives [14].

Assessing the teaching quality is one part of a course evaluation strategy used to help inform the quality cycle necessary for review, improvement and program accreditation. Student ratings are already widely used to explore the quality of clinical teaching [33,34]. For that reason, there are a large and growing number of clinical teaching quality questionnaires in the literature with systematic reviews of available questionnaires by Fluit et al. [35] and Beckman et al. [33]. As with any performance measure, the validity, reliability and feasibility of a questionnaire are important to investigate and establish [14,36], particularly where the results of the questionnaire are used for employment decisions or performance appraisals. Ideally questionnaires should be convenient for the student to complete with the results providing motivation for clinical educators to continue to improve their teaching [22].

Clinical teaching in osteopathic education outside of the United States typically takes place in out-patient or on-campus clinics where students manage and treat patients under the supervision of qualified osteopaths (the osteopathic clinical educator). Senior students in the osteopathy program take on the responsibility of conducting the entire patient consultation including taking a clinical/medical history, physical examination, manual therapy treatment, and provision of advice related to exercise and lifestyle factors as part of the management of the patients' presenting complaint. Supervision of the student is provided by qualified osteopaths in a ratio of 1 educator for every 5–6 students. This ratio is different to other professions such as physiotherapy [37,38] and occupational therapy [39] where 1:1 or 1:2 ratios are common. The role of the osteopathic clinical educator is fourfold: 1) to ensure that the student is performing a safe and effective consultation; 2) support the student through the experience of managing patients with a variety of musculoskeletal and concomitant psychosocial issues; 3) encourage the student to reflect on their patient management; and 4) propose alternative patient management strategies. These roles are consistent with idea of 'supported participation' as a model for learning in a clinical environment as described by Dornan et al. [6]. The role may also occasionally require the clinical educator to perform aspects of the examination or treatment, and this provides a limited opportunity to role model patient management skills. Literature regarding osteopathic clinical education and clinical educators in Australia, New Zealand and the United Kingdom is beginning to emerge [8]. However, research into osteopathic clinical education is required. Further, there is a clear need to investigate the students' perception of the quality of clinical teaching in an osteopathic student-led teaching clinic. The current paper reports on the development of a questionnaire to assess clinical teaching in osteopathic clinical education in on-campus, student-led clinics.

A number of authors [9,40] contend that questionnaires should be specific for the environment in which the clinical teaching is taking place. Therefore, when exploring on-campus, student-led clinics the use of

previously developed validated questionnaires, particularly those developed for in-patient or ambulatory environments, are considered unsuitable. The current paper reports on the development of a purposefully designed questionnaire to evaluate clinical teaching in an on-campus, student-led osteopathic teaching clinics at one Australian university.

Methods

The current study is the first in a series of studies using Kane's validity perspective [41] to develop a fit for purpose evaluation tool, to identify clinical educators knowledge, skills and abilities by students in a student-led, on campus ambulatory clinic. The current study sought to begin developing the validity argument for the evaluation tool. Kane [41] contends that it is not possible for a measurement in itself to be 'valid' however it is possible to develop and mount an argument that the score itself is 'valid' based on multiple sources of evidence [42]. Cook [43] defines this as "…degree to which the interpretations of scores resulting from an assessment [measurement] activity are 'well grounded or justifiable'". Kane's validity perspective was used as the framework for the current study and sought to provide evidence for the 'observation' to 'target domain' components of the argument (Figure 1). The study was undertaken in four phases and was approved by the Victoria University Human Research Ethics Committee.

Phase 1 - literature review

The questionnaires identified in the systematic reviews by Fluit [35] and Beckman et al. [33] were retrieved in the first instance. To ensure the literature review for the current study was up-to-date, a further search, using the search terms outlined by Fluit [35] was undertaken from the end of the Fluit [35] review (end of March 2010) to 1st January 2013. Medline and CINAHL were searched as per the Fluit [35] review and English language studies only were retrieved. Articles were retrieved where the title and/or abstract suggested the development or validation of a measure of clinical teaching quality. An overview of the search is presented in Figure 2 and Additional file 1.

Questionnaires identified from both systematic reviews, and those located during the updated search were independently reviewed by the author and an academic colleague. Items from each of the questionnaires were extracted where they were deemed to be relevant to a questionnaire evaluating clinical teaching quality in a student-led, on campus clinic. Where there was disagreement between the authors, a consensus was reached. The authors agreed on the extraction of eighty-three (83) items from twenty-three (23) questionnaires.

Phase 2 – item review

Utilising the process employed by Roff et al. [44] to develop their clinical teaching questionnaire, 5 osteopathic clinical educators and academics, and 3 students in the VU osteopathy program responded to an invitation to review each of the 83 items. Using a 5-point Likert scale - 1 (strongly disagree) to 5 (strongly agree), the respondents were asked to rate whether the item should be included in a questionnaire about osteopathic student-led, on-campus teaching clinics. Once the respondents had completed their review of the items, the author (BV) collated the responses. Items where 6 or more of the respondents provided a rating of 4 or 5 on the Likert Scale were retained. This provided a list of 56 items and 2 global rating items. An additional global item was suggested by one of the academics and was subsequently included in the Phase 3 questionnaire. This third global rating is similar to that used in the patient satisfaction literature where a patient would recommend the particular facility to another person [45]. On the draft clinical teaching quality questionnaire, the student is asked whether they would

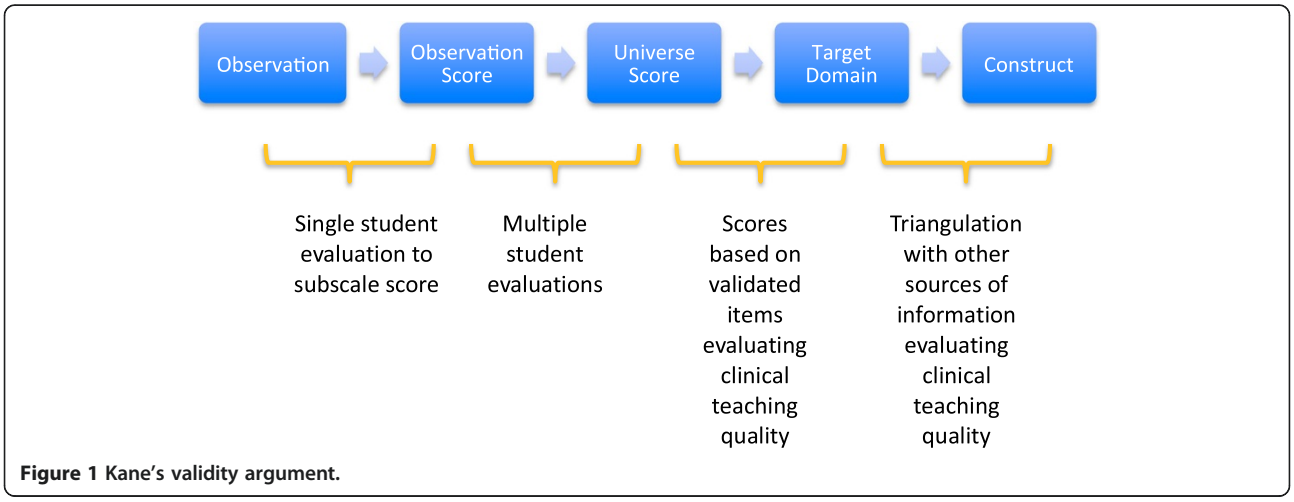


Figure 1 Kane's validity argument.

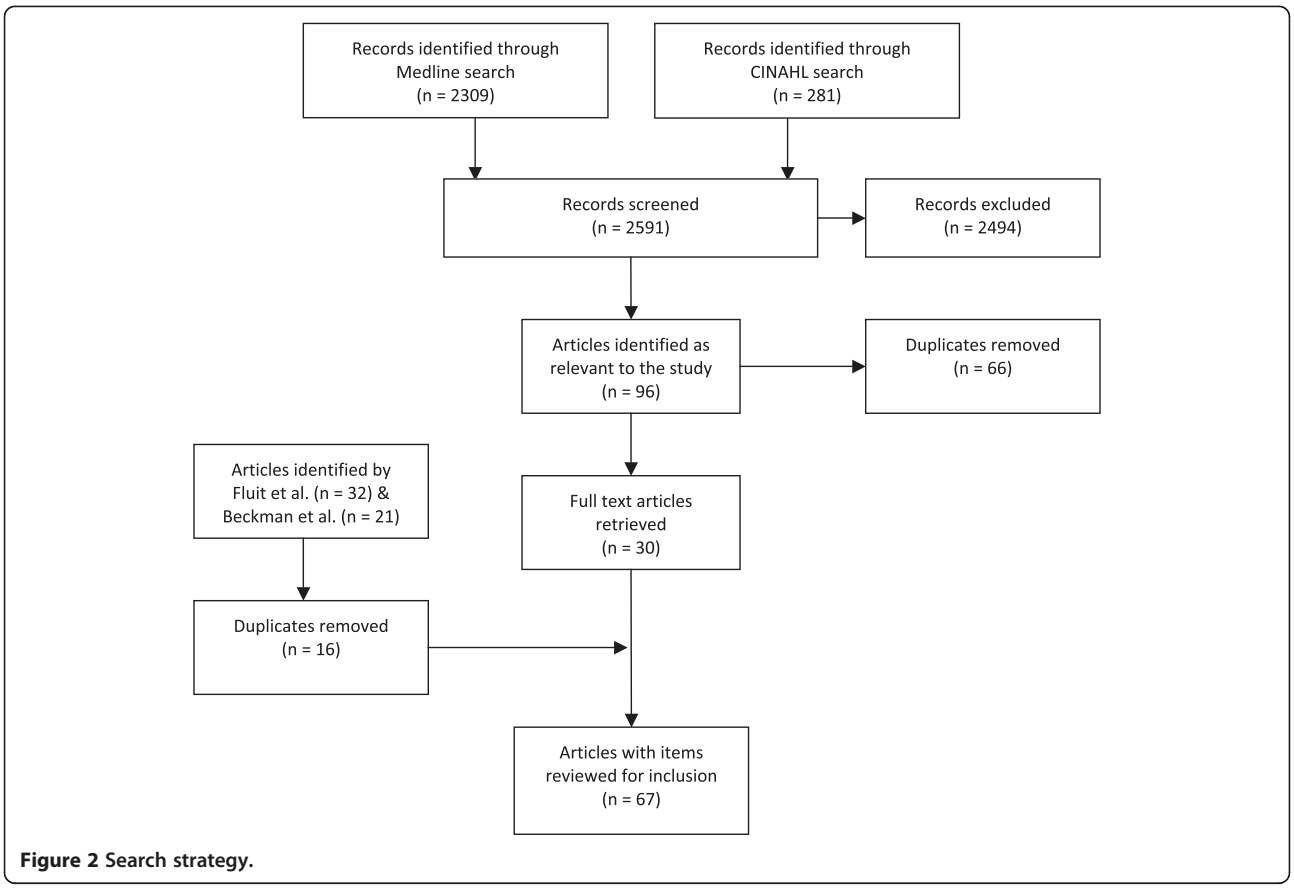


Figure 2 Search strategy.

recommend the clinical educator to other students and provides the contrast to the first global rating asking the student whether they would work with the clinical educator in the future.

Phase 3 - questionnaire pilot testing

The draft clinical teaching quality questionnaire was distributed as a paper-based questionnaire to all students in year 4 and 5 of the VU osteopathy program. Each student was asked to complete the questionnaire and rate two of their clinical educators who had supervised them over semester 1, 2013 (March 2013 - May 2013). The students were asked to name the clinical educator on the survey, however they were not required to identify themselves. Each item in the draft clinical teaching quality questionnaire was rated on a 5-point Likert scale: 5 being 'strongly agree' and 1 being 'strongly disagree'. Previous research suggests between 4 and 7 ordinal responses is best [46], as it may allow for neutral responses, and a sufficient range of responses to each item [47]. Student responses to the questionnaire were made available to each of the clinical educators who received a rating(s) from any student. The results were used for feedback purposes only and were not used as a basis for employment decisions or reward.

Phase 4 - data analysis

Data from each completed questionnaire were entered into Microsoft Excel for quantitative analysis. Many of the questionnaires used to evaluate the quality of clinical teaching that have been published in the literature, and where the items for the current questionnaire were extracted from, have used a principal components analysis (PCA). Numerous authors have discussed the pros and cons of using a PCA [48,49], and it is now accepted that it is more appropriate to use an exploratory factor analysis (EFA) over a PCA [48], particularly where confirmatory factor analysis is to be used in the future [49,50].

There is also a move away from the use of Pearson correlations with EFAs to polychoric correlations. The polychoric correlation is more appropriate for ordinal data as Pearson correlations assume that the data has been measured on an interval scale [50-52]. Determining the number of factors to extract is traditionally based on the K1 criteria (eigenvalues greater than 1) and visual inspection of the Scree plot are both problematic - K1 has a tendency to overestimate the number of factors to be extracted [53]. Authors are now reporting the use of other methods to determine the number of factors to extract including PA [54], Velicer's MAP [55,56], VSS [57],

OC and AF [58], although such techniques have existed in the literature for many years. Of these methods, the most accurate is PA using the polychoric matrix [59]. Readers are directed towards other authors for further discussion of the factor extraction methods [48,52,59].

The exploratory factor analysis (EFA) was conducted with R [60] using the *psych* [61], *GPArotation* [62], *polycor* [63] and *nFactors* [64] packages. Data were screened and determined to be non-normally distributed. Initially a polychoric correlation matrix was generated. Polychoric correlations are more appropriate than Pearson correlations for ordinal data as they are based on the concept that the ordinal categories are bivariate normal [59].

Multiple methods were employed to determine the number of factors to extract. Parallel analysis (PA) [54], mean average partial (MAP) [55,56], eigenvalues, Very Simple Structure (VSS) [57], acceleration factor (AF) [58] and optimal coordinate (OC) [58] were all undertaken, each using the previously generated polychoric correlations. Both PA and OC have been reported to provide similar results, albeit using Pearson correlations [53].

An EFA was performed on the polychoric correlation [52] using the ordinary least squares (OLS) extraction method [48]. The questionnaire data were not normally distributed and ordinal in nature therefore the OLS extraction method should be used with the polychoric matrix [48]. Further, two rotation criteria were employed as the choice of criteria may produce different results [65,66]. Orthogonal rotations (i.e. Varimax) are commonly employed and assume that there is no correlation between the factors extracted [48]. Conversely, where the factors are expected to correlate (as in the present study) an oblique rotation is more appropriate [48]. The Geomin and Oblimin rotations were selected in the present study to reduce the cross-loadings between factors [48], and anticipating that each factor would correlate with the others. Items were retained if they loaded greater than 0.45 on a factor [67,68], had a communality of greater than 0.6 [69], and demonstrated a cross-loading of less than 0.32 [68]. After an item was removed, the EFA was conducted again (iteration) [68]. The Kaiser-Myer-Olkin (KMO) statistic and Bartlett's test of sphericity were also calculated to determine factorability of the data.

Once the factor analysis was completed, descriptive statistics were generated for each retained item, and internal consistency of each of the factors was calculated using ordinal reliability alpha [70]. Ordinal reliability alpha is the most appropriate internal consistency statistic for ordinal data as it uses the polychoric correlation rather than the Pearson correlation [70,71]. Descriptive statistics were also generated for the three global ratings items. Descriptive statistics for the total questionnaire score and internal consistency for the whole questionnaire were not calculated as dimensionality of the questionnaire was not assessed. Dimensionality of the questionnaire will be the subject of future research.

Results

One hundred and seventy two ratings of all 27 clinical educators employed at the time of study were received. All clinical educators received more than one rating. Data were incomplete on one questionnaire and was subsequently removed from the analysis; 171 questionnaires were analysed. The results of the PA, MAP, VSS, eigenvalue, OC and AF are presented in Figures 3 and 4. The MAP suggested extracting two factors and the VSS suggested extracting four. OLS factor analyses were conducted extracting between 3–6 factors in order to identify an appropriate structure, consistent with recommendations from previous authors [59]. Eight analyses were conducted; four using the Geomin rotation and four using the Oblimin rotation. Extracting four and five factors using the Oblimin rotation provided the most appropriate solutions for (Table 1).

Four factor solution

The 4-factor solution initially demonstrated minimal cross-loadings and slightly lower communalities than the 5 factor solution. KMO was 0.6 and Bartlett's test was $p < 0.01$ ($\chi^2 = 3468.40$) indicating a minimally-acceptable level of factorability. Twenty-four iterations were performed; the final solution collapsed into a 3-factor structure containing 19 items explaining 77% of the variance (Table 2). The alpha scores were high (0.93 or greater) and the correlations between the three factors were 0.57 or higher.

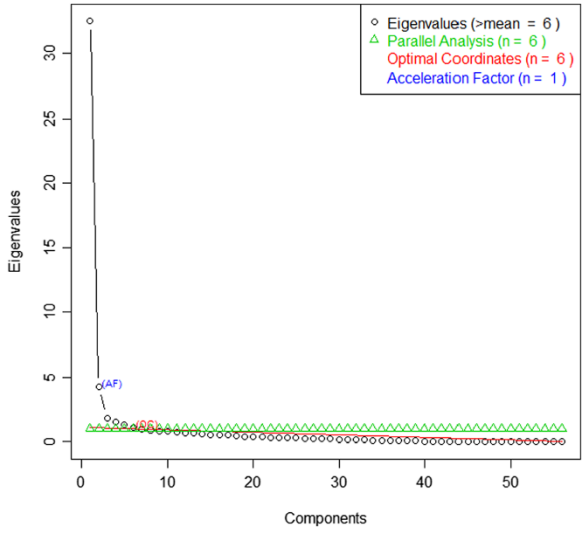


Figure 3 Number of factors to extract (part 1).

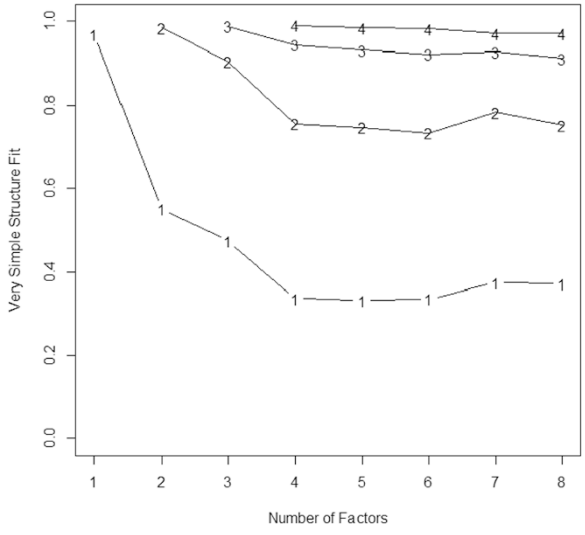


Figure 4 Number of factors (part 2). VSS plots the goodness of fit statistic as a function of the number of factors to extract. The number of factors to extract is demonstrated when the goodness of fit value no longer changes. In the above graph, the VSS fit statistic does not change when extracting four factors.

Retained items loaded greater than 0.54 on a factor and had communalities (h^2) of greater than 0.63.

The descriptive statistics for each of the items in the 3-factor solution are presented in Table 3.

Five factor solution

The 5-factor solution initially demonstrated minimal cross-loadings and higher communalities than the 4-factor solution. KMO was 0.79 and Bartlett's test was $p < 0.001$ ($\chi^2 = 31046.67$) indicating acceptable factorability. The 5-factor structure was maintained after 18 iterations and contained 30 items explaining 80% of the variance (Table 4). Retained items loaded greater than 0.51 on a factor with communalities (h^2) greater than 0.67. The alpha scores for each factor were above the acceptable level of 0.70 and correlations between the factors were greater than 0.24.

The descriptive statistics for each of the items in the 5-factor solution are presented in Table 5.

Global ratings

The descriptive statistics for the three global rating items are presented in Table 6.

Discussion

The aim of the current paper was to develop a questionnaire to evaluate the quality of clinical teaching in an osteopathic student-led, on-campus teaching clinic at one Australian university. A systematic search of the clinical teaching evaluation literature identified questionnaires from which 83 possible items for inclusion on the new questionnaire were extracted. The extraction of these items was based on their perceived applicability to a student-led, on-campus teaching clinic environment. Items were drawn from published questionnaires designed for a range of clinical teaching environments. No questionnaire assessing clinical teaching quality in student-led clinics or ambulatory, on-campus clinics was identified. As a questionnaire should be designed for the environment in which it is to be used [9,40], drawing on these previously published items is appropriate for developing the questionnaire for a student-led, on-campus clinic. By employing the same method as Roff et al. [44], osteopathy academics, clinical educators and students refined the list of 83 items to 56 items. During this process items that related to the conduct of formative and summative assessments by the clinical educators were removed as the content reviewers felt this was not a role that the students could provide constructive ratings for nor were the assessments a major role undertaken by the clinical educators. The resulting draft clinical teaching quality questionnaire was then completed by the cohort of osteopathy students who were managing patients in the student-led clinic at the time of the study.

Table 1 Factor solution choice

Number of factors	Extraction	Rotation	Comment
3	OLS	Geomin	12 items with communalities less than 0.6, removing these items would also remove one factor
4	OLS	Geomin	9 items with communalities less than 0.6, strong cross-loadings
5	OLS	Geomin	One item on factor 5, and strong cross-loadings between factors 1 and 4
6	OLS	Geomin	Single items on factors 5 and 6, factor 6 item cross-loads with factor 1
3	OLS	Oblimin	Failed to converge
4	OLS	Oblimin	More than 5 items on each factor, minimal number of cross-loading items, logical items grouping for each factor
5	OLS	Oblimin	Logical item grouping for each factor
6	OLS	Oblimin	Failed to converge

OLS – ordinary least squares.

Table 2 3 factor solution

Item	F1	F2	F3	h^2
Treated me with respect	1.01			0.96
Maintained a positive attitude towards me	1.01			0.91
Fostered an environment of respect in which I felt comfortable participating	0.85			0.95
Showed genuine concern for my professional well-being	0.76			0.75
Established a good learning environment (approachable, focused, nonthreatening, professional and enthusiastic	0.76			0.87
Had reasonable expectations of students	0.75			0.69
Demonstrated humanistic attitudes in relating to patients (integrity, compassion and respect)	0.7			0.66
Has good communication skills	0.63			0.68
Gave me the opportunity to offer opinions on patient problems or treatment	0.54			0.69
Gave timely feedback to me		0.92		0.79
Gave me regular, useful feedback about my knowledge and performance		0.88		0.87
Offered me suggestions for improvement when required		0.87		0.86
Identified my strengths		0.66		0.63
Explained to me why I was correct or incorrect		0.64		0.77
Encouraged me to think			0.81	0.78
Asked questions that promote learning (clarifies, probes, reflective questions etc.)			0.8	0.74
Encouraged questions and active participation			0.79	0.87
Stimulates me to learn independently			0.79	0.73
Asked questions to enhance my learning			0.68	0.72
Variance explained	35%	22%	20%	
Internal consistency	0.96	0.93	0.94	

Table 3 Descriptive statistics for 3-factor solution

Item descriptives	Mean	SD	Median	Min	Max	Range
Treated me with respect	4.4	0.97	5	1	5	4
Maintained a positive attitude towards me	4.37	0.92	5	1	5	4
Fostered an environment of respect in which I felt comfortable participating	4.16	1.1	5	1	5	4
Showed genuine concern for my professional well-being	3.24	0.89	3	1	4	3
Established a good learning environment (approachable, focused, nonthreatening, professional and enthusiastic	4.19	1.07	5	1	5	4
Had reasonable expectations of students	3.23	0.87	3	1	4	3
Demonstrated humanistic attitudes in relating to patients (integrity, compassion and respect)	4.41	0.81	5	1	5	4
Has good communication skills	4.19	1.02	5	1	5	4
Gave me the opportunity to offer opinions on patient problems or treatment	4.34	0.89	5	1	5	4
Gave timely feedback to me	3.95	1.03	4	1	5	4
Gave me regular, useful feedback about my knowledge and performance	3.79	1.03	4	1	5	4
Offered me suggestions for improvement when required	4.14	0.89	4	1	5	4
Identified my strengths	3.72	1.04	4	1	5	4
Explained to me why I was correct or incorrect	3.09	0.93	3	1	4	3
Encouraged me to think	4.24	0.86	4	1	5	4
Asked questions that promote learning (clarifies, probes, reflective questions etc.)	4.12	0.93	4	1	5	4
Encouraged questions and active participation	4.17	0.88	4	1	5	4
Stimulates me to learn independently	4.03	0.9	4	1	5	4
Asked questions to enhance my learning	4.18	0.93	4	1	5	4

Table 4 5 factor solution

Item	F1	F2	F3	F4	F5	h^2
Treated me with respect	0.99					0.93
Maintained a positive attitude towards me	0.97					0.90
Fostered an environment of respect in which I felt comfortable participating	0.88					0.90
Established a good learning environment (approachable, focused, nonthreatening, professional and enthusiastic)	0.85					0.88
Demonstrated humanistic attitudes in relating to patients (integrity, compassion and respect)	0.80					0.73
Was approachable for discussion	0.78					0.87
Showed genuine concern for my professional well-being	0.77					0.79
Had reasonable expectations of students	0.76					0.73
Has good communication skills	0.69					0.77
Is open to student questions and alternative approaches to patient management	0.68					0.75
Gave me the opportunity to offer opinions on patient problems or treatment	0.59					0.72
Adjusted teaching to my needs (experience, competence, interest)	0.52					0.71
Is an effective clinical teacher	0.51					0.87
Encouraged me to think		0.86				0.84
Promoted reflection on clinical practice		0.83				0.67
Emphasises a problem-solving approach rather than solutions		0.78				0.68
Asked questions that promote learning (clarifies, probes, reflective questions etc.)		0.75				0.77
Asked questions to enhance my learning		0.70				0.75
Encouraged questions and active participation		0.65				0.82
Stimulates me to learn independently		0.62				0.70
Gave timely feedback to me			0.89			0.85
Gave me regular, useful feedback about my knowledge and performance			0.77			0.86
Offered me suggestions for improvement when required			0.76			0.86
Identified areas needing improvement			0.69			0.71
Identified my strengths			0.65			0.67
Explained to me why I was correct or incorrect			0.58			0.80
Promoted keeping of medical records in a way that is thorough, legible, efficient and organised				0.96		0.92
Encouraged me to assume responsibility for patient care				0.76		0.84
Demonstrates knowledge of current medical and manual therapy literature					0.89	0.85
Demonstrated osteopathic, clinical examination and rehabilitation knowledge and skill(s)					0.57	0.74
Variance explained	39%	24%	19%	9%	8%	
Internal consistency	0.97	0.94	0.93	0.82	0.73	

To determine the most appropriate items and factor structure for a questionnaire to assess the quality of clinical teaching in the student-led on-campus clinic, results from the 56-item questionnaire were analysed with an EFA. Multiple methods were used to determine the number of factors to extract withthe PA, MAP, and VSS all suggesting different numbers of factors be extracted. Where the number of factors to be extracted differs between methods, Courtney and Gordon [59] suggest that multiple factors be extracted guided by the results of the PA and MAP. In the present study PA suggested extracting six factors and MAP four factors. To maximise the ability to identify an appropriate factor structure, three, four, five and six factors were extracted using both the Geomin and Oblimin oblique rotations. A number of analyses were undertaken with the most appropriate factor structures being a 4-factor and 5-factor solution. The VSS suggested extracting four factors however none of the methods suggested extracting five factors. This outcome supports the assertion of Courtney and Gordon [59] that a multiple numbers of factors should be extracted where the different methods are not in agreement.

Further analysis of the 4 and 5 factor solutions, including item removal based on multiple criteria, produced a 3-factor and 5-factor structure respectively. It is of note that the 4-factor solution collapsed to a 3-factor solution

Table 5 Descriptive statistics for the 5-factor solution

Item	Mean	SD	Median	Min	Max	Range
Treated me with respect	4.40	0.97	5	1	5	4
Maintained a positive attitude towards me	4.37	0.92	5	1	5	4
Fostered an environment of respect in which I felt comfortable participating	4.16	1.10	5	1	5	4
Established a good learning environment (approachable, focused, nonthreatening, professional and enthusiastic)	4.19	1.07	5	1	5	4
Demonstrated humanistic attitudes in relating to patients (integrity, compassion and respect)	4.41	0.81	5	1	5	4
Was approachable for discussion	4.38	0.94	5	1	5	4
Showed genuine concern for my professional well-being	3.24	0.89	3	1	4	3
Had reasonable expectations of students	3.23	0.87	3	1	4	3
Has good communication skills	4.19	1.02	5	1	5	4
Is open to student questions and alternative approaches to patient management	4.21	0.99	5	1	5	4
Gave me the opportunity to offer opinions on patient problems or treatment	4.34	0.89	5	1	5	4
Adjusted teaching to my needs (experience, competence, interest)	4.05	0.97	4	1	5	4
Is an effective clinical teacher	4.23	1.04	5	1	5	4
Encouraged me to think	4.24	0.86	4	1	5	4
Promoted reflection on clinical practice	4.19	0.85	4	1	5	4
Emphasises a problem-solving approach rather than solutions	4.11	0.93	4	1	5	4
Asked questions that promote learning (clarifies, probes, reflective questions etc.)	4.12	0.93	4	1	5	4
Asked questions to enhance my learning	4.18	0.93	4	1	5	4
Encouraged questions and active participation	4.17	0.88	4	1	5	4
Stimulates me to learn independently	4.03	0.9	4	1	5	4
Gave timely feedback to me	3.95	1.03	4	1	5	4
Gave me regular, useful feedback about my knowledge and performance	3.79	1.03	4	1	5	4
Offered me suggestions for improvement when required	4.14	0.89	4	1	5	4
Identified areas needing improvement	3.93	0.95	4	1	5	4
Identified my strengths	3.72	1.04	4	1	5	4
Explained to me why I was correct or incorrect	3.09	0.93	3	1	4	3
Promoted keeping of medical records in a way that is thorough, legible, efficient and organised	4.20	0.82	4	1	5	4
Encouraged me to assume responsibility for patient care	3.39	0.69	3.5	1	4	3
Demonstrates knowledge of current medical and manual therapy literature	4.23	0.86	4	1	5	4
Demonstrated osteopathic, clinical examination and rehabilitation knowledge and skill(s)	4.39	0.82	5	1	5	4

due to the removal of some of the items. The initial factorability of the 4-factor solution was minimally acceptable, and may have been an indicator as to the potential for the factor structure to collapse during the analysis. The decision was made to use the 5-factor, 30-item questionnaire as it displayed characteristics of previously validated questionnaires [27,72], and also incorporated modelling behaviours (e.g. interacting with patients and professional practice) that were not being examined by items remaining in the 3-factor structure. The 5-factor,

Table 6 Descriptive statistics for the total score and global ratings

	I would do more clinics with this Clinical Educator	Rate the overall effectiveness of this Clinical Educator as an educator/supervisor	I would recommend other students to work with this Clinical Educator
Mean	4.09	4.10	4.09
SD	1.15	0.98	1.12
Median	4.50	4.00	4.00
Minimum	1	1	1
Maximum	5	5	5

30-item questionnaire was called the Osteopathy Clinical Teaching Questionnaire (OCTQ). The five factors identified in the present study were labelled: Learning Environment (Factor 1), Reflective Practice (Factor 2), Feedback (Factor 3) and Patient Management (Factor 4) and Modelling (Factor 5).

Factor 1 - learning environment

The clinical learning environment is a confluence of factors. This includes those listed previously (i.e. patient census) as well as system-based considerations such as the requirements of accrediting bodies, university requirements (i.e. graduate attributes), operational issues (i.e. physical clinic environment, clinic operating procedures), and interpersonal issues (i.e. patients, administrative staff, peers, clinical educators) that students must learn to cope with and manage. These system-based influences expose the student to issues that they will experience in the workplace through their training program or upon graduation. Managing such influences is part of becoming a capable health professional [73].

Griffith III et al. [74] indicate that the learning environment, managed and/or facilitated by the clinical educator, improves student learning more than the clinical educator imparting information. Furthermore the learning environment effects the overall judgement of the clinical teaching as rated by a student [1,27,30,31,72]. Boerboom et al. [72] noted that over 20% of the variance in the Maastricht Clinical Teaching Questionnaire scores was accounted for by the modelling, coaching and learning climate domains – factors directly related to the clinical teacher and the learning environment. In the present study, Learning Environment (Factor 1) was the strongest factor and accounted for just over a third of the variance in the data (39%). This factor also demonstrated a strong ordinal alpha value indicating that it is internally consistent, and contained items that are measuring a similar construct.

Students feel that a positive relationship with the clinical educator contributes to a favourable learning environment [75] and this is reflected in a higher score on the OCTQ. The potential for this relationship or 'halo effect' to influence item responses should be investigated further. Respect is also a key component of this factor. Whilst the focus of this factor was on the interaction between clinical educator and student, it also addressed the interaction between the clinical educator and patient (item 5). This interaction is an important part of the role of an effective clinical educator [10] and provides an opportunity for the educator to role model patient communication and management skills.

Factor 2 - reflective practice

This factor addressed a range of areas including reflective skills (items 14, 15 & 17), and the use of questioning to promote learning (item 17, 18 & 19). A number of items on the OCTQ address reflective practice and reinforce the importance of the clinical educator stimulating self-directed learning. Litzelman et al. [76] have reported a strong positive correlation between clinical educators who stimulate self-directed learning and higher clinical teaching quality ratings. These authors [76] suggest that such a relationship is an indication of the clinical educators knowledge. However, items that require the student to actively participate in the clinical education process, *i.e. item 14 - Encouraged me to think* may not be a true reflection of the quality of clinical education provided by the clinical educator. These items are potentially susceptible to differences in students willingness to engage with the clinical educator rather than differences in approaches clinical educators use to stimulate thinking [72]. Given the potential positive impact of the stimulation of self-directed learning on clinical teaching quality ratings [76], items such *7. Asked questions that promote learning (clarifies, probes, reflective questions etc.)* should be retained. Institutions could use this information to design professional development activities for their clinical educators to help them work with students to develop their reflective practice and self-directed learning skills.

Factor 3 - feedback

Feedback to the student about their performance is a strong theme in the clinical teaching literature [15,27,77]. Timely feedback to the student was the strongest loading item on this factor. This result is consistent with the literature that suggests feedback should be provided to the student in a timely fashion [14,21,78,79]. The provision of both positive and negative feedback to the student are captured in the OCTQ. Further, item 26 - *Explained to me why I was correct or incorrect* allows the student to report their perception of the ability of the clinical educator to provide constructive feedback. Feedback provided to a student can be positive or negative, informal or continuous, and formative, or based around summative assessments such as the mini Clinical Examination (mini-CEX) [80,81]. What is not captured in this factor is the quality of the feedback provided by the clinical educator, and whether this had an impact on the future performance of the student. This is an area that could be explored in the future.

Factor 4 - patient management

The role of the clinical educator in the student-led clinic is to oversee the student writing the clinical history, and an expectation is that the educator will promote best practice in relation to record keeping. This is captured in item 27 - *Promoted keeping of medical records in a way that is thorough, legible, efficient and organised*. This type of item is not common in clinical teaching

quality questionnaires, and given the importance of case notes for practitioner communication and medicolegal reasons, it is felt that this item is a valuable addition. Developers of clinical teaching evaluation questionnaires should consider the inclusion of the same or similar item in the future. It is also noteworthy that this item loaded strongly onto the factor.

Traditionally, patient care in the early stages of the students' clinical education is scaffolded from observation through to autonomous patient care. In the student-led clinic environment, students will often have greater patient care responsibilities compared to the hospital setting [82]. In the osteopathy program at VU, students who completed the questionnaire in the present study were already responsible for patient care, under supervision. The inclusion of item 28 - *Encouraged me to assume responsibility for patient care* is relevant for the student-led clinic environment, as lower scores for this item indicates that the student may feel that the supervisor 'takes over' or directs the treatment and as such, the student may perceive that their responsibility for patient care has reduced. Students need to feel as though they are supported by the clinical educator when managing a patient but they have substantial autonomy in conducting the treatment.

Factor 5 - modelling

Modelling (Factor 5) is the clinical educator taking on the responsibility of professional role model and the demonstration of the skills and knowledge that are expected of a capable health professional. Modelling was identified by Stalmeijer et al. [27] as an important determinant of the effectiveness of the clinical educator. Students in the present study appear to value the knowledge and technical skills of the clinical educator. It is important for the clinical educator to undertake professional development and reading outside of their education role in order to inform their clinical teaching. The VU osteopathy program emphasises evidence informed practice [83,84] including it in the mission statement for the program. Therefore incorporating such an item in the OCTQ is important to ensure the clinical educators are modelling appropriate behaviours. Modelling extends to the demonstration of osteopathic examination and technique, clinical examination skills (e.g. performing a cranial nerve examination) as well as rehabilitation and advice to the patient; all parts of the typical osteopathic consultation. Demonstration of these physical and clinical skills has previously been demonstrated to be behaviours of an effective clinical educator [24,26].

Global ratings

The global ratings provide a way for the student to rate the overall performance and quality of the clinical educator. All three global ratings demonstrated mean scores greater than 4. An issue with the use of such a rating is the possibility of a 'ceiling effect' [22] therefore the rating needs to be interpreted in conjunction with the individual OCTQ items. With the 'ceiling effect' it may be difficult to differentiate between high quality clinical educators, although this may not be of great concern given they are already achieving high ratings. Working with the results of individual items may assist the clinical educator and their supervisor/manager to develop targeted professional development activities or assist with promotion decisions.

Psychometric properties

From a statistical viewpoint, Factors 1, 2 and 3 had high alpha values whereas Factors 4 and 5 had moderate (<0.7) alpha values. These moderate values are likely due to the fact that the factors contained two items. Whilst alpha values of 0.7 above are generally considered to be the minimum acceptable [85], these factors will be retained given the relevance of the items (items 22–26). Future studies into the OCTQ will employ modern test theory approaches such as Rasch analysis in order to strength the psychometric properties of the questionnaire.

The results of the present study provide evidence for the content and face validity of the OCTQ as the items were drawn from published clinical teaching questionnaires, and examined by both the clinical educators and students who will be using the questionnaire. A unique aspect of the OCTQ is that it contains items that are specific to osteopathic clinical education (items 21, 22 and 24) and this has not been reported in the literature previously. The factors generated in the present study are generally consistent with the clinical education literature and therefore, the questionnaire could be generalisable to other student-lead, on-campus clinics.

Limitations

Larger student populations and multiple institutions should be used in future studies to improve the generalisability of the questionnaire. A limitation of the present study was that each student only rated two clinical educators. There is the potential for bias to occur in that students may have rated those educators they wished to rate based on a positive or negative perception leading to ratings at the extremes of the scale options. Further research is required to confirm the factor structure of the questionnaire, establish the test-retest reliability, undertake a generalisability analysis to determine the number of ratings to generate a reliable result as well as examining the concurrent validity.

Conclusions

This study has developed a questionnaire - the Osteopathy Clinical Teaching Questionnaire - to assess the quality of clinical education in a student-led teaching clinic in a pre-

registration osteopathic teaching program at one Australian university. The items were identified in the literature and then tested with students in the clinical education component of the program. The OCTQ contains 30 items, and 3 global items, which address a range of behaviours and roles that students perceive to be important for an osteopathic clinical educator. The evidence-informed approach to the EFA employed in the present study helps to strengthen the construct validity of the questionnaire. This paper provides evidence for the 'observation' through to 'target' domain components of Kane's perspective on validity. Further evidence will be sought to provide a justification of the validity of the scores derived from the OCTQ and the questionnaire will now be the subject of further investigation to establish its psychometric properties and generalisability - these will be reported on in subsequent papers. Questionnaires like the OCTQ have the potential to improve the clinical learning experience for the student and the ensuing positive impact on patient care.

Additional file

Additional file 1: Systematic literature search.

Competing interests
The author declares that they have no competing interests.

Authors' contributions
BV designed the study, developed the literature review, undertook the statistical analysis and wrote the manuscript.

Authors' information
Brett Vaughan is a lecturer in the College of Health & Biomedicine, Victoria University, Melbourne, Australia and a Professional Fellow in the School of Health & Human Sciences at Southern Cross University, Lismore, New South Wales, Australia. His professional interests include competency and fitness-to-practice assessments, educational evaluation, and clinical education in allied health.

Acknowledgements
The author would like to thank Dr. Patrick McLaughlin, Dr. Keri Moore, Dr. Jane Mulcahy, Chris Macfarlane and Prue Florentine for their comments on the manuscript.

Received: 8 August 2014 Accepted: 30 March 2015
Published online: 08 April 2015

References
1. Levy LS, Sexton P, Willeford KS, Barnum M, Guyer S, Gardner G, et al. Clinical Instructor Characteristics, Behaviors and Skills in Allied Health Care Settings: A Literature Review. Athl Train Educ J. 2009;4(1):8–13.
2. Rogers J, Lautar C, Dunn L. Allied health students' perceptions of effective clinical instruction. Health Care Manag. 2010;29(1):63–7.
3. Hoffman KG, Donaldson JF. Contextual tensions of the clinical environment and their influence on teaching and learning. Med Educ. 2004;38(4):448–54.
4. Stalmeijer RE, Dolmans D, Snellen-Balendong H, van Santen-Hoeufft M, Wolfhagen H, Scherpbier A. Clinical teaching based on principles of cognitive apprenticeship: views of experienced clinical teachers. Acad Med. 2013;88(6):861–5.
5. Patton N, Higgs J, Smith M. Using theories of learning in workplaces to enhance physiotherapy clinical education. Physiother Theory Pract. 2013;29(7):493–503.
6. Dornan T, Boshuizen H, King N, Scherpbier A. Experience-based learning: a model linking the processes and outcomes of medical students' workplace learning. Med Educ. 2007;41(1):84–91.
7. Dornan T, Tan N, Boshuizen H, Gick R, Isba R, Mann K, et al. How and what do medical students learn in clerkships? Experience based learning (ExBL). Advances in Health Sciences Education. 2014:1–29
8. Vaughan B, Macfarlane C, Florentine P. Clinical education in the osteopathy program at Victoria University. Int J Osteopath Med. 2013;17(3):199–205.
9. Conigliaro RL, Stratton TD. Assessing the quality of clinical teaching: a preliminary study. Med Educ. 2010;44(4):379–86.
10. Parsell G, Bligh J. Recent perspectives on clinical teaching. Med Educ. 2001;35(4):409–14.
11. Scheepers RA, Lombarts KM, van Aken MA, Heineman MJ, Arah OA. Personality Traits Affect Teaching Performance of Attending Physicians: Results of a Multi-Center Observational Study. PLoS One. 2014;9(5):e98107.
12. Cross V. Perceptions of the ideal clinical educator in physiotherapy education. Physiotherapy. 1995;81(9):506–13.
13. Jahan F, Sadaf S, Kalia S, Khan A, Hamza HB. Attributes of an effective clinical teacher: a survey on students' and teachers' perceptions. J Coll Physicians Surg Pak. 2008;18(6):357.
14. Snell L, Tallett S, Haist S, Hays R, Norcini J, Prince K, et al. A review of the evaluation of clinical teaching: new perspectives and challenges*. Med Educ. 2000;34(10):862–70.
15. Sutkin G, Wagner E, Harris I, Schiffer R. What makes a good clinical teacher in medicine? A review of the literature. Acad Med. 2008;83(5):452–66.
16. Kilminster S, Jolly B. Effective supervision in clinical practice settings: a literature review. Med Educ. 2001;34(10):827–40.
17. Tang F, Chou S, Chiang H. Students' perceptions of effective and ineffective clinical instructors. Journal of Nurse Education. 2005;44(4):187–92.
18. Curtis N, Helion JG, Domsohn M. Student athletic trainer perceptions of clinical supervisor behaviors: a critical incident study. J Athl Train. 1998;33(3):249.
19. Dolmans D, Wolfhagen IH, Essed GG, Scherpbier AJ, van der Vleuten CP. The impacts of supervision, patient mix, and numbers of students on the effectiveness of clinical rotations. Acad Med. 2002;77(4):332–5.
20. Ramani S, Leinster S. AMEE Guide no. 34: teaching in the clinical environment. Med Teach. 2008;30(4):347–64.
21. Ernstzen D, Bitzer E, Grimmer-Somers K. Physiotherapy students' and clinical teachers' perceptions of clinical learning opportunities: a case study. Med Teach. 2009;31(3):102–15.
22. Copeland HL, Hewson MG. Developing and testing an instrument to measure the effectiveness of clinical teaching in an academic medical centre. Acad Med. 2000;75(2):161–6.
23. Kernan WN, Hershman W, Alper EJ, Lee MY, Viscoli CM, Perry JR, et al. Disagreement between students and preceptors regarding the value of teaching behaviors for ambulatory care settings. Teach Learn Med. 2008;20(2):143–50.
24. Prideaux D, Alexander H, Bower A, Dacre J, Haist S, Jolly B, et al. Clinical teaching: maintaining an educational role for doctors in the new health care environment. Med Educ. 2000;34(10):820–6.
25. Elnicki DM, Cooper A. Medical students'perceptions of the elements of effective inpatient teaching by attending physicians and housestaff. J Gen Intern Med. 2005;20(7):635–9.
26. Gerzina TM, McLean T, Fairley J. Dental clinical teaching: perceptions of students and teachers. J Dent Educ. 2005;69(12):1377–84.
27. Stalmeijer RE, Dolmans D, Wolfhagen H, Muijtjens AM, Scherpbier A. The Maastricht Clinical Teaching Questionnaire (MTCQ) as a valid and reliable instrument for the evaluation of clinical teachers. Acad Med. 2010;85(11):1732–8.
28. Laurent T, Weidner TG. Clinical instructors' and student athletic trainers' perceptions of helpful clinical instructor characteristics. J Athl Train. 2001;36(1):58–61.
29. Hewson MB, Jensen N. An inventory to improve clinical teaching in the general internal medicine clinic. Med Educ. 1990;24:518–27.
30. Boor K, Scheele F, Van Der Vleuten CP, Teunissen PW, Den Breejen EM, Scherpbier AJ. How undergraduate clinical learning climates differ: a multi-method case study. Med Educ. 2008;42(10):1029–36.
31. Dolmans D, Wolfhagen I, Heineman E, Scherpbier A. Factors adversely affecting student learning in the clinical learning environment: A student perspective. Education for Health. 2008;20(3):e1–10.

32. Abey S, Lea S, Callaghan L, Cotton D, Shaw S. The development of a scale to assess practitioner capacity to engage in clinical education. Journal of Further and Higher Education. 2015;39(2):22–240.
33. Beckman TJ, Ghosh AK, Cook DA, Erwin PJ, Mandrekar JN. How reliable are assessments of clinical teaching? J Gen Intern Med. 2004;19(9):971–7.
34. Irby DM. Teaching and learning in ambulatory care settings: a thematic review of the literature. Acad Med. 1995;70(10):898–931.
35. Fluit CR. Assessing the quality of clinical teachers. J Gen Intern Med. 2010;25(12):1337–45.
36. van der Leeuw R, Lombarts K, Heineman MJ, Arah O. Systematic evaluation of the teaching qualities of obstetrics and gynecology faculty: reliability and validity of the SETO tools. PLoS One. 2011;6(5):e19142.
37. Currens JB, Bithell CP. The 2: 1 clinical placement model: perceptions of clinical educators and students. Physiotherapy. 2003;89(4):204–18.
38. Moore A, Morris J, Crouch V, Martin M. Evaluation of physiotherapy clinical educational models: comparing 1: 1, 2: 1 and 3: 1 placements. Physiotherapy. 2003;89(8):489–501.
39. Martin M, Morris J, Moore A, Sadlo G, Crouch V. Evaluating practice education models in occupational therapy: Comparing 1: 1, 2: 1 and 3: 1 placements. The British Journal of Occupational Therapy. 2004;67(5):192–200.
40. Beckman TJ, Cook DA, Mandrekar JN. Factor instability of clinical teaching assessment scores among general internists and cardiologists. Med Educ. 2006;40(12):1209–16.
41. Kane MT. Current concerns in validity theory. J Educ Meas. 2001;38(4):319–42.
42. Kane MT. Validation. In: Brennan R, editor. Educational measurement. 4th ed. Westport, USA: Praeger Publishers; 2006. p. 17–64.
43. Cook DA. When I, say… validity. Med Educ. 2014;48(10):948–9.
44. Roff S, McAleer S, Skinner A. Development and validation of an instrument to measure the postgraduate clinical learning and teaching environment for hospital-based junior doctors in the UK. Med Teach. 2005;27(4):326–31.
45. Hudak PL, Wright JG. The Characteristics of Patient Satisfaction Measures. Spine. 2000;25(24):3167–77.
46. Lozano LM, García-Cueto E, Muñiz J. Effect of the number of response categories on the reliability and validity of rating scales. Methodology. 2008;4(2):73–9.
47. Oppenheim AN. Questionnaire design, interviewing and attitude measurement: Continuum International Publishing Group. 2000.
48. Gaskin CJ, Happell B. On exploratory factor analysis: A review of recent evidence, an assessment of current practice, and recommendations for future use. Int J Nurs Stud. 2014;51:511–21.
49. Schmitt TA. Current methodological considerations in exploratory and confirmatory factor analysis. J Psychoeduc Assess. 2011;29(4):304–21.
50. Baglin J. Improving your exploratory factor analysis for ordinal data: a demonstration using FACTOR. Practical Assessment, Research & Evaluation. 2014;19(5):2.
51. Holgado–Tello FP, Chacón–Moscoso S, Barbero–García I, Vila–Abad E. Polychoric versus Pearson correlations in exploratory and confirmatory factor analysis of ordinal variables. Qual Quant. 2010;44(1):153–66.
52. Basto M, Pereira JM. An SPSS R-Menu for ordinal factor analysis. J Stat Softw. 2012;46(4):1–29.
53. Ruscio J, Roche B. Determining the number of factors to retain in an exploratory factor analysis using comparison data of known factorial structure. Psychol Assess. 2012;24(2):282.
54. Garrido LE, Abad FJ, Ponsoda V. A new look at Horn's parallel analysis with ordinal variables. Psychol Methods. 2013;18(4):454.
55. Velicer WF. Determining the number of components from the matrix of partial correlations. Psychometrika. 1976;41(3):321–7.
56. Zwick WR, Velicer WF. Comparison of five rules for determining the number of components to retain. Psychol Bull. 1986;99(3):432.
57. Revelle W, Rocklin T. Very simple structure: An alternative procedure for estimating the optimal number of interpretable factors. Multivariate Behav Res. 1979;14(4):403–14.
58. Raîche G, Walls TA, Magis D, Riopel M, Blais J-G. Non-graphical solutions for Cattell's scree test. Methodology. 2013;9(1):23.
59. Courtney R, Gordon M. Determining the Number of Factors to Retain in EFA: Using the SPSS R-Menu v2. 0 to Make More Judicious Estimations. Practical Assessment, Research & Evaluation. 2013;18(8).
60. R Core Team. R: A language and environment for statistical computing Vienna, Austria2014. Available from: http://www.R-project.org/.
61. Revelle W. psych; Procedures for Personality and Psychological Research Evanston, Illinois, USA: Northwestern University; 2014. 1.4.5: Available from: http://CRAN.R-project.org/package=psych.
62. Bernaards CA, Jennrich RI. Gradient projection algorithms and software for arbitrary rotation criteria in factor analysis. Educ Psychol Meas. 2005;65(5):676–96.
63. Fox J. Polycor: polychoric and polyserial correlations. 2007. Available from: http://cran.R-project.org/package=polycor.
64. Raiche G, Magis D. nFactors: Parallel analysis and non graphical solutions to the Cattell Scree Test 2010 [cited 2 2]. 60]. Available from: http://CRAN.R-project.org/package=nFactors.
65. Sass DA, Schmitt TA. A comparative investigation of rotation criteria within exploratory factor analysis. Multivariate Behav Res. 2010;45(1):73–103.
66. Schmitt TA, Sass DA. Rotation criteria and hypothesis testing for exploratory factor analysis: Implications for factor pattern loadings and interfactor correlations. Educ Psychol Meas. 2011;71(1):95–113.
67. Hair JF, Anderson RE, Tatham RL, William C. Multivariate data analysis. Prentice Hall: London, UK; 1998.
68. Tabachnick BG, Fidell LS. Using multivariate statistics. 5th ed. Boston: Allyn and Bacon; 2007.
69. MacCallum RC, Widaman KF, Zhang S, Hong S. Sample size in factor analysis. Psychol Methods. 1999;4(1):84.
70. Zumbo BD, Gadermann AM, Zeisser C. Ordinal versions of coefficients alpha and theta for Likert rating scales. J Mod Appl Stat Methods. 2007;6(1):4.
71. Gadermann AM, Guhn M, Zumbo BD. Estimating ordinal reliability for Likert-type and ordinal item response data: A conceptual, empirical, and practical guide. Practical Assessment, Research & Evaluation. 2012;17(3):1–13.
72. Boerboom TB, Mainhard T, Dolmans DH, Scherpbier AJ, Van Beukelen P, Jaarsma AC. Evaluating clinical teachers with the Maastricht clinical teaching questionnaire: How much'teacher'is in student ratings? Med Teach. 2012;34(4):320–6.
73. Rethans JJ, Norcini J, Barón-Maldonado M, Blackmore D, Jolly BC, LaDuca T, et al. The relationship between competence and performance: implications for assessing practice performance. Med Educ. 2002;36:901–19.
74. Griffith III CH, Georgeson JC, Wilson JF. Six-year documentation of the association between excellent clinical teaching and improved students' examination performances. Acad Med. 2000;75(10):S62–4.
75. Bates J, Konkin J, Suddards C, Dobson S, Pratt D. Student perceptions of assessment and feedback in longitudinal integrated clerkships. Med Educ. 2013;47(4):362–74.
76. Litzelman DK, Stratos G, Marriott D, Skeff K. Factorial validation of a widely disseminated educational framework for evaluating clinical teachers. Acad Med. 1998;73(6):688–95.
77. Kikukawa M, Nabeta H, Ono M, Emura S, Oda Y, Koizumi S, et al. The characteristics of a good clinical teacher as perceived by resident physicians in Japan: a qualitative study. BMC Med Educ. 2013;13(1):100.
78. Kelly C. Student's perceptions of effective clinical teaching revisited. Nurse Educ Today. 2007;27(8):885–92.
79. Van de Ridder J, Stokking KM, McGaghie WC, Ten Cate OTJ. What is feedback in clinical education? Med Educ. 2008;42(2):189–97.
80. Bennett D, Kelly M, O'Flynn S. Framework for feedback: the peer mini-clinical examination as a formative assessment tool. Med Educ. 2012;46(5):512.
81. Pelgrim E, Kramer A, Mokkink H, Van der Vleuten C. Quality of written narrative feedback and reflection in a modified mini-clinical evaluation exercise: an observational study. BMC Med Educ. 2012;12(1):97.
82. Schweitzer PJ, Rice TR. The student-run clinic: a new opportunity for psychiatric education. Acad Psychiatry. 2012;36(3):233–6.
83. Fryer G. Muscle energy technique: An evidence-informed approach. Int J Osteopath Med. 2011;14(1):3–9.
84. Fryer G. Teaching critical thinking in osteopathy–Integrating craft knowledge and evidence-informed approaches. Int J Osteopath Med. 2008;11(2):56–61.
85. Peterson R. A meta-analysis of Cronbach's coefficient alpha. J Consum Res. 1994;21:381–91.

CHAPTER 3

Exploring the measurement properties of the osteopathy clinical teaching questionnaire using Rasch analysis

The measure for the evaluation of clinical teaching in the osteopathy student-led clinical learning environment described in chapter 2 was tested with osteopathy students from across Australia, New Zealand and the United Kingdom. Data collected for chapter 3 were analyzed using Rasch analysis to produce a feasible and acceptable measure of clinical teaching quality. The 12-item measure developed in this chapter demonstrates the properties of measurement invariance in addition to being feasible in this learning environment.

Vaughan *Chiropractic & Manual Therapies* (2018) 26:13
https://doi.org/10.1186/s12998-018-0182-2

Chiropractic & Manual Therapies

RESEARCH Open Access

Exploring the measurement properties of the osteopathy clinical teaching questionnaire using Rasch analysis

Brett Vaughan

Abstract

Background: Clinical teaching evaluations are common in health profession education programs to ensure students are receiving a quality clinical education experience. Questionnaires students use to evaluate their clinical teachers have been developed in professions such as medicine and nursing. The development of a questionnaire that is specifically for the osteopathy on-campus, student-led clinic environment is warranted. Previous work developed the 30-item Osteopathy Clinical Teaching Questionnaire. The current study utilised Rasch analysis to investigate the construct validity of the Osteopathy Clinical Teaching Questionnaire and provide evidence for the validity argument through fit to the Rasch model.

Methods: Senior osteopathy students at four institutions in Australia, New Zealand and the United Kingdom rated their clinical teachers using the Osteopathy Clinical Teaching Questionnaire. Three hundred and ninety-nine valid responses were received and the data were evaluated for fit to the Rasch model. Reliability estimations (Cronbach's alpha and McDonald's omega) were also evaluated for the final model.

Results: The initial analysis demonstrated the data did not fit the Rasch model. Accordingly, modifications to the questionnaire were made including removing items, removing person responses, and rescoring one item. The final model contained 12 items and fit to the Rasch model was adequate. Support for unidimensionality was demonstrated through both the Principal Components Analysis/t-test, and the Cronbach's alpha and McDonald's omega reliability estimates. Analysis of the questionnaire using McDonald's omega hierarchical supported a general factor (quality of clinical teaching in osteopathy).

Conclusion: The evidence for unidimensionality and the presence of a general factor support the calculation of a total score for the questionnaire as a sufficient statistic. Further work is now required to investigate the reliability of the 12-item Osteopathy Clinical Teaching Questionnaire to provide evidence for the validity argument.

Keywords: Item response theory, Reliability estimation, Medical education, Clinical education

Background

Clinical teaching influences the development of clinical and patient management skills students need for competent, safe and effective practice. At present, little is known about clinical education in osteopathy in non-United States teaching programs beyond the commentary on one Australian osteopathy program by Vaughan et al. [1]. These authors postulated that Collins' cognitive apprenticeship model [2, 3] could account for a number of aspects of the student-clinical teacher interaction within a student-led clinical environment.

Osteopathy students undertaking their clinical education in Australia, New Zealand and United Kingdom are in their final years of training and are responsible for the management of patients under the supervision of a qualified osteopath ('clinical teacher'). Clinical education in osteopathy is typically undertaken in a student-led, on-campus clinic environment – in the Australian context. Allan et al. [4] referred to these as 'university clinics'. These clinics provide students with an opportunity to

Correspondence: brett.vaughan@vu.edu.au
College of Health & Biomedicine, Victoria University, Melbourne, Australia

© The Author(s). 2018 **Open Access** This article is distributed under the terms of the Creative Commons Attribution 4.0 International License (http://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. The Creative Commons Public Domain Dedication waiver (http://creativecommons.org/publicdomain/zero/1.0/) applies to the data made available in this article, unless otherwise stated.

develop their work-readiness, and practice the application of skills and knowledge acquired in the classroom in a supervised environment. In osteopathy clinical education, each clinical teacher typically supervises between 5 and 7 students at any one time [1] however this may be up to 10 students in some instances [5]. The evaluation of this teaching is important to ensure students receive appropriate clinical skills education and development.

Systematic reviews of questionnaires to evaluate clinical teaching have been undertaken [6, 7]. These reviews have identified a substantial number of questionnaires with varying degrees of evidence of their validity or reliability. The statistical approaches to the development of these questionnaires are also variable. Both systematic reviews reported 'factor analysis' was used in the development of many of the questionnaires. It was not clear what methods were employed in all instances however it appears that Principal Components Analysis (PCA) was typically used. This was potentially due to convenience [8] (e.g. PCA is the default analysis in SPSS), or following how other researchers have developed clinical teaching evaluations [8], or a genuine desire to retain explained variance. Whilst PCA can be an effective approach to retain the least number of items to explain a substantial portion of the variance [9–12], the models produced often do not fit those generated by more advanced statistical techniques [13]. Extraction methods such as Principal Axis Factoring and ordinary (unweighted) least squares (OLS) are more appropriate than PCA [11], the latter (OLS) being particularly suitable for ordinal data that is typical of self-report questionnaires. In the last five years, researchers have used these Exploratory Factor Analysis (EFA) and Confirmatory Factor Analysis (CFA) approaches in the development of questionnaires to evaluate the quality of clinical teaching [14–16].

Previous work has developed the Osteopathy Clinical Teaching Questionnaire (OCTQ) [16]. The OCTQ is a 33-item questionnaire (30 items and 3 global rating items) designed to evaluate the quality of clinical teaching in on-campus, student-led osteopathy teaching clinics. Work undertaken thus far has provided evidence for the validity argument for the OCTQ through item development and EFA. The use of modern test theory [17] approaches for questionnaire development is particularly relevant in health sciences education and clinical research [18–21], where there is a desire to measure attitudes and abilities. The current study is the second to employ Rasch analysis in the ongoing development of a clinical teaching quality questionnaire, the other being that by Winstanley and White [22] in a revision of the Manchester Clinical Supervision Scale (MCSS). The aim of the current study was to explore the construct validity of the OCTQ by using Rasch analysis. Consistent with Kane's approach to the development of a validity argument [23], the present study also aims to provide further evidence for the validity of the scores derived from the OCTQ.

Methods

The study received ethics approval from Victoria University (VU) (Australia), Southern Cross University (SCU) (Australia), Unitec Institute of Technology (New Zealand), and the British School of Osteopathy (BSO) (United Kingdom). Participation in the study was voluntary and did not impact on the ability of the students to receive their grades nor graduate from their program of study. Results from this study were not used for employment or promotion decisions nor made available to the clinical teachers' supervisor, however copies of the anonymous student responses were provided to the individual clinical teacher upon request. Consent to participate in the study was implied upon the return of a completed questionnaire.

Participants

Students in the final two years of the programs at VU, SCU, Unitec, and the BSO were invited to participate in the study. These students were completing the clinical practice requirement of their respective programs and were at similar stages in their clinical training. Students received an email via their university email address. They were informed of the study and encouraged to complete a questionnaire for each of the clinical teachers whom they had worked with in the period July 2014 – December 2014 (VU and SCU), and March 2015 – July 2015 (Unitec and BSO). The clinical teachers at each institution also received an email informing them the study was taking place. Students were not required to identify themselves on the questionnaire.

Data collection

Students completed version 2 of the Osteopathy Clinical Teaching Questionnaire (OCTQ) [16] (Additional file 1) during their scheduled clinic placement time. The OCTQ (version 2) is a 33-item questionnaire that contains 30 items evaluating different aspects of the clinical teachers' performance across 5 factors: learning environment; modelling; feedback; patient management; and reflective practice. There are also 3 global rating items (Additional file 1). Each item is anchored with the statement "This Clinical Educator…" and rated on a scale of 1 (strongly disagree) to 5 (strongly agree) with a neutral category (option 3). The students were asked to complete the OCTQ (version 2) for each of the clinical teachers they had worked with, basing their responses on the entirety of their interaction with the teacher for the relevant teaching period and not focusing on a single

positive and/or negative encounter with the teacher. Responses were anonymous – neither the student nor clinical teacher being rated were identified. The student was asked to indicate their gender and the gender of the clinical teacher being rated as previous research has identified that student and teacher gender can influence responses to clinical teaching questionnaire items [24]. The institution where the questionnaire was completed was also noted.

Data analysis

Descriptive statistics were generated in *R* [25] using the *psych* package [26].

Rasch analysis

Data were entered into Microsoft Excel for Mac then exported to RUMM2030 [27] for Rasch analysis (RA). The target construct in the present study is the quality of clinical teaching provided by osteopathy clinical teachers. As each OCTQ (version 2) item was scored on a 1–5 scale the polytomous Rasch model was used for the analysis. Each step in the RA informed the next. Within each step, a number of statistical analyses were undertaken to determine the most appropriate action for the next step. Figure 1 presents the analyses undertaken within each step.

Rasch model fit

Overall model fit was first evaluated using the chi-square statistic and Bonferonni-adjusted *p*-value. Fit residual standard deviations (SD) were then used to evaluate the fit of the items and persons respectively to the Rasch model [18] along with Bonferonni-adjusted chi-square probabilities. The Person Separation Index (PSI) was calculated at each step. Differential item function was then evaluated for each item using the person factors institution, clinical educator gender and student gender. Person fit was evaluated using the fit residual statistic, with responses from misfitting persons removed from subsequent steps. Correlations between each of the items were evaluated to identify item combinations with residuals greater than 0.20 suggesting 'local dependence' [28]. To determine if the local dependence was impacting on the PSI, a subtest analysis was performed in RUMM2030. A reduction in the PSI with the subtest suggests the item combination is inflating this value and requires the removal of one of the items. The information from each of these analyses informed the next step (e.g. remove an item/person response, rescore an item). Once fit to the Rasch model was achieved, the PSI informed the number of possible strata that could be identified [29]. Given clinical teaching evaluation data are unlikely to be normally distributed [7], the method described by Wright [30] to identify each strata was used. This method included the addition of 10% to the standard error for each logit "…to allow for the unmodeled noise encountered in real data" (p. 786).

Dimensionality testing, reliability estimates and descriptive statistics

Multiple approaches were employed to evaluate the dimensionality of the questionnaire to ascertain whether the items were measuring the same latent construct. These approaches were PCA of the standardised residuals, and evaluating the number of factors to extract using methods for EFA.

Principal components analysis

Once a fit to the Rasch model was achieved through each of the analyses, a PCA of the standardised residuals was undertaken to derive the 'Rasch factor' or 'Rasch dimension'. An independent t-test was used to evaluate the difference between the items that loaded positively and negatively onto the 'Rasch factor'. The binomial confidence interval for the t-test was calculated in R [25] using the binom package [31].

Number of factors to extract Parallel analysis (PA) [32], eigenvalues, acceleration factor (AF) [33] and optimal coordinates (OC) [33] were the methods used to confirm the number of factors to extract. These procedures were performed using the *psych* (version 1.5.4) [26] and *nFactors* (version 2.3.3) [34] packages in *R* utilising the polychoric correlation generated using the *polycor* package (version 0.7–8) [35].

Reliability estimates

Three reliability estimates were calculated using a variety of statistics in the *psych* package [26] in *R* [25]: Cronbach's alpha (α); and McDonald's omega hierarchical (ω_h) and total (ω_t) [36–38]. High ω_h values suggest that general factor accounts for the total score variance supporting unidimensionality [39], and values greater than 0.5 have been suggested to support the calculation of a total score for all scale items [40]. Omega subscale (ω_s) was also calculated for the subfactors identified when calculating the ω coefficient. Each of the reliability estimates were calculated using the polychoric correlation given the underlying data were ordinal in nature [41, 42], and also calculated based on the raw data. The explained common variance (ECV) was also calculated to further evaluate unidimensionality. Higher ECV values support unidimensionality [39, 43] however there is no guidance as to an acceptable value [39].

Results

Four hundred questionnaires were received. One questionnaire was not completed therefore 399 were available for analysis. Demographic data are presented in Table 1.

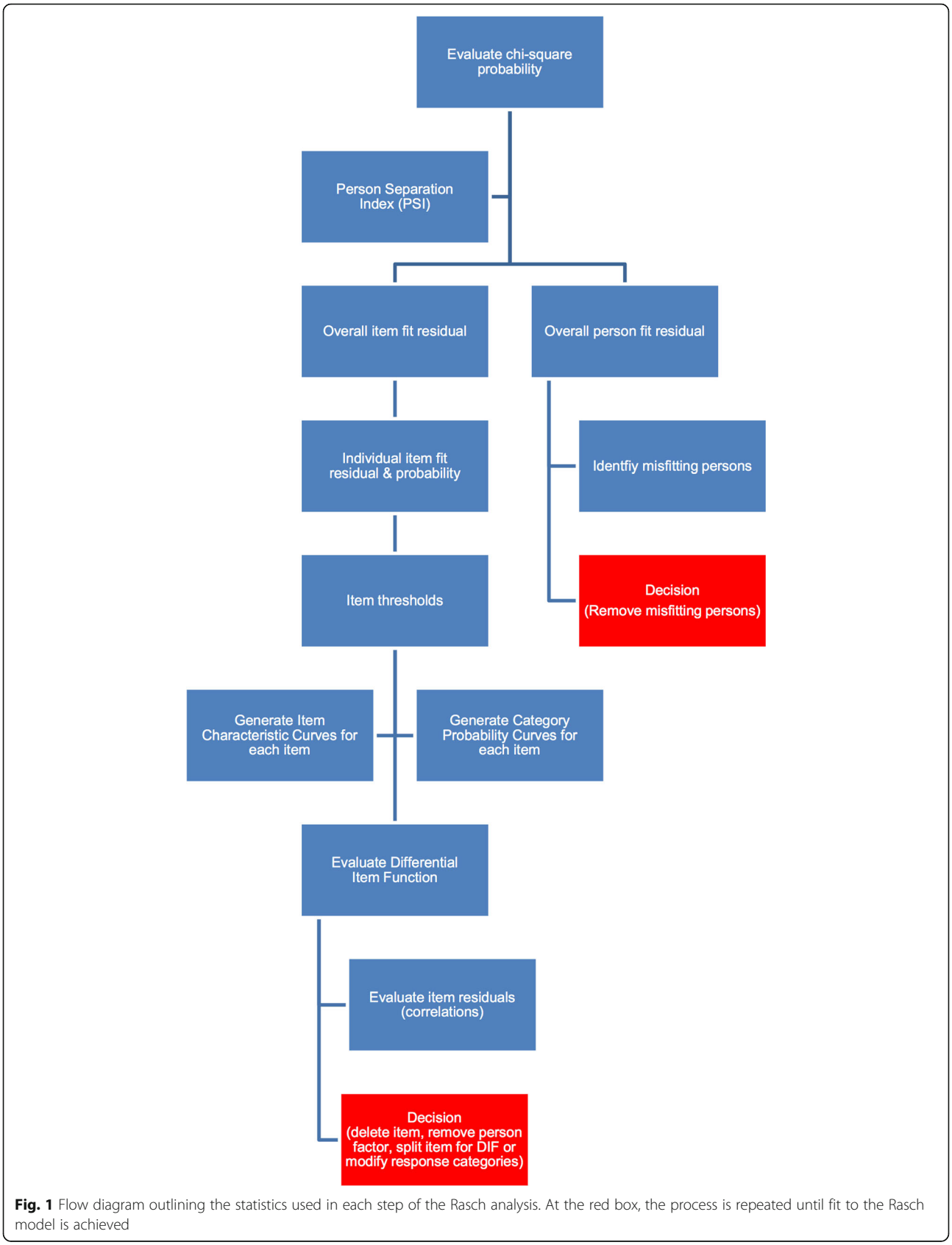


Fig. 1 Flow diagram outlining the statistics used in each step of the Rasch analysis. At the red box, the process is repeated until fit to the Rasch model is achieved

Table 1 Demographic data

			Institution			
		Total	Victoria University	Southern Cross University	British School of Osteopathy	Unitec
Total responses		399	149	119	42	89
Student gender	Male	150	44	58	14	34
	Female	224	98	49	26	51
Clinical Educator gender	Male	261	78	108	31	44
	Female	123	65	9	8	41

Note: some participants did not indicate either their gender or the gender of the clinical educator being rated

The category response frequencies were negatively skewed, however responses were observed for each item across the five response categories. The neutral *Neither agree nor disagree* response category was used, on average, 12% of the time suggesting the usefulness of this category.

Rasch analysis

Overall Rasch model fit

The Likelihood ratio test was statistically significant ($p < 0.001$), subsequently the partial credit model was used for the Rasch analysis. Overall fit was significantly different to the Rasch model ($\chi^2(150, N = 399) = 407.42$, $p < 0.001$) with a PSI of 0.910, item fit residual mean of – 0.32 (SD 2.34) and person fit residual mean of – 0.57 (SD 2.03). Item fit statistics and the threshold map for the initial analysis are presented in Table 2 and Fig. 2 respectively. Sixty responses were identified as *extreme* by RUMM and a further 62 responses demonstrated person fit residual SDs of greater than 2.5 suggesting misfit to the Rasch model.

Modifications to the model

Extreme and misfitting responses ($n = 122$) were removed. Once completed the overall fit remained significantly different to that of the Rasch model ($\chi^2(150, N = 277) = 315.98$, $p < 0.001$), the PSI (0.927), however the item fit residual mean – 0.34 (SD 2.09) and person fit residual mean – 0.27 (SD 1.11) all improved. Local dependence was observed for a number of items (Additional file 2). Two iterations of the analysis were undertaken to produce a fit to the Rasch model. The steps to produce this model are at Additional file 3 and the fit statistics were reviewed after each modification in order to determine the next step in the analysis.

Osteopathy Clinical Teaching Questionnaire (OCTQ) items were removed where they demonstrated misfit to the Rasch model, differential item functioning, or local dependence. Additional file 3 identifies the reasons for the removal of the items during the analysis. Item 30 *Demonstrated osteopathic, clinical examination and rehabilitation knowledge and skill(s)* required rescoring in order to improve its fit to the Rasch model, as

Table 2 Item fit statistics for the full 30-item Osteopathy Clinical Teaching Questionnaire

Item	Location	Fit Residual	χ^2	Probability
1	−0.338	−2.592[a]	9.220	0.101
2	−0.292	−1.144	3.301	0.658
3	0.115	−2.670[a]	9.115	0.104
4	0.104	−2.747[a]	11.628	0.040
5	−0.360	−0.501	5.967	0.309
6	−0.426	−1.513	7.659	0.176
7	0.063	−1.644	4.684	0.457
8	−0.056	1.023	5.296	0.381
9	−0.129	−1.866	7.391	0.193
10	−0.117	0.085	3.599	0.608
11	−0.272	−2.476	10.194	0.070
12	0.167	−1.909	10.186	0.070
13	−0.079	−2.155	17.007	0.005
14	−0.205	−1.581	4.161	0.526
15	0.092	1.082	4.790	0.442
16	−0.244	−1.347	2.679	0.749
17	−0.101	−2.614[a]	13.235	0.023
18	−0.263	−1.744	7.138	0.211
19	−0.091	−2.881[a]	14.378	0.014
20	−0.090	0.885	5.992	0.307
21	0.358	0.662	7.751	0.170
22	0.778	0.357	10.095	0.072
23	0.268	−0.109	8.925	0.112
24	0.467	4.217[a]	43.774	0.000[b]
25	0.978	3.859[a]	35.118	0.000[b]
26	0.529	1.770	15.103	0.009
27	0.395	7.145[a]	105.105	0.000[b]
28	−0.412	0.477	2.452	0.783
29	−0.327	1.384	15.011	0.010
30	−0.511	−1.280	6.476	0.262

[a]Item fit residual greater than 2.5. Large negative residuals suggest item redundancy, large positive residuals

[b]Statistically significant chi-square probability (Bonferonni-adjusted $p = 0.0003$)

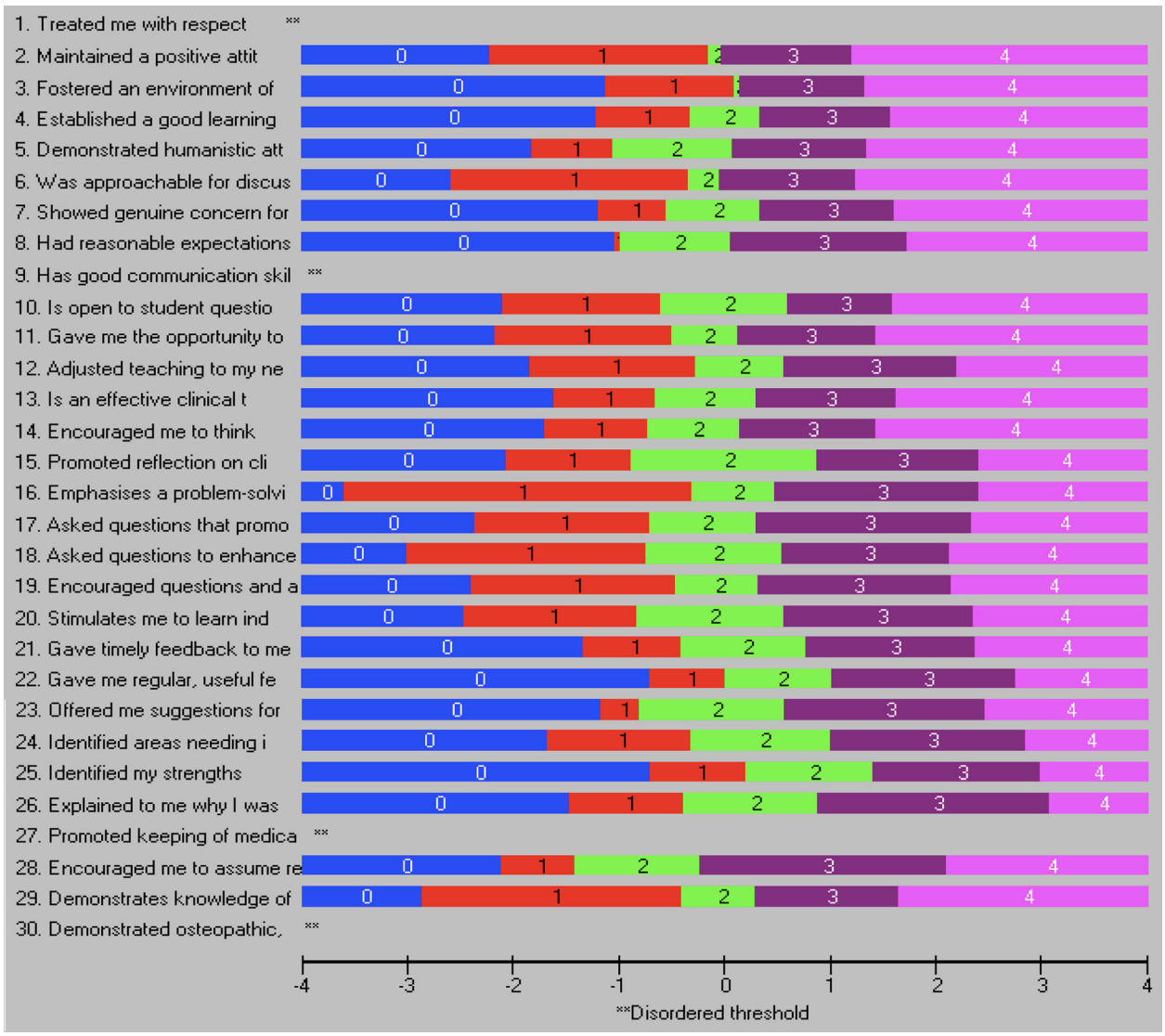


Fig. 2 Threshold map for the Osteopathy Clinical Teaching Questionnaire (version 2) items

respondents did not appear to be using the *Strongly Disagree* and *Disagree* responses in the manner predicted by the Rasch model. Figure 3 demonstrate the category probability curves pre- and post-rescoring.

Differential item function analysis

Uniform differential item function (DIF) was observed for institution (Fig. 4), clinical teacher gender (Fig. 4), and student gender (Fig. 4) for item 14 *Encouraged me to think*. For gender, female clinical teachers received scores that were systematically lower than that expected by the Rasch model, and were significantly lower when compared to males across all of the class intervals. With regard to student gender, systematic differences existed between males and females across the class intervals however whether males or females selected higher responses was not consistent.

Other items that demonstrated DIF through each iteration included item 19 *Encouraged questions and active participation* and item 28 *Encouraged me to assume responsibility for patient care*. In order to ensure that the items on the modified OCTQ were applicable to a range of teaching institutions and free from gender influence, any item demonstrating DIF was removed. Those items demonstrating local dependence were analysed using a subtest to examine whether they were inflating the PSI value, and one item removed (Additional file 3).

Final Rasch model

Fit to the Rasch model was achieved by removing 18 items, rescoring item 30, and removing misfitting 153 responses in total comprising 122 at the initial analysis and 31 misfitting responses identified during the subsequent

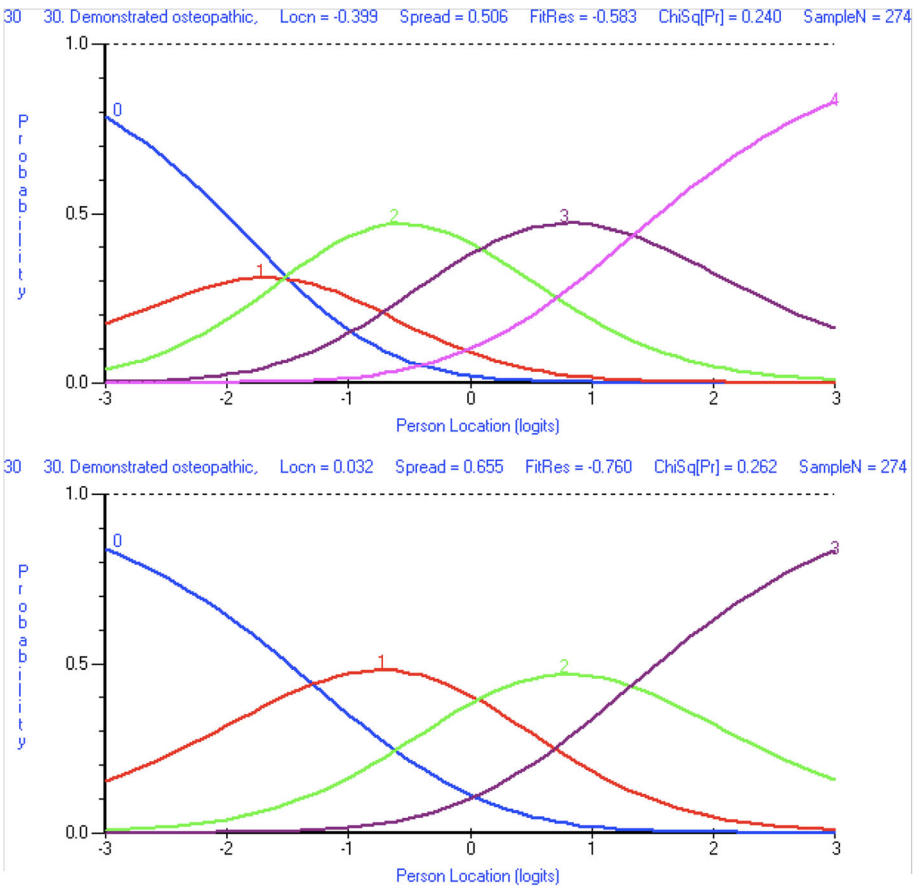


Fig. 3 Category probability curve for Osteopathy Clinical Teaching Questionnaire item 30 *Demonstrated osteopathic, clinical examination and rehabilitation knowledge and skill(s)*. Before item rescoring (top image). After item rescoring (bottom image)

analysis (Additional file 3 and Additional file 4). The final model contained 12 items and 246 responses. Overall fit to the Rasch model was demonstrated ($\chi^2(60, N = 246) = 65.26$, $p = 0.298$). The item and person fit residual means were – 0.34 (SD 1.18) and – 0.20 (SD 0.82). These fit residual SDs were both within the acceptable range. The item fit statistics are presented in Table 3 and the threshold map at Fig. 5. There is a spread of item location values that represent different levels of a single latent construct (Table 3).

The mean person location logit was 2.44 suggesting that respondents used the higher options of the 1–5 scale for each OCTQ item. The item-person map and the item-threshold distribution are shown in Figs. 6 and 7 respectively. The item-threshold distribution shows that the OCTQ item scale covers a range of possible options on the 1–5 scale, and largely covers the responses provided in the present study. The questionnaire may be subject to a 'ceiling-effect' however.

No item in the final model demonstrated DIF or local dependence. The PSI was 0.827 suggesting that approximately 83% of the variance in the observed score on the final 12-item questionnaire is due to the true variance in a students' perceived quality of teaching provided by a clinical teacher. The remaining 17% is classified as error variance. The PSI also indicates that 3–4 strata could be identified [29].

The 'Rasch factor' or first component of the PCA accounted for 17.13% of the variance, suggesting the questionnaire is unidimensional. The PCA/t-test identified twelve responses where the t-tests were significantly different between the OCTQ items that loaded positively and negatively (Table 4) onto the 'Rasch factor' ($p = 0.04$, 95% CI 0.021–0.085). As the 95%CI (calculated using the 11 statistics in the *binom R* package) contains the value of $p = 0.05$, this provides further evidence to support the unidimensionality of the 12-item questionnaire.

Factor extraction

Additional support for the unidimensionality of the 12-item questionnaire was obtained from the four methods used in EFA to determine the number of factors to extract. Using the data from the 399 completed questionnaires and the subsequent polychoric correlation of this

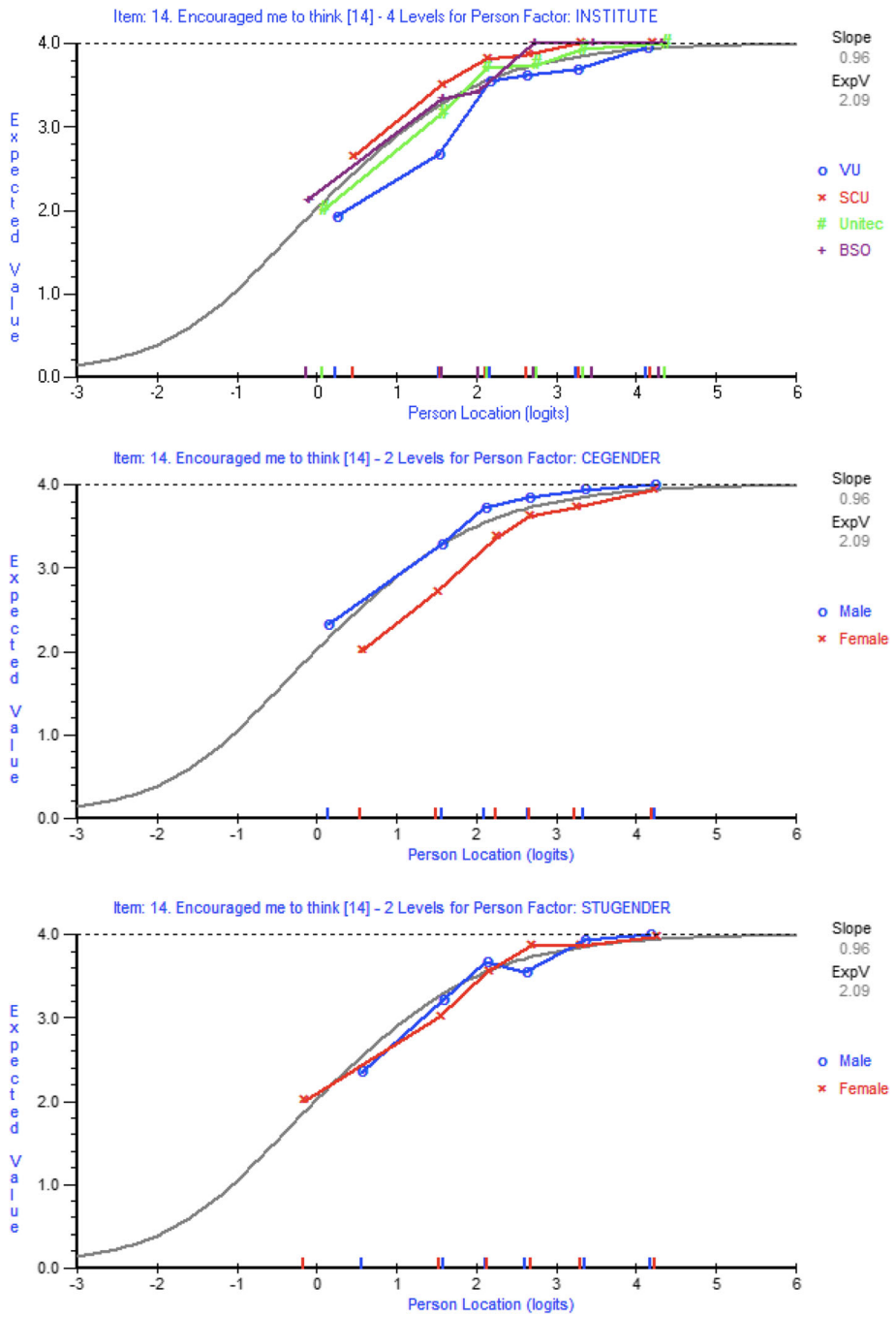


Fig. 4 Item characteristic curves for Osteopathy Clinical Teaching Questionnaire item 14 *Encouraged me to think*. Differential item function for institution (top image). Differential item function for clinical teacher gender (middle image). Differential item function for student gender (bottom image)

data, all four methods suggested one factor should be extracted (Fig. 8). All valid responses were used in this analysis to ensure the accuracy of the result.

Reliability estimates

Reliability estimates for the 12-item OCTQ (Table 5) using the polychoric correlation were slightly larger compared to the raw data, consistent with Revelle's [44] suggestion that ω_h can be overestimated with polychoric correlations. All of the reliability estimates, regardless of whether the raw data or polychoric correlation was used, were well above an acceptable level of 0.80 suggesting less than 20% of the variance is measurement error. This level of measurement error is consistent with the PSI.

With regard to ω, the ECV for the general factor (g) was 82% (raw data) and 84% (polychoric correlation),

Table 3 Item fit statistics for the 12-item Osteopathy Clinical Teaching Questionnaire

Item	Location	SE	Fit Residual	χ^2	Probability
This Clinical Educator:					
2. Maintained a positive attitude towards me	− 0.283	0.104	0.311	1.293	0.935
5. Demonstrated humanistic attitudes in relating to patients (integrity, compassion and respect)	−0.118	0.106	−1.559	4.633	0.462
7. Showed genuine concern for my professional well-being	0.232	0.099	−0.816	4.057	0.541
9. Has good communication skills	0.225	0.099	−1.479	7.065	0.216
10. Is open to student questions and alternative approaches to patient management	−0.856	0.098	−0.391	6.938	0.225
12. Adjusted teaching to my needs (experience, competence, interest)	0.236	0.096	−1.675	9.890	0.078
15. Promoted reflection on clinical practice	0.292	0.096	1.073	2.665	0.751
16. Emphasises a problem-solving approach rather than solutions	−0.714	0.097	−1.589	3.700	0.593
18. Asked questions to enhance my learning	−0.224	0.100	−0.917	12.669	0.027
20. Stimulates me to learn independently	0.304	0.097	0.415	3.068	0.689
23. Offered me suggestions for improvement when required	0.723	0.093	1.900	4.945	0.423
30. Demonstrated osteopathic, clinical examination and rehabilitation knowledge and skill(s)	0.183	0.117	0.591	4.339	0.501

SE Standard error

again consistent with the PSI. All items correlated substantially higher with the general factor (0.70 or higher) than with the two subfactors (0.4 or less) (Additional file 5) thereby supporting unidimensionality and the appropriateness of the calculation of a total score.

Descriptive statistics

The descriptive statistics for the 12 items in the modified OCTQ are presented in Table 6. All items had a mean value greater than 4 (except for item 30 as it was rescored) and the median value for each item was 4 or 5. As the modified OCTQ was unidimensional, the calculation of a total score for the questionnaire is permissible (Table 6). The total possible score for the 12-item OCTQ is 59. The scoring structure and strata for each score is presented in Additional file 6.

Discussion

The present study was designed to further evaluate the construct validity of the Osteopathy Clinical Teaching Questionnaire [16] using Rasch analysis. Tennant et al. [20] advocate the use of Rasch analysis in the development of unidimensional measures in the health sciences, and this type of analysis has previously been utilised to

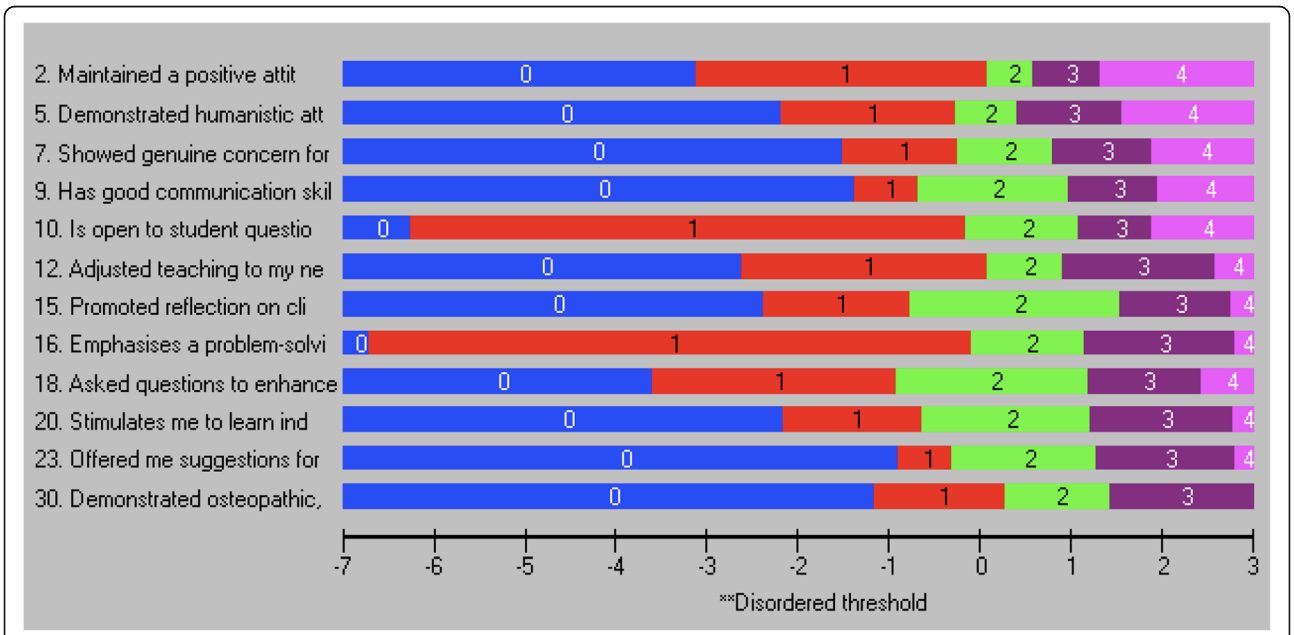


Fig. 5 Threshold map for the 12-item Osteopathy Clinical Teaching Questionnaire

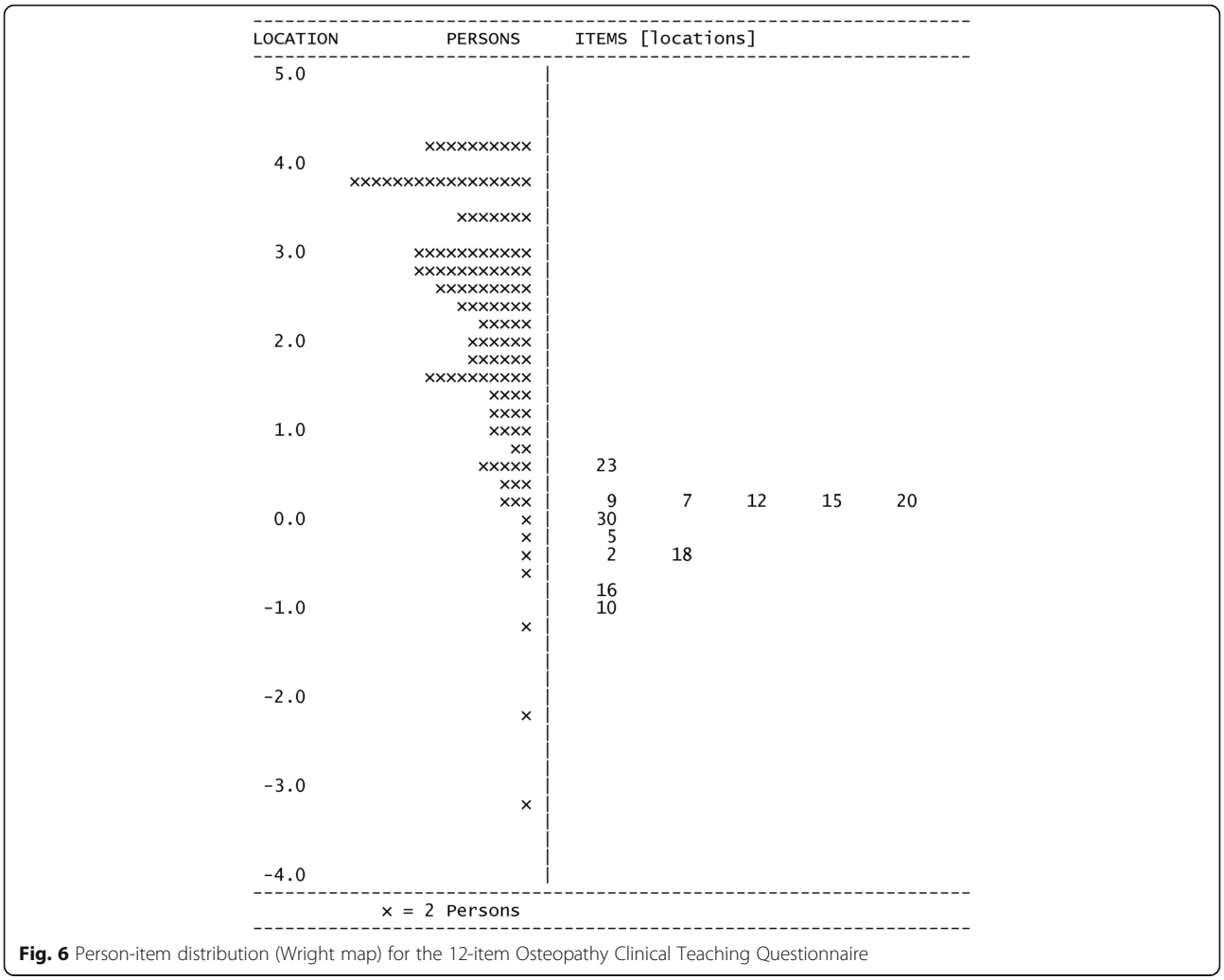


Fig. 6 Person-item distribution (Wright map) for the 12-item Osteopathy Clinical Teaching Questionnaire

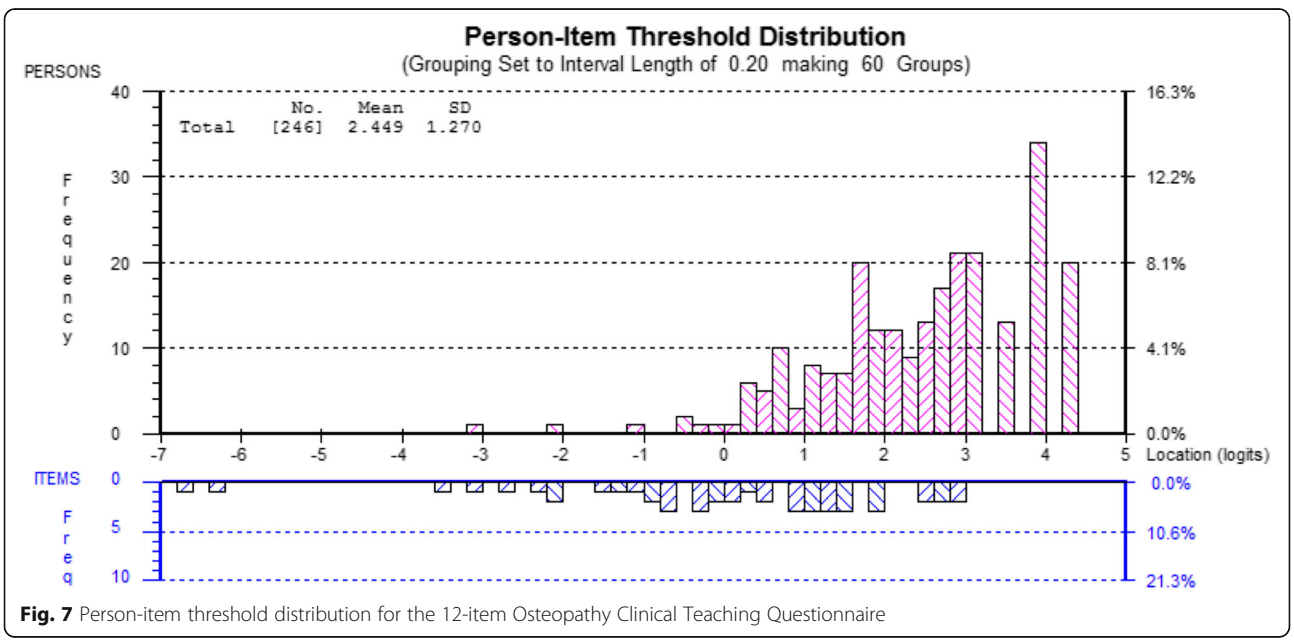


Fig. 7 Person-item threshold distribution for the 12-item Osteopathy Clinical Teaching Questionnaire

Table 4 Principal Component Analysis of the residuals for the 12-item Osteopathy Clinical Teaching Questionnaire

Item	Loading on 'Rasch factor'
2. Maintained a positive attitude towards me	**0.515**
5. Demonstrated humanistic attitudes in relating to patients (integrity, compassion and respect)	**0.396**
7. Showed genuine concern for my professional well-being	0.172
9. Has good communication skills	**0.538**
10. Is open to student questions and alternative approaches to patient management	**0.504**
12. Adjusted teaching to my needs (experience, competence, interest)	**0.351**
15. Promoted reflection on clinical practice	**−0.538**
16. Emphasises a problem-solving approach rather than solutions	**−0.508**
18. Asked questions to enhance my learning	**−0.464**
20. Stimulates me to learn independently	**−0.333**
23. Offered me suggestions for improvement when required	−0.242
30. Demonstrated osteopathic, clinical examination and rehabilitation knowledge and skill(s)	−0.066

Items in bold were used in the t-test

examine a questionnaire related to the clinical education of health profession students [22].

Overall Rasch model fit

Data presented here suggest that the 12-item OCTQ satisfies the requirements of the Rasch model, that is, invariant measurement [45]. Rasch measurement "can be viewed as a psychometric model that can meet the requirements of IM [invariant measurement] when there is acceptable model-data fit" ([45], p.1375). Given an acceptable fit to the Rasch model was achieved after modifications, it is reasonable to conclude that these the 12-item OCTQ demonstrates the properties of invariant measurement. Further, the calculation of a total or summed score for the OCTQ is valid [46–48].

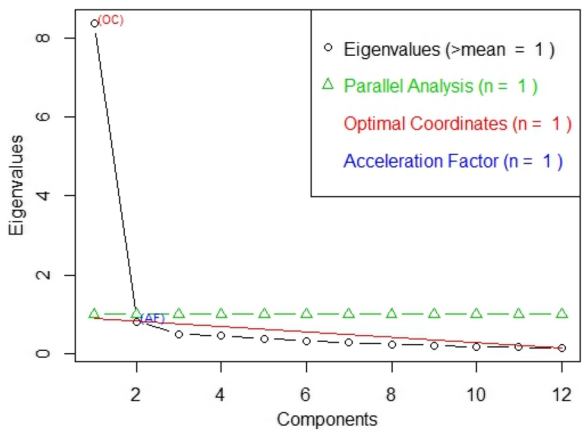


Fig. 8 Number of factors to extract for the 12-item Osteopathy Clinical Teaching Questionnaire

Modifications to fit the Rasch model

Person misfit

Work by Curtis [49] has demonstrated that "…the inclusion of responses that underfit the Rasch measurement model, and that may reflect carelessness in responding, increase the standard errors of item estimates, reduce the range of item locations on the scale, and reduce the inter-threshold range within items" (p. 141). Further, Curtis's work suggests that approximately 30% of respondents to an attitude survey may misfit (over- or under-fit) the Rasch model and require removal. This is consistent with the present study where 38% of the responses misfit the Rasch model. One example of why a response may be removed is that the participant circled all 5 s (strongly agree) for each item albeit that they are unlikely to strongly agree with each item [49]. By removing these responses in a systematic manner (Additional file 3), a fit to the Rasch model can be achieved.

Item misfit

Eighteen items did not fit the Rasch model at various stages during each iteration and were removed from the analysis in order to improve fit (Additional file 3). Both poor fit residuals and significant Bonferonni-adjusted chi-square values were observed initially for items 24 Identified areas needing improvement, and 25 Identified my strengths (Additional file 3). The wording of these

Table 5 Reliability estimates for the 12-item Osteopathy Clinical Teaching Questionnaire

Reliability estimate	Raw data	Polychoric correlation
Cronbach's α	0.94	0.96
Guttman λ_6	0.94	0.97
McDonald $\omega_{hierarical}$	0.86	0.87
McDonald ω_{total}	0.95	0.97

Table 6 Descriptive statistics for the 12-item Osteopathy Clinical Teaching Questionnaire

Item	Mean	SD	Median	Minimum	Maximum	Standard error
2. Maintained a positive attitude towards me	4.46	0.89	5	1	5	0.04
5. Demonstrated humanistic attitudes in relating to patients (integrity, compassion and respect)	4.47	0.83	5	1	5	0.04
7. Showed genuine concern for my professional well-being	4.35	0.95	5	1	5	0.05
9. Has good communication skills	4.35	0.90	5	1	5	0.04
10. Is open to student questions and alternative approaches to patient management	4.34	0.93	5	1	5	0.05
12. Adjusted teaching to my needs (experience, competence, interest)	4.22	0.98	5	1	5	0.05
15. Promoted reflection on clinical practice	4.17	0.93	4	1	5	0.05
16. Emphasises a problem-solving approach rather than solutions	4.21	0.92	4	1	5	0.05
18. Asked questions to enhance my learning	4.27	0.91	5	1	5	0.05
20. Stimulates me to learn independently	4.23	0.89	4	1	5	0.04
23. Offered me suggestions for improvement when required	4.19	0.94	4	1	5	0.05
30. Demonstrated osteopathic, clinical examination and rehabilitation knowledge and skill(s)	3.54	0.72	4	1	4	0.04
Total score	50.83	8.40	53	17	59	0.42

items may have contributed to the misfit in that they are not specific to the feedback being provided by the clinical teacher as a part of their workplace learning, and rewording may improve their fit. Item 27 Promoted keeping of medical records in a way that is thorough, legible, efficient and organised demonstrated fit issues as observed on the ICC and the category probability curve, suggesting the item is measuring a construct that is inconsistent with the other OCTQ items. Respondents are being asked to make a judgment about the clinical teacher on a number of aspects of record keeping within the one item (i.e. thoroughness, organisation, efficiency), and this may be contributing to the misfit.

Item modification to achieve model fit

A strength of the Rasch measurement approach is the ability to rescore then reanalyse the fit of the item to the model [17, 50]. This approach can ensure that items that measure the latent construct are not removed when they can be modified to ensure a fit to the Rasch model. Respondents in the present study did not appear to be using the Strongly disagree and Disagree categories for item 30 Demonstrated osteopathic, clinical examination and rehabilitation knowledge and skill(s) – only 9 respondents (2%) used these categories. Although there was only a small number in each category, the responses to this item were inconsistent with that expected by the Rasch model. This item did not demonstrate misfit even though the thresholds were not being utilised in an ordered manner, a possibility identified by Hagquist [51]. Threshold disordering can introduce "…noise into the measurement" ([52], p.4733) and needs to be resolved either through removing the item or rescoring so that "… the threshold estimates located on the latent trait must appear in the same order as the manifest categories" ([51], p.514). It was appropriate to collapse the Strongly disagree and Disagree categories together as respondents are providing a negative response to the item regardless of which of the two categories they select. Item fit statistics improved with the rescoring of this item [18]. There are two important elements to note: 1) the item still has five categories on the questionnaire itself, it is only the scoring of the item during an analysis that changes (Additional file 6); and 2) it is not possible to collapse two categories together that do not make sense (i.e. collapse disagree with the neutral response category).

Differential item function

The presence of differential item function (DIF) violates the assumption of invariant measurement. Multiple items demonstrated DIF for institution, clinical teacher gender, and student gender in the present study. It is possible for one item demonstrating DIF to influence the fit of other items to the Rasch model [47, 51]. Item 14 Encouraged me to think demonstrated systematic differences for the same level of the underlying trait with regard to the three person-factors investigated. This systematic difference is termed uniform-DIF and given its presence across the three person-factors (institution, clinical teacher gender, and student gender). Pallant and Tennant [18] suggest such items are candidates for removal. None of the items in the present study demonstrated 'artificial DIF' [47]. No item in the 12-item OCTQ demonstrated either uniform or non-uniform DIF, meeting one of the assumptions of invariant measurement [47]. The analysis in the present study demonstrates that some items used in clinical teaching evaluations are affected by DIF and authors of subsequent

evaluations should consider investigating the presence of DIF in the items contained within their questionnaire.

Person separation

The OCTQ PSI is acceptable and is sufficient to separate different levels of the underlying trait as perceived by the respondents [53]. A PSI of 0.85 or greater is reported to indicate a questionnaire is appropriate for decision-making about individuals (clinical teachers in the present study) [54] and the value of 0.827 in the present study suggests that the OCTQ could be used to make these decisions. This PSI value indicates there are likely four strata for the OCTQ [29] - this information could provide a degree of certainty in the decision making process. Such support is valuable, particularly where lower performing clinical teachers are identified for remediation by providing a statistical basis for the decision. Clinical teachers with a total OCTQ score of more than 32 are likely performing at an appropriate level as this value relates to the 4th strata (Additional file 6). Conversely, educators with a total OCTQ score of 25–32 (3rd strata) could be assisted with formal professional development activities or mentoring.

Binomial dimensionality testing & factor extraction

One of the strengths of the current study is the evidence-informed approach to the testing of the dimensionality of the OCTQ. Following the suggestions of Hagell [55, 56] and Engelhard Jr. [45], multiple methods were utilised to investigate the dimensionality of the OCTQ. The 95% confidence interval for the t-test between the items that loaded positively and negatively on the 'Rasch factor' contained the *p*-value of 0.05. Further, the number of factors to extract using the four extraction calculations was one, supporting the argument that the 12-item OCTQ is unidimensional.

Reliability estimates

Further evidence for the unidimensionality of the OCTQ is provided through the ω and α values being well over the accepted value of 0.80. Although it has been suggested that the upper limit for α should be 0.90 [57] and values greater than this may indicate item redundancy [58], the fact that none of the 12 OCTQ items demonstrate local dependency ($r < 0.20$) suggests item redundancy is unlikely to be an issue.

McDonald's ω_h [59] was also calculated for the OCTQ in order to investigate whether the items correlated more strongly with a general factor versus subfactors, and this was the case as evidenced by the path diagram at Additional file 5. ω_t is the estimate of the total reliability of a questionnaire including both the general factor and subfactors [44, 60]. The ω_t value in the present study is consistent with the Cronbach's α value. Both the α and ω values suggest that over 94% of the total questionnaire score variance is due to all the factors in the model (both general and subfactors). ω_h on the other hand has been reported to be the most appropriate reliability estimation method [37] and represents the total questionnaire score variance due to the general factor or latent trait being measured [60]. In the present study over 85% of the total OCTQ score variance is due to the general factor as evidenced by ω_h values of 0.86 and 0.89 for the raw data and polychoric correlation respectively. These values are well above the 0.50 suggested by Revelle [40] and 0.70 suggested by Hermsen et al. [60], supporting the argument for unidimensionality of the OCTQ. Further support is provided by a large explained common variance (ECV) of 0.82 or 82% for the general factor using the raw data, and 84% for the polychoric correlation. The present study utilises multiple methods to provide evidence for the unidimensionality of a Rasch-derived measure.

Targeting

The targeting of the thresholds of the OCTQ items covers a range of levels on the latent trait, particularly in the middle and lower ends of the scale. This targeting potentially allows the OCTQ to be used to identify clinical teachers who are perceived by the students to be performing suboptimally [61]. That said, respondents in the present study typically rated their clinical teachers highly, and this is consistent with reviews by Beckman et al. [62] and Fluit [7] on the validity evidence of clinical teaching evaluations. This potential ceiling effect is demonstrated by the mean person location value of 2.45. Whether this ceiling effect could be, or should be, reduced through modification of the response options is a matter for debate, as some of the clinical teachers in the present study could already be performing highly [61].

Support for the accuracy of the item and person location values is provided by the fact the initial 399 responses, and final 246 responses that demonstrated fit to the Rasch model, are greater than the sample size suggested by Linacre [63] and Pallant and Tennant [18]. Work by Linacre [63] suggests that a sample size of 243 will provide item and person location values that are accurate, regardless of the targeting of the scale.

Developing the validity argument

The framework proposed by Kane [23] covers four stages: *scoring*, *generalisation*, *extrapolation* and *implications* and requires an initial definition of the latent construct under consideration. In the present study the latent construct is quality of clinical teaching provided in osteopathy on-campus, student-led clinics. Previous work has provided evidence to argue for the validity of the scores derived from version 2 of the OCTQ [16],

particularly the *scoring* and *generalisation* arguments. The present study strengthens the *scoring* argument by evaluating the fit of the items and responses to the Rasch model, producing a questionnaire that meets the assumptions of invariant measurement. The unidimensionality of the OCTQ also provides a total score to estimate the latent construct thereby satisfying the requirements of a sufficient statistic, and provides further evidence for the *scoring* argument. The method by which the OCTQ is scored, along with the raw score-to-Rasch score conversion, is presented in Additional file 6. The total score (calculated by adding up each of the 12 items on the OCTQ) can be converted from an ordinal level raw score to a Rasch-derived interval level score that can be used in parametric statistical analyses. This data can then be used to evaluate the impact of faculty development activities, or track changes in clinical teacher performance over time.

The *generalisation* argument is also strengthened in that responses were collected from multiple students, rating multiple clinical teachers, at multiple institutions, in multiple countries. Initial development of version 2 of the OCTQ was focused on one institution, and the inclusion of institutions from New Zealand and the United Kingdom, in addition to Australia, progresses the *generalisation* argument. The evaluation of DIF, and subsequent removal of items that demonstrated this feature, provides evidence for the generalisation argument in that no item in the 12-item OCTQ produces different responses according to student gender, clinical teacher gender or institution.

Initial evidence for the *extrapolation* argument is also provided in the form of the OCTQ total score and item thresholds. The total score can be used to make judgements about the performance of a clinical teacher based on a students' perception, and given the fit of the items to the Rasch model their thresholds can be used to differentiate between levels of clinical teacher performance at item level. These inferences are also supported by a PSI of over 0.80 for the OCTQ. Support for the *implications* argument is also presented in the form of the four statistically discrete strata that separate clinical teacher performance. By applying these strata, program administrators may be able to identify clinical teachers who would benefit from professional development activities or mentoring, as well as identifying those performing at the required level.

It is important to note that those elements described above are only parts of the validity argument, and not the argument as a whole. Further work is required to provide evidence for other aspects of Kane's argument, particularly *generalisation* and *implications*, and this will be the subject of subsequent investigations using the 12-item OCTQ.

Limitations of the study

Although the number of responses received was sufficient to undertake a Rasch analysis, the generalisability of the OCTQ is potentially limited to Australia, New Zealand and United Kingdom osteopathy teaching institutions. Further work would be required to argue for its use in teaching institutions in continental Europe, particularly around the validity of translations. There is also likely to be a degree of profession specificity in that the OCTQ has only been tested in the osteopathy profession. That said, it is possible that the questionnaire could be applied in other on-campus, student-led clinical teaching environments in professions such as chiropractic and podiatry, with only minor modifications. This assumption requires further testing. Item removal and modification was based both on the various fit statistics produced by RUMM, and the opinion of the author. Possible reasons for the removal of the 18 items from the OCTQ could have been explored through the use of a qualitative approach (asking students why they answered items in a particular way), either confirming the removal of the item or suggesting modifications for further testing. Further research is also required to investigate the relationship between the OCTQ and student age, and year level in their respective programs. Additionally, the influence of clinical educator demographics on OCTQ scores provides another avenue for investigation.

Conclusion

The preceding analysis and discussion has provided further evidence to support the developing validity argument for the scores derived from the OCTQ, consistent with Kane's approach to validity [23]. The present study has provided evidence to argue for the construct validity of a 12-item version of the OCTQ. The OCTQ is the first clinical teaching evaluation questionnaire to be developed using Rasch analysis during its initial stages, ensuring that it meets the assumptions of invariant measurement. Fit of the OCTQ items to the Rasch model and unidimensionality were achieved. Further evidence of unidimensionality was demonstrated through the omega hierarchical reliability estimate. The use of five response categories (*Strongly disagree* to *Strongly agree*) for 11 of the 12 items in the final version of the OCTQ is also supported by their fit to the Rasch model. Together this information supports the validity of using the total OCTQ score as a sufficient statistic representing the latent construct of clinical teaching quality in osteopathy.

Items were included in 4 of the 5 factors identified by Vaughan [16] in the initial development of the OCTQ. The learning environment, feedback, reflective practice and modelling factors all contributed items to the 12-item

OCTQ however no item was drawn from the Patient Management factor.

The OCTQ has a number of uses. Firstly the questionnaire can be used as part of a quality assurance strategy in the clinical education component of a teaching program. Secondly, the results obtained from the OCTQ questionnaire can be used to inform faculty development or professional development activities to improve the clinical education experience for students and the educators, potentially improving patient care. Thirdly, the questionnaire has the potential to provide a focus for professional development activities. Finally, there is the potential for the questionnaire to be evaluated for use in allied health student-led clinics (or university clinics), including other non-United States osteopathy programs.

Further research is now required to evaluate the reliability of the 12-item OCTQ to strengthen the validity argument and determine how many evaluations need to be completed by students in order to obtain a reliable indication as to the quality of clinical teaching provided by a clinical teacher in osteopathy on-campus, student-led clinics.

Additional files

Additional file 1: Osteopathy Clinical Teaching Questionnaire. (PDF 67 kb)

Additional file 2: Residual correlations. (PDF 35 kb)

Additional file 3: Steps in the Rasch analysis of the Osteopathy Clinical Teaching Questionnaire. (PDF 127 kb)

Additional file 4: Rationale for items retained or removed. (PDF 28 kb)

Additional file 5: McDonald's *omega* path diagrams for the 12-item Osteopathy Clinical Teaching Questionnaire. (PDF 117 kb)

Additional file 6: Scoring structure for the 12-item Osteopathy Clinical Teaching Questionnaire. (PDF 93 kb)

Acknowledgements
The author would like to thank the students and clinical teachers from each of the participating institutions. The author would also like to thank Stephen Dullard (Southern Cross University), Associate Professor Sandra Grace (Southern Cross University), Paul Orrock (Southern Cross University), Raymond Blaich (Southern Cross University), Sue Palfreyman (Unitec Institute of Technology), Rob Moran (Unitec Institute of Technology), Fran Wiggins (British School of Osteopathy), and Simeon London (British School of Osteopathy) for their assistance with data collection. In addition, thanks are extended to Associate Professor Patrick McLaughlin, Dr. Keri Moore and Dr. Jane Mulcahy for their constructive comments on the manuscript.

Funding
This work was supported by an Australian Government Research Training Program Scholarship.

Availability of data and materials
The data set(s) supporting the results of this article are available in the *figshare* repository, https://doi.org/10.6084/m9.figshare.c.3816553.

Authors' contributions
The author designed the study, undertook the data analysis, developed the manuscript and approved the final version.

Authors' information
Brett Vaughan is a lecturer in the College of Health & Biomedicine, Victoria University, Melbourne, Australia and a Professional Fellow in the School of Health & Human Sciences at Southern Cross University, Lismore, New South Wales, Australia. His interests centre on competency and fitness-to-practice assessments, and clinical education in allied health.

Ethics approval and consent to participate
Ethics approval was provided by the Victoria University Human Research Ethics Committee (HRE15–238). Consent to participate was implied by the return of a completed questionnaire(s).

Consent for publication
Not applicable.

Competing interests
The author declares that they have no competing interests.

Publisher's Note
Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Received: 19 October 2017 Accepted: 23 March 2018
Published online: 03 May 2018

References
1. Vaughan B, Macfarlane C, Florentine P. Clinical education in the osteopathy program at Victoria University. Int J Osteopath Med. 2014;17(3):199–205.
2. Collins A. Cognitive apprenticeship: making things visible. Am Educ. 1991; 15(3):6–11. 38-46
3. Collins A. Cognitive apprenticeship. In: Sawyer RK, editor. The Cambridge handbook of the learning sciences. Cambridge: Cambridge University Press; 2005. p. 47–60.
4. Allan J, O'Meara P, Pope R, et al. The role of context in establishing university clinics. Health Soc Care Community. 2011;19(2):217–24.
5. Quality Assurance Agency for Higher Education. Subject Benchmark Statement: Osteopathy. London: United Kingdom Quality Code for Higher Education; 2015.
6. Beckman TJ, Ghosh AK, Cook DA, et al. How reliable are assessments of clinical teaching? J Gen Intern Med. 2004;19(9):971–7.
7. Fluit C. Assessing the quality of clinical teachers. J Gen Intern Med. 2010; 25(12):1337–45.
8. Fabrigar LR, Wegener DT, Maccallum RC, et al. Evaluating the use of exploratory factor analysis in psychological research. Psychol Methods. 1999;4(3):272.
9. Conway JM, Huffcutt AI. A review and evaluation of exploratory factor analysis practices in organizational research. Organ Res Methods. 2003; 6(2):147–68.
10. Park HS, Dailey R, Lemus D. The use of exploratory factor analysis and principal components analysis in communication research. Hum Commun Res. 2002; 28(4):562–77.
11. Gaskin CJ, Happell B. On exploratory factor analysis: a review of recent evidence, an assessment of current practice, and recommendations for future use. Int J Nurs Stud. 2014;51:511–21.
12. Worthington RL, Whittaker TA. Scale development research a content analysis and recommendations for best practices. Couns Psychol. 2006;34(6):806–38.
13. Schmitt TA. Current methodological considerations in exploratory and confirmatory factor analysis. J Psychoeduc Assess. 2011;29(4):304–21.
14. Stalmeijer RE, Dolmans D, Wolfhagen H, et al. The Maastricht clinical teaching questionnaire (MTCQ) as a valid and reliable instrument for the evaluation of clinical teachers. Acad Med. 2010;85(11):1732–8.
15. Fluit C, Bolhuis S, Grol R, et al. Evaluation and feedback for effective clinical teaching in postgraduate medical education: validation of an assessment instrument incorporating the CanMEDS roles. Med Teach. 2012;34(11):893–901.
16. Vaughan B. Developing a clinical teaching quality questionnaire for use in a university osteopathic pre-registration teaching program. BMC Med Educ. 2015;15(1):70.
17. da Rocha NS, Chachamovich E, de Almeida Fleck MP, et al. An introduction to Rasch analysis for psychiatric practice and research. J Psychiatr Res. 2013; 47(2):141–8.

18. Pallant JF, Tennant A. An introduction to the Rasch measurement model: an example using the hospital anxiety and depression scale (HADS). Br J Clin Psychol. 2007;46(1):1–18.
19. Tennant A, Conaghan PG. The Rasch measurement model in rheumatology: what is it and why use it? When should it be applied, and what should one look for in a Rasch paper? Arthritis Care Res. 2007;57(8):1358–62.
20. Tennant A, Mckenna SP, Hagell P. Application of Rasch analysis in the development and application of quality of life instruments. Value Health. 2004; 7(s1:S22–6.
21. Prieto L, Alonso J, Lamarca R. Classical test theory versus Rasch analysis for quality of life questionnaire reduction. Health Qual Life Outcomes. 2003;1(1):27.
22. Winstanley J, White E. The MCSS-26©: revision of the Manchester clinical supervision scale© using the Rasch measurement model. J Nurs Meas. 2011; 19(3):160–78.
23. Kane MT. Validation. In: Brennan R, editor. Educational measurement. 4th ed. Westport, USA: Praeger Publishers; 2006. p. 17–64.
24. Fluit CR, Feskens R, Bolhuis S, et al. Understanding resident ratings of teaching in the workplace: a multi-Centre study. Adv Health Sci Educ. 2014;20(3):691–707.
25. R Core Team. R: A language and environment for statistical computing Vienna, Austria: R Foundation for Statistical Computing; 2015; Version 3.2.2: Available from: https://www.R-project.org/. Accessed 9 Sept 2015.
26. Revelle W. Psych; procedures for personality and psychological research. 1.5. 4. Evanston, Illinois, USA: Northwestern University; 2015.
27. Andrich D, Sheridan B, Luo G. Rasch models for measurement: RUMM2030. Perth, Western Australia: RUMM Laboratory Pty Ltd.; 2010.
28. Baghaei P. Local dependency and Rasch measures. Rasch Measurement Transactions. 2008;21(3):1105–6.
29. Fisher W. Rating scale instrument quality criteria. Rasch Measurement Transactions. 2007;21(1):1095.
30. Wright BD. Separation, reliability and skewed distributions: statistically different levels of performance. Rasch Measurement Transactions. 2001; 14(4):786.
31. Doari-Raj S. binom. Binomial Confidence Intervals For Several Parameterizations 2015; Version 1.1–1: Available from: http://cran.r-project.org/web/packages/binom/index.html. Accessed 25 Sept 2015.
32. Garrido LE, Abad FJ, Ponsoda V. A new look at Horn's parallel analysis with ordinal variables. Psychol Methods. 2013;18(4):454.
33. Raîche G, Walls TA, Magis D, et al. Non-graphical solutions for Cattell's scree test. Methodology. 2013;9(1):23.
34. Raiche G, Magis D. nfactors: Parallel analysis and non graphical solutions to the Cattell Scree Test 2010; Version 2.3.3: Available from: http://cran.R-project.org/package=nFactors. Accessed 25 Sept 2015.
35. Fox J. Polycor: polychoric and polyserial correlations 2007; Version 0.7–8: Available from: http://cran.Rproject.org/package=polycor. Accessed 25 Sept 2015.
36. Revelle W, Zinbarg RE. Coefficients alpha, beta, omega, and the glb: comments on Sijtsma. Psychometrika. 2009;74(1):145–54.
37. Zinbarg RE, Revelle W, Yovel I, et al. Cronbach's α, Revelle's β, and McDonald's ω H: their relations with each other and two alternative conceptualizations of reliability. Psychometrika. 2005;70(1):123–33.
38. Zinbarg RE, Yovel I, Revelle W, et al. Estimating generalizability to a latent variable common to all of a scale's indicators: a comparison of estimators for ωh. Appl Psychol Meas. 2006;30(2):121–44.
39. Reise SP. The rediscovery of bifactor measurement models. Multivariate Behav Res. 2012;47(5):667–96.
40. Revelle W. Hierarchical cluster analysis and the internal structure of tests. Multivariate Behav Res. 1979;14(1):57–74.
41. Gadermann AM, Guhn M, Zumbo BD. Estimating ordinal reliability for Likert-type and ordinal item response data: A conceptual, empirical, and practical guide. Practical Assessment, Research & Evaluation. 2012;17(3):1–13.
42. Zumbo BD, Gadermann AM, Zeisser C. Ordinal versions of coefficients alpha and theta for Likert rating scales. J Mod Appl Stat Methods. 2007;6(1):4.
43. Rios J, Wells C. Validity evidence based on internal structure. Psicothema. 2014;26(1):108–16.
44. Revelle W Calculate McDonald's omega estimates of general and total factor saturation 2015 [cited 4 Aug 2015]. Available from: http://www.personality-project.org/r/html/omega.html
45. Engelhard G Jr. Rasch measurement theory and factor analysis. Rasch Measurement Transactions. 2012;26(3):1375.
46. Chachamovich E, Fleck MP, Trentini CM, et al. Development and validation of the Brazilian version of the attitudes to aging questionnaire (AAQ): an example of merging classical psychometric theory and the Rasch measurement model. Health Qual Life Outcomes. 2008;6(5):1–10.
47. Andrich D, Hagquist C. Real and artificial differential item functioning. J Educ Behav Stat. 2012;37(3):387–416.
48. Meads DM, Bentall RP. Rasch analysis and item reduction of the hypomanic personality scale. Personal Individ Differ. 2008;44(8):1772–83.
49. Curtis DD. Person misfit in attitude surveys: influences, impacts and implications. Int Educ J. 2004;5(2):125.
50. Lamoureux EL, Pallant JF, Pesudovs K, et al. The impact of vision impairment questionnaire: an assessment of its domain structure using confirmatory factor analysis and Rasch analysis. Invest Ophthalmol Vis Sci. 2007;48(3):1001.
51. Hagquist C. Psychometric properties of the PsychoSomatic problems scale: a Rasch analysis on adolescent data. Soc Indic Res. 2008;86(3):511–23.
52. Lamoureux EL, Pallant JF, Pesudovs K, et al. The impact of vision impairment questionnaire: an evaluation of its measurement properties using Rasch analysis. Invest Ophthalmol Vis Sci. 2006;47(11):4732.
53. Hagquist C, Andrich D. Is the sense of coherence-instrument applicable on adolescents? A latent trait analysis using Rasch-modelling. Personal Individ Differ. 2004;36(4):955–68.
54. Mills R, Young C, Nicholas R, et al. Rasch analysis of the fatigue severity scale in multiple sclerosis. Mult Scler. 2008. https://doi.org/10.1177/1352458508096215.
55. Hagell P. Testing rating scale unidimensionality using the principal component analysis (PCA)/t-test protocol with the Rasch model: the primacy of theory over statistics. Open Journal of Statistics. 2014;4. https://doi.org/10.4236/ojs.2014.46044.
56. Hagell P. Testing unidimensionality using the PCA/t-test protocol with the Rasch model: a cautionary note. Rasch Measurement Transactions. 2015; 28(4):1487–8.
57. Streiner DL. Starting at the beginning: an introduction to coefficient alpha and internal consistency. J Pers Assess. 2003;80(1):99–103.
58. Tavakol M, Dennick R. Making sense of Cronbach's alpha. International Journal of Medical Education. 2011;2:53–5.
59. Dunn TJ, Baguley T, Brunsden V. From alpha to omega: a practical solution to the pervasive problem of internal consistency estimation. Br J Psychol. 2014;105(3):399–412.
60. Hermsen LA, Leone SS, Smalbrugge M, et al. Exploring the aggregation of four functional measures in a population of older adults with joint pain and comorbidity. BMC Geriatr. 2013;13(1):119.
61. Copeland HL, Hewson MG. Developing and testing an instrument to measure the effectiveness of clinical teaching in an academic medical Centre. Acad Med. 2000;75(2):161–6.
62. Beckman TJ, Cook DA, Mandrekar JN. What is the validity evidence for assessments of clinical teaching? J Gen Intern Med. 2005;20(12):1159–64.
63. Linacre JM. Sample size and item calibration stability. Rasch Measurement Transactions. 1994;7:28.

Submit your next manuscript to BioMed Central and we will help you at every step:

- We accept pre-submission inquiries
- Our selector tool helps you to find the most relevant journal
- We provide round the clock customer support
- Convenient online submission
- Thorough peer review
- Inclusion in PubMed and all major indexing services
- Maximum visibility for your research

Submit your manuscript at
www.biomedcentral.com/submit

CHAPTER 4

Evaluation of teaching in a student-led clinic environment: assessing the reliability of a questionnaire

Using the 12-item measure described in chapter 3, chapter 4 explores elements of the reliability of the questionnaire. A significant portion of the investigation in this chapter relates to the identification of the number of questionnaires to be completed by individual students in order to produce a reliable decision about the clinical teaching quality of an individual educator. Additional data to support the measurement properties of the questionnaire, including test-retest and intra-rater reliability, are also presented.

International Journal of Osteopathic Medicine 31 (2019) 28–35

Contents lists available at ScienceDirect

International Journal of Osteopathic Medicine

journal homepage: www.elsevier.com/locate/ijosm

Original Article

Evaluation of teaching in a student-led clinic environment: Assessing the reliability of a questionnaire

Brett Vaughan*

College of Health & Biomedicine, Victoria University, Melbourne, Australia

ARTICLE INFO

Keywords:
Educational measurement
Generalizability
validity
Osteopathic medicine
Medical education

ABSTRACT

Introduction: The Osteopathy Clinical Teaching Questionnaire (OCTQ) has been designed to evaluate the quality of clinical teaching in osteopathy student-led clinics. Previous research has provided evidence for the scoring, generalisation and implications components of the validity argument for the OCTQ. The aim of the present study was to assess the reliability of the OCTQ providing further evidence for the validity argument.

Method: Senior students in the final years of two Australian osteopathy programs completed the OCTQ for each clinical educator with whom they had worked with during a semester. Generalisability analysis, test-retest reliability and internal structure estimation were used to investigate the reliability.

Results: Each of the forty-one clinical educators received an average of 5.97 evaluations resulting in a G-coefficient of 0.63. D-studies demonstrated eight evaluations are required for a coefficient of 0.7 and fourteen for a coefficient of 0.8. Test-retest reliability demonstrated substantial to almost perfect agreement for all but one OCTQ item. Internal structure estimations using Cronbach's alpha and McDonald's omega were both 0.93.

Conclusion: The results suggest that the OCTQ is a reliable tool to provide feedback to clinical educators, and potentially, used to inform decisions to reward clinical educators for their performance.

Introduction

The Osteopathy Clinical Teaching Questionnaire (OCTQ) has been developed as a student evaluation of their learning experience with an individual clinical teacher during student-led clinics in the on-campus university clinic environment. Like other evaluations of clinical teaching [1,2], the OCTQ explores the students' opinion of the quality of teaching provided by their clinical educator(s). Work undertaken thus far to develop the OCTQ has been guided by current recommendations with respect to exploration of the questionnaire measurement properties, including the use of classical test [3] and item response theory approaches [4]. The current paper reports on an extension of the development work - to explore the reliability of the OCTQ.

Student evaluations of teaching such as the OCTQ are often used to inform faculty and professional development activities, and potentially curricula changes. There are also numerous examples of where these are used for employment, tenure and promotion decisions by program administrators [2,5]. Where they are used for such decisions, the validity of interpretation of the score derived from the evaluation(s) needs to be established or argued [2]. This argument allows for defensible decision-making and potentially greater engagement with, and trust in, the evaluation process by students, clinical educators, and program administrators. The reliability of a questionnaire also contributes to trustworthy and defensible decision-making when combined with other evidence to develop a validity argument.

The reliability of a questionnaire contributes valuable information to the overall understanding of its measurement properties. Providing such evidence allows stakeholders to make a judgement with respect to the dependability of the scores derived from the questionnaire. Multiple forms of reliability have been reported in the clinical teaching evaluation literature, including internal consistency (typically calculated as Cronbach's alpha) in addition to inter-rater reliability, and test-retest reliability [2,6,7]. More recently, generalisability theory has been applied in the development of clinical teaching evaluation questionnaires [8–15]. However, few studies utilise multiple approaches to the exploration of reliability to support the validity argument [16]. Each reliability statistic has its own limitation, primarily related to the data-driven nature of the calculation of these statistics. Therefore, there is great value in using multiple forms of reliability calculation as this will ensure "… ample evidence be [is] accumulated to establish the reliability of scores before using an instrument in practice" [16].

Kane [17,18] has provided numerous commentaries on the use of an argument-based approach to validity. Kane's framework is predicated

* Brett Vaughan College of Health & Biomedicine, City Flinders Lane Campus, Victoria University, PO Box 14428, Melbourne, VIC, 8001, Australia.
E-mail address: brett.vaughan@vu.edu.au.

https://doi.org/10.1016/j.ijosm.2018.11.001
Received 20 August 2018; Accepted 2 November 2018
1746-0689/ © 2018 Elsevier Ltd. All rights reserved.

on the idea that multiple sources of evidence are required in order to demonstrate validity – one develops an argument for the validity of the interpretation of a score. The premise is that an assessment or evaluation tool in itself cannot be valid, it is the interpretation of the score derived from the tool that may demonstrate validity. Kane outlined a sequential approach to developing and presenting relevant evidence for the validity argument: *scoring*, *generalisation*, *extrapolation* and *implications*. Importantly Kane's approach does not dictate the type of evidence that can be presented for each part of the argument. This allows for research into different contexts to contribute evidence to building an argument, and the flexibility to utilise qualitative, quantitative and mixed methods research designs as part of this approach.

Evidence for the validity argument for the Osteopathy Clinical Teaching Questionnaire (OCTQ) has been presented previously [3,4]. This evidence to date provides support for the *scoring*, *generalisation* and *implication* arguments for the interpretation of the score derived from the OCTQ however the *extrapolation* argument requires further development. A key component of the validity argument that requires addressing from the outset is the proposed interpretation of the score derived from the questionnaire. In the context of the OCTQ, the proposed interpretation is the *quality of clinical teaching provided by a clinical educator in the osteopathy student-led, on-campus clinic environment*. The aim of the current study is to present a suite of reliability statistics to evaluate the generalisability, retest reliability and criterion validity of the Osteopathy Clinical Teaching Questionnaire (OCTQ). This information provides further evidence for the validity argument for the interpretation of the score derived from the OCTQ.

Method

This study was approved by the Victoria University and Southern Cross University Human Research Ethics Committees (HRE15-238 and ECN15-301).

Participants

Students enrolled in year 4 and year 5 of the osteopathy programs at Victoria University (VU) (Melbourne, Australia) and Southern Cross University (SCU) (Lismore, Australia) were invited to complete a paper version of the Osteopathy Clinical Teaching Questionnaire during October and November 2015.

Questionnaire

The Osteopathy Clinical Teaching Questionnaire (OCTQ) [3,4] is a 15-item measure of clinical teaching quality in osteopathy student-led, on-campus clinics – 12 items measuring clinical teaching quality and 3 global rating items. Each of the 12 clinical teaching quality items are scored on a 1 (strongly disagree) to 5 (strongly agree) Likert-type scale and takes approximately 2 min to complete. The reliability estimates for the OCTQ have been calculated as greater than 0.85 and the calculation of a total score (not including the 3 global items) is a sufficient statistic [4]. When calculating the total score, item 12 requires rescoring - the *strongly disagree/disagree* options are collapsed to a 1 with the other three options scored as 2, 3 and 4 respectively. The total available score for the 12 items on the OCTQ is 59 and can be converted to an interval-level score using the table at Appendix 1. Each of the three global items are rated on their own five-point Likert-type scale.

Data collection

At both Victoria University and Southern Cross University, students were asked to complete one questionnaire for each of the osteopathy clinical educators they had worked with in the on-campus, student-led clinics during the second-half of the 2015 academic year. When completing the OCTQ, the student was asked to identify the clinical educator being rated and their own gender however they were not required to personally identify themselves and could complete the questionnaires at a time of their choosing. Completion of an evaluation was not a requirement to receive a grade or satisfactorily complete a clinical unit in either osteopathy program.

At Victoria University, in order to evaluate the test-retest reliability of the OCTQ, students in year 4 only were asked to write a self-generated code on each questionnaire they completed at both time 1 (T1) and time 2 (T2) in order to match responses. Students completed the OCTQ at T1 then undertook their normal one-and-a-half-hour practical skills class. This process was used to ensure that the students were not exposed to any clinical teaching in between the administrations of the questionnaire. At the end of the class (T2), the students were asked to complete the OCTQ again for each educator they rated at T1.

Data analysis

Data were entered into SPSS (version 21) to organise the export to other programs for the data analysis. Item 12 was rescored as per Vaughan [4] where the *strongly disagree* and *disagree* options were collapsed into one. Data were then exported to *R* [19] to generate the descriptive statistics using the *psych* package [20].

Generalisability analysis

Background. Generalisability theory (GT) is based on an extension of the analysis of variance (ANOVA) in classical test theory and is a useful analysis where there are multiple sources of variance (i.e. students, examiners, items, cases, patients) in the total score derived from a measure [21]. The measurement score is decomposed into these multiple sources, or 'facets' in GT, allowing for the evaluation of how the individual facet and its interactions with other facets, contribute to the score variance. The *object of the measurement* [21] or *facet of differentiation* [22] is the entity being measured and the score for this entity is thought to represent the universe score [23]. The other entities are termed either fixed or random *facets of generalization*. A fixed facet is one that, for the purposes of the measurement, does not change (i.e. number of items on the measure) [23]. Conversely, a random facet is one where the number can change as these facets are typically drawn from a universe of entities (i.e. raters, examiners, students). In GT, consideration is also given to whether a facet is crossed or nested within another facet [21]. A facet is crossed with another where each level of the facet interacts with the other (i.e. all students are assessed by all examiners). Conversely a facet is nested when one level of the facet interacts with a portion of the other facet (i.e. some students are assessed by one examiner, and other students by another examiner).

Together this information is used to design the generalisability analysis. Bloch and Norman [21] summarise the desired outcome of a generalisability analysis as

> "… to what extent can we extrapolate the results achieved on a limited sample of test tasks, measured under unique test conditions to a universe of tasks and conditions, from which the specific test set has been drawn more or less arbitrarily".

The generalisability analysis produces variance component values and two reliability coefficients, $E\rho^2$ (relative error) and Φ (absolute error). The variance components can be converted to percentages by dividing them by the total score variance in order to assist in their interpretation [23]. The decision about which coefficient to report is based on whether the desire is to make a comparison with other individuals within the *facet of generalization* ($E\rho^2$) (i.e. normative decision making) or the interest is only in the individual (Φ) [21,23].

The previous is a description of a G study – the derivation of a reliability coefficient by estimating the variance attributable to different facets [23]. GT allows for the variation of the *facets of generalization* to model the impact of changing the number or level of the facet on the reliability coefficient. This is referred to as a decision (D) study and is

designed to develop "… a measurement that minimizes error for a particular purpose" [24]. A more in-depth review of GT, its mathematical basis and the practical applications beyond that described in the present study is provided by Bloch and Norman [21].

Analysis. A generalisability analysis was performed to determine the variance components and reliability of the OCTQ and ascertain how many questionnaires are required for a reliable decision. The design of the present study was i x (e:s) where 'i' represents the 12 OCTQ items, 'e' represents the clinical educators and 's' represents the students. This design is consistent with other G studies of clinical teaching evaluations [8,12–15], and is succinctly described by Bloch and Norman [21]:

> "Teacher [clinical educator] is the facet of differentiation, and here student (rater), a facet of generalization, is nested in teacher since each student belongs to only one section. But item (on the scale) is crossed with teacher, since all teachers are rated on the same items"

The design was unbalanced in that each clinical educator received a different number of ratings from the students. The Φ coefficient was reported as the performance of an individual clinical educator was not compared to that of their peers (an *absolute* rather than *relative* decision in GT terminology).

Using the variance components, a D study was performed to determine how many questionnaires would be required to achieve a generalisability coefficient of 0.80 [23]. This value is widely considered to be acceptable for high-stakes decision making. Data were imported into *G_string IV* [25] for the generalisability analysis (G-study) and decision study (D-study). The number of *Items* was fixed at 12 as these were the only items of interest in the present study [7,13].

Temporal stability (test-retest reliability)

Temporal stability was evaluated using weighted kappa in *R* [19] using the *psych* package [20]. Norman and Streiner [26] have previously reported that weighted kappa is equivalent to an intraclass correlation coefficient ($ICC_{2,1}$) and can be used interchangeably. Weighted kappa was chosen as the data were ordinal in nature. The interpretation of the weighted kappa was based on Landis and Koch [27]: < 0 *Less than chance agreement*; 0.01–0.20 *Slight agreement*; 0.21–0.40 *Fair agreement*; 0.41–0.60 *Moderate agreement*; 0.61–0.80 *Substantial agreement*; and 0.81–0.99 Almost *perfect agreement*.

Relationship between global ratings and total scores

The relationship between the three global ratings and total Rasch-converted OCTQ score was evaluated using Spearman's rho (ρ) and interpreted according to Hinkle, Wiersma [28]: 0–0.30 (negligible); 0.30–0.50 (low); 0.50–0.70 (moderate); 0.70–0.90 (high); 0.90–1.00 (very high). The total Rasch-converted OCTQ score [4] was used as the data are interval-level rather than ordinal potentially allowing for greater sensitivity in the difference between total questionnaire scores.

Reliability estimation

Cronbach's alpha (α) and McDonald's omega total (ωt) (and their confidence intervals) were calculated as the reliability estimates using the *MBESS* package [29] in *R* [19]. Bootstrapping was applied to confirm the CI's using 1000, 5000 and 10000 iterations. Only the responses from T1 were used for the analysis and as the OCTQ has been shown to be unidimensional [4] calculating both reliability estimates using all 12 OCTQ items is appropriate. Item-total correlations and alpha if item deleted statistics were generated using the *psych* package [20]. Cronbach's alpha is a widely reported reliability estimate [30] however numerous authors have highlighted issues with its use as a measure of internal consistency [30–32]. That said, there is little agreement in the literature about which estimates to calculate. Dunn, Baguley [30] suggest that McDonald's omega (ω) is a suitable alternative particularly where the tau-equivalence assumption for α is violated.

The test-retest Cronbach's alpha was calculated [33] in order to evaluate the effects of transient error in the test-retest administration of the OCTQ. Effects of item recall on alpha is reported to be less for test-retest alpha than the test-retest correlation, and an underestimate of reliability [33]. Tau-equivalency is an assumption underlying the use of test-retest alpha however Green [33] suggests this assumption is reported to be largely met in measures that demonstrate unidimensionality. Previous research has provided evidence for the unidimensionality of the OCTQ [4] therefore use of Cronbach's alpha is acceptable.

McDonald's omega hierarchical (ωh) provides an indication as to the reliable variance attributable to the general factor or latent construct being measured. Values greater than 0.50 [34] and 0.70 [35] have been suggested as acceptable. ωh was calculated using the *psych* package [20] in *R* [19].

Table 1
Overview of clinical educator and student responses.

		Total	Institution	
			Victoria University	Southern Cross University
Total responses		304	199 (65.4%)	105 (34.6%)
Responses by student gender	Male	156	96 (48.2%)	60 (57.1%)
	Female	136	95 (47.7%)	41 (39.0%)
Responses by clinical educator gender	Male	196	117 (58.8%)	79 (75.2%)
	Female	107	81 (40.7%)	26 (24.8%)

Note: some participants did not indicate either their gender or the gender of the clinical educator being rated.

Results

Three-hundred and four questionnaires were received evaluating 41 of the 44 clinical educators across the two institutions (Table 1). Descriptive statistics for the OCTQ are presented in Table 2.

A G-coefficient of 0.63 was obtained based on an average of 5.97 questionnaires per clinical educator (range - 2 to 15 questionnaires per educator). The variance components are presented in Table 3 and the D-study coefficients are presented in Fig. 1. To provide feedback to an individual clinical educator that is reliable, 8 questionnaires are required to achieve a Φ coefficient of 0.70.

Test-retest reliability results are presented in Table 1. *Substantial* to *almost perfect* agreement was observed for all but item 12 where *fair* agreement was observed. *Almost perfect* agreement was observed for all three global ratings. Test-retest alpha was 0.90. These results suggest the OCTQ items, except for item 12, are stable across a short-term administration. The relationships between the total Rasch-converted OCTQ score and global rating one (I would do more clinics with this Clinical Educator) and two (I would recommend other students to work with this Clinical Educator) were both *moderate* ($\rho = 0.68$, $p < 0.01$). A *high* correlation was observed between the OCTQ total score and global rating three (Rate the overall effectiveness of this Clinical Educator as an educator/supervisor) ($\rho = 0.72$, $p < 0.01$).

Cronbach's α and McDonald's ωt were 0.93 [95%CI 0.91–0.94] and 0.93 [95%CI 0.92–0.94] respectively. These values suggest that approximately 7% of the variance in the OCTQ score is measurement error, and these values did not change in relation to the number of bootstrapping iterations. Item-total correlations were *moderate* or greater (range 0.63–0.83), the inter-item correlation average was 0.51 (range 0.50–0.53) (Table 1), and alpha did not improve if any OCTQ item was deleted. ωh was acceptable at 0.73 and suggests that 73% variance in the total OCTQ score is accounted for by the latent construct of clinical teaching quality in osteopathy student-led, on-campus clinics.

Table 2
Descriptive, intra-examiner and internal structure correlation statistics for the Osteopathy Clinical Teaching Questionnaire.

Item	Mean	St Dev	Median	Minimum	Maximum	Weighted Kappa	95% Confidence Interval	Item-total Correlation	Average Inter-item Correlation
1. Maintained a positive attitude towards me	4.64	0.66	5	1	5	0.86	0.77–0.96	0.70	0.52
2. Demonstrated humanistic attitudes in relating to patients (integrity, compassion and respect)	4.70	0.57	5	2	5	0.78	0.66–0.89	0.64	0.52
3. Showed genuine concern for my professional well-being	4.61	0.65	5	2	5	0.74	0.59–0.89	0.75	0.51
4. Has good communication skills	4.53	0.71	5	2	5	0.72	0.58–0.86	0.81	0.50
5. Is open to student questions and alternative approaches to patient management	4.46	0.81	5	1	5	0.83	0.74–0.92	0.72	0.52
6. Adjusted teaching to my needs (experience, competence, interest)	4.36	0.79	5	1	5	0.81	0.73–0.90	0.83	0.50
7. Promoted reflection on clinical practice	4.38	0.76	5	1	5	0.73	0.59–0.86	0.74	0.51
8. Emphasises a problem-solving approach rather than solutions	4.45	0.76	5	2	5	0.71	0.60–0.82	0.77	0.51
9. Asked questions to enhance my learning	4.45	0.80	5	1	5	0.75	0.64–0.86	0.75	0.51
10. Stimulates me to learn independently	4.39	0.84	5	1	5	0.80	0.69–0.91	0.80	0.50
11. Offered me suggestions for improvement when required	4.55	0.71	5	1	5	0.72	0.60–0.84	0.75	0.51
12. Demonstrated osteopathic, clinical examination and rehabilitation knowledge and skill(s)	3.70	0.55	4	2	4	0.50	0.26–0.75	0.63	0.53
Total score	53.29	6.42	55	28	59				
G1. I would do more clinics with this Clinical Educator	4.57	0.78	5	1	5	0.88	0.80–0.95		
G2. I would recommend other students to work with this Clinical Educator	4.59	0.74		1	5	0.95	0.90–0.99		
G3. Rate the overall effectiveness of this Clinical Educator as an educator/supervisor	4.49	0.76	5	1	5	0.97	0.94–1.00		

Table 3
Variance components for the generalisability study of the Osteopathy Clinical Teaching Questionnaire.

Effect	Degrees of freedom	Sum of Squares	Variance Component	Percentage Variance Component
Educator	39	327.78	0.06	10.5
Student:Educator	256	688.36	0.20	35.1
Items	11	214.99	0.06	10.5
Educator x Items	429	165.35	0.02	3.6
Items x Student:Educator[a]	2816	652.73	0.23	40.3

[a] Residual error.

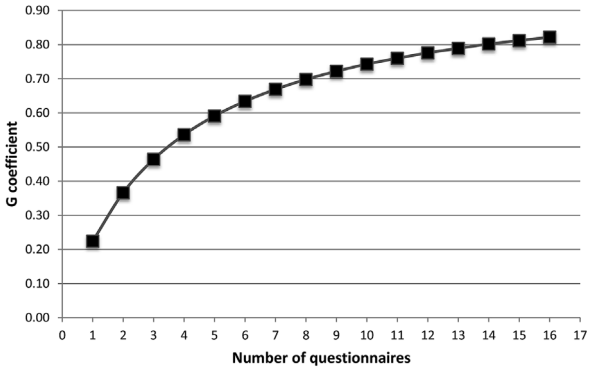


Fig. 1. Number of Osteopathy Clinical Teaching Questionnaires for different reliability coefficients.

Discussion

The present study extends on previous work by evaluating the reliability of the OCTQ to support trustworthy decision-making based on the results of the questionnaire. That is, program administrators and clinical educators can have a sense of 'trust' in the data based on the evidence supporting the validity of the score derived from the questionnaire described here.

Generalisability analysis

The G-study demonstrated that an average of 5.97 questionnaires achieved a Φ coefficient of 0.63. This is consistent with the initial analysis of a 9-item clinical teaching behaviours questionnaire by Keely, Oppenheimer [15]. A D-study using the OCTQ data suggests that questionnaires completed by eight different students on a single clinical educator provides a Φ coefficient of 0.7. This reliability coefficient provides evidence to support the use of the OCTQ as a feedback tool [36]. The information and scores could be used to guide program administrators about faculty development activities targeted to an individual clinical educator for example. In order to utilise the scores derived from the questionnaire for higher-stakes decisions such as promotion, tenure or teaching awards, a single clinical educator would require 14 individual student questionnaires. This number of questionnaires provides a Φ coefficient of 0.8, and is consistent with other clinical teaching evaluations published by Keely, Oppenheimer [15] and Hindman, Dexter [14]. Beyond this number of evaluations the reliability would not substantially increase (Fig. 1).

Calculation of the variance components that underpin these Φ coefficients provides further information about the psychometric properties of the OCTQ. There are no universal guidelines by which to determine the acceptability of a variance component value [23] and they need to be interpreted within the context of the individual study. This can also be a limitation applied to any study of reliability as the calculated coefficients may change, albeit minimally. When combined

with other reliability data, as in the present study, the impact of this limitation is reduced but does not negate the need for other researchers to evaluate reliability in their own clinical education environment.

The decision to report the Φ rather than the Ep^2 coefficient is supported by the small variance component for Educator (10.5%). The variance attributable to the differences between individual clinical educator OCTQ scores is approximately one-tenth [37]. The clinical educators in the present study are likely to already be performing at a high level, as evidenced by the high mean OCTQ score. Therefore, there is little difference between the mean scores for each Educator evaluated in the present study therefore there is little ability to rank the clinical educators based on their total OCTQ scores. Further comparisons between educators are not appropriate, and from a practical standpoint, would represent a normative decision that is unlikely to be defensible. de Oliveira Filho, Dal Mago [12] and Haider, Johnson [13], using the same G-study design as the present study, reported variance components of 56% and 28% respectively for the instructor/teacher [Educator] in their studies. The substantial difference in the variance component is likely due to the fact that de Oliveira Filho, Dal Mago [12] utilised a balanced design, whereas an unbalanced design that is more reflective of the clinical education environment was used in the present study.

The Student:Educator interaction in the present study accounts for the second largest variance (35.1%) and is consistent with other studies [12,14,38]. This result suggests that each student has a different perception of the quality of teaching provided by the clinical educator [12,15,23]. de Oliveira Filho, Dal Mago [12] also suggest that this interaction term could reflect differences in the students' interpretation of the questionnaire items, real differences in teaching quality between clinical educators, and/or a combination of all of these factors. Haider, Johnson [13] demonstrated a 4% variance component for this interaction in their study. These authors argued that stringency (all raters agreed on the same level for the teaching performance) may have accounted for this small variance component. The substantial difference between the variance components for Student:Educator in the present study and those presented by Haider, Johnson [13] provides support for the argument that students are rating each clinical educator differently, whilst using the same items on the OCTQ. This appears to be a strength of the twelve items that comprise the OCTQ.

It is not possible to deconstruct this Student:Educator variance component further using the current design, however it does provide an avenue for future research. The substantial contribution of this interaction to the score variance supports the need to obtain a suitable number of responses to make a trustworthy *absolute* decision, and avoid making *relative* decisions [39].

The Item facet provides an indication as the internal consistency of the measure used in the study [23]. The variance accounted for by Item was 10.5% suggesting the OCTQ items generated relatively similar mean scores (Table 1) with some variation in the level of difficulty of each item [14]. Vaughan [4] demonstrated that the OCTQ items evaluate different levels of clinical educator performance therefore variation in item difficulty is expected. The Item variance component is relatively consistent with the other reliability estimation methods in the present study (α and ωt = 0.93), and those described by Copeland and Hewson [8] (11%) and Haider, Johnson [13] (18%) using alternative questionnaires to evaluate clinical teaching.

The Educator x Item term accounted for 3.6% of the variance, and is somewhat less than the 23% demonstrated by Copeland and Hewson [8] and 16% by Haider, Johnson [13] but more consistent with Hindman, Dexter [14]. Again, this difference could be attributed to the respective study designs. Although these studies treated Student as a random factor, Copeland and Hewson [8] and Haider, Johnson [13] used 5 randomly selected responses whereas the present study utilised all responses, regardless of how many a clinical educator received. Given the random nature of the selected responses, the relative ranking of the clinical educators would change if different students rated the educators in the studies by Copeland and Hewson [8] and Haider, Johnson [13]. In contrast, there would be very little change in the relative rankings in the present study with the small Educator x Item variance component.

Items crossed with Student nested in Educator reflects both random and systematic error variance [10]. This was the largest contribution to the variance of the total OCTQ score at just over 40%, and is consistent with studies by Copeland and Hewson [8] (46%), Hindman, Dexter [14] (43%), and Haider, Johnson [13] (36%). Because of the nested design of the G-study it is not possible to deconstruct this variance component further [10]. This result does however provide some support for the notion that the students may be interpreting the OCTQ items differently [8,13] - a 3-way interaction between the student-clinical educator-OCTQ item [14]. As highlighted previously, this difference in interpretation could be due to an actual difference in the interpretation of the items between students, or the influence of other factors (e.g. age, level of training, clinical learning environment, personality) on item interpretation [10].

A number of studies have investigated the inter-rater reliability of clinical teaching evaluations [12,40–42]. Such studies make the implicit assumption that different students will have the same (or similar) perception of a clinical educator's teaching quality. The G-study presented here demonstrated an inter-rater reliability coefficient of 0.63 suggesting that approximately 6 OCTQ responses would provide a moderate level of inter-rater reliability. Students in the present study did not appear to have the same perception of a single clinical educator. It is known that, amongst other factors, the interpersonal relationship between the clinical educator and an individual student [10,14,43], and the clinical learning environment [44] influence responses to clinical teaching evaluations. These factors may account for the moderate inter-rater reliability. Moderate inter-rater reliability may also be a valuable property of clinical educator evaluations. If all students are consistent in their evaluation of a single clinical educator the utility of the information that can be derived from the questionnaire becomes somewhat limited.

Temporal stability

To ascertain an accurate indication of temporal stability of the OCTQ items, weighted Kappa was used in a subset of the student population reported here. Temporal stability of OCTQ items 1 to 11 was *substantial to almost perfect*. Importantly, no student was exposed to their clinical educator in the period during the data collection. Therefore, it is not possible for any interaction between educator and student to influence these results. Given students were required to rate multiple clinical educators they are less likely to remember the previous rating, reinforcing the temporal stability results. The high test-retest alpha (0.90) and narrow 95% confidence intervals for each of these items further supports the accuracy of these results.

Item 12 was the only item that did not demonstrate substantial to almost perfect agreement. This item evaluates the students' perception of demonstration of skills related to manual therapy practice. Previous work to develop the OCTQ demonstrated that this item was the only one that required rescoring [4]. The *strongly disagree* and *disagree* options were collapsed together to create the OCTQ total score. Whilst this item did demonstrate fit to the Rasch model in a previous study [4], these two issues (item rescoring and moderate agreement) suggest that the item may require further investigation.

A possible reason could be that suggested by Mackillop, Parker-Swift [45] in their study. These authors identified that compound items on a multisource feedback form were less reliable than non-compound items. Item 12 *Demonstrated osteopathic, clinical examination and rehabilitation knowledge and skill(s)* could be classified as a compound item in that it asks the respondent to evaluate the clinical educator on multiple aspects of their performance within the one item. Splitting item 12 may assist in improving its temporal stability. Future

work could investigate two options for the item:

1 Split the item into Demonstrated osteopathic, clinical examination and rehabilitation knowledge and Demonstrated osteopathic, clinical examination and rehabilitation skills; or
2 Split the item into Demonstrated osteopathic knowledge and skills, Demonstrated clinical examination knowledge and skills, and Demonstrated rehabilitation knowledge and skills.

The cognitive load associated with this item may be too high for the student when evaluating their clinical educator. That is, they are being asked to evaluate the clinical educator on multiple areas – the compound item. Adding to the increased cognitive load is the student rating multiple clinical educators in a single administration in the present study. Whether the temporal stability would improve if the student was asked to repeatedly evaluate only one clinical educator requires further investigation.

Relationship with global ratings

Bierer and Hull [46] suggest that clinical teaching evaluations that are to be used for summative decisions should contain a global rating item, and it is posited here that this should also apply to formative decisions. The OCTQ has three global rating items: one to capture a students' overall perception of effectiveness; and two 'satisfaction' items.

Rate the overall effectiveness of this Clinical Educator as an educator/supervisor demonstrated a high correlation with the total Rasch-converted OCTQ score ($\rho = 0.72$) and provides some evidence for criterion validity [11,47]. This high correlation provides support for the argument that the total OCTQ score captures a substantial proportion of the effectiveness of the teaching and supervision provided by the clinical educator [11,48] and is within the 0.40–0.80 range suggested by van der Leeuw, Lombarts [47].

Satisfaction could provide an indicator of quality processes [49]. *I would do more clinics with this Clinical Educator* and *I would recommend other students to work with this Clinical Educator* are the two global satisfaction items on the OCTQ. Both were *moderate* correlations but only slightly less than the effectiveness global rating described above. The correlations suggest that the interpersonal and teaching domains covered by the OCTQ [4] relate to satisfaction with the performance of the clinical educator. However, given the *moderate* correlations, it is likely there are other factors beyond that captured by the OCTQ that influence satisfaction with an educators' performance. Examples of these factors could be level of supervision provided [50], clinical learning environment, and the student-educator personality interaction [10,14,43].

Reliability estimations

Reliability estimates for the OCTQ were also calculated. Consistent with previous work [4] both α and ωt were both above 0.92. In the case of ωt the value of 0.93 indicates the general factor (latent construct) and group factors account for 93% of the variance in the total OCTQ score. The group factors have been described previously [4] and potentially represent the *interpersonal* and *clinical teaching* domains [6]. ωh provides an indication as to the variance attributable to the general factor (latent construct of quality of clinical teaching in osteopathy) [51]. In the present study, 73% of the total score variance is attributable to the general factor only. This is lower than the 86% observed in previous work on the OCTQ [4] but above an acceptable level [35]. This finding supports the previous assertion that reliability coefficients are data driven, will vary between datasets and require calculation with each administration. The current study is also the first to utilise McDonald's omega as a reliability estimation as part of the evidence for the validity of a clinical teaching evaluation, and other authors are encouraged to report the same when using the OCTQ or other clinical teaching evaluation tool.

Cronbach's α was also acceptable and removing an item did not improve the α score. Item-total correlations were also acceptable (0.63–0.81). The α value in the present study was over 0.90 suggesting item redundancy [52]. However removal of an item does not improve the α value, the item-total correlations were not high [15], and the OCTQ was constructed in such a way as to limit the inter-item relationships [4]. Therefore, item redundancy is unlikely to be an issue. Further, other authors of clinical teaching evaluations have demonstrated α values over 0.90 [7,8,12]. The α and ω statistics presented here provide further evidence to support the valid calculation of a total score for the OCTQ as it measures one latent construct – quality of clinical teaching in osteopathy student-led clinics [4].

Validity argument

Kane's validity argument approach [17] suggests evidence be provided for four aspects in order to argue for the validity of the interpretation of a score. Evidence for the *scoring* argument continues to be provided in the form of the reliability estimations. McDonald's omega hierarchal (ωh) [51] was above 0.70 in the present study and a previous study investigating the construct validity of the OCTQ [4]. When combined, both studies provide evidence to support the calculation of a total score for the OCTQ.

The statistical approach employed in the present study provides evidence for the *generalisation* argument. The students and clinical educators are drawn from the possible 'universe' [21,23] of stakeholders in the evaluation of clinical teaching in osteopathy student-led, on-campus clinics. This allows the results of the present study to be generalised to Australian osteopathy programs. Given the development of the OCTQ in a previous study [4] also included evaluations from osteopathy programs in New Zealand and the United Kingdom, it may be possible to argue for the *generalisation* to these programs. However, further evidence is required to support this assertion.

Reliable feedback can be achieved with eight individual students completing an evaluation on a single clinical educator. This information provides evidence for the implications argument in that the OCTQ can be used to provide feedback based on the student's perception of clinical educator performance. Temporal stability of the OCTQ items provides further support however the evidence is tempered by the *fair* intra-rater reliability for item 12. Further work is required to ensure the temporal stability of item 12 before it can be used for high-stakes decision-making – its use for guiding feedback and professional development is acceptable however.

Limitations

There are a number of limitations in the present study. Firstly, the study was undertaken in two Australian teaching institutions which may limit the generalisability of the results to other non-United States osteopathy teaching programs.

The anonymous nature of the responses to the OCTQ prevents a definitive statement about the response rates in the present study. However, there were approximately 130 students eligible to participate across both teaching programs at the time of the study. Therefore more than 300 responses provides an indication that the responses are representative of the 'universe' [23] from which the data could possibly be drawn at the time of the study. Each Clinical Educator also received six evaluations on average which represents the number of students the typical educator would supervise in an Australian osteopathy student-led, on-campus clinical setting [53].

Whilst the average number of questionnaires completed for each clinical educator was six, there was a range from two to fifteen. It is possible that a student may have only evaluated one or two clinical educators when they worked with more over the period of the study. This may have also led to the situation where the student chose to

evaluate only the clinical educators that they had a strong opinion about (whether positive or negative). Therefore, those clinical educators performing at the expected level may not have been evaluated.

The temporal stability of the OCTQ items could have been evaluated within the G-study. The G-study design would have included an Occasions facet representing the multiple administrations of the OCTQ [10]. This was not possible for logistical reasons as the ability to sample students between their clinic times was limited by timetabling. Such a study design is important to ensure that the students are not exposed to the clinical educators whom they are evaluating in the period between administrations [10].

Conclusion

The present study sought to investigate the reliability and number of questionnaires that would be required to be completed by students for a single clinical educator in two Australian osteopathy programs. This information can then be used to provide the clinical educator with feedback about their clinical teaching performance from the standpoint of the students. Eight questionnaires will provide a reliable indication as to the educator's performance. The information derived from these 8 questionnaires could be used to provide feedback to an Australian osteopathy clinical educator to assist them to identify their strengths, improve their performance and guide professional development. All but one OCTQ item, and the three global rating items demonstrated *substantial* to *almost perfect* agreement. This issue with the test-retest stability of one item requires further investigation prior to the OCTQ being used for promotion, tenure and teaching award decisions. The reliability estimations are consistent with previous work and support the validity of the calculation of a total score for the OCTQ. These results provide evidence for all four aspects of Kane's validity argument: scoring, generalisation, extrapolation and implications. Further work is now required to address the issue with one OCTQ item. Future studies will also explore the relationship of the OCTQ to clinical educator self-evaluations and self-efficacy, and how other clinical education environmental factors relate to OCTQ scores to continue building evidence for the validity argument.

Conflicts of interest

Brett Vaughan is a section editor for the International Journal of Osteopathic Medicine. He was not involved in review or editorial decisions regarding this manuscript.

Funding

This work was supported by an Australian Government Research Training Program Scholarship.

Ethical approval

This study was approved by the Victoria University and Southern Cross University Human Research Ethics Committees (HRE15-238 and ECN15-301).

Availability of data and material

The data set(s) supporting the results of this article are available in the *figshare* repository: 10.6084/m9.figshare.6667913.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.ijosm.2018.11.001.

References

[1] Beckman TJ, Cook DA, Mandrekar JN. What is the validity evidence for assessments of clinical teaching? J Gen Intern Med 2005;20(12):1159–64.
[2] Fluit C. Assessing the quality of clinical teachers. J Gen Intern Med 2010;25(12):1337–45.
[3] Vaughan B. Developing a clinical teaching quality questionnaire for use in a university osteopathic pre-registration teaching program. BMC Med Educ 2015;15(1):70.
[4] Vaughan B. Exploring the measurement properties of the osteopathy clinical teaching questionnaire using Rasch analysis. Chiropr Man Ther 2018;26(1):13.
[5] Oermann MH, Conklin JL, Rushton S, Bush MA, editors. Student evaluations of teaching (SET): guidelines for their use. Nurs Forum (Auckl). Wiley Online Library; 2018.
[6] Beckman TJ, Ghosh AK, Cook DA, Erwin PJ, Mandrekar JN. How reliable are assessments of clinical teaching? J Gen Intern Med 2004;19(9):971–7.
[7] Nation JG, Carmichael E, Fidler H, Violato C. The development of an instrument to assess clinical teaching with linkage to CanMEDS roles: a psychometric analysis. Med Teach 2011;33:e290–6.
[8] Copeland HL, Hewson MG. Developing and testing an instrument to measure the effectiveness of clinical teaching in an academic medical centre. Acad Med 2000;75(2):161–6.
[9] Der Hem-Stokroos V, Van der Vleuten C, Daelmans H, Haarman H, Scherpbier A. Reliability of the clinical teaching effectiveness instrument. Med Educ 2005;39(9):904–10.
[10] Conigliaro RL, Stratton TD. Assessing the quality of clinical teaching: a preliminary study. Med Educ 2010;44(4):379–86.
[11] Boerboom T, Dolmans D, Jaarsma A, Muijtjens A, Van Beukelen P, Scherpbier A. Exploring the validity and reliability of a questionnaire for evaluating veterinary clinical teachers' supervisory skills during clinical rotations. Med Teach 2011;33(2):e84–91.
[12] de Oliveira Filho GR, Dal Mago AJ, Garcia JHS, Goldschmidt R. An instrument designed for faculty supervision evaluation by anesthesia residents and its psychometric properties. Anesth Analg 2008;107:1316–22.
[13] Haider SI, Johnson N, Thistlethwaite JE, Fagan G, Bari MF. WATCH: warwick assessment insTrument for clinical teacHing: development and testing. Med Teach 2015;37(3):289–95.
[14] Hindman BJ, Dexter F, Kreiter CD, Wachtel RE. Determinants, associations, and psychometric properties of resident assessments of anesthesiologist operating room supervision. Anesth Analg 2013;116(6):1342–51.
[15] Keely E, Oppenheimer L, Woods T, Marks M. A teaching encounter card to evaluate clinical supervisors across clerkship rotations. Med Teach 2010;32(2):e96–100.
[16] Cook DA, Beckman TJ. Current concepts in validity and reliability for psychometric instruments: theory and application. Am Med J 2006;119(2). 166.e7-.e16.
[17] Vaalidation Kane MT. Brennan R, editor. Educational measurement. fourth ed.Westport, USA: Praeger Publishers; 2006. p. 17–64.
[18] Kane MT. Validating the interpretations and uses of test scores. J Educ Meas 2013;50(1):1–73.
[19] R Core Team. R: A language and environment for statistical computing Vienna. Austria: R Foundation for Statistical Computing; 2015 Available from: https://www.R-project.org/, Version 3.2.2.
[20] Revelle W. Psych; procedures for personality and psychological research 1.5 eighth ed. Evanston, Illinois, USA: Northwestern University; 2015.
[21] Bloch R, Norman G. Generalizability theory for the perplexed: a practical introduction and guide: AMEE Guide No. 68. Med Teach 2012;34(11):960–92.
[22] Streiner DL, Norman GR, Cairney J. Health measurement scales: a practical guide to their development and use. fifth ed. Oxford University Press; 2014.
[23] Briesch AM, Swaminathan H, Welsh M, Chafouleas SM. Generalizability theory: a practical guide to study design, implementation, and interpretation. J Sch Psychol 2014;52(1):13–35.
[24] Shavelson RJ, Webb NM. Generalizability theory: a primer. Sage Publications; 1991.
[25] Bloch R, Norman G. G_String: a Windows wrapper for urGENOVA. 6.1.1 ed. Hamilton, Canada: The Program for Educational Research and Development (PERD), Faculty of Health Sciences, McMaster University; 2015.
[26] Norman GR, Streiner DL. Biostatistics: the bare essentials. Bc Decker Hamilton; 2008.
[27] Landis JR, Koch GG. The measurement of observer agreement for categorical data. Biometrics 1977:159–74.
[28] Hinkle DE, Wiersma W, Jurs SG. Applied statistics for the behavioral sciences. Boston, Mass., [etc.]: Houghton Mifflin; 2003.
[29] Kelley K, Lai K. MBESS. 2015. 3.3.3 ed.
[30] Dunn TJ, Baguley T, Brunsden V. From alpha to omega: a practical solution to the pervasive problem of internal consistency estimation. Br J Psychol 2014;105(3):399–412.
[31] Sijtsma K. On the use, the misuse, and the very limited usefulness of Cronbach's alpha. Psychometrika 2009;74(1):107–20.
[32] Green SB, Yang Y. Commentary on coefficient alpha: a cautionary tale. Psychometrika 2009;74(1):121–35.
[33] Green SB. A coefficient alpha for test-retest data. Psychol Methods 2003;8(1):88.
[34] Revelle W. Hierarchical cluster analysis and the internal structure of tests. Multivariate Behav Res 1979;14(1):57–74.
[35] Hecimovich MD, Styles I, Volet SE. Development and psychometric evaluation of scales to measure professional confidence in manual medicine: a Rasch measurement approach. BMC Res Notes 2014;7(1):338.

[36] Downing SM. Reliability: on the reproducibility of assessment data. Med Educ 2004;38(9):1006–12.
[37] Ziegler M, Poropat A, Mell J. Does the length of a questionnaire matter? J Indiv Differ 2015;35:250–61.
[38] Mazor K, Clauser B, Cohen A, Alper E, Pugnaire M. The dependability of students' ratings of preceptors. Acad Med 1999;74(10):S19–21.
[39] Hoyt WT. Rater bias in psychological research: when is it a problem and what can we do about it? Psychol Methods 2000;5(1):64.
[40] Solomon DJ, Speer AJ, Rosebraugh CJ, DiPette DJ. The reliability of medical student ratings of clinical teaching. Eval Health Prof 1997;20(3):343–52.
[41] Smith CA, Varkey AB, Evans AT, Reilly BM. Evaluating the performance of inpatient attending physicians. J Gen Intern Med 2004;19:766–71.
[42] Beckman TJ, Lee MC, Rohren CH, Pankratz VS. Evaluating an instrument for the peer review of inpatient teaching. Med Teach 2003;25(2):131–5.
[43] Lases SL, Arah OA, Pierik ER, Heineman E, Lombarts MK. Residents' engagement and empathy associated with their perception of faculty's teaching performance. World J Surg 2014;38(11):2753–60.
[44] Brown T, Williams B, Lynch M. Relationship between clinical fieldwork educator performance and health professional students' perceptions of their practice education learning environments. Nurs Health Sci 2013;15(4):510–7.
[45] Mackillop L, Parker-Swift J, Crossley J. Getting the questions right: non-compound questions are more reliable than compound questions on matched multi-source feedback instruments. Med Educ 2011;45(8):843–8.
[46] Bierer SB, Hull AL. Examination of a clinical teaching effectiveness instrument used for summative faculty assessment. Eval Health Prof 2007;30(4):339–61.
[47] van der Leeuw R, Lombarts K, Heineman MJ, Arah O. Systematic evaluation of the teaching qualities of obstetrics and gynecology faculty: reliability and validity of the SETO tools. PloS One 2011;6(5):e19142.
[48] McGrath C, Yeung RWK, Comfort M, McMillan A. Development and evaluation of a questionnaire to evaluate clinical dental teachers (ECDT). Br Dent J 2005;198(1):45–8.
[49] Saarikoski M. Mentor relationship as a tool of professional development of student nurses in clinical practice. Int J Psychiatr Nurs Res 2003;9(1):1014–24.
[50] Ko CY, Escarce JJ, Baker L, Sharp J, Guarino C. Predictors of surgery resident satisfaction with teaching by attendings: a national survey. Ann Surg 2005;241(2):373.
[51] Zinbarg RE, Yovel I, Revelle W, McDonald RP. Estimating generalizability to a latent variable common to all of a scale's indicators: a comparison of estimators for ωh. Appl Psychol Meas 2006;30(2):121–44.
[52] Tavakol M, Dennick R. Making sense of Cronbach's alpha. Int J Med Educ 2011;2:53–5.
[53] Vaughan B, Macfarlane C, Florentine P. Clinical education in the osteopathy program at Victoria University. Int J Osteopath Med 2014;17(3):199–205.

CHAPTER 5

Clinical educator self-efficacy, self-evaluation and its relationship with student evaluations of clinical teaching

Using the 12-item measure developed in the preceding chapters, chapter 5 explores the relationship between self-evaluation of clinical teaching quality and the ratings provided by learners. This study is consistent with the notion that multiple sources of evidence be collected to inform decisions about clinical teaching quality. The study in this chapter also explored the self-efficacy of osteopathy clinical educators and identified that those educators with high self-efficacy were more likely to receive lower ratings from learners compared to educators who reported lower levels of self-efficacy.

BMC Medical Education

RESEARCH ARTICLE Open Access

Clinical educator self-efficacy, self-evaluation and its relationship with student evaluations of clinical teaching

Brett Vaughan

Abstract

Background: In a whole-of-system approach to evaluation of teaching across any degree, multiple sources of information can help develop an educators' understanding of their teaching quality. In the health professions, student evaluations of clinical teaching are commonplace. However, self-evaluation of teaching is less common, and exploration of clinical educators' self-efficacy even less so. The aim of the study was to evaluate how a clinical educator's self-evaluation of teaching intersects with their self-efficacy, to ascertain if that matches student evaluation of their teaching. This information may assist in facilitating targeted professional development to improve teaching quality.

Methods: Clinical educators in the osteopathy program at Victoria University (VU) were invited to complete: a) self-evaluation version of the Osteopathy Clinical Teaching Questionnaire (OCTQ); and b) the Self-Efficacy in Clinical Teaching (SECT) questionnaire. Students in the VU program completed the OCTQ for each of the clinical educators they worked with during semester 2, 2017.

Results: Completed OCTQ and SECT were received from 37 clinical educators. These were matched with 308 student evaluations (mean of 6 student ratings per educator). Three possible educator cohorts were identified: a) high clinical eductor self-OCTQ with low student evaluation; b) low clinical educator self-evaluation and high student evaluations; and, c) no difference between self- and student evaulations. Clinical educators in the first cohort demonstrated significantly higher SECT subscale scores (effect size > 0.42) than their colleagues. Age, gender, teaching qualification, and years practicing or years as a clinical educator were not associated with clinical educator OCTQ scores or the SECT subscales.

Conclusions: Targeted professional development directed towards fostering self-efficacy may provide an avenue for engaging those clinical educators whose self-efficacy is low and/or those who did not receive high student evaluations. Given there is no gold standard measure of clinical teaching quality, educators should engage with multiple sources of feedback to benchmark their current performance level, and identify opportunities to improve. Student and self-evaluations using the OCTQ and evaluation of self-efficacy using the SECT, are useful tools for inclusion in a whole-of-system approach to evaluation of the clinical learning environment.

Keywords: Clinical education, Medical education, Osteopathic medicine, Evaluation, Measurement, Educational environment

Correspondence: brett.vaughan@unimelb.edu.au
College of Health & Biomedicine, Victoria University, Melbourne, Australia

© The Author(s). 2020 **Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit http://creativecommons.org/licenses/by/4.0/. The Creative Commons Public Domain Dedication waiver (http://creativecommons.org/publicdomain/zero/1.0/) applies to the data made available in this article, unless otherwise stated in a credit line to the data.

Background

A more comprehensive picture of clinical educator teaching quality and performance can be developed through the collection and triangulation of data from multiple sources, including students, peers, program administrators and self-evaluation [1–4]. In clinical education, students will typically evaluate their clinical educators at the end of a clinical placement or rotation offering one perspective of teaching quality. Widely used in clinical education, these evaluations serve to provide feedback to the educators, as well as faculty and program administrators [3, 5] in order to maintain and improve teaching quality [6]. Faculty and program administrators are typically interested in this information for the purposes of professional development activities, remediation, teaching awards, promotion, and potentially ongoing employment decisions [4]. Student evaluations of teaching are used extensively in higher education however authors have highlighted significant challenges with their interpretation (i.e. poor construct definition, gender bias, low reponse rates) and use of the results [7–11], particularly when the student perspective is used in isolation. This collective literature suggests data from student evaluations be limited to formative decision-making that is informed by data collected longitudinally and triangulated with other measures of teaching quality [8, 12, 13].

When data about teaching quality are drawn from multiple sources, it is anticipated that the clinical educator will use this data to assist them to improve their teaching. Gathering this data may also stimulate the clinical educator to reflect on their performance, and institute changes to their education practice to improve teaching quality. The 'self-regulated professional' [14] engages in this reflective practice cycle as part of daily clinical practice. However, if or how they use self-evaluation in their practice as a clinical educator is less clear with few examples in the literature [1, 6]. Whilst self-evaluation has been shown to have limitations when used in isolation [14–16], if combined with data from external sources [2, 17–20] it can be regarded as *informed self-assessment* [21] and this combined data can be valuable to improve performance. Self-assessment judgements appear to be multifactorial [21], with contextual factors and "underlying tensions" (p. 1212) influencing the use of data from one source over another.

Our understanding of self-assessment is better informed by exploring the external and internal information individuals draw on to inform this judgement [20–22], whilst also acknowledging that this information can be of varying quality [20]. The current study draws on the definition of self-assessment by Eva and Regehr [16] who describe this construct as:

> " … a pedagogical process by which one takes personal responsibility for looking outward, seeking feedback and explicit information from external sources, then using these externally generated sources of assessment data to direct performance improvements" (p.15).

Several studies have investigated the relationship of clinical education self-evaluation data to that generated by learners [1, 3, 6]. These studies suggest there is limited concordance between self- and student evaluations, inferring potential use of differing standards when making quality judgements [23]. This difference in student and self-evaluation appears to stimulate reflection on performance [1], typically for those who under- rather than over-estimate their own performance [6]. Notwithstanding the aforementioned research, feedback from students appears to stimulate self-evaluation [2, 20].

A potential influence or mediator of self-evaluation of performance, amongst other processes, is self-efficacy [14]. Self-efficacy as a construct stems from the work of Bandura [24] and is defined as the self-perceived ability to perform a task, self-monitoring, and to an extent, motivation to persevere when faced with challenges or difficulties with said task. Self-efficacy, however, is task and context-specific [25], and develops through experience with task outcomes, observation of successful or positive performances, feedback and reflection on task performance [26]. In the clinical teaching context, self-efficacy could be considered to be the beliefs of the educator in their ability to facilitate student learning through engaging with alternative educational approaches, tolerance to mistakes and student-centred learning [27]. Various meta-analyses from the wider educational literature have demonstrated small positive, and significant, relationships between self-efficacy and teaching effectiveness [28], self-assessment and self-efficacy [29], and self-efficacy with a commitment to teaching [30] in teachers. However, we know little about the self-efficacy of clinical educators in the health professions context, and how this construct correlates with teaching quality.

Although the construct of *quality of clinical teaching* has not been agreed on in the literature [31] – likely due to its context-specific nature [32] - it broadly incorporates the interpersonal attributes, and teaching approaches utilised, by clinical educators [33], and is a term widely used in the literature [31, 34–38]. Drawing on Beckman, Ghosh, Cook, Erwin and Mandrekar [33], *quality of clinical teaching* in the current study was defined as 'the interpersonal attributes exhibited, and teaching approaches used by osteopathy clinical educators in a student-led clinical learning environment'. The present study continues developing the validity argument of a measure of quality of clinical teaching – the Osteopathy Clinical Teaching

Questionnaire (OCTQ) [39]. Specifically, the study evaluates clinical educator self-efficacy, in context, and its relationship to self- and student perception of quality of clinical teaching using the OCTQ. The current study is also part of a larger program of research to develop a validity argument for the tools that might be used in a whole-of-system approach to evaluation of clinical teaching and quality assurance of clinical education in the student-led clinic context. The work presented here explores the intersection of student and self-evaluation data about clinical teaching quality, and its relationship to self-efficacy as one factor that may influence this data.

Methods

This study was approved by the Victoria University (VU) Human Research Ethics Committee.

Participants

Students enrolled in year 4 ($n = 80$) and 5 ($n = 55$) of the VU were introduced to the study in a practical skills class (outside of the clinic environment) and provided with copies of the OCTQ. Those students interested in participating were encouraged to, prior to their next clinic session [4], complete the OCTQ for each clinical educator with whom they had worked during the July 2017 to November 2017 teaching period and return it to a secure box in the teaching clinic. Student responses were anonymous, and participation in the study was not a requirement of any academic subject in their programs. The student was not required to identify themselves however they were required to write the name of the clinical educator being rated at the top of the form. Consent to participate was implied by return of the questionnaire.

Clinical educators ($n = 42$) employed in the osteopathy program at VU during the same period were invited to complete the questionnaires (OCTQ and SECT), in their own time, in November 2017. Those who chose to participate in the study were asked to identify themselves by name in order to match their self-evaluation data with that obtained from the students. Each clinical educator returned the completed questionnaires to a locked box with consent implied by return of the questionnaire. Only the author had access to the identifiable data and had no role in employment or promotion decisions for clinical educators in the program. The participating clinical educator cohort data summary was made available to the academic clinic coordinator – no data identifying an individual clinical educator was included in this summary.

Measures

Students

Students were asked to complete the Osteopathy Clinical Teaching Questionnaire (OCTQ) for each clinical educator they had worked with during the July to November 2017 period. The Osteopathy Clinical Teaching Questionnaire (OCTQ) was developed to evaluate student perceptions of the quality of clinical teaching in their respective programs' in student-led, on-campus clinics [40], or university clinics [41]. Previous work provided evidence for the validity argument for the interpretation of scores derived from the OCTQ, including reliability (internal structure, test-retest, inter-rater), content validity, and structural validity [39, 42, 43]. The questionnaire uses a Likert-type scale (strongly disagree (1) to strongly agree (5), with a neutral category) to allow students to respond to each statement. Questionnaires were completed in early November 2017.

Clinical educators

The clinical educators were asked to complete:

- a self-evaluation version of the OCTQ containing the same 12 items and 1 global rating item. The anchor for each item was "As a Clinical Educator I …" and items were rephrased to reflect self-rating [1].
- the Self Efficacy of Clinical Teachers (SECT) tool. The SECT tool was developed by McArthur [44] to evaluate self-efficacy of Australian general practice clinical educators, however, the items appear to be suitable for measurement of self-efficacy in the student-led clinical learning environment. The tool contains 22 items across three domains of clinical teaching practice: Customising Teaching to Learning Needs; Teaching Prowess; and, Impact on Learner's Development, with a total score created for each domain. The Cronbach's alpha for the 22-item SECT is reported at 0.95 [44].
- a brief demographic questionnaire asking their age, years of practice as an osteopath, years as a clinical educator and whether they had completed a formal university qualification in teaching and learning and/or clinical education.

Data analysis

Data were entered into SPSS (IBM Corp, USA) for analysis. Total scores were generated for the student evaluations (the OCTQ) and a total score for the clinical educator's self-evaluation (OCTQ) and also for each of the SECT subscales. Descriptive statistics were generated for the OCTQ completed by the students and the clinical educators, and for the SECT completed by the clinical educators. A difference score was calculated between the student OCTQ scores and the educators self-evaluation OCTQ for both total score and mean. This resulted in a range of scores whereby higher difference scores represented the clinical educator having a higher self-evaluation score than that reported by the students on the OCTQ. Difference scores were then

categorized as *higher*, *neutral* and *lower*. Non-parametric tests were used to investigate differences between the demographic variables and the educators' self-evaluation OCTQ total score, global rating (5-point Likert-type scale) and their SECT subscale scores. Non-parametric effect sizes (r) [6] were calculated where relevant.

Relationship between student and clinical educator ratings

Correlations between the student's OCTQ and educators self-OCTQ ratings were explored with Spearman's rho (ρ) using the median values for each item, and the global rating item. The relationship between the the educator's self-evaluation OCTQ, the SECT and the global rating were explored using Spearman's rho (ρ) and interpreted according to Hinkle, Wiersma and Jurs [45]: 0–0.30 (negligible); 0.30–0.50 (low); 0.50–0.70 (moderate); 0.70–0.90 (high); 0.90–1.00 (very high).

Reliability estimates

Reliability estimates for the student OCTQ evaluations were calculated in R [46] using the the MBESS package [47]. Cronbach's alpha (α) and McDonald's omega total (ωt), and their respective confidence intervals were calculated consistent with Vaughan [42].

Results

Three hundred and eight student ratings of a cohort of 42 out of 43 clinical educators who had worked in the student-led, on-campus clinic during the July 2017 to November 2017 teaching period were received. Of the 43 clinical educators, 37 chose to participate in the study including one educator who did not receive student evaluations.

Demographics

Table 1 presents demographic data for the clinical educators who chose to participate. Table 2 presents descriptive statistics for the student and clinical educator's self-evaluation versions of the OCTQ for comparison. The mean number of student ratings per educator was 6.75 ± 4.06 with a median of 6 (range 1–14). Clinical educators demonstrated lower means and the same or lower median values for most items when compared to the students. Figure 1 presents clinical educators self-evaluation of overall teaching quality with over 75% rating their effectiveness as *very good* or *excellent*. Table 3 presents the descriptive statistics for the SECT. No significant difference ($p > 0.05$) was identified for any gender, age, years in clinical practice, years clinical teaching, and qualifications for the OCTQ self-evaluation total score, global rating or SECT subscale scores suggesting these variables were not associated with teaching or self-efficacy scores. The reliability estimations for the OCTQ were: ωt = 0.93 [95%CI 0.92–0.95]; and, α = 0.93 [0.91–0.95]. For the SECT, the reliability estimations were: Customising Teaching to Learning Needs subscale (ωt = 0.88 [95%CI 0.82–0.95], α = 0.87 [95%CI 0.79–0.93]); Teaching Prowess subscale (ωt = 0.86 [95%CI 0.80–0.91], α = 0.85 [95%CI 0.79–0.90]; and, Impact on Learner's Development subscale (ωt = 0.83 [95%CI 0.73–0.92], α = 0.82 [95%CI 0.71–0.89]).

Table 1 Demographic characteristics of the clinical educators

Age	
25–34 years	18 (48.6%)
35–40 years	11 (29.7%)
41–50 years	6 (16.2%)
51–60 years	1 (2.7%)
Greater than 60 years	1 (2.7%)
Gender	
Male	14 (37.8%)
Female	23 (62.2%)
Years practicing as an osteopath	
0–4 years	0
5–9 years	20 (54.1%)
10–14 years	9 (24.3%)
15–19 years	6 (16.2%)
20 or more years	2 (5.4%)
Years as a clinical educator	
0–4 years	26 (70.4%)
5–9 years	8 (21.6%)
10–14 years	1 (2.7%)
15–19 years	1 (2.7%)
20 or more years	1 (2.7%)
University teaching and learning qualification	
Yes	11 (29.7%)
Currently completing	8 (20.0%)
No	18 (48.6%)
University clinical education qualification	
Currently completing	1 (2.7%)
No	36 (97.3%)

Difference score

Twenty-four educators (66.7%) had a *lower* difference score (i.e. clinical educator self-OCTQ was less than student OCTQ score) and eleven (30.6%) had a *higher* difference score with one educator (2.8%) demonstrating equal scores. The median difference score was − 1.95 (range − 12 to 16) and no significant difference was identified for the number of student ratings per educator and the difference score category. Age, gender, years practicing as an osteopath, years as a clinical educator and university education qualification were not significantly

Table 2 Descriptive statistics for the Osteopathy Clinical Teaching Questionnaire completed by the students and clinical educators

Item	Student			Clinical Educator		
	Mean	St Dev	Median	Mean	St Dev	Median
1. Maintain a positive attitude towards students	4.66	0.64	5	4.43	0.60	4
2. Demonstrate humanistic attitudes in relating to patients (integrity, compassion and respect)	4.70	0.57	5	4.73	0.45	5
3 Show genuine concern for my students professional well-being	4.60	0.71	5	4.57	0.50	5
4. Have good communication skills	4.59	0.66	5	4.43	0.50	4
5. Am open to student questions and alternative approaches to patient management	4.55	0.74	5	4.43	0.60	4
6. Adjust teaching to my student's needs (experience, competence, interest)	4.38	0.82	5	4.14	0.58	4
7. Promote reflection on clinical practice	4.38	0.82	5	4.19	0.66	4
8. Emphasise a problem-solving approach rather than solutions	4.46	0.80	5	4.19	0.87	5
9. Ask questions to enhance my students learning	4.39	0.81	5	4.30	0.66	4
10. Stimulate student's to learn independently	4.33	0.81	5	4.14	0.79	4
11. Offer my student's suggestions for improvement when required	4.48	0.82	5	4.32	0.58	4
12. Demonstrated osteopathic, clinical examination and rehabilitation knowledge and skill(s)[a]	3.66	0.67	4	4.27	0.51	4
Total score	53.14	6.99	55	51.34	4.11	50
Global rating	4.39	0.85	5	3.80	0.55	4

[a] rescored for students only according to Vaughan [42]

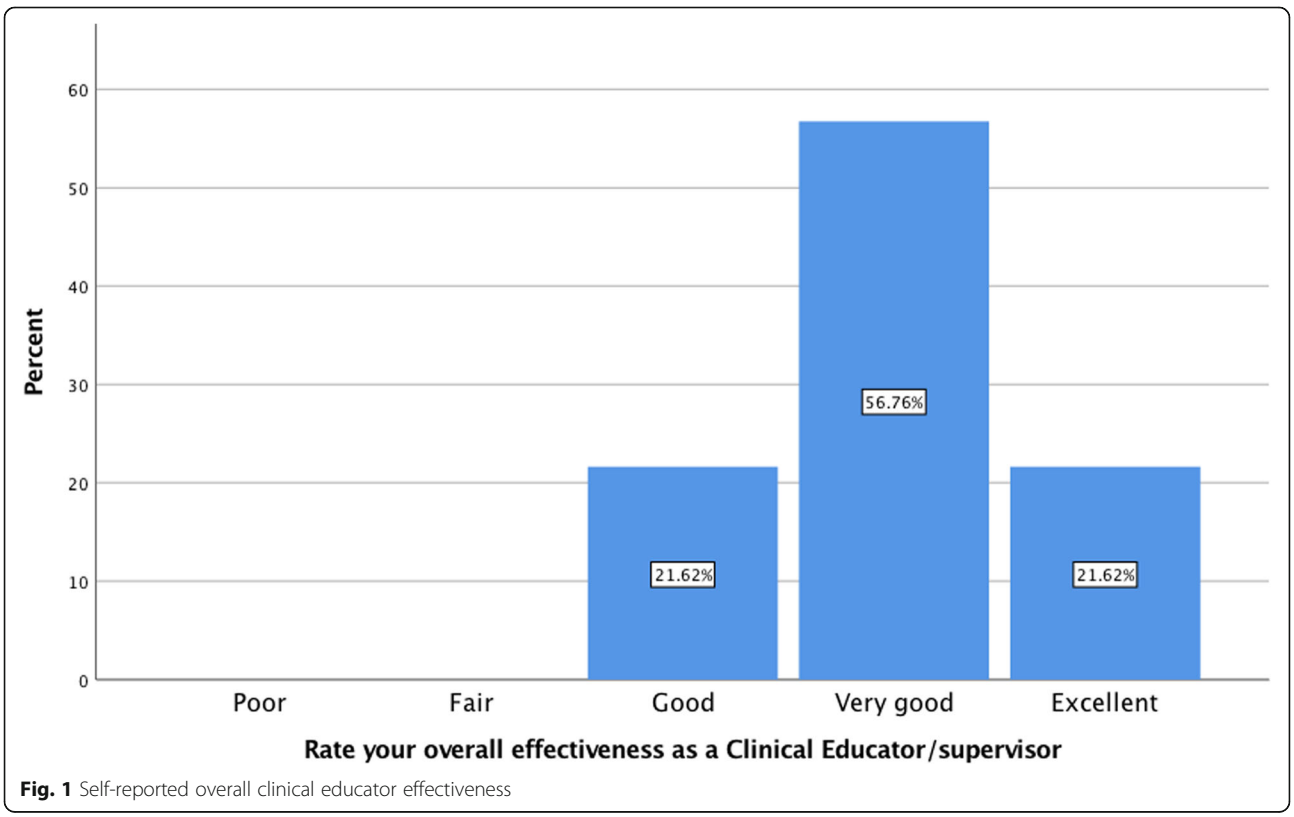


Fig. 1 Self-reported overall clinical educator effectiveness

Table 3 Descriptive statistics for the Self-efficacy in Clinical Teachers (SECT) tool

SECT item	Mean	St Dev	Median
1. I can correctly appraise the learning needs of students	5.16	0.83	5
2. I can write individualised learning objectives based on a student's unique situation	4.97	1.04	5
3. I can provide appropriate instructional content, based on a student's learning need.	5.27	0.77	5
4. I can select appropriate teaching strategies when encountering different student's needs.	5.35	0.98	6
5. I can refine teaching content and methods based on a student's learning needs and confounding factors	5.05	0.97	5
6. I can teach what the student needs to know	5.62	0.72	6
7. I am effective in my clinical training	5.57	0.80	6
8. I am well organised and prepared for the in-practice teaching	5.46	0.87	5
9. I can provide clinical instruction in a clear manner that students can understand	5.68	0.82	6
10. I can correctly demonstrate clinical skills such as management of the patient consultation/interaction	6.08	0.79	6
11. I have the ability to evaluate the effectiveness of a student's clinical and consulting efforts through direct observation.	5.43	0.76	5
12. I can teach registrars to determine their professional boundaries	5.54	0.80	6
13. I can handle most difficult student questions or situations	5.59	0.83	6
14. I give clear explanations to questions around clinical scenarios	5.70	0.89	6
15. I can tailor my feedback to be constructive and developmental	5.62	0.92	6
16. I am concerned for my students wellbeing	6.27	0.80	6
17. I have the ability to change the attitude/values of a student	4.95	0.91	5
18. I can design teaching plans for students	4.84	0.99	5
19. I can prepare learning objectives across a student's area of development	4.86	0.95	5
20. I can give instruction on strategies and resources in a student's area of development	5.19	0.99	5
21. I can stimulate the student to learn areas of curriculum that don't interest them	4.92	1.04	5
22. I can provide appropriate support for helping students learn and sustain work/life/family balance and personal wellbeing	5.73	1.02	6
SECT Domain 1 - Customising Teaching to Learning Needs	42.46	5.15	43
SECT Domain 2 - Teaching Prowess	45.92	4.67	47
SECT Domain 3 - Impact on Learner's Development	30.49	4.28	30

different for the difference score category. A significant difference was identified between those who had/had not completed a university clinical teaching qualification and difference score ($\chi^2 = 35.0$, $p < 0.01$). This result suggests that completion of a university teaching qualification may be associated with higher student evaluations compared to those who haven't completed the qualification. Of note is that there is only one educator currently completing a university clinical education qualification, and this individual educator also demonstrated no difference score, that is, their self and student OCTQ evaluations were equal supporting the aforementioned observation.

As only one educator had no difference score they were excluded from the following analyses. Those educators with a *higher* difference score demonstrated significantly higher total scores for all three SECT domains (Customising Teaching to Learning Needs (Domain 1) – $p = 0.01$, $z = -2.49$, $r = 0.42$; Teaching Prowess (Domain 2) – $p < 0.01$, $z = -2.83$, $r = 0.48$; Impact on Learner's Development (Domain 3) – $p < 0.01$, $z = -2.68$, $r = 0.46$). These educators were also more likely to rate their global effectiveness as an educator significantly higher with a large effect size ($p < 0.01$, $z = -3.43$, $r = 0.58$).

Relationship between student evaluations and clinical educator self-evaluations

Table 4 presents the relationship between the student and self-evaluation responses to the OCTQ items. Most of the relationships were *negligible*. The relationship between the mean values for item 8 "Emphasises a problem-solving approach rather than solutions" was *low*. The shared common variance for each item ranged from 0.01 to 11.6% suggesting there is little concordance between student evaluations and clinical educator evaluations. Figure 2 shows the associations between student global rating of clinical teaching effectiveness and SECT domains, all of which were trivial and again supporting the limited concordance assertion. Associations

Table 4 Association between Osteopathy Clinical Teaching Questionnaire student and self-evaluation

Self-evaluation	Student	Common Variance
1. Maintain a positive attitude towards students	−0.04	0.1%
2. Demonstrate humanistic attitudes in relating to patients (integrity, compassion and respect)	−0.01	0.01%
3 Show genuine concern for my students professional well-being	0.12	1.4%
4. Have good communication skills	−0.12	1.4%
5. Am open to student questions and alternative approaches to patient management	0.22	4.8%
6. Adjust teaching to my student's needs (experience, competence, interest)	−0.19	3.6%
7. Promote reflection on clinical practice	−0.20	4%
8. Emphasise a problem-solving approach rather than solutions	0.34*	11.6%
9. Ask questions to enhance my students learning	−0.25	6.2%
10. Stimulate student's to learn independently	−0.15	2.2%
11. Offer my student's suggestions for improvement when required	−0.15	2.2%
12. Demonstrated osteopathic, clinical examination and rehabilitation knowledge and skill(s)	0.14	2.0%
Global - Rate your overall effectiveness as a Clinical Educator/supervisor	−0.21	4.4%

between clinical educator completed measures are described in Table 5 with most being moderately correlated except for SECT Domain 3 - Impact on Learner's Development and the OCTQ self-evaluation total score where a small correlation was observed.

Discussion

A whole-of-system approach to evaluation of clinical education quality is one aspect of the wider quality assurance program in any health professions education course. One challenge in implementing this approach is the lack of a gold standard measure of clinical teaching quality. Consequently, clinical educators should be encouraged to engage with multiple sources of feedback to benchmark their current performance level [4, 6], and identify opportunities to improve their performance. For that reason this study explored the intersection between clinical educators' self-evaluation of clinical teaching

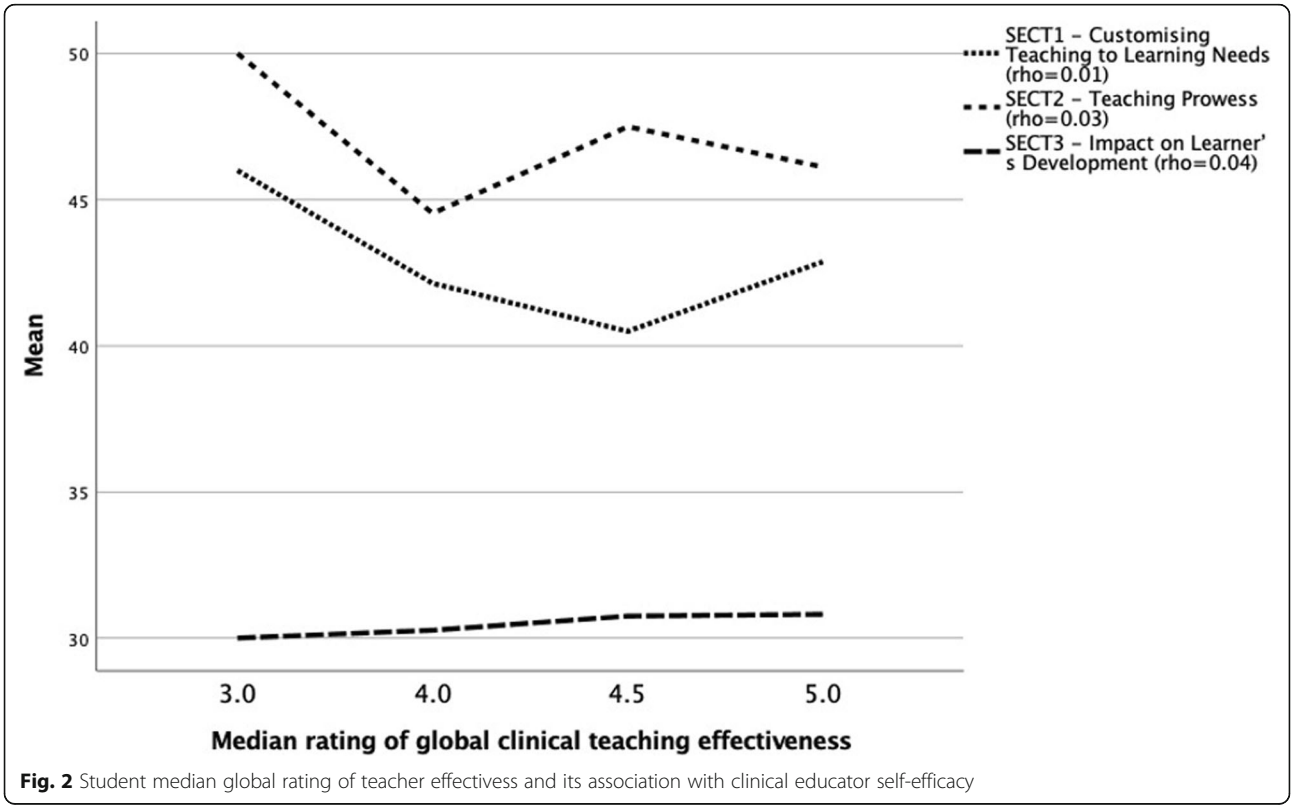


Fig. 2 Student median global rating of teacher effectivess and its association with clinical educator self-efficacy

Table 5 Associations between measures completed by the clinical educators

	OCTQ Total	OCTQ Global	SECT Domain 1	SECT Domain 2	SECT Domain 3
OCTQ Total	1				
OCTQ Global	0.73*	1			
SECT Domain 1	0.52*	0.51*	1		
SECT Domain 2	0.62*	0.65*	0.72*	1	
SECT Domain 3	0.28	0.46*	0.65*	0.56*	1

*$p < 0.001$; Customising Teaching to Learning Needs (Domain 1); Teaching Prowess (Domain 2); Impact on Learner's Development (Domain 3

quality and self-efficacy, and student perceptions of clinical teaching quality. The current study also extends the work of Stalmeijer, Dolmans, Wolfhagen, Peters, van Coppenolle and Scherpbier [1] on clinical educator self-assessment through the inclusion of self-efficacy, given its relationship to teaching effectiveness measures [28].

Self- and student evaluation

In the current study, three distinct groups of clinical educators were identified:

Group 1. Those with student evaluations that were higher than the educator's self-evaluation;
Group 2. Those with student evaluations that were lower than the educator's self-evaluations; and,
Group 3. Those with student evaluations that were consistent with educator self-evaluation.

In relation to clinical educators' own views of their performance, the disconnect between self- and external evaluation is not new [1, 3, 6], and this trend appears to be the case in the current clinical educator cohort. The trivial to small relationships at item level between the student- and clinical educator OCTQ self-evaluations suggests the educators may be interpreting the items differently to the students, have differing conceptions of clinical teaching quality, or that the OCTQ is not a suitable self-evaluation measure.

Over- and under-estimation of clinical teaching performance in the current work was similar to that of Boerebach et al. [6]. These authors concluded that there were groups who over- and under-estimated their teaching performance, and that in subsequent evaluation rounds, these differences were ameliorated. As these authors highlighted, whether this was due to enacting feedback received in prior rounds, or matching their self-evaluation to previous resident (student) evaluations, could be debated. The results of Boerebach et al. [6] also support the collection of longitudinal teaching quality data [13], affording the educator an opportunity to enact strategies to improve their teaching in response to previous feedback.

Whilst some of the clinical educator cohort in the present study have received ad-hoc formal or informal feedback on their performance, this did not occur on a consistent basis over the study period. The current study was also the first time clinical educators were asked to formally self-evaluate their clinical teaching. Without feedback, it can be challenging for clinical educators to accurately gauge the effectiveness of their clinical teaching performance [1, 48], and this appears to be borne out in the findings of the current study. How clinical educators use this self- and student-derived performance effectiveness information may be mediated by educators' clinical teaching self-efficacy.

Self-efficacy

Those clinical educators who were in group 1 (self-evaluation scores higher than student evaluations) demonstrated significantly higher self-efficacy across all three of the SECT domains. This group of clinical educators self-reported they were able to successfully manage the varying demands of clinical supervision and education in the student-led clinical learning environment. This result may also reflect a level of self-confidence with their own performance as a clinical educator. Less experienced clinical educators, both in a clinical and education sense, have been shown to have less confidence in their performance as a clinical educator [49]. However, experience as an osteopathy clinical educator did not appear to be related to higher self-efficacy in the current work. Self-efficacy is both context- and task-specific and when related to self-confidence, a subset of clinical educators in a clinical teaching context may be more likely to display this confidence through their perceived self-efficacy. However, some students in the current study rated clinical educators with low self-efficacy higher than the educator rated themselves (group 2), potentially suggesting this group of clinical educators may be less confident in their performance in this educational context.

Within Bandura's framework [24], mastery learning is likely to drive confidence with a task (through success or failure) and therefore higher self-efficacy. In the group of clinical educators that demonstrated high self-efficacy, it may be that they have had more perceived successes, and potentially place increased demands on students beyond the students' *zone of proximal development*. This may have resulted in lower student evaluation scores - an assertion that requires further investigation. Self-efficacy

across the three SECT domains was also moderately positively associated with overall self-evaluated teaching effectiveness, further supporting the self-confidence assertion described previously. Self-efficacy accounted for between 21 and 42% of the overall variance in self-evaluated global teaching effectiveness suggesting self-efficacy plays a role in self-evaluation. The significant variation in self-efficacy in our clinical educator cohort, suggests that self-efficacy could be developed in some educators and tempered in others, potentially through professional/faculty development. Thus the current study provides an argument for the use of clinical teaching self-efficacy evaluation as a basis for developing faculty/professional development programs.

Arah et al. [50] demonstrated that those educators who attend training programs are likely to obtain higher student ratings than those who do not, however, participation in formal education programs did not result in higher ratings in the current study. Participating in a generalist post-graduate university teaching qualification may not be the most suitable program for those wanting to undertake more formal education in the clinical education context. This qualification did not appear to be associated with any of the OCTQ completed by the students and clinical educators, nor the SECT. Conversely, the study identified that the one educator who was completing their formal qualification in clinical education demonstrated a self-evaluation score that is consistent with the students' ratings, although they were not the highest rated educator in the current population. Whether this clinical educator was more accurate at self-assessing due to their clinical education qualification would require additional exploration. It is also important to note that historically, very little clinical education-specific professional development (beyond workplace orientation) has been made available to the educators in the current work.

Limitations

It is important to be cognizant of the limitations of the current work and the ability to generalize the results to other osteopathy teaching programs, student-led clinics and clinical education more broadly. Defining the construct of 'clinical teaching quality' has reported to be challenging [31], and although a definition is provided in the context of the current work, there is no agreed one defined in the literature [31] and the OCTQ may in fact measure 'satisfaction'. This may also be an additional limit on the generalizability of the study. There are a number of limitations associated with the cross-sectional design of the study including the data being wholly self-report, recall biases, and potential response biases on the part of the students and educators. The student responses were anonymous and therefore less susceptible to social desirability [51], however clinical educator responses were identifiable, and the high self-efficacy and self-evaluations may be due to this bias.

Additional limitations of the work include the study taking place at a single educational institution, there was no question on the demographic form exploring participation in non-award faculty development in clinical education, and the assumption that the SECT captures the breadth of self-efficacy of clinical teaching in the university-based clinical learning environment. The SECT has only been published within a doctoral thesis and the current study is the first to publish data on its use in the peer-review literature. Additional testing of the SECT will strengthen the argument for its use as a measure of self-efficacy for clinical teachers.

The low number of ratings received by some clinical educators may also bias the results in that the student responses may have been more towards one end of the scale providing a biased picture of performance. That said, a single clinical educator receiving a low number of ratings is reflective of the reality of the learning environment in the current study where the educator-student ratio may be small. Statistically this appeared to have minimal impact but larger numbers would be preferable to provide stronger support for the assertions in this work. The difference in self- and student evaluations could be associated with a differing interpretation of the meaning of the OCTQ items. This provides an interesting avenue for further work to understand how the different stakeholders interpret individual items. The small number of educators participating in the study limited the use of regression models that may have assisted in shedding light on the influence of the demographic variables, particularly the influence of gender, on over- or under-estimation of performance [6].

Conclusions

A whole-of-system approach to evaluation of clinical education is one aspect of quality assurance in any health profession's education program. Conceptions of clinical teaching quality are likely to be different between the various stakeholders within the clinical education process: student; educator; patient; faculty; peer; and administrator. This study evaluated how clinical educator's self-evaluation of teaching intersects with their self-efficacy to ascertain if that matches student evaluation of their teaching. Results identified three possible cohorts: a) low student evaluations with high self-evaluation; b) high student evaluations and low self-evaluation; and c) equal student and self-evaluations. Of note was the relationship of the former two groups to self-efficacy - educators self-evaluating their clincal teaching higher than student ratings reported significantly higher self-efficacy. Professional development may be a valuable means of empowering clinical educators, whose self-efficacy is low

or those who did not receive high student evaluations. Those educators who have high self-efficacy and low student evaluations may also be tempered through such activities.

Given there is no gold standard measure of clinical teaching quality, clinical educators should engage with multiple feedback sources to benchmark their current performance level and identify opportunities for improvement. Program administrators are also encouraged to consider longitudinal data collected from multiple data sources when making decisions about teaching quality and performance. To further enhance a whole-of-system approach to evaluation of clinical education, future research will explore patient views of clinical educator effectiveness. Such research may illuminate other factors that could assist clinical educators to improve their practice. The complexity of the potential influences on clinical educator performance and teaching quality, requires multiple data sources to inform formative decisions and professional development.

Abbreviations

OCTQ: Osteopathy Clinical Teaching Questionnaire; SECT: Self-efficacy of Clinical Teachers

Acknowledgements

Thanks are extended to Dr. Lawrie McArthur for permission to use the SECT. The author wishes to acknowledge Associate Professor Patrick McLaughlin, Dr. Keri Moore and Dr. Jane Mulcahy for their assistance with review and comments on initial versions of the manuscript.

Author's contributions

BV was responsible for the design of the study, data collection, data analysis and all stages of the development of the manuscript. The author(s) read and approved the final manuscript.

Funding

This work was supported by an Australian Government Research Training Program Scholarship. This work was supported by an Australian Government Research Training Program Scholarship. The funders played no role in the design or interpretation of the study.

Availability of data and materials

The datasets generated and/or analysed during the current study are available in the *figshare* repository, https://doi.org/10.6084/m9.figshare.7963823

Ethics approval and consent to participate

Ethics approval was provided by the Victoria University Human Research Ethics Committee (HRE15–238). Consent to participate was implied by the return of a completed questionnaire(s) for students and clinical educators provided written informed consent.

Consent for publication

Not applicable.

Competing interests

The author declares that they have no competing interests.

Received: 25 May 2020 Accepted: 1 October 2020
Published online: 07 October 2020

References

1. Stalmeijer RE, Dolmans DHJM, Wolfhagen IHAP, Peters WG, van Coppenolle L, Scherpbier AJJA. Combined student ratings and self-assessment provide useful feedback for clinical teachers. Adv Health Sci Educ. 2010;15:315–28. https://doi.org/10.1007/s10459-009-9199-6.
2. Boerboom TBB, Jaarsma D, Dolmans DHJM, Scherpbier AJJA, Mastenbroek NJJM, Van Beukelen P. Peer group reflection helps clinical teachers to critically reflect on their teaching. Med Teach. 2011;33:e615–23. https://doi.org/10.3109/0142159X.2011.610840.
3. Boerebach BC, Arah OA, Busch OR, Lombarts KM. Reliable and valid tools for measuring surgeons' teaching performance: Residents' vs. self evaluation. J Surg Educ. 2012;69:511–20. https://doi.org/10.1016/j.jsurg.2012.04.003.
4. Berk RA. Top five flashpoints in the assessment of teaching effectiveness. Med Teach. 2013;35:15–26. https://doi.org/10.3109/0142159X.2012.732247.
5. Lombarts KM. A (good) look at the rating of teaching effectiveness: towards holistic and programmatic assessment. Med Educ. 2014;48:744–7. https://doi.org/10.1111/medu.12491.
6. Boerebach BCM, Arah OA, Heineman MJ, Busch ORC, Lombarts KMJMH. The impact of resident- and self-evaluations on Surgeon's subsequent teaching performance. World J Surg. 2014;38:2761–9. https://doi.org/10.1007/s00268-014-2655-3.
7. Spooren P, Brockx B, Mortelmans D. On the validity of student evaluation of teaching: the state of the art. Rev Educ Res. 2013;83:598–642. https://doi.org/10.3102/0034654313496870.
8. Stark P, Freishtat R. An evaluation of course evaluations. Sci Open Res. 2014. https://doi.org/10.14293/S2199-1006.1.SOR-EDU.AOFRQA.v1.
9. Boring A, Ottoboni K, Stark P. Student evaluations of teaching (mostly) do not measure teaching effectiveness. Sci Open Res. 2016;0. https://doi.org/10.14293/S2199-1006.1.SOR-EDU.AETBZC.v1.
10. Garger J, Jacques PH, Gastle BW, Connolly CM. Threats of common method variance in student assessment of instruction instruments. High Educ Eval Develop. 2019;13:2–17. https://doi.org/10.1108/HEED-05-2018-0012.
11. Hornstein HA. Student evaluations of teaching are an inadequate assessment tool for evaluating faculty performance. Cogent Educ. 2017;4:1304016. https://doi.org/10.1080/2331186X.2017.1304016.
12. Ching G. A literature review on the student evaluation of teaching. High Educ Eval Develop. 2018;12:63–84. https://doi.org/10.1108/HEED-04-2018-0009.
13. Sulis I, Porcu M, Capursi V. On the use of student evaluation of teaching: a longitudinal analysis combining measurement issues and implications of the exercise. Soc Indic Res. 2019;142:1305–31. https://doi.org/10.1007/s11205-018-1946-8.
14. Regehr G, Eva K. Self-assessment, self-direction, and the self-regulating professional. Clin Orthop Relat Res. 2006;449:34–8. https://doi.org/10.1097/01.blo.0000224027.85732.b2.
15. Davis DA, Mazmanian PE, Fordis M, Van Harrison R, Thorpe KE, Perrier L. Accuracy of physician self-assessment compared with observed measures of competence: a systematic review. JAMA. 2006;296:1094–102. https://doi.org/10.1001/jama.296.9.1094.
16. Eva KW, Regehr G. "I'll never play professional football" and other fallacies of self-assessment. J Contin Educ Heal Prof. 2008;28:14–9. https://doi.org/10.1002/chp.150.
17. Boud D. Avoiding the traps: seeking good practice in the use of self assessment and reflection in professional courses. Soc Work Educ. 1999;18:121–32. https://doi.org/10.1080/02615479911220131.
18. Colthart I, Bagnall G, Evans A, Allbutt H, Haig A, Illing J, McKinstry B. The effectiveness of self-assessment on the identification of learner needs, learner activity, and impact on clinical practice: BEME guide no. 10. Med Teach. 2008;30:124–45. https://doi.org/10.1080/01421590701881699.
19. Galbraith RM, Hawkins RE, Holmboe ES. Making self-assessment more effective. J Contin Educ Heal Prof. 2008;28:20–4.
20. Lockyer J, Armson H, Chesluk B, Dornan T, Holmboe E, Loney E, Mann K, Sargeant J. Feedback data sources that inform physician self-assessment. Med Teach. 2011;33:e113–20. https://doi.org/10.3109/0142159X.2011.542519.
21. Sargeant J, Armson H, Chesluk B, Dornan T, Eva K, Holmboe E, Lockyer J, Loney E, Mann K, van der Vleuten C. The processes and dimensions of informed self-assessment: a conceptual model. Acad Med. 2010;85:1212–20. https://doi.org/10.1097/ACM.0b013e3181d85a4e.
22. Mann K, van der Vleuten C, Eva K, Armson H, Chesluk B, Dornan T, Holmboe E, Lockyer J, Loney E, Sargeant J. Tensions in informed self-assessment: how the desire for feedback and reticence to collect and use it can conflict. Acad Med. 2011;86:1120–7. https://doi.org/10.1097/ACM.0b013e318226abdd.

23. Kunter M, Baumert J. Who is the expert? Construct and criteria validity of student and teacher ratings of instruction. Learn Environ Res. 2006;9:231–51. https://doi.org/10.1007/s10984-006-9015-7.
24. Bandura A. Self-efficacy: the exercise of control: wH freeman New York; 1997.
25. Mavis B. Self-efficacy and OSCE performance among second year medical students. Adv Health Sci Educ. 2001;6:93–102.
26. Hemmings BC. Strengthening the teaching self-efficacy of early career academics. Issues Educ Res. 2015;25:1.
27. Schiefele U, Schaffner E. Teacher interests, mastery goals, and self-efficacy as predictors of instructional practices and student motivation. Contemp Educ Psychol. 2015;42:159–71. https://doi.org/10.1016/j.cedpsych.2015.06.005.
28. Klassen RM, Tze VMC. Teachers' self-efficacy, personality, and teaching effectiveness: a meta-analysis. Educ Res Rev. 2014;12:59–76. https://doi.org/10.1016/j.edurev.2014.06.001.
29. Panadero E, Jonsson A, Botella J. Effects of self-assessment on self-regulated learning and self-efficacy: four meta-analyses. Educ Res Rev. 2017;22:74–98. https://doi.org/10.1023/A:1011404132508.
30. Chesnut SR, Burley H. Self-efficacy as a predictor of commitment to the teaching profession: a meta-analysis. Educ Res Rev. 2015;15:1–16. https://doi.org/10.1016/j.edurev.2015.02.001.
31. Conigliaro RL, Stratton TD. Assessing the quality of clinical teaching: a preliminary study. Med Educ. 2010;44:379–86. https://doi.org/10.1111/j.1365-2923.2009.03612.x.
32. Beckman TJ, Cook DA, Mandrekar JN. Factor instability of clinical teaching assessment scores among general internists and cardiologists. Med Educ. 2006;40:1209–16. https://doi.org/10.1111/j.1365-2929.2006.02632.x.
33. Beckman TJ, Ghosh AK, Cook DA, Erwin PJ, Mandrekar JN. How reliable are assessments of clinical teaching? J Gen Intern Med. 2004;19:971–7. https://doi.org/10.1111/j.1525-1497.2004.40066.x.
34. Bruijn M, Busari JO, Wolf BHM. Quality of clinical supervision as perceived by specialist registrars in a university and district teaching hospital. Med Educ. 2006;40:1002–8. https://doi.org/10.1111/j.1365-2929.2006.02559.x.
35. Busari JO, Koot BG. Quality of clinical supervision as perceived by attending doctors in university and district teaching hospitals. Med Educ. 2007;41:957–64. https://doi.org/10.1111/j.1365-2923.2007.02837.x.
36. Busari JO, Weggelaar NM, Knottnerus AC, Greidanus PM, Scherpbier AJJA. How medical residents perceive the quality of supervision provided by attending doctors in the clinical setting. Med Educ. 2005;39:696–703. https://doi.org/10.1111/j.1365-2929.2005.02190.x.
37. Fluit C. Assessing the quality of clinical teachers. J Gen Intern Med. 2010;25:1337–45. https://doi.org/10.1007/s11606-010-1458-y.
38. Der Hem-Stokroos V, Van der Vleuten C, Daelmans H, Haarman H, Scherpbier A. Reliability of the clinical teaching effectiveness instrument. Med Educ. 2005;39:904–10. https://doi.org/10.1111/j.1365-2929.2005.02245.x.
39. Vaughan B. Developing a clinical teaching quality questionnaire for use in a university osteopathic pre-registration teaching program. BMC Med Educ. 2015;15:70. https://doi.org/10.1186/s12909-015-0358-6.
40. Vaughan B, Macfarlane C, Florentine P. Clinical education in the osteopathy program at Victoria University. Int J Osteopath Med. 2014;17:199–205. https://doi.org/10.1016/j.ijosm.2013.10.010.
41. Allan J, O'Meara P, Pope R, Higgs J, Kent J. The role of context in establishing university clinics. Health Soc Care Community. 2011;19:217–24. https://doi.org/10.1111/j.1365-2524.2010.00971.x.
42. Vaughan B. Exploring the measurement properties of the osteopathy clinical teaching questionnaire using Rasch analysis. Chiro Man Thera. 2018;26:13.
43. Vaughan B. Evaluation of teaching in a student-led clinic environment: assessing the reliability of a questionnaire. Int J Osteopath Med. 2019;31:28–35. https://doi.org/10.1016/j.ijosm.2018.11.001.
44. McArthur L. Developing clinical teacher's self-efficacy in Australian general practice. Doctoral dissertion. University of Adelaide: University of Adelaide; 2016.
45. Hinkle DE, Wiersma W, Jurs SG. Applied statistics for the behavioral sciences. Houghton Mifflin: Boston, United States of America; 2003.
46. R Core Team. R: A language and environment for statistical computing. 2015. https://www.R-project.org/. Accessed 18 July 2015.
47. Kelley K, Lai K. MBESS. 3.3.3. 2015. http://CRAN.R-project.org/package=MBESS.
48. Boerboom TBB, Stalmeijer RE, Dolmans DHJM, Jaarsma DADC. How feedback can foster professional growth of teachers in the clinical workplace: a review of the literature. Stud Educ Eval. 2015;46:47–52. https://doi.org/10.1016/j.stueduc.2015.02.001.
49. Newstead C, Johnston CL, Nisbet G, McAllister L. Physiotherapy clinical education in Australia: an exploration of clinical educator characteristics, confidence and training requirements. Aust Health Rev. 2018;43:696-705.
50. Arah OA, Heineman MJ, Lombarts KM. Factors influencing residents' evaluations of clinical faculty member teaching qualities and role model status. Med Educ. 2012;46:381–9. https://doi.org/10.1111/j.1365-2923.2011.04176.x.
51. Dodaj A. Social desirability and self-reports: testing a content and response-style model of socially desirable responding. Eur J Psychol. 2012;8:651–66. https://doi.org/10.5964/ejop.v8i4.462.

Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Ready to submit your research? Choose BMC and benefit from:

- fast, convenient online submission
- thorough peer review by experienced researchers in your field
- rapid publication on acceptance
- support for research data, including large and complex data types
- gold Open Access which fosters wider collaboration and increased citations
- maximum visibility for your research: over 100M website views per year

At BMC, research is always in progress.

Learn more biomedcentral.com/submissions

CHAPTER 6

Concluding remarks and future directions

"What one student might perceive as being helpful supervision, another might view as unhelpful interference" (Billett & Sweet, 2018, p. 201).

This thesis set out to develop a measure to allow learners to evaluate the quality of the clinical teaching they receive in a student-led, on-campus clinical learning environment. The discussion presented in this chapter will begin by providing a summary of the validity evidence for the questionnaire within the framework described by Kane (1992). A narrative review of the key literature underpinning each of the items comprising the Osteopathy Clinical Teaching Questionnaire (OCTQ) will then be described. The limitations of the questionnaire are then outlined, along with how these limitations could be addressed within a quality assurance framework. How the OCTQ could be utilised within a clinical teaching and learning quality assurance framework will also be explored. Finally, future research possibilities following this thesis will be proposed.

The validity framework

"Our conception of validity has evolved over the last century, and it continues to evolve, but a consistent thread in this ongoing evolution has been the need to justify the claims that are being based on the test scores" (Kane, 2016, p. 209).

The conceptual framing of this thesis is the validity framework described by Kane (1992). Kane's framework has been used to summarise and argue for the validity of scores derived from educational assessments, including in medical education (e.g. Cook, 2014; Cook et al., 2015; Hatala et al., 2015; Ilgen et al., 2015). However, the use of Kane's framework has not previously been used to describe work in an educational evaluation context, particularly in the health professions, albeit the framework is suitable for application to an evaluation context (LeBaron Wallace, 2011). The structure of the framework proposed by Kane (1992) outwardly lends itself to the development and testing of educational evaluation questionnaires, hence its use in this thesis. Further, the use of Kane's framework provides for a degree of flexibility in the order in which evidence is collected and presented, and also flexibility in the type(s) of evidence to support the validity argument. This is in

contrast to the COSMIN framework (Consensus-based Standards for the Selection of Health Measurement Instruments, 2010) that describes specific evidence to be presented (i.e. convergent validity, divergent validity) in support of validity. Not all types of validity described in the COSMIN framework are necessary for an educational evaluation tool.

Educational evaluation questionnaires are often designed to describe the outcome of an intervention or program as indicated by a score. It is the interpretation of this score that Kane (2013a) contends should have evidence to support it, and that the questionnaire or measurement tool itself cannot be 'valid'. It is argued throughout chapters 2 to 5 that the OCTQ score represents the construct of clinical teaching quality in osteopathy student-led clinics. Specifically, the thesis draws on the work of Beckman et al. (2004) to define quality of clinical teaching as *the interpersonal attributes exhibited, and teaching approaches used by osteopathy clinical educators in a student-led clinical learning environment*. In the initial stages of the development of an *interpretation and use* argument, the interpretation and use of the score derived from the questionnaire needs to be explicitly described. For the OCTQ, the higher the total score, the higher the quality of clinical teaching as perceived by the learner (*interpretation*). Although the *use* of the score could be for employment decision making, for example (Snell et al., 2000), the *use* was limited to the identification of educators providing high quality clinical teaching for the purpose of recognition, and to guide clinical educator self-learning and professional development. This *use* argument reduces the requirement to provide 'strong' evidence for the *implication* inference, for example, as the stakes associated with the decision-making are reduced.

In order to build evidence for the validity of the interpretation of the score, Kane (2013b) contends that evidence should be provided for each inference: scoring; generalisation; extrapolation; and implication/inference; although the strength of evidence required to be presented for each inference is determined by the score *use*. The description presented below is intended to be a synopsis of the evidence for each inference drawing on the outcomes described in chapters 2 through to 5.

Scoring

As highlighted in chapter 1, the *scoring* inference is the most commonly presented evidence for assessments and educational measures. This is due to the ease of collecting such data and it being the natural basis for the assessments and evaluations where a score is generated. However, it is important to consider the 'step' preceding the scoring inference: *observation*. The OCTQ relies on the observations of clinical educators by learners. In the context of osteopathy clinical education, the learners are in an ideal position to be able to observe clinical educator performance and quality because of the longitudinal nature of placements in this environment (Vaughan et al., 2014b). This is consistent with the contention of Clauser et al. (2012) that there needs to be sufficient opportunity to observe the behaviour/task to make a credible observation and judgment. Here the learners are in a position to be able to make a credible judgment of the quality of clinical teaching they receive from an individual clinical educator.

Chapter 2 describes the process to identify items relevant to the evaluation of clinical teaching quality in the osteopathy student-led clinical learning environment. Relevant stakeholders (learners, academics, clinical educators) were involved in the item review process to ensure that the items were relevant to this learning environment and captured the construct of clinical teaching quality. This process provided additional evidence for the *scoring* inference and potential acceptability of the OCTQ items to these stakeholders.

Each of the studies in chapters 2 to 5 have relied on learner observation of individual clinical educators to generate a score for each OCTQ item. A summary of the studies is set out below.

- In chapter 2, the justification for the use of Likert-type response options for OCTQ items was described, allowing learners to agree or disagree with the statements identified from the literature and composed by consensus between learners, academics and clinical educators.
- In chapter 3, these Likert-type response options for the OCTQ items were evaluated as part of the Rasch analysis to ensure that each response option was

being utilised by respondents appropriately. That is, higher levels of clinical teaching quality should be reflected in a learners' choice of a higher Likert-type response option. The study in chapter 3 supports the use of the five-point Likert-type scale for the OCTQ, including the use of a neutral category. Although neutral categories are reported to be problematic as a response option, there is also evidence supporting their use (Chyung et al., 2017) and this category was retained as a response option for the OCTQ items.

- Chapter 3 also identified the need to rescore item 12 (Demonstrated osteopathic, clinical examination and rehabilitation knowledge and skill(s)). An advantage of using the Rasch model for analysis of the OCTQ was the ability to identify whether rescoring of an item was required, where respondents were not using the response options as expected by the model. For item 12, respondents were not using the agree and strongly agree options as expected, so the scoring of these items is collapsed into the one option. It should be noted that the learner can still select the agree or strongly agree option – it is only the scoring of the item that changes.
- As suggested in chapters 3 and 4, additional research could be directed towards splitting item 12 into two new items: one addressing 'osteopathic and clinical examination knowledge and skills', and the other addressing 'rehabilitation knowledge and skills'.
- Acquiescence is also reported to be an issue with student evaluations of teaching (Valencia, 2020). However, this has not been observed with the OCTQ, as shown in chapter 3. All five response options appear to be utilised by learners when completing the OCTQ providing additional evidence in support of the *scoring* inference.

Student-evaluations of teaching and clinical teaching evaluations such as the OCTQ are reported to be affected or biased by factors such as educator gender (Boring, 2017; Mengel et al., 2019; Spooren et al., 2013). In chapter 3, the influence of student gender, educator gender and institution on each of the OCTQ items was evaluated using differential item function analysis. The outcome of this analysis suggests that the OCTQ items are not systematically affected or biased by student gender, educator gender and institution. This finding may also contribute to the defensibility of the

interpretation of the OCTQ scores, that is, the score is not biased by these factors. Additionally, the lack of differential item function across teaching institutions in the United Kingdom, New Zealand and Australia supports the use of the OCTQ in different osteopathy clinical learning environments and provides evidence for the *scoring* inference. Overall, the contention presented in this thesis is that there is strong evidence supporting the *scoring* inference for the OCTQ items.

Generalisation

The generalisation inference takes the collective observations of an individual clinical educator (referred to in section the Scoring above) to make an argument about their expected performance across "a universe of possible observations" (Kane, 2013b, p. 10) or performance in the future. Cook et al. (2015) posits that the *generalisation* inference should be comprised of evidence supporting the items being selected from the "universe of possibilities" (p. 567) for the construct being measured. These authors suggest the evidence should be derived from the representativeness of the content and the reproducibility of scores.

Acceptability of a measure is, in part, influenced by its length. However, this also needs to be balanced with the need to ensure that content representativeness is achieved. Chapter 2 describes the process to identify items that could comprise a measure of clinical teaching quality in the osteopathy student-led clinical learning environment. Rather than reproducing previous systematic reviews of clinical teaching evaluation tools, an update of the reviews by Fluit (2010) and Beckman et al. (2004) was undertaken to ensure that content representativeness was consistent with the literature and also took into account newer measures that had been developed. Utilising this strategy provides evidence for the *generalisation* inference by ensuring that OCTQ items were drawn from the breadth of the clinical teaching evaluation literature.

Reproducibility is the second arm of the *generalisation* inference. Cook et al. (2015) describe the use of classical test theory approaches to demonstrating reliability and providing evidence for the *generalisation* inference. Chapter 4 presents the substantive evidence for the reproducibility and reliability of the OCTQ. A range of

reliability statistics for the OCTQ are also presented in chapters 3, 4 and 5, including item level reliability, retest reliability, reliability estimation statistics and generalisability analysis. Key features of the work through these chapters are the latter two analyses. However, all of the aforementioned statistics contribute to the evidence for the *generalisation* inference.

Further, the *reliability estimation* statistics presented throughout chapters 3 to 5 are consistent with contemporary approaches advocated by numerous authors (Dunn et al., 2014; Green & Yang, 2009; Zinbarg et al., 2005). Whilst Cronbach's *alpha* is widely used and reported in the clinical teaching evaluation literature, these authors have described limitations with its use, particularly the very limited scenarios where the assumptions underlying its use are met. McDonald's *omega* is proposed as a more appropriate alternative reliability estimation, and this statistic was used through chapters 3 to 5. In each of these chapters, the *omega* statistics were above acceptable levels, supporting the internal structure of the OCTQ, as well as the calculation of a total score for the questionnaire that represents the latent construct of *quality of clinical teaching in osteopathy student-led clinics*.

Generalisability theory can be used to provide additional evidence for the *generalisation* inference (Tavares et al., 2018). Classical test theory suggests that a score comprises the 'true score' and 'measurement error' (Brennan, 2010). Generalisability theory extends this to evaluate the influence of facets on the 'true score' (Bloch & Norman, 2012) providing a greater depth to our understanding of the score derived from a measure. Chapter 4 describes the use of generalisability theory for the analysis of OCTQ scores. An advantage of generalisability analysis is the ability to estimate the number of responses required to support reliable decision making. Consistent with the proposed *use* of the OCTQ for professional development and identification of clinical educators providing high quality clinical teaching, eight questionnaires for an individual educator would be required. This number is feasible across most osteopathy student-led clinical learning environments and provides evidence for the *generalisation* inference.

Additional evidence for the *generalisation* inference is also provided by the acceptable retest reliability statistics as described in chapter 4. Retest reliability of 11 of the 12 OCTQ items was supported; however, item 12 demonstrated statistics that were below an acceptable level. The potential issue with this item capturing multiple elements of clinical teaching quality was highlighted in chapter 4 and in the *scoring* inference discussion in the Scoring section above. This issue may impact on its retest reliability and further investigation after splitting the item could test this assertion. That said, the evidence for the retest reliability presented in chapter 4 provides additional support for the *generalisation* inference. Inter-rater reliability of the OCTQ was also examined and reported in chapter 4. Whilst high inter-rater reliability is desired in most measurement scenarios, it was argued in chapter 4 that moderate inter-rater reliability should be acceptable for evaluations of clinical teaching. The quote at the start of the current chapter highlights why two learners may not necessarily agree on the quality of clinical teaching provided by the same clinical educator, thereby impacting learners' inter-rater reliability. The retest and inter-examiner reliability statistics provide evidence in support of the *generalisation* inference.

Extrapolation

The "observed score might be used to draw conclusions about the [level of measurement] on some trait or construct that is assumed to explain the observed performances" (Kane, 2013b, p. 11), an inference based on credible theory or theories (Kane, 1992). As highlighted in Chapter 1, there have been no theoretical explorations of clinical education as undertaken in osteopathy. However, the contention throughout chapters 2 to 4 is that the OCTQ evaluates the *interpersonal* and *clinical teaching* domains described by Beckman et al. (2004) and demonstrates the properties of measurement invariance and measures the construct of the *quality of clinical teaching in osteopathy student-led clinics*.

Cook et al. (2015) describes the *extrapolation* inference as the link between the score observed and real-world performance. The OCTQ was designed to allow learners to provide an evaluation of the quality of clinical teaching provided by a clinical educator; hence, the real-world performance of the educator is the behaviour that is

rated. In the case of the OCTQ, evidence for the *extrapolation* inference comes from multiple sources. The first of those sources is the differential item function (Cook et al., 2015) and this has been described in chapter 3 in addition to the Scoring section. Again, the OCTQ items do not demonstrate differential item function. The factor structure can provide additional evidence for the extrapolation inference. Chapter 2 described the use of a contemporary statistical approach to the exploratory factor analysis of an early version of the OCTQ. This work supported a factor structure consistent with the cognitive apprenticeship model (Collins, 1991) – a model used by Stalmeijer et al. (2010a) for the development of the Maastricht Clinical Teaching Questionnaire and proposed as a model that could underpin osteopathy clinical education (Vaughan et al., 2014b).

Work described in chapter 4 using the Rasch model also provides evidence for the *extrapolation* inference. The results of the Rasch analysis demonstrated the OCTQ is unidimensional, that is, it measures a single latent construct. Although this is different to the five-factor structure described in chapter 2, the fact that the items are drawn from the same pool and are consistent with the interpersonal attributes and teaching approaches structure suggested by Beckman et al. (2004), provides evidence for the *extrapolation* inference.

Additional evidence for this inference is also presented in chapter 5 where self-assessment of clinical teaching quality and its relationship to learner outcomes was described. Although there are limitations with the use of self-assessment (Eva & Regehr, 2008; Regehr & Eva, 2006), when combined with other data sources it can be valuable, particularly over time where clinical educators have the ability to reflect and refine their teaching approach. The data in chapter 5 contributes, in a small way given the inherent limitation of self-assessment, to the evidence for the *extrapolation* inference.

Responsiveness is the ability of a measure to demonstrate change where change has occurred and is another form of evidence that can be provided for the *extrapolation* inference. Whilst none of the work in the current thesis directly addresses responsiveness of the OCTQ, the thesis does provide a basis on which this could be

explored. As part of the outcome of the Rasch analysis in chapter 3, interval-level scores for the questionnaire are presented. An advantage of the Rasch model is the ability to convert ordinal-level total scores into interval-level scores (Pallant & Tennant, 2007) that are potentially more suited to parametric statistical analysis. It is also argued that these interval-level scores are more appropriate for detecting change in clinical teaching quality than ordinal level scores. This assertion requires testing and the calculation of the minimum educationally important difference score for the questionnaire (Steiner et al., 2003) presents an avenue for further research.

Interpretation or implications

The *interpretation/implication* inference is based on evidence about the intended and unintended consequences of the evaluation (Kane, 2016) and is often under-explored due to the challenge of developing evidence to support it (Cook et al., 2015). The intended consequence argued throughout this thesis is to identify quality clinical teaching as perceived by the students, and where additional support (i.e. professional development) could be used to improve teaching quality. Unintended consequences in the current work were explored, particularly those related to educator/learner gender bias that may influence learner evaluations as described in chapter 3. The potential for gender bias was also identified in chapters 1 and 5 when describing SETs.

In the context of the OCTQ, one source of evidence in support of this *interpretation* inference is the ability to use the scores to group clinical educators based on clinical teaching quality. Chapter 3 describes the process by which this was demonstrated through a Rasch analysis of the OCTQ. OCTQ total scores can be used to place a clinical educator in one of four groups and these groups relate to the proposed *use* (i.e. professional development, recognition). Clinical educators demonstrating lower OCTQ scores may benefit from professional development and support in their teaching role, and those in the highest OCTQ score group could receive some form of recognition for their teaching.

There is also the potential for the OCTQ to be used to measure change in clinical teaching quality following participation in professional development activities, providing additional evidence in support of the *implication* inference. As described in chapter 3 and the Extrapolation section of the current chapter, the development of the OCTQ using Rasch analysis allowed for the calculation of interval-level scores. This will allow for small, meaningful changes in clinical educator teaching quality to be reflected in OCTQ scores observed over time. Measurement of this change will be the subject of future research in order to strengthen the evidence for the *implication* inference.

Literature underpinning Osteopathy Clinical Teaching Questionnaire items

The items comprising the OCTQ were derived from published measures of clinical teaching quality in health professions education. The process to identify these items was described in Chapter 2 and was primarily informed by the systematic reviews of clinical teaching questionnaires by Fluit (2010) and Beckman et al. (2004), with an additional literature search used to identify measures published after these reviews (Chapter 2). The following discussion will provide an evidence base for each of the items included in the final version of the OCTQ. This discussion draws on key literature for the concept(s) reflected in each item and considers the *current* state of the literature. Further, it should be appreciated that this review is not meant to be exhaustive. Table 3 maps how these items relate to the seven roles of the Australian Capabilities for Osteopathic Practice (Osteopathy Board of Australia, 2019).

The items comprising the OCTQ focus on the clinical educator's ability to:

1. Maintain a positive attitude towards students;
2. Demonstrate humanistic attitudes in relating to patients;
3. Show genuine concern for student's professional well-being;
4. Communicate well;
5. Remain open to student questions and alternative approaches to patient management;
6. Adjust their teaching to meet student's learning needs;

7. Promote student's reflection on clinical practice;
8. Use a problem-based teaching style;
9. Ask questions that enhance learning;
10. Stimulate students to learn independently;
11. Offer suggestions for improvement; and,
12. Demonstrate osteopathic, clinical examination and rehabilitation knowledge and skill(s).

The three global rating items allow learners to provide a quantitative evaluation of their overall impressions with respect to the quality of clinical teaching they have received from an individual clinical educator.

Item 1. Maintain a positive attitude toward students.

The importance of knowing that students come away from clinical education experiences having a sense that their clinical educator had a positive attitude toward them, and their learning journey, cannot be underestimated. This item was derived from the work by Shellenberger and Mahan (1982) who developed a measure of clinical teaching quality in the medical general practice setting. That the clinical educator ought to display a positive attitude towards the learner is a consistent theme in the clinical teaching quality literature (AlHaqwi et al., 2010; Gibson et al., 2019; Hanson & Stenvig, 2008; Ramani & Leinster, 2008). This positive educator attitude has also been described in the context of role-modelling by an educator (Cruess et al., 2008; Jochemsen-van der Leeuw et al., 2013; Knight & Bligh, 2006), reducing the influence of the hidden curriculum on learners (Knight & Bligh, 2006), and improving intrinsic motivation of learners (Bengtsson & Ohlsson, 2010; Gibson et al., 2019). The importance of a clinical educator's enthusiasm for teaching is frequently identified in the literature and is regarded highly by learners (AlHaqwi et al., 2010; Buchel & Edwards, 2005; Parsell & Bligh, 2001; Ramani & Leinster, 2008; Sutkin et al., 2008). This is exemplified by the following quote from Ramani and Leinster (2008, p. 354):

"The starting point for any good teacher must be enthusiasm for the subject being taught. This has to be complemented by an eagerness to transmit this enthusiasm to others, which will necessarily result in a positive attitude to learners."

Duration of a placement has also been identified as another rationale for the educator to display positive attitudes (Smedley et al., 2010). Short-term placements, and placements with multiple supervisors require the educator to display a positive attitude towards the learner in order to maximise the educational relationship (Smedley et al., 2010). Learners working with multiple clinical educators on a placement is commonplace in osteopathy education (Vaughan et al., 2014b). Therefore, educators should be encouraged to demonstrate a positive attitude from their initial interactions with learners through to the conclusion of the placement.

Smedley et al. (2010) also suggest clinical educators with higher levels of preceptor (educator) self-efficacy were more likely to self-report a more positive attitude towards students and that this positive attitude can be fostered through professional/faculty development that influences educator self-efficacy. Data from the current work (Chapter 5) captured both self-reported attitude towards students and self-efficacy, using the Self-Efficacy of Clinical Teachers scale (McArthur, 2016). This data suggests there is are small to moderate positive relationship between self-efficacy and quality of clinical teaching reported by learners. Such an outcome presents an opportunity to design faculty development based on self-efficacy (Bray-Clark & Bates, 2003).

Arah et al. (2012) suggests female clinical educators are more likely to receive higher ratings on items related to 'communicative' aspects of clinical education practice such as attitudes towards learners. However, it should be noted that this finding is by no means consistent across the literature (Ladha et al., 2017). In the development of the OCTQ, differential item function analysis (Andrich & Hagquist, 2012) was used to reduce the systematic influence of both educator and learner gender and responses to OCTQ items. That is, the item functions in the same manner regardless of gender. However, this does not preclude individual educator differences being identifiable for this OCTQ item. Where differences are identified, it is more likely that the difference

is a true reflection of gender difference rather than the item interpretation and response resulting in the difference.

Finally, there is also a reported effect where the positive attitude of a clinical educator can influence a student's choice of medical speciality (Budd et al., 2011). Although this is unlikely to be a consideration in osteopathy, it is possible that students may be 'put off' or develop negative impressions of the profession where clinical educators do not maintain a positive attitude.

Item 2. Demonstrate humanistic attitudes in relating to patients

Throughout the health professions, the clinical educator's ability to demonstrate humanistic attitudes toward patients – that is, to demonstrate integrity, compassion and respect in patient care is critical. This item's inclusion in the OCTQ was inspired by the work of Hewson and Jensen (1990) in their exploration of the internal medicine clinical environment. It is generally considered that humanistic approaches to 'whole person' patient-care (des Ordons et al., 2018) are paramount for quality care (Branch Jr et al., 2001; Gracey et al., 2005) and can be defined "as being patient-centered and integrating the psychosocial with the biomedical aspects of care" (Kern et al., 2005, p. 8).

Consistent with the discussion of OCTQ item 1 above, item 2 likely reflects the students' perception of educator role-modelling of humanistic and patient-centred care (Gibson et al., 2019; Hekelman & Blase, 1996; Ramani & Orlander, 2013; Yazigi et al., 2006). This role-modelling in the consultation room (Ramani, 2003; Ramani & Orlander, 2013) may be one of the few ways that we can instil humanistic behaviours in our learners, given it is difficult to 'teach' clinical skills in other learning environments (Branch Jr et al., 2001; Hanna & Fins, 2006). Branch Jr et al. (2001) also identify the influence of the hidden curriculum on the demonstration (or lack thereof) of these attitudes towards patients. The educator may be able to display these humanistic attitudes towards patients through their communication, engagement of the patient in care discussions with the learner (Ramani & Orlander, 2013), and attention to patient comfort and modesty (Branch Jr et al., 2001).

This OCTQ item asks the learner to rate the humanistic attitudes towards individual patients displayed by their educator and possibly how the clinical educator engages the patient with their care and the education of the learner in the treatment room. In the student-led clinical learning environment, the opportunities for the learner to observe these educator/patient interactions will be minimal in the context of the whole patient consultation as the student is 'leading' the consultation. It is also likely that the educator does not want to influence the learner/patient interaction substantially to ensure that the learner makes the most of this pedogogically rich activity (Billett, 2009). Gracey et al. (2005) suggest that educators can exhibit these humanistic attitudes by exploring the patient psychosocial issues relevant to their health concern, and also through observation of the learner/patient interaction. Furthermore, the success or otherwise of this display of humanistic behaviours may require the educator to explicitly 'prime' or orient the learner, prior to the consultation, to observe these behaviours during a patient interaction (Branch Jr et al., 2001), then co-reflect subsequent to the consultation (Gracey et al., 2005; Kern et al., 2005). This process can enhance the chances of the clinical interaction being a learning 'moment'. It is likely that communication strategies role-modelled by the educators are being observed by the learner (Passi et al., 2013) and, in the context of the current work, it is possible that the learners' response to this OCTQ item is a reflection of this role-modelling.

Item 3. Show genuine concern for student's professional well-being

This item in the OCTQ was taken from the work of Silber et al. (2006) who developed this item using a critical incident methodology, drawing on the experiences of a range of medical residents (learners) and their clinical educators at a single institution. The work of Blakey et al. (2019) suggests that exemplary clinical educators were often identified as being those who enquired about learner well-being. The well-being of health professions learners has been a substantial focus of the peer-review literature, particularly in medical education (Raj, 2016; Rotenstein et al., 2016; Wasson et al., 2016). Beckman et al. (2010) explored the role of resident (learner) well-being and whether this played a role in their evaluations of faculty. These

authors found no associations between measures of well-being (quality of life, burnout, depression) and faculty clinical teaching evaluations. Whilst this result is a valuable aspect of the validity evidence of clinical teacher evaluations (i.e. these concerns do not influence evaluations), educator concern for the well-being of learners appears to have a positive impact on the learning environment.

Monitoring of learner well-being and its interplay with the learning environment was identified as a key role of the general practice educator/supervisor in the review by Wearne et al. (2012). These authors highlighted that part of the concern for learner well-being was related to ensuring patient safety and monitoring of learning. Clinical educators can also influence and facilitate intrinsic motivation, which has been shown to positively influence learner well-being (Orsini et al., 2015; Walker et al., 2016). Others have also suggested that well-being strategies be taught within the clinical learning environment (Rich & Lamiman, 2019) and this may be a role that could be undertaken by a clinical educator. Additionally, this regard for learner well-being may create an environment where the learner feels psychologically-safe to articulate their weaknesses and challenges with practice (Wearne et al., 2012). Conversely, the 'less active' clinical educator can negatively impact on the well-being of the learner causing additional stress (Buery-Joyner et al., 2019), or display a complete lack of concern about the learners' well-being, impacting on their perception of relatedness, or sense of belonging, in the clinical learning environment (Knight, 2018).

Item 4. Communicates well

This item in the OCTQ was developed by Roff et al. (2005) who included the item in the Postgraduate Hospital Educational Environment Measure (PHEEM). Unquestionably, good communication skills are vital in the health professions (Denniston et al., 2019) - whether it be for patient care, work with colleagues, or engaging with learners. Effective communication may underpin "teaching excellence" (Prideaux et al., 2000) and appropriate communication skills are consistently rated as one of the most significant behaviours demonstrated by effective clinical educators (Kelly, 2007; Prideaux et al., 2000; Sutkin et al., 2008). Learner's "most certainly recognize the importance of their teachers' communication skills"

(Prideaux et al., 2000, p. 822). In the context of the current work, the term 'communication skills' refers to the clinical educators' interactions with the patient and the learner. Thus, the inclusion of this important item could be considered a standard item in any teaching evaluation.

Item 5. Remain open to student questions and alternative approaches to patient management

The stem for this item was derived from the work of Solomon et al. (1997), and was modified to meet the student-led clinical learning environment in this thesis. The item described by Solomon et al. (1997) used the term 'disagreement'. However, it was felt this was too strong for a pre-registration learning environment and subsequently modified to 'alternative approaches'.

A willingness to answer learner questions has been identified as a characteristic of effective clinical educators (Forbes et al., 2018; Sandhu et al., 2016) and the wider clinical learning environment (Kandiah, 2017; Stalmeijer et al., 2009). Asking questions of the educator is also built into clinical teaching models such as the SNAPPS[1] – probing the preceptor (Pascoe et al., 2015; Wolpaw et al., 2003). Learners utilise questions as part of their learning strategy (Edgecombe & Bowden, 2009) but will likely only do so when they feel psychologically safe to do so (Finn et al., 2016; Perry et al., 2018) and feel as though the educator is a 'partner' in their learning (Perry et al., 2018). Although the learner may feel comfortable enough to ask a question of the educator, work by Stalmeijer et al. (2009) with senior clinical learners indicated that frustration could develop where the response to the learner was to 'look it up'. Educators may have used this response in an effort to encourage the learner to direct their own learning. However, this response may be perceived negatively by learners.

Chapter 2 highlights the idea of 'supported participation' described by Dornan et al. (2007) and how this may contribute to a learners' willingness to proffer alternative

[1] Summarise, Narrow, Analyse, Probe, Plan, Select (Wolpaw et al., 2003)

patient management strategies. In the early stages of their clinical learning, the learner is likely to need assistance and support with their patient management. In the osteopathy student-led clinical learning environment, the learner has been able to learn in the classroom and observe others managing patients. However, involvement in a patient care role at this point in time has been limited (Vaughan et al., 2014b).

In the latter stages of the curriculum as the learner gains more clinical experience and exposure, it is anticipated the learner will be more confident (or comfortable) in their own competence and capability to manage patients (Polyzois et al., 2010). It is in these latter stages of training that the learner should be in a position to offer alternative management approaches. In the dental clinical learning environment, learners identified that disagreement with the proposed management approach was the least helpful characteristic of a clinical educator (Polyzois et al., 2010). Further, openness to student input in patient care and alternative management strategies, has been shown to be a behaviour of effective clinical teachers (Gibson et al., 2019; Wijbenga et al., 2019) and is strongly associated with overall clinical teaching satisfaction (Field et al., 2019).

A significant advantage of the osteopathy student-led clinical learning environment is that it allows the learner to lead the patient management, within the confines of the oversight of the clinical educator, particularly patient safety (Moore et al., 2018). In this context, the 'alternative approach' to patient care is one that the clinical educator may not necessarily use themselves or agree with, however if there is little risk to patient safety then the implementation of the proposed care provides an experiential learning opportunity.

Item 6. Adjust their teaching to meet student's learning needs

Van der Hem-Stokroos et al. (2005) included this item in the Clinical Teaching Effectiveness Instrument. The literature suggests that those who are regarded as excellent and/or effective clinical teachers will adjust their teaching strategies to match learners' level of competence, and sequence learning activities to extend and build on this competence (Busari et al., 2006; Busari et al., 2005; Chen et al., 2015;

Gibson et al., 2019; Goldie et al., 2015). Effective clinical teachers also appear to adjust their teaching to support patient safety (Chen et al., 2015) and are self-aware (Sutkin et al., 2008). Frith et al. (2015) also note that guided individual and/or group reflection on clinical teaching can assist in developing clinical educator competence to adjust their teaching. This approach to the use of reflective practice to encourage educators to adjust their teaching to learner needs and competence could form part of curricula for clinical educators.

Moreover, within the Cognitive Apprenticeship Model (Collins, 1991) described in Chapter 2 is the concept of *scaffolding*. Stalmeijer et al. (2009, p. 537) suggest "that support from teachers for students' learning must be tailored to students' individual knowledge levels. As students become more competent support can be gradually reduced" and this reflects an adjustment in teaching to learner needs captured in this OCTQ item. The notion of scaffolding takes account of the changing needs of the learner and of the type and amount of supervision required. As the learner progresses through their clinical learning, the type of supervision required will evolve whilst still ensuring patient safety (Olmos-Vega et al., 2015).

Item 7. Promote student's reflection on clinical practice

Hewson and Jensen (1990) developed this item for a questionnaire to assist in improving the quality of teaching within internal medicine. Reflective practice continues to be espoused as a fundamental behaviour of effective health professionals (Mann et al., 2009) and is consistent with the use of the Cognitive Apprenticeship Model to structure clinical learning (Stalmeijer et al., 2013).

Most health professions pre- and post-professional curricula include elements of reflective practice and literature supports the desire for both learners and clinical educators to reflect on their practice within the clinical learning environment (Frith et al., 2015; Parsell & Bligh, 2001; Snell et al., 2000; Steinert et al., 2016; Van Lierop et al., 2018). Role-modelling of reflective practice by the clinical educator can assist in promoting learner reflection and establishing this behaviour as a part of everyday practice (Cruess et al., 2008; McSparron et al., 2019). Promotion of reflection can be

achieved through exhibiting trust in the learner, displaying interest in patients, and developing a skillset to facilitate reflection (Branch Jr & Paranjape, 2002), including developing an understanding of the appropriate question(s) to ask (McSparron et al., 2019).

In osteopathy student-led clinics, like in other clinical education environments, reflection by the learner can also be promoted through the clinical educator displaying a genuine interest in furthering the learners' competence (Fluit et al., 2012; Gibson et al., 2019). Thus the importance of the inclusion of this question in the OCTQ stems from the fact that developing the skills to facilitate reflection should form part of professional development for clinical educators (Steinert et al., 2016).

Item 8. Use a problem-based teaching style

As far back as 1975, the clinical teaching literature emphasised the fostering of problem-solving skills in clinical learners (Jarski et al., 1990; Smith & Lane, 2015; Spencer, 2003; Stritter et al., 1975) and its value as a co-learning activity (Strand et al., 2015). This OCTQ item was identified from the measure of clinical teaching quality developed by Nation et al. (2011) based on the CanMEDS framework. This item related to the Scholar domain in the CanMEDS and was in the *Teaching skills* factor of the Clinical Teaching Assessment Instrument. Unquestionably, appropriate question phrasing can assist the learner through this problem solving, by linking the current patient care experience with their knowledge-base (Parsell & Bligh, 2001; Rich & Lamiman, 2019). There is a small volume of literature on the use of case-based/problem-based learning approaches in the pre-clinical curricula. As such, learners have exposure to this teaching and learning approach prior to their clinical learning (Johnston & Vaughan, 2020; Lalonde, 2013). Again, the importance of including this item in the OCTQ stems from the fact that we also know professional development has been demonstrated to improve the clinical educators' ability to engage the learner with problem-solving (Delany et al., 2020; Liao et al., 2019).

Item 9. Asked questions that enhance learning

This question was derived from the MedIQ developed by James and Osbourne (1999) as a measure of teaching quality in the ambulatory clinic context. Questions posed by the clinical educator are consistently used throughout a learners' clinical placement (Lo & Regehr, 2017; Steinert et al., 2017; Wear et al., 2005), with questioning scaffolded from closed-questions in the early stages to more open-questions as the learner progresses (Steinert et al., 2017). Questioning is reported to be used by clinical educators to establish a learners knowledge level ('diagnose the learner' (Beckman & Lee, 2009)) and highlight the skills and knowledge that are important for the learner to know (Steinert et al., 2017).

Clinical educators should be cognisant that the choice of questioning may either assist or hinder the learner. Wear et al. (2005) explored learner conceptions of 'good' and 'bad' questioning, with the latter typically being the 'humiliation'. Continuing to draw on the Cognitive Apprenticeship Model (Collins, 1991), the *articulation* domain involves the clinical educator asking questions of the learner as a strategy to improve their competence (Stalmeijer et al., 2009). These authors identified the strategy as being useful right throughout a learners' clinical education - clinical educators were described as using questioning to 'stimulate' learning or 'humiliate', that is, expose gaps in a learners' knowledge.

The latter approach is described in the literature in the context of the Socratic questioning method (Stoddard & O'Dell, 2016) – a common questioning approach in clinical education environments. Recent work using the Socratic approach, where a series of difficult questions is posed to a clinical learner or 'pimping', has highlighted where this strategy is unlikely to be productive unless there is a climate of psychological safety (Stoddard & O'Dell, 2016). Regardless of the learning environment, this strategy may be perceived as mistreatment (McEvoy et al., 2019) and should be avoided where possible.

Lo and Regehr (2017) also highlight the strategies that learners use in response to various questioning strategies. In qualitative interviews, these authors identified that students tailor their responses to educator preferences and perceived questioning intent by adopting strategies to retain a positive image in the eyes of the educator (Lo

& Regehr, 2017). Where the questioning took place (i.e. in front of a patient, away from the consultation room) also influenced the choice of strategy adopted by learner. However, learners also appeared to appreciate the role educator questioning had in projecting a positive image of them to the patient.

To highlight the value of questions to enhance learning, Hausmann and Schwartzstein (2018, p. 2) suggest that with:

"Mastering the art of questioning, faculty can improve a learner's ability to understand and remember the material, apply knowledge in new settings, overcome cognitive biases, support an environment of inquiry, and develop metacognition, even under uncertainty".

Item 10. Stimulate students to learn independently

Cox and Swanson (2002) included this item in their questionnaire exploring teaching excellence in the surgical and clinical context. Fostering independent learning is consistent with adult learning theory – the notion that learners have a desire to direct their own learning, drawing on previous experience and incorporating this with new knowledge, as well as actively participating in work related to future practice (Burns et al., 2006). Further, an outcome, or attribute, desired of graduates from university level education is the attainment of skills to ensure independence as a lifelong learner, whilst also developing discipline competency (Huttly et al., 2003; ten Cate et al., 2004).

Fostering independent learning as a clinical teaching strategy (Busari et al., 2005; Gibson et al., 2019; Parsell & Bligh, 2001; Valiee et al., 2016), while at the same time being supported by the clinical educator (Van Gelderen et al., 2018), is one of the identified behaviours of effective clinical teaching. We know that in the clinical context learners are likely to rely on the clinical educator in their early clinical learning and progress to more independent practice, or require less supervision requirements as they become more experienced (Beach, 2017; Burns et al., 2006; Parsell & Bligh, 2001). Thus, achieving independence has also been suggested as a

signal the learner is 'learning' in the clinical setting (Burns et al., 2006). DaRosa et al. (1997) also argue that independent learning may be a time saving teaching strategy for the clinical educator, in addition to developing learner critical thinking skills. As clinical educators have significant dual responsibilities for patient care and student education thus, having capabilities to develop students as independent learners is crucial to the success of both.

Item 11. Offer suggestions for improvement

This item was derived from Beckman et al. (2003) who utilised it in a questionnaire for peer evaluation of teaching in the clinical learning environment. The OCTQ only explores the feedback provided for improvement in performance and not where the feedback may be corrective or otherwise – this improvement orientation may assist in fostering the learners' intrinsic motivation (Johnson et al., 2016). This was explored in Chapter 3 where many of the other 'feedback' items were removed from the analysis to create the final version of the OCTQ.

To improve the quality of their work learner's require feedback (Boud & Molloy, 2012) and in the clinical learning environment this feedback can be obtained from numerous sources, including the clinical educator (Burgess & Mellis, 2015). In this context the clinical educator may be viewed as a credible source (Bing-You et al., 1997; Eva et al., 2012), particularly where feedback is based on direct observation of performance (Kraut et al., 2015; Schlair et al., 2017; Schopper et al., 2016).

Again, this OCTQ item focuses specifically on 'improvement'; therefore, the feedback should be coupled with an action plan to enact in the near future (Burgess & Mellis, 2015; Farrell et al., 2016). However, educator practice in this regard appears to be variable (Hamburger et al., 2011; Pelgrim et al., 2012). Educators may provide such feedback when implementing clinical teaching models, for example, the One Minute Preceptor (Farrell et al., 2016). Where the learner is in a position to be able to identify their own performance gaps, this enables the clinical educator and learner to co-construct a plan for improving performance (Johnson et al., 2016). Finally, the ability and willingness to provide learners with feedback is consistently identified as a

key skill of clinical educators in medicine (Cantillon & Sargeant, 2008) and in allied health (Gibson et al., 2019), supporting the inclusion of this feedback item in the OCTQ.

Item 12. Demonstrate osteopathic, clinical examination and rehabilitation knowledge and skill(s)

There is often great interest on the part of the learner in the clinical educator demonstrating the psychomotor and technical skills required of the profession (Fluit et al., 2012; Hesketh et al., 2001; Laurent & Weidner, 2001). Schum et al. (1993) included this item in the measurement of teaching effectiveness in a paediatric clinical learning environment.

Learners often rate their clinical educators highly on such skill based items (Cox & Swanson, 2002; Ramsey et al., 1988). As an important characteristic of an effective clinical educator (Gat et al., 2016; Gibson et al., 2019; Irby et al., 1987; Laurent & Weidner, 2001), learners appear to appreciate when skills are demonstrated repeatedly (Stalmeijer et al., 2009), pitched at their level of learning (Burgess et al., 2015) and within the patient care context (Gat et al., 2016). Furthermore, demonstrating these skills at the bedside has also been shown to positively influence patient satisfaction (Lehmann et al., 1997).

Within the Cognitive Apprenticeship Model, the demonstration of clinical skills and knowledge is described within the *modelling* domain (Stalmeijer et al., 2009). Laurent and Weidner (2001) also located this teaching skill within the modelling subgroup of effective behaviours for athletic training clinical educators. However, the excellent clinician may not make an effective educator given the differing skill sets (Summers, 2017). Further, it may be difficult for the novice clinical educator, regardless of their clinical experience, to explain what has likely become subconscious in their own practice (Wallace & Infante, 2008) although this think-aloud approach can be taught through professional development (Delany et al., 2020).

It may also be possible that learners are rating their clinical educators' overall clinical competency using this item as a surrogate measure. The OCTQ does not contain an item specifically addressing the learners perception of educator clinical competency even though it is consistently identified as being a characteristic of an effective clinical educator (Buchel & Edwards, 2005; Fluit, 2010; Kelly, 2007; Snell et al., 2000). Given learners are still in the formative stages of their role as a health professional, it may be unreasonable for learner to comment on the 'competency' of a qualified practitioner (Spooren & Christiaens, 2017). Such an issue was raised in Chapters 1 and 5 in relation to student evaluations of teaching – learner perceptions of competence may not be a true reflection.

Global rating items

Global ratings are widely used in measurement of perceptions, with SETs (Zhao & Gallant, 2012) and clinical teaching evaluations being no exception. Three global rating items were included on the OCTQ, two derived from the literature and one developed through discussion between the researcher and one of the supervisors (JM).

The first of the global ratings '*Rate the overall effectiveness of this Clinical Educator'* was described in the work by Afonso et al. (2005) and the second '*I would do more clinics with this Clinical Educator'* is from Zuberi et al. (2007) in the context of outpatient clinical teaching. The final global rating '*I would recommend other students to work with this Clinical Educator'* was derived from similar items in the patient satisfaction measurement literature (Batbaatar et al., 2015; Batbaatar et al., 2017; Tung & Chang, 2009).

The main purpose of including a global rating scale (GRS) is to capture aspects of a construct that may not be obtained from items in the questionnaire, and single-item GRSs have been shown to be a suitable measure of constructs such as general health (Macias et al., 2015) and life satisfaction (Cheung & Lucas, 2014). There is a some literature that supports the use of global ratings as an alternative to checklists in health professions assessment (Cömert et al., 2016; Hatala et al., 2015; Ilgen et al., 2015; Read et al., 2015). However, this literature is less developed in clinical teaching

evaluations. This assessment literature provides validity evidence supporting the use of a GRS over checklists (Cunnington et al., 1996; Hodges & McIlroy, 2003) and this validity evidence may also be generalisable to clinical teaching evaluations.

Work by Williams et al. (2002) provides some of the strongest evidence supporting the use of GRSs in clinical teaching evaluations. These authors demonstrated strong correlations between the domains of the Stanford Faculty Development questionnaire and a global rating item with between 74-96% shared variance. With such substantial levels of shared variance, Cashin and Downey (1992) posit that "a short and economical form could capture much of the information needed for summative evaluation and longer diagnostic forms could be reserved for teaching improvement" (p. 563). However, others using clinical teaching evaluations have demonstrated lower levels of shared variance (Arah et al., 2011; Lombarts et al., 2009; Van der Leeuw et al., 2011; Zibrowski et al., 2011) more in line with that expected in health measurement for example (between 16-64%) (Streiner et al., 2014). That said, there appears to be value in the use of GRSs in a clinical teaching evaluation (Zibrowski et al., 2011) with respect to ease of administration and stakeholder acceptability. However, it does not allow for identification of specific educator performance considerations that may guide future professional development (Mintz et al., 2015). Together, this literature suggests a combination of items targeted towards specific educator characteristics and global rating items is likely to produce an accurate representation of the quality of clinical teaching provided by an educator from the student perspective.

What the Osteopathy Clinical Teaching Questionnaire does not measure

Throughout this thesis, reference has been made to the *method* strategy within Cognitive Apprenticeship Model (CAM) as the underpinning educational framework for the OCTQ. The *method* strategy within the CAM has been proposed as a model for clinical education in medicine (Stalmeijer et al., 2009). However, its use as a framework for osteopathy clinical education is only hypothesised. How teaching and learning in this osteopathy student-led environment occurs is not known and this thesis has used a data-driven approach to the identification of a possible framework,

that being the CAM. Brown et al. (2019) suggest that one use of theory is to assist with the interpretation of data, and it is this use that the thesis applies.

In chapter 2, the exploratory factor analysis identified the item groupings that were consistent with those of the CAM and described by Stalmeijer et al. (2008). Reduction of the number of items in the OCTQ through a Rasch analysis described in chapter 3 suggested that the coverage of the domains within the CAM was reduced. However, additional analysis using bifactor modelling in the McDonald's omega calculation suggested that the 12 items comprising the OCTQ were consistent with the *interpersonal* and *clinical teaching* domains described by Beckman et al. (2004). In discussing what the OCTQ is not measuring, the thesis uses the systematic review by Fluit (2010) and the mapping of published clinical teaching evaluation items to the CanMEDS roles. In this work, Fluit et al. (2012) identified that whilst a significant number of the CanMEDS roles were covered, aspects such as assessment and planning were not well represented. With respect to the OCTQ, a number of areas of educator practice were not included: assessing learning in the workplace; planning learning; development of teaching resources; and health record keeping.

Assessing learners in the workplace

One of the limitations of the OTCQ is related to the assessment of learners in the workplace setting. In the development of the OCTQ through chapter 2, neither educators nor learners identified workplace-based assessments as a role for the osteopathy clinical educator. Numerous authors (Hays, 2008; Hesketh et al., 2001; Ramani & Leinster, 2008) have articulated that assessment should form part of the role of the clinical educator. However, this was not included as an item in the OCTQ. As highlighted earlier, Fluit (2010) also identified that assessment of learners was broadly underrepresented in evaluations of clinical teaching. Why this might be the case is an avenue to explore in further work.

Assessment in the non-United States osteopathy clinical learning environment has received scant attention in the peer-review literature beyond a small number of commentaries (London, 2008; Moore & Vaughan, 2015; Vaughan et al., 2013;

Vaughan et al., 2014b) and empirical pieces (Orrock et al., 2014; Vaughan & Moore, 2015a, 2015b). The limited literature may be a reflection of the level of interest in workplace-based assessment in this learning environment, or a genuine reflection that assessment of students is not a role for the osteopathy clinical educator. Other work has suggested that educators do not see assessment as part of their role, given the lack of time to undertake such an activity in addition to providing a patient-care role and attending to learner needs (Al-Kadri et al., 2013; Koh, 2008; Seabrook, 2003; Shayne et al., 2002), or considering that it may negatively influence the educator-learner relationship (Hays, 2008; Meyer et al., 2019). Regardless of the reason(s) for its non-inclusion in any evaluation tool, learner evaluation of assessment in this environment may be better situated in a whole learning environment measure, rather than in a measure of clinical teaching quality of a single educator.

Planning for, and of, learning

Although it has been identified in other measures of clinical teaching quality (i.e. Wormley et al., 2017), learners' perceptions of the clinical teachers' planning for teaching is not evaluated in any of the OCTQ items. In both the in-patient and outpatient teaching settings, it is incumbent on the clinical educator to identify opportunities for the learner to participate in clinical work and activities (Fluit et al., 2012). In the student-led clinical learning environment, the learners are already participating in clinical work through their supervised, patient management role (Vaughan et al., 2014b). As such, this planning on the part of the educator is not necessary *per se*, rather it is opportunistic. Planning of structured clinical education activities in the student-led environment is typically undertaken by clinical academics and placement administrators who work within the governance procedures and policies of the institution. Given what can be a broad remit, there are always opportunities for clinical educators in this environment to structure their teaching to take advantage of the unique features of this learning environment. Billett's concept of *pedagogically rich activities* (Billett et al., 2013) could prove useful as a lens through which an educator can view their daily work and identify learning opportunities. Consistent with the assertion in the 'Assessing learners in the

workplace' discussion above, planning of academic activities in the student-led clinic environment may be best measured as part of a global learning environment measure.

Resource developer

Development of educational resources for learners has been identified as a role of the clinical educator (Fluit, 2010). How this applies in the student-led clinical learning environment is unknown and requires exploration. There is an opportunity for clinical educators to develop resources guided by that day's patient list or learner experiences. However, many of the resources a learner would access and utilise are developed by the academic staff within the course as part of their planning as described in the 'Planning for, and of, learning' section above.

Health record keeping

Clinical teaching evaluation items that explore health record keeping practices are rarely included in questionnaires designed to evaluate the quality of clinical teaching. Hewson and Jensen (1990) included one such item in their questionnaire, "*Promoted keeping of medical records that is thorough, legible, efficient and organised",* suggesting that this practice should be role-modelled by a clinical educator (Atwater et al., 2016; Ellaway et al., 2013; Varpio et al., 2015), as it is best learnt in the clinical learning environment (Atwater et al., 2016; Ellaway et al., 2013). Although this item was published in 1990, no other questionnaire in the reviews by Fluit (2010) or Beckman et al. (2004), nor questionnaires developed subsequent to that time have included such an item. However, it has been identified as a characteristic of effective bedside clinical educators in an Omani medical school (Alweshahi et al., 2007). The lack of inclusion of a health record keeping item may stem from the practice not being explicitly described in any of the published clinical teaching models or frameworks (Kilminster & Jolly, 2001). Further, in the current thesis, chapters 2 and 3 describe both the inclusion and removal of this item by Hewson and Jensen (1990) from the OCTQ.

Given that electronic medical records (EMRs) are the dominant method by which clinical information is recorded in osteopathy (Adams et al., 2018) and other health professions, in addition to the legal requirements for quality, contemporaneous record keeping across the health professions, it would be prudent to include reference to EMRs within clinical teaching models (Wald et al., 2014; Warboys et al., 2014). From an evaluation standpoint, inclusion of EMR practices may be better placed within a more global learning environment measure - given that both the individual educator (Heiman et al., 2014) and the system influence record keeping practices (Daugherty et al., 1998; Solarte & Könings, 2017). Supporting this environmental influence assertion is recent work on implementation of EMRs, identifying its significant negative influence on time spent on, and enthusiasm for, clinical teaching (Spencer et al., 2012), in addition to decreased appreciation of care continuity and increased cognitive load (Varpio et al., 2015). Work by Schultz et al. (2004) also identified 'teaching of medical record keeping skill' as a characteristic of the learning site, rather than one of the individual educator.

Although the aforementioned studies relate to the implementation of EMRs, these works highlight potential challenges with role modelling EMR practices in the clinical learning environment. As Ellaway et al. (2013) describe "their [clinical educator's] teaching might be expected to be more about accommodating EHRs [electronic health records] in clinical practice rather than embracing them" (p. 283) from an educational perspective.

Summary

The aforementioned aspects that are not explored in the OCTQ have the potential to be included in other measures that form part of the quality assurance framework, namely a measure of the learning environment. Identification, modification or development of a clinical learning environment measure suitable for use in the student-led context, inclusive of the aspects of practice described above, would be valuable future work.

Quality assurance in clinical education

"The validity of the system as a whole needs to be evaluated in terms of its effects on improving instruction and student learning, and therefore a fundamental aspect of validation is to examine whether these benefits are an outcome of the use of the system" (Lane, 2014, p. 129).

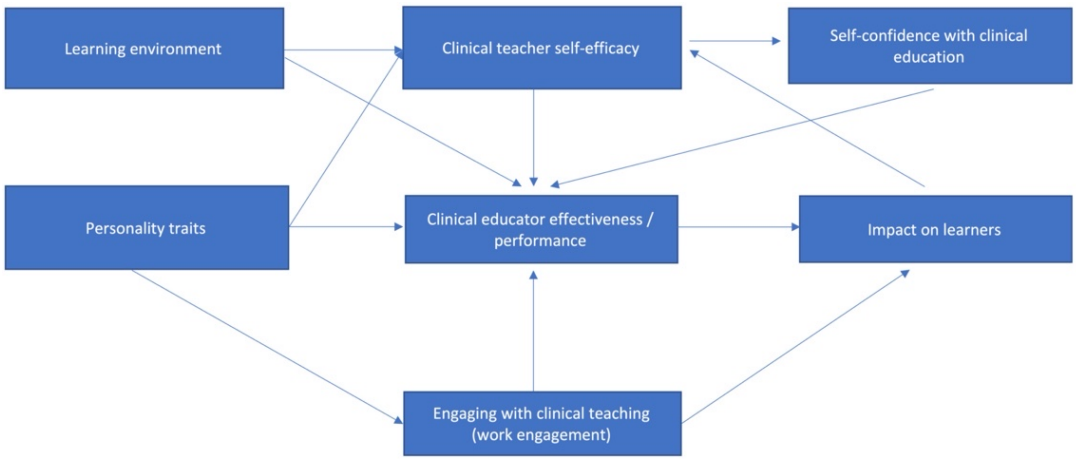
The overarching purpose of a quality assurance program or system in the clinical education context is to maintain, develop, monitor and improve the stakeholder experience, that is, the experience of the clinical educator, learner, patient, faculty and administrative staff (Barrow & McKimm, 2010; Cross, 1995; Ragsdale et al., 2020; Snell et al., 2000). Barrow and McKimm (2010) also include external agencies such as accreditors as another stakeholder in the quality assurance system, and this is pertinent to osteopathy clinical education particularly in Australia, New Zealand and the United Kingdom, where the profession is regulated by government. Descriptions of whole of system approaches to assurance of clinical education quality in the literature are scarce (Stachura et al., 2000) and often limited to commentaries (Barrow & McKimm, 2010). This scarcity may be associated with the challenge of defining *quality* in the context of clinical education (McCallum et al., 2013; Stachura et al., 2000) and differences in the delivery of clinical education across sites or institutions (Jette et al., 2014; McCallum et al., 2013) and countries.

However, where quality assurance 'systems' are described, the focus is typically on the individual clinical educator and the clinical placement site or clinical learning environment. In chapter 5, it was identified that a number of issues have been raised with the use of student-evaluations of teaching, particularly where the data derived from them are used in isolation of other measures of quality and performance (Boring et al., 2016; Garger et al., 2019; Hornstein, 2017; Spooren et al., 2013; Stark & Freishtat, 2014). The purpose of this section is to highlight the role of the OCTQ within a broader quality assurance system for clinical education in osteopathy.

A discussion of individual elements of the clinical education quality assurance system (i.e. learning environment, system influences, patient experience and satisfaction) is

beyond the remit of this thesis. Figure 1 draws together the findings of this thesis detailed in chapters 2 to 5 and the wider literature on evaluation of clinical teaching, to summarise the significant range of potential influences on outcomes related to teaching quality that should be considered within a clinical education analysis. This figure is an extension of the model presented by McArthur (2016), combined with the work of Scheepers et al. (2016). It is important to note that the figure is not a statistical path analysis, but a demonstration of the potential complexity of, and influences on, evaluations of clinical teaching quality.

Figure 1. Potential influences on clinical educator performance, effectiveness and teaching quality.



A quality assurance mechanism should allow individual stakeholders to input data into the system about the elements on which they are able to evaluate and comment. For example, a clinical educator could input data on their own performance, their clinical education peers and contribute to learner assessments (as surrogate measure of teaching and learning outcomes). The patient voice is also invaluable in any quality assurance system not only because of their central role as the receiver of care and participant in the clinical education process, but ultimately the goal of clinical education is to improve patient outcomes (Ragsdale et al., 2020). However, this voice is not common in the broader clinical education literature (Ford et al., 2016). Here the patient can provide data on satisfaction with their care. An individual clinical

educator's peers, other health professionals and administrators can contribute through multisource (360º) evaluations of clinical educators (Emke et al., 2017). Students and educators can also input data on the clinical learning environment, and the former, input data on individual clinical educators through tools such as the OCTQ. It is the clinical learning environment and individual clinical educator measures that will be described below.

Clinical learning environment

The clinical learning environment has been reported to contribute to students' perceived effectiveness of the teaching provided by their individual clinical educator (Bruijn et al., 2006; Lombarts et al., 2014; Phillips et al., 2017). For example, Brown et al. (2013) report that clinical teaching may only contribute up to 40% of the variance in education environment measures, suggesting there are other factors, beyond teaching, that influence the clinical learning environment. Faculty and administrators should include an education environment measure (e.g. Rusticus et al., 2019; Strand et al., 2013) within their clinical education quality assurance system to ensure these additional factors are explored. The OCTQ could play a role alongside a clinical learning environment measure to ensure that a range of factors contributing to quality within the clinical education system are captured. There is a small volume of literature that has explored the broader education environment in osteopathy (Luciani et al., 2018; Luciani et al., 2015; Vaughan et al., 2014a). However, there is no literature that addresses the clinical learning environment specifically. This presents an opportunity for additional research into the osteopathy student-led clinical learning environment, particularly when combined with the OCTQ as the measure of clinical teaching quality.

Individual clinical educator

Literature describing quality assurance systems in clinical education focuses mainly on the individual clinical educator (Recker-Hughes et al., 2014). Throughout the previous five chapters in this thesis, a consistent theme is the use of student evaluations of clinical teaching quality as a way to identify opportunities for faculty

or professional development. The outcomes of the reliability investigations in chapter 4 suggest that eight OCTQ evaluations on a single clinical educator can provide a reliable indication about the quality of teaching provided by that educator in the osteopathy student-led clinical learning environment. This data could then be used to inform the professional development of the individual clinical educator, or the professional development of the clinical educator cohort (Notzer & Abramovitz, 2008). Albanese (1999) contends that whilst global measures of clinical educator performance (i.e. the global rating items on the OCTQ) may be useful for quality assurance, they are less useful as a guide for required professional development. Choice of areas of clinical education quality that could be improved or reinforced through professional development could be determined at either the individual OCTQ item level and/or through analysis of the qualitative comments provided by learners.

When professional development is used to improve clinical education quality, learner outcomes may be negatively influenced due to the clinical educators' behaviour change (Breckwoldt et al., 2014). It is also reasonable to expect that clinical educator evaluations by students may also become more negative through this transition phase. Therefore, a longitudinal approach to evaluation of individual clinical educator teaching quality as part of the system is required, along with triangulation of student evaluations with other data sources such as educator self and peer-evaluation (Ching, 2018; Stark & Freishtat, 2014; Sulis et al., 2019). It is also important to ensure that a sufficient volume of evaluations are collected (particularly from learners), in order to observe changes in teaching quality, and to make reliable, informed recommendations about areas for improvement in the future.

Recognition of quality clinical teaching

Whilst a quality assurance system in clinical education should identify aspects of the system that require improvement, there is also a need to recognise where aspects are performing well. It may be that individual clinical educator performance should be recognised as part of this system, and this recognition may also increase acceptability of clinical teaching evaluations by clinical educators (Snell et al., 2000). As described in chapter 4, the OCTQ presents sufficient reliability for formative decision making

(i.e. professional development); and chapter 3 suggests that OCTQ total scores can be used to reliably identify those educators who are performing well. Together, these outcomes provide support for the OCTQ scores being used to recognise a clinical educators' teaching quality over the previous teaching period, or over a sustained period (Blue et al., 1999; Snell et al., 2000). Such recognition has been shown to be a significant motivator for clinical educators, particularly where learners nominate the educator (Wheeler & Gill, 2010). What this recognition comprises (e.g. letter, award, gift) would be for the individual institution to decide (Woolliscroft et al., 2002). However, it is not recommended that OCTQ scores be the sole determinant of promotion or academic tenure processes.

OCTQ scores may also feed into wider osteopathy profession recognition of quality clinical teaching. Recker-Hughes et al. (2014) have advocated for the recognition of expert clinical educators in United States physical therapy, and there is the potential for a similar model of recognition to be explored in the osteopathy profession world-wide. We know little about clinical education in osteopathy, and we know less about who the osteopathy clinical educator is – this picture is beginning to emerge through the data in chapter 5 and that described by Vaughan et al. (2020). This lack of visibility to both the academic and professional communities could be addressed through more formal recognition at the professional level (i.e. advanced practice/credentialing) and additional research (Recker-Hughes et al., 2014). The OCTQ scores, and qualitative comments provided by learners, could be used by individual clinical educators as evidence towards such recognition.

Future directions

The construction, testing and review of the OCTQ described in chapters 2 through to 5 has led to a number of possibilities for future work, in addition to those included in Figure 1 and those described throughout the previous sections of this chapter. This thesis did not seek to gather evidence for the validity of the score interpretation in multiple contexts but sought to provide initial evidence allowing others to ascertain whether the questionnaire is suitable for their learning environment and educator evaluation needs. Further, this thesis has outlined how other researchers, academics

and administrators may be able to gather validity evidence to support score interpretations in their own context.

A significant shift in the osteopathy profession

A significant shift in the Australian osteopathy profession has occurred since the OCTQ has been developed and tested. There are three key areas where this shift has occurred, as follows:

- There is a significant number of learner's training to be an osteopath (n=1456) across three institutions – equivalent to 60% of the practicing profession in the year 2017/18 (Osteopathy Board of Australia, 2018).
- The models and duration of courses leading to registration as an osteopath in Australia have changed.[2] For example, Southern Cross University (Lismore, Australia) now offers a 4-year program leading to registration and Victoria University (Melbourne, Australia) has instituted a single unit completed sequentially approach to their program. These changes will likely influence clinical education in pre-registration training programs in Australia, potentially with reductions in hours devoted to education in a clinical learning environment, or exploration of other educational strategies to supplement or in part replace aspects of learners' clinical education (Fitzgerald et al., 2017; Moore & Field, 2017).
- The Osteopathy Board of Australia, the registration body for Australian osteopaths, published an updated version of the Capabilities for Osteopathic Practice (Osteopathy Board of Australia, 2019) drawing on the Canadian Medical Educational Directives for Specialists (CanMEDS) framework (Frank & Danoff, 2007). The Capabilities for Osteopathic Practice (2019) are the practice standards for the Australian profession.

[2] There have been no changes to the structure of the RMIT University osteopathy course. The course is offered as a 5-year, double Bachelor qualification.

Mapping to a new practice framework

This change in the basis of the Australian osteopathy practice standards to a CanMEDS-style framework provides the first opportunity for research emanating from the thesis. Throughout the development of the OCTQ, there has been an emphasis on the items measuring constructs consistent with those in the Cognitive Apprenticeship Model (Collins, 1991). Further, the items also appear to map to the interpersonal and clinical teaching domains described by Beckman et al. (2004). However, there is an opportunity to draw on the work of Prideaux et al. (2000) who described how clinical education could be incorporated within each CanMEDS role.

With the Australian osteopathy profession moving to a CanMEDS framework, there is an opportunity to evaluate the OCTQ within this framework and potentially reconceptualise it, or to concurrently evaluate the OCTQ with those clinical teaching evaluations previously developed using this framework. There are a number of examples of such questionnaires where the CanMEDS is either incorporated into the questionnaire (Fluit et al., 2012), or was used to frame their development (Nation et al., 2011). Table 3 presents an initial proposal for how the OCTQ items may be mapped to the 'roles' within the Capabilities for Osteopathic Practice (Osteopathy Board of Australia, 2019).

Table 3. Mapping of the Osteopathy Clinical Teaching Questionnaire (OCTQ) items to the Australian Capabilities for Osteopathic Practice roles.

Role	**Osteopathy Clinical Teaching Questionnaire (OCTQ) item**
1. Osteopath	12. Demonstrated osteopathic, clinical examination and rehabilitation knowledge and skill(s)
2. Professional and ethical practitioner	1. Maintained a positive attitude towards me

	2. Demonstrated humanistic attitudes in relating to patients (integrity, compassion and respect) 3. Showed genuine concern for my professional well-being
3. Communicator	4. Has good communication skills 5. Is open to student questions and alternative approaches to patient management
4. Critical reflective practitioner and lifelong learner	7. Promoted reflection on clinical practice 10. Stimulates me to learn independently
5. Educator and health promoter	6. Adjusted teaching to my needs (experience, competence, interest) 9. Asked questions to enhance my learning
6. Collaborative practitioner	8. Emphasises a problem-solving approach rather than solutions
7. Leader and manager	11. Offered me suggestions for improvement when required

Future research possibilities

This section presents a discussion of potential research and evaluation possibilities based on the work described in this thesis and drawing on pertinent literature from health professions education. The possibilities described below address aspects

including the development of learners as health professionals, the relationship between learner evaluations of clinical teaching quality and assessment outcomes, the role personality may play in quality clinical teaching, how clinical educators engage in their work in the student-led clinical learning environment, and how professional development and learner attributes influence clinical teaching evaluations.

Impact on learners

One goal of clinical teaching is to improve learner outcomes (Snell et al., 2000). The impact of clinical teaching on learners can be evaluated through their satisfaction with the clinical teaching they receive (Snell et al., 2000) and the outcomes of assessments, particularly those at the *Shows How* and *Does* levels of Miller's competency pyramid (Miller, 1990). There may also be an opportunity to explore the impact of clinical teaching on the development of learners at the *Is* level of Miller's pyramid as proposed by Cruess et al. (2016). These authors proposed the *Is* level to relate to professional identity formation. The influence of clinical educator role-modelling on learner identity formation could be explored using the OCTQ and qualitative approaches.

The relationship between clinical teaching quality and assessment of learner clinical competence presents another opportunity for research. Although there are likely to be a range of influences on learning outcomes (Huang et al., 2019), it is possible that the quality of clinical teaching received by a learner may be one of these. The workplace-based assessment practices of other health professions have been described, particularly the use of standardised assessment tools (Dalton et al., 2011; Dalton et al., 2012; McAllister et al., 2010; Rodger et al., 2016). However, a significant challenge in investigating the relationship between learning outcome and clinical teaching evaluations in osteopathy is the lack of research on the assessment practices of the profession. The first stage in exploring this relationship will need to be a more thorough understanding of workplace-based assessment in the profession (Moore & Vaughan, 2015), and potentially the development and/or implementation of assessment tools that enable valid and reliable measurement of learning outcomes.

There is also an opportunity to evaluate role-modelling specifically and explore its relationship with the OCTQ scores. One example of a tool to evaluate role-modelling in the clinical teaching context is that described by Jochemsen-van der Leeuw et al. (2014), with additional validity evidence described by Said et al. (2019). Professionalism of the clinical educator has also been described in the health professions education literature and presents another research opportunity. Young et al. (2014) have developed a tool that allows learners to evaluate the professionalism of their clinical educators. Exploring the relationship between OCTQ scores and tools to evaluate role-modelling and professionalism will provide additional validity evidence for the former, and potentially guide professional development.

Clinical educator personality

It may be that personality contributes to clinical teaching effectiveness, both self- and learner-evaluated. Work by Scheepers et al. (2014) with attending physicians suggests that personality interactions between learner and educator play a role in students' clinical teaching effectiveness evaluations. Big Five personality traits (extraversion, agreeableness, openness, conscientiousness, neuroticism) may "provide one of the best foundations for a relatively coherent framework for a more comprehensive psychological theory of teacher personality" (Göncz, 2017, p. 76). With respect to the Big Five in the clinical education context, Scheepers et al. (2014) identified extraversion as being associated with learner-reported overall clinical teaching effectiveness, and was more likely to be displayed by non-surgical versus surgical physicians. This work suggests that personality factors in clinical education may be context specific (Scheepers et al., 2014). Clinical education may attract those health professionals who demonstrate extraversion traits – engaging in think-aloud and active learning strategies as part of developing a positive learning environment.

Scheepers et al. (2016) also suggest that professional development could focus on the personality traits of effective clinical educators. However, this may be more challenging given the stability of these traits. Could these traits play a role in our recruitment of effective clinical educators? Or could it be related to the personality of the learner and the compatibility of the educator's personality with the learner? These

two questions provide interesting prospective study ideas and one where the OCTQ could be used to monitor clinical teaching quality. Further consideration could also be given to other intra- and interpersonal factors that may influence teaching evaluations including leadership (Rieck et al., 2015), emotional intelligence (Cruz, 2011), conflict management skills (Jones, 2001) and a willingness to trust learners (Hauer et al., 2014).

Work engagement

The concept of work engagement has received some attention in health professions education (Van den Berg et al., 2017; Van den Berg et al., 2018) and has also been described in the context of clinical teaching (Scheepers et al., 2016). Being 'engaged' in one's work has been associated with personal well-being and improved work performance (Van den Berg et al., 2017), and work engagement may also be positively associated with quality clinical teaching (Scheepers et al., 2016; Van den Berg et al., 2017). In the work of Scheepers et al. (2014) on clinical educator personality highlighted previously, extraverted educators are also reported to be more engaged in their teaching activities, with a positive flow-on effect to their student-reported teaching effectiveness. It is here that the OCTQ may be useful as a measurement tool to measure teaching quality in concert with measures of work engagement.

Providing strategies to develop an individual's work engagement (Scheepers et al., 2016) may be a suitable initial faculty development strategy, as this will likely lead to improvements in the clinical learning environment and subsequent (potential) improvement in clinical teaching quality. There is no literature on work engagement of clinical educators in osteopathy, and limited work in health professions education more broadly (i.e. Scheepers et al., 2016) and this presents a possible future research opportunity. Such work could be valuable as osteopathy clinical educators are only 'engaged' in clinical education in the student-led clinical learning environment, rather than balancing administrative, academic, and clinical roles as their hospital-based colleagues are required to do.

Impact of professional development

A rationale for the work described in this thesis was the need to be able to measure clinical teaching quality in the osteopathy student-led clinical environment in order to maintain and improve the learner experience. Measuring the students' perception of their learning experiences at regular intervals provides feedback to all stakeholders, in particular the clinical educators, whose effectiveness as educators is reviewed by the process. In this way, the OCTQ can be used as a measure of change in teaching quality. Chapter 3 describes the measurement properties of the OCTQ that would allow for the identification of changes in clinical teaching quality.

We know from studies of other health professions that clinical teaching quality may change following faculty development (Steinert et al., 2016) or changes in education practices enacted by clinical educators following feedback (Stalmeijer et al., 2010b). There is no literature describing the professional development of osteopathy clinical educators, and as identified in chapter 5, very few clinical educators have undertaken formal studies in clinical education. Even fewer have undertaken studies in teaching and learning.

Without these formal studies, and anecdotally, very little professional development about clinical education, there is an opportunity to explore how osteopathy clinical educators conceive learning and teaching. Jacobs et al. (2012) developed the Conceptions of Learning and Teaching questionnaire and following that identified five teacher profiles that could be used to inform professional development activities (Jacobs et al., 2014). Subsequent work has also identified a sixth potential teacher profile (Jacobs et al., 2020). Additional research within the osteopathy profession could be directed towards understanding if the same five or six teacher profiles are consistent in the osteopathy clinical education context and explore their relationship to OCTQ scores. These profiles could be used to inform professional development activities and may identify teacher profiles that are best suited to the clinical education milieu.

Further, the OCTQ scores at both individual item and overall score levels could be used to evaluate the impact of both targeted professional development for the individual clinical educator, and/or to for professional development of the clinical educator group. Chapter 5 proposed the use of professional development focused on fostering self-efficacy, a position supported by Dybowski et al. (2017). In chapter 5, differences in self-efficacy were noted in the clinical educator cohort where those with low self-efficacy with respect to their clinical teaching were often rated highly by students, and the converse for educators with high self-efficacy. Professional development guided by self-efficacy is widely described in the K-12 education literature (Althauser, 2015; Bray-Clark & Bates, 2003; Klassen & Tze, 2014). Together, the OCTQ and Self Efficacy of Clinical Teachers tool (McArthur, 2016) scores could be used to evaluate the effect of professional development activities, providing an opportunity for future research.

Learner attributes and perspectives, and their influences on clinical teaching evaluation scores

The OCTQ has been designed to ensure that factors such as age, gender and country of training do not have a systematic influence on the scores provided on the questionnaire. How these factors are accounted for is described in chapter 3 and, further as part of the *scoring* inference in the current chapter. However, there may be additional influences brought to bear by the learner when completing clinical teaching quality evaluations (Dybowski et al., 2017). Learner academic motivation (Dybowski et al., 2017) and approaches to their own learning (Zoghi et al., 2010) may be related to the scores that a learner provides on the OCTQ. Learner academic motivation could be evaluated through measures such as the Motivated Strategies for Learning Questionnaire (Pintrich et al., 1993), whilst learning approach could be explored through the Study Process Questionnaire (Biggs et al., 2001) or similar – the latter having some evidence for its utility in an osteopathy student population (Vaughan, 2016, 2018). Consideration should also be given to the context in which the clinical learning occurs. For example, there may be differences in learner experiences between the student-led and private practice placement environments – although the latter is seldom reported in the osteopathy literature (Moore & Field, 2017). Studies

that explore the individual learner and environmental influences could provide additional insight into factors that may impact learner performance and clinical teaching evaluations.

Conclusion

This thesis has developed a measure of clinical teaching quality for the osteopathy student-led clinical learning environment and presented initial validity evidence supporting its use. The development of the measure involved reviewing the seminal literature; engaging with relevant stakeholders; examining different stages of learning; testing at multiple domestic and overseas institutions; and utilised, contemporary, evidence-based statistical approaches. This process has provided evidence supporting the use of the OCTQ in the osteopathy clinical learning environment, and potentially adapted to student-led clinical learning environments in other health professions. The thesis also provides the first example of how the validity framework described by Kane (1992) can be used beyond the educational assessment context, consistent with LeBaron Wallace (2011).

Grillo et al. (2016) surmise that "student evaluations of teaching (SETs) should be completed by all students for all courses, be short, provide opportunities for open-ended comments, and be administered in class to improve response rate" (p. 439). The OCTQ has been designed to meet these recommendations within the constraints of the student-led clinical learning environment. The OCTQ, like other student evaluation of teaching measures, should not be used in isolation for decision-making but be combined with other measures from across the learning environment. However, it is advocated for use in a formative sense - to inform clinical educator induction and preparatory activities, guide professional development, and identify (and possibly reward) those clinical educators who learners have identified as being excellent clinical educators. It is anticipated that the OCTQ will be used as one tool to improve the quality of clinical teaching in osteopathy student-led clinics, and to assist in improving learner and patient outcomes.

References

Adams, J., Sibbritt, D., Steel, A., & Peng, W. (2018). A workforce survey of Australian osteopathy: analysis of a nationally-representative sample of osteopaths from the Osteopathy Research and Innovation Network (ORION) project. *BMC Health Services Research, 18*(1), 352. https://doi.org/10.1186/s12913-018-3158-y

Afonso, N. M., Cardozo, L. J., Mascarenhas, O. A., Aranha, A. N., & Shah, C. (2005). Are anonymous evaluations a better assessment of faculty teaching performance? A comparative analysis of open and anonymous evaluation processes. *Family Medicine, 37*(1), 43-47. https://www.stfm.org/familymedicine/vol37issue1/Afonso43

Al-Kadri, H. M., Al-Kadi, M. T., & Van Der Vleuten, C. P. M. (2013). Workplace-based assessment and students' approaches to learning: a qualitative inquiry. *Medical Teacher, 35*(Supplement 1), S31-S38. https://doi.org/10.3109/0142159X.2013.765547

Albanese, M. (1999). Rating educational quality: factors in the erosion of professional standards. *Academic Medicine, 74*(6), 652-658. https://doi.org/10.1097/00001888-199906000-00009

AlHaqwi, A. I., van der Molen, H. T., Schmidt, H., & Magzoub, M. (2010). Determinants of effective clinical learning: a student and teacher perspective in Saudi Arabia. *Education for Health, 23*(2), 369. https://www.educationforhealth.net/text.asp?2010/23/2/369/101490

Althauser, K. (2015). Job-embedded professional development: Its impact on teacher self-efficacy and student performance. *Teacher Development, 19*(2), 210-225. https://doi.org/10.1080/13664530.2015.1011346

Alweshahi, Y., Harley, D., & Cook, D. A. (2007). Students' perception of the characteristics of effective bedside teachers. *Medical Teacher, 29*(2-3), 204-209. https://doi.org/10.1080/01421590701271818

Andrich, D., & Hagquist, C. (2012). Real and artificial differential item functioning. *Journal of Educational and Behavioral Statistics, 37*(3), 387-416. https://doi.org/10.3102/1076998611411913

Arah, O. A., Heineman, M. J., & Lombarts, K. M. (2012). Factors influencing residents' evaluations of clinical faculty member teaching qualities and role model status. *Medical Education, 46*(4), 381-389. https://doi.org/10.1111/j.1365-2923.2011.04176.x

Arah, O. A., Hoekstra, J. B. L., Bos, A. P., & Lombarts, K. M. (2011). New tools for systematic evaluation of teaching qualities of medical faculty: results of an ongoing multi-center survey. *PLoS ONE, 6*(10), e25983. https://doi.org/10.1371/journal.pone.0025983

Atwater, A. R., Rudd, M., Brown, A., Wiener, J. S., Benjamin, R., Lee, W. R., & Rosdahl, J. A. (2016). Developing teaching strategies in the EHR era: a survey of GME experts. *Journal of Graduate Medical Education, 8*(4), 581-586. https://doi.org/10.4300/JGME-D-15-00788.1

Barrow, M., & McKimm, J. (2010). Assuring and maintaining quality in clinical education. *British Journal of Hospital Medicine (2005), 71*(4), 224-228. https://doi.org/10.12968/hmed.2010.71.4.47518

Batbaatar, E., Dorjdagva, J., Luvsannyam, A., & Amenta, P. (2015). Conceptualisation of patient satisfaction: a systematic narrative literature review. *Perspectives in Public Health, 135*(5), 243-250. https://doi.org/10.1177/1757913915594196

Batbaatar, E., Dorjdagva, J., Luvsannyam, A., Savino, M. M., & Amenta, P. (2017). Determinants of patient satisfaction: a systematic review. *Perspectives in Public Health, 137*(2), 89-101. https://doi.org/10.1177/1757913916634136

Beach, R. A. (2017). Strategies to maximise teaching in your next ambulatory clinic. *The Clinical Teacher, 14*(2), 85-89. https://doi.org/10.1111/tct.12638

Beckman, T. J., Ghosh, A. K., Cook, D. A., Erwin, P. J., & Mandrekar, J. N. (2004). How reliable are assessments of clinical teaching? *Journal of General Internal Medicine, 19*(9), 971-977. https://doi.org/10.1111/j.1525-1497.2004.40066.x

Beckman, T. J., & Lee, M. C. (2009). Proposal for a collaborative approach to clinical teaching. *Mayo Clinic Proceedings, 84*(4), 339-344. https://doi.org/10.1016/S0025-6196(11)60543-0

Beckman, T. J., Lee, M. C., Rohren, C. H., & Pankratz, V. S. (2003). Evaluating an instrument for the peer review of inpatient teaching. *Medical Teacher, 25*(2), 131-135. https://doi.org/10.1080/0142159031000092508

Beckman, T. J., Reed, D. A., Shanafelt, T. D., & West, C. P. (2010). Impact of resident well-being and empathy on assessments of faculty physicians. *Journal of General Internal Medicine, 25*(1), 52-56. https://doi.org/10.1007/s11606-009-1152-0

Bengtsson, M., & Ohlsson, B. (2010). The nursing and medical students motivation to attain knowledge. *Nurse Education Today, 30*(2), 150-156. https://doi.org/10.1016/j.nedt.2009.07.005

Biggs, J., Kember, D., & Leung, D. Y. P. (2001). The revised two-factor study process questionnaire: R-SPQ-2F. *British Journal of Educational Psychology, 71*(1), 133-149. https://doi.org/10.1348/000709901158433

Billett, S. (2009). Realising the educational worth of integrating work experiences in higher education. *Studies in Higher Education, 34*(7), 827-843. https://doi.org/10.1080/03075070802706561

Billett, S., & Sweet, L. (2018). Understanding and appraising medical students' learning through clinical experiences. In S. Loo (Ed.), *Multiple Dimensions of Teaching and Learning for Occupational Practice* (pp. 200). Routledge.

Billett, S., Sweet, L., & Glover, P. (2013). The curriculum and pedagogic properties of practice-based experiences: the case of midwifery students. *Vocations and Learning, 6*(2), 237-257. https://doi.org/10.1007/s12186-012-9094-9

Bing-You, R. G., Paterson, J., & Levine, M. A. (1997). Feedback falling on deaf ears: residents' receptivity to feedback tempered by sender credibility. *Medical Teacher, 19*(1), 40-44. https://doi.org/10.3109/01421599709019346

Blakey, A. G., Smith-Han, K., Anderson, L., Collins, E., Berryman, E., & Wilkinson, T. (2019). "They cared about us students:" learning from exemplar clinical teaching environments. *BMC Medical Education, 19*(1), 119. https://doi.org/10.1186/s12909-019-1551-9

Bloch, R., & Norman, G. (2012). Generalizability theory for the perplexed: A practical introduction and guide: AMEE Guide No. 68. *Medical Teacher, 34*(11), 960-992. https://doi.org/10.3109/0142159X.2012.703791

Blue, A. V., Griffith III, C. H., Wilson, J., Sloan, D. A., & Schwartz, R. W. (1999). Surgical teaching quality makes a difference. *The American Journal of Surgery, 177*(1), 86-89. https://doi.org/10.1080/02602938.2013.875984

Boring, A. (2017). Gender biases in student evaluations of teaching. *Journal of Public Economics, 145*, 27-41. https://doi.org/10.1016/j.jpubeco.2016.11.006

Boring, A., Ottoboni, K., & Stark, P. (2016). Student evaluations of teaching (mostly) do not measure teaching effectiveness. *ScienceOpen Research*. https://doi.org/10.14293/S2199-1006.1.SOR-EDU.AETBZC.v1

Boud, D., & Molloy, E. (2012). What is the problem with feedback? In D. Boud & E. Molloy (Eds.), *Feedback in higher and professional education* (pp. 11-20). Routledge.

Branch Jr, W. T., Kern, D., Haidet, P., Weissmann, P., Gracey, C. F., Mitchell, G., & Inui, T. (2001). Teaching the human dimensions of care in clinical settings. *JAMA, 286*(9), 1067-1074. https://doi.org/10.1001/jama.286.9.1067

Branch Jr, W. T., & Paranjape, A. (2002). Feedback and reflection: teaching methods for clinical settings. *Academic Medicine, 77*(12 Part 1), 1185-1188. https://doi.org/10.1097/00001888-200212000-00005

Bray-Clark, N., & Bates, R. (2003). Self-efficacy beliefs and teacher effectiveness: Implications for professional development. *Professional Educator, 26*(1), 13-22.

Breckwoldt, J., Svensson, J., Lingemann, C., & Gruber, H. (2014). Does clinical teacher training always improve teaching effectiveness as opposed to no teacher training? A randomized controlled study. *BMC Medical Education, 14*(1), 6. https://doi.org/10.1186/1472-6920-14-6

Brennan, R. L. (2010). Generalizability Theory and Classical Test Theory. *Applied Measurement in Education, 24*(1), 1-21. https://doi.org/10.1080/08957347.2011.532417

Brown, J., Bearman, M., Kirby, C., Molloy, E., Colville, D., & Nestel, D. (2019). Theory, a lost character? As presented in general practice education research papers. *Medical Education, 53*(5), 443-457. https://doi.org/doi:10.1111/medu.13793

Brown, T., Williams, B., & Lynch, M. (2013). Relationship between clinical fieldwork educator performance and health professional students' perceptions of their practice education learning environments. *Nursing and Health Sciences, 15*(4), 510-517. https://doi.org/10.1111/nhs.12065

Bruijn, M., Busari, J. O., & Wolf, B. H. M. (2006). Quality of clinical supervision as perceived by specialist registrars in a university and district teaching hospital. *Medical Education, 40*(10), 1002-1008. https://doi.org/10.1111/j.1365-2929.2006.02559.x

Buchel, T. L., & Edwards, F. D. (2005). Characteristics of effective clinical teachers. *Family Medicine, 37*(1), 30-35. https://www.stfm.org/familymedicine/vol37issue1/Buchel30

Budd, S., Kelley, R., Day, R., Variend, H., & Dogra, N. (2011). Student attitudes to psychiatry and their clinical placements. *Medical Teacher, 33*(11), e586-e592. https://doi.org/10.3109/0142159X.2011.610836

Buery-Joyner, S. D., Ryan, M. S., Santen, S. A., Borda, A., Webb, T., & Cheifetz, C. (2019). Beyond mistreatment: Learner neglect in the clinical teaching environment. *Medical Teacher, 41*(8), 949-955. https://doi.org/10.1080/0142159X.2019.1602254

Burgess, A., Goulston, K., & Oates, K. (2015). Role modelling of clinical tutors: a focus group study among medical students. *BMC Medical Education, 15*(1), 17. https://doi.org/10.1186/s12909-015-0303-8

Burgess, A., & Mellis, C. (2015). Feedback and assessment for clinical placements: achieving the right balance. *Advances in Medical Education and Practice, 6*, 373. https://doi.org/10.2147/AMEP.S77890

Burns, C., Beauchesne, M., Ryan-Krause, P., & Sawin, K. (2006). Mastering the preceptor role: challenges of clinical teaching. *Journal of Pediatric Health Care, 20*(3), 172-183. https://doi.org/10.1016/j.pedhc.2005.10.012

Busari, J. O., Scherpbier, A. J. J. A., Van der Vleuten, C. P. M., & Essed, G. G. M. (2006). A two-day teacher-training programme for medical residents: Investigating the impact on teaching ability. *Advances in Health Sciences Education, 11*(2), 133-144. https://doi.org/10.1007/s10459-005-8303-9

Busari, J. O., Weggelaar, N. M., Knottnerus, A. C., Greidanus, P. M., & Scherpbier, A. J. J. A. (2005). How medical residents perceive the quality of supervision provided by attending doctors in the clinical setting. *Medical Education, 39*(7), 696-703. https://doi.org/10.1111/j.1365-2929.2005.02190.x

Cantillon, P., & Sargeant, J. (2008). Giving feedback in clinical settings. *British Medical Journal, 337*, a1961. https://doi.org/10.1136/bmj.a1961

Cashin, W. E., & Downey, R. G. (1992). Using global student rating items for summative evaluation. *Journal of Educational Psychology, 84*(4), 563. https://doi.org/10.1037/0022-0663.84.4.563

Chen, H. C., O'Sullivan, P., Teherani, A., Fogh, S., Kobashi, B., & Ten Cate, O. (2015). Sequencing learning experiences to engage different level learners in the workplace: An interview study with excellent clinical teachers. *Medical Teacher, 37*(12), 1090-1097. https://doi.org/10.3109/0142159X.2015.1009431

Cheung, F., & Lucas, R. E. (2014). Assessing the validity of single-item life satisfaction measures: Results from three large samples. *Quality of Life Research, 23*(10), 2809-2818. https://doi.org/10.1007/s11136-014-0726-4

Ching, G. (2018). A literature review on the student evaluation of teaching. *Higher Education Evaluation and Development, 12*(2), 63-84. https://doi.org/10.1108/HEED-04-2018-0009

Chyung, S. Y., Roberts, K., Swanson, I., & Hankinson, A. (2017). Evidence-based survey design: The use of a midpoint on the Likert scale. *Performance Improvement, 56*(10), 15-23. https://doi.org/10.1002/pfi.21727

Clauser, B. E., Margolis, M. J., Holtman, M. C., Katsufrakis, P. J., & Hawkins, R. E. (2012). Validity considerations in the assessment of professionalism. *Advances in Health Sciences Education, 17*(2), 165-181. https://doi.org/10.1007/s10459-010-9219-6

Collins, A. (1991). Cognitive Apprenticeship: Making Things Visible. *American Educator, 15*(3), 6-11, 38-46.

Cömert, M., Zill, J. M., Christalle, E., Dirmaier, J., Härter, M., & Scholl, I. (2016). Assessing communication skills of medical students in objective structured clinical examinations (OSCE)-a systematic review of rating scales. *PLoS ONE, 11*(3), e0152717. https://doi.org/10.1371/journal.pone.0152717

Consensus-based Standards for the Selection of Health Measurement Instruments. (2010). *COSMIN Taxonomy of Measurement Properties*. Retrieved August 20th, 2019, from https://www.cosmin.nl/tools/cosmin-taxonomy-measurement-properties/

Cook, D. A. (2014). When I say… validity. *Medical Education, 48*(10), 948-949. https://doi.org/10.1111/medu.12401

Cook, D. A., Brydges, R., Ginsburg, S., & Hatala, R. (2015). A contemporary approach to validity arguments: a practical guide to Kane's framework. *Medical Education, 49*(6), 560-575. https://doi.org/10.1111/medu.12678

Cox, S. S., & Swanson, M. S. (2002). Indentification of teaching excellence in operating room and clinic settings. *The American Journal of Surgery, 183*, 251-255. https://doi.org/10.1016/s0002-9610(02)00787-0

Cross, V. (1995). Perceptions of the ideal clinical educator in physiotherapy education. *Physiotherapy, 81*(9), 506-513. https://doi.org/10.1016/S0031-9406(05)66680-1

Cruess, R. L., Cruess, S. R., & Steinert, Y. (2016). Amending Miller's pyramid to include professional identity formation. *Academic Medicine, 91*(2), 180-185. https://doi.org/10.1097/ACM.0000000000000913

Cruess, S. R., Cruess, R. L., & Steinert, Y. (2008). Role modelling—making the most of a powerful teaching strategy. *British Medical Journal, 336*(7646), 718-721. https://doi.org/10.1136/bmj.39503.757847.BE

Cruz, S. S. S. M. S. (2011). Clinical supervision in nursing: effective pathway to quality. *Procedia-Social and Behavioral Sciences, 29*, 286-291. https://doi.org/10.1016/j.sbspro.2011.11.240

Cunnington, J. P. W., Neville, A. J., & Norman, G. R. (1996). The risks of thoroughness: reliability and validity of global ratings and checklists in an OSCE. *Advances in Health Sciences Education, 1*(3), 227-233. https://doi.org/10.1007/BF00162920

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. https://doi.org/10.1016/S1836-9553(11)70054-6

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. https://doi.org/10.1016/S1836-9553(12)70072-3

DaRosa, D. A., Dunnington, G. L., Stearns, J., Ferenchick, G., Bowen, J. L., & Simpson, D. E. (1997). Ambulatory teaching" lite": less clinic time, more

educationally fulfilling. *Academic Medicine, 72*(5), 358-361. https://doi.org/10.1097/00001888-199705000-00013

Daugherty, S. R., Baldwin Jr, D. C., & Rowley, B. D. (1998). Learning, satisfaction, and mistreatment during medical internship: a national survey of working conditions. *JAMA, 279*(15), 1194-1199. https://doi.org/10.1001/jama.279.15.1194

Delany, C., Kameniar, B., Lysk, J., & Vaughan, B. (2020). "Starting from a higher place": linking Habermas to teaching and learning clinical reasoning in the emergency medicine context. *Advances in Health Sciences Education, 25*, 809-824. https://doi.org/10.1007/s10459-020-09958-x

Denniston, C., Molloy, E. K., Ting, C. Y., Lin, Q. F., & Rees, C. E. (2019). Healthcare professionals' perceptions of learning communication in the healthcare workplace: an Australian interview study. *BMJ Open, 9*(2), e025445. https://doi.org/10.1136/bmjopen-2018-025445

des Ordons, A. L. R., de Groot, J. M., Rosenal, T., Viceer, N., & Nixon, L. (2018). How clinicians integrate humanism in their clinical workplace—'Just trying to put myself in their human being shoes'. *Perspectives on Medical Education, 7*(5), 318-324. https://doi.org/10.1007/s40037-018-0455-4

Dornan, T., Boshuizen, H., King, N., & Scherpbier, A. (2007). Experience-based learning: a model linking the processes and outcomes of medical students' workplace learning. *Medical Education, 41*(1), 84-91. https://doi.org/10.1111/j.1365-2929.2006.02652.x

Dunn, T. J., Baguley, T., & Brunsden, V. (2014). From alpha to omega: A practical solution to the pervasive problem of internal consistency estimation. *British Journal of Psychology, 105*(3), 399-412. http://onlinelibrary.wiley.com/doi/10.1111/bjop.12046/abstract

Dybowski, C., Sehner, S., & Harendza, S. (2017). Influence of motivation, self-efficacy and situational factors on the teaching quality of clinical educators. *BMC Medical Education, 17*(1), 84. https://doi.org/10.1186/s12909-017-0923-2

Edgecombe, K., & Bowden, M. (2009). The ongoing search for best practice in clinical teaching and learning: A model of nursing students' evolution to

proficient novice registered nurses. *Nurse Education in Practice, 9*(2), 91-101. https://doi.org/10.1097/00001888-199705000-00013

Ellaway, R. H., Graves, L., & Greene, P. S. (2013). Medical education in an electronic health record-mediated world. *Medical Teacher, 35*(4), 282-286. https://doi.org/10.3109/0142159X.2013.773396

Emke, A. R., Cheng, S., Chen, L., Tian, D., & Dufault, C. (2017). A novel approach to assessing professionalism in preclinical medical students using multisource feedback through paired self- and peer evaluations. *Teaching and Learning in Medicine, 29*(4), 402-410. https://doi.org/10.1080/10401334.2017.1306446

Eva, K. W., Armson, H., Holmboe, E., Lockyer, J., Loney, E., Mann, K., & Sargeant, J. (2012). Factors influencing responsiveness to feedback: on the interplay between fear, confidence, and reasoning processes. *Advances in Health Sciences Education, 17*(1), 15-26. https://doi.org/10.1007/s10459-011-9290-7

Eva, K. W., & Regehr, G. (2008). "I'll never play professional football" and other fallacies of self-assessment. *Journal of Continuing Education in the Health Professions, 28*(1), 14-19. https://doi.org/10.1002/chp.150

Farrell, S. E., Hopson, L. R., Wolff, M., Hemphill, R. R., & Santen, S. A. (2016). What's the evidence: a review of the one-minute preceptor model of clinical teaching and implications for teaching in the emergency department. *The Journal of Emergency Medicine, 51*(3), 278-283. https://doi.org/10.1016/j.jemermed.2016.05.007

Field, J. C., Zhang, E., Milke, V., McCance, S., Worts, C., Stone, S. J., & Edwards, D. C. (2019). Patterns of verbal interaction and student satisfaction within a clinical setting: A video-enhanced observational study. *European Journal of Dental Education, 23*(1), e45-e52. https://doi.org/10.1111/eje.12399

Finn, D. J., Taylor, K. H., & Balmer, M. C. (2016). Dental core trainee teaching of dental students in a clinical setting: a pilot study. *Oral Surgery, 9*(1), 25-30. https://doi.org/10.1111/ors.12165

Fitzgerald, K., Denning, T., & Vaughan, B. (2017). Simulated learning activities as part replacement of clinical placements in osteopathy: A case study. *International Journal of Osteopathic Medicine, 26*, 44-48. https://doi.org/https://doi.org/10.1016/j.ijosm.2017.10.003

Fluit, C. (2010). Assessing the quality of clinical teachers. *Journal of General Internal Medicine, 25*(12), 1337-1345. https://doi.org/10.1007/s11606-010-1458-y

Fluit, C., Bolhuis, S., Grol, R., Ham, M., Feskens, R., Laan, R., & Wensing, M. (2012). Evaluation and feedback for effective clinical teaching in postgraduate medical education: Validation of an assessment instrument incorporating the CanMEDS roles. *Medical Teacher, 34*(11), 893-901. http://informahealthcare.com/doi/abs/10.3109/0142159X.2012.699114

Forbes, J., Sierra, T., Trapp, J., & Parish, T. (2018). Characteristics of effective teaching in physician assistant programs from the students' perspective. *Internet Journal of Allied Health Sciences and Practice, 16*(3), 3. https://nsuworks.nova.edu/ijahsp/vol16/iss3/3/

Ford, K., Courtney-Pratt, H., Marlow, A., Cooper, J., Williams, D., & Mason, R. (2016). Quality clinical placements: The perspectives of undergraduate nursing students and their supervising nurses. *Nurse Education Today, 37*, 97-102. https://doi.org/10.1016/j.nedt.2015.11.013

Frank, J. R., & Danoff, D. (2007). The CanMEDS initiative: implementing an outcomes-based framework of physician competencies. *Medical Teacher, 29*(7), 642-647. https://doi.org/10.1080/01421590701746983

Frith, C., Cowan, S., & Delany, C. (2015). Reflection training as a form of professional development for physiotherapy clinical educators. *Focus on Health Professional Education: A Multi-Disciplinary Journal, 16*(2), 88. https://doi.org/10.11157/fohpe.v16i2.77

Garger, J., Jacques, P. H., Gastle, B. W., & Connolly, C. M. (2019). Threats of common method variance in student assessment of instruction instruments. *Higher Education Evaluation and Development, 13*(1), 2-17. https://doi.org/10.1108/HEED-05-2018-0012

Gat, I., Pessach-Gelblum, L., Givati, G., Haim, N., Paluch-Shimon, S., Unterman, A., Bar-Shavit, Y., Grabler, G., Sagi, D., Achiron, A., & Ziv, A. (2016). Innovative integrative bedside teaching model improves tutors' self-assessments of teaching skills and attitudes. *Medical Education Online, 21*(1), 30526. https://doi.org/10.3402/meo.v21.30526

Gibson, S. J., Porter, J., Anderson, A., Bryce, A., Dart, J., Kellow, N., Meiklejohn, S., Volders, E., Young, A., & Palermo, C. (2019). Clinical educators' skills and qualities in allied health: a systematic review. *Medical Education, 53*(5), 432-442. https://doi.org/doi:10.1111/medu.13782

Goldie, J., Dowie, A., Goldie, A., Cotton, P., & Morrison, J. (2015). What makes a good clinical student and teacher? An exploratory study. *BMC Medical Education, 15*(1), 40. https://doi.org/10.1186/s12909-015-0314-5

Göncz, L. (2017). Teacher personality: a review of psychological research and guidelines for a more comprehensive theory in educational psychology. *Open Review of Educational Research, 4*(1), 75-95. https://doi.org/10.1080/23265507.2017.1339572

Gracey, C. F., Haidet, P., Branch, W. T., Weissmann, P., Kern, D. E., Mitchell, G., Frankel, R., & Inui, T. (2005). Precepting humanism: strategies for fostering the human dimensions of care in ambulatory settings. *Academic Medicine, 80*(1), 21-28. https://doi.org/10.1097/00001888-200501000-00007

Green, S. B., & Yang, Y. (2009). Commentary on coefficient alpha: A cautionary tale. *Psychometrika, 74*(1), 121-135. https://doi.org/10.1007/s11336-008-9098-4

Grillo, A. C., Murdoch-Kinch, C. A., Ramaswamy, V., & Inglehart, M. R. (2016). Student evaluations of teaching: dental and dental hygiene students' and faculty members' perspectives. *Journal of Dental Education, 80*(4), 439-451. https://doi.org/10.1002/j.0022-0337.2016.80.4.tb06102.x

Hamburger, E. K., Cuzzi, S., Coddington, D. A., Allevi, A. M., Lopreiato, J., Moon, R., Yu, C., & Lane, J. L. (2011). Observation of resident clinical skills: outcomes of a program of direct observation in the continuity clinic setting. *Academic Pediatrics, 11*(5), 394-402. https://doi.org/10.1016/j.acap.2011.02.008

Hanna, M., & Fins, J. J. (2006). Power and communication: why simulation training ought to be complemented by experiential and humanist learning. *Academic Medicine, 81*(3), 265-270. https://doi.org/10.1097/00001888-200603000-00016

Hanson, K. J., & Stenvig, T. E. (2008). The good clinical nursing educator and the baccalaureate nursing clinical experience: Attributes and praxis. *Journal of*

Nursing Education, 47(1), 38-42. https://doi.org/10.3928/01484834-20080101-04

Hatala, R., Cook, D. A., Brydges, R., & Hawkins, R. (2015). Constructing a validity argument for the Objective Structured Assessment of Technical Skills (OSATS): A systematic review of validity evidence. *Advances in Health Sciences Education, 20*(5), 1149-1175. https://doi.org/10.1007/s10459-015-9593-1

Hauer, K. E., ten Cate, O., Boscardin, C., Irby, D. M., Iobst, W., & O'Sullivan, P. S. (2014). Understanding trust as an essential element of trainee supervision and learning in the workplace. *Advances in Health Sciences Education, 19*(3), 435-456. https://doi.org/10.1007/s10459-013-9474-4

Hausmann, J. S., & Schwartzstein, R. M. (2018). Using questions to enhance rheumatology education. *Arthritis Care and Research, 71*(10), 1304-1309. https://doi.org/10.1002/acr.23753

Hays, R. (2008). Assessment in medical education: roles for clinical teachers. *The Clinical Teacher, 5*(1), 23-27. https://doi.org/10.1111/j.1743-498X.2007.00165.x

Heiman, H. L., Rasminsky, S., Bierman, J. A., Evans, D. B., Kinner, K. G., Stamos, J., Martinovich, Z., & McGaghie, W. C. (2014). Medical students' observations, practices, and attitudes regarding electronic health record documentation. *Teaching and Learning in Medicine, 26*(1), 49-55. https://doi.org/10.1080/10401334.2013.857337

Hekelman, F. P., & Blase, J. R. (1996). Excellence in clinical teaching: the core of the mission. *Academic Medicine, 71*(7), 738-742. https://journals.lww.com/academicmedicine/Abstract/1996/07000/Excellence_in_clinical_teaching__the_core_of_the.5.aspx

Hesketh, E. A., Bagnall, G., Buckley, E. G., Friedman, M., Goodall, E., Harden, R. M., Laidlaw, J. M., Leighton-Beck, L., McKinlay, P., Newton, R., & Oughton, R. (2001). A framework for developing excellence as a clinical educator. *Medical Education, 35*(6), 555-564. https://doi.org/10.1046/j.1365-2923.2001.00920.x

Hewson, M. B., & Jensen, N. (1990). An inventory to improve clinical teaching in the general internal medicine clinic. *Medical Education, 24*, 518-527. https://doi.org/10.1111/j.1365-2923.1990.tb02668.x

Hodges, B., & McIlroy, J. H. (2003). Analytic global OSCE ratings are sensitive to level of training. *Medical Education, 37*(11), 1012-1016. https://doi.org/10.1046/j.1365-2923.2003.01674.x

Hornstein, H. A. (2017). Student evaluations of teaching are an inadequate assessment tool for evaluating faculty performance. *Cogent Education, 4*(1), 1304016. https://doi.org/10.1080/2331186X.2017.1304016

Huang, P.-H., Haywood, M., O'Sullivan, A., & Shulruf, B. (2019). A meta-analysis for comparing effective teaching in clinical education. *Medical Teacher, 41*(10), 1129-1142. https://doi.org/10.1080/0142159X.2019.1623386

Huttly, S., Sweet, J., & Taylor, I. (2003). *Effective learning and teaching in medical, dental and veterinary education*. Routledge.

Ilgen, J. S., Ma, I. W., Hatala, R., & Cook, D. A. (2015). A systematic review of validity evidence for checklists versus global rating scales in simulation-based assessment. *Medical Education, 49*(2), 161-173. https://doi.org/10.1111/medu.12621

Irby, D. M., Gillmore, G. M., & Ramsey, P. G. (1987). Factors affecting ratings of clinical teachers by medical students and residents. *Journal of Medical Education, 62*(1). https://doi.org/10.1097/00001888-198701000-00001

Jacobs, J. C. G., van Luijk, S. J., Galindo-Garre, F., Muijtjens, A. M. M., van der Vleuten, C. P. M., Croiset, G., & Scheele, F. (2014). Five teacher profiles in student-centred curricula based on their conceptions of learning and teaching. *BMC Medical Education, 14*(1), 220. https://doi.org/10.1186/1472-6920-14-220

Jacobs, J. C. G., Van Luijk, S. J., Van Berkel, H., Van der Vleuten, C. P. M., Croiset, G., & Scheele, F. (2012). Development of an instrument (the COLT) to measure conceptions on learning and teaching of teachers, in student-centred medical education. *Medical Teacher, 34*(7), e483-e491. https://doi.org/10.3109/0142159X.2012.668630

Jacobs, J. C. G., Wilschut, J., van der Vleuten, C., Scheele, F., Croiset, G., & Kusurkar, R. A. (2020). An international study on teachers' conceptions of

learning and teaching and corresponding teacher profiles. *Medical Teacher*, 1-5. https://doi.org/10.1080/0142159X.2020.1772465

James, P. A., & Osbourne, J. (1999). A measure of medical instructional quality in ambulatory settings: the MedIQ. *Family Medicine, 31*(4), 263-269.

Jarski, R. W., Kulig, K., & Olson, R. E. (1990). Clinical teaching in physical therapy: student and teacher perceptions. *Physical Therapy, 70*(3), 173-178. https://doi.org/10.1093/ptj/70.3.173

Jette, D. U., Nelson, L., Palaima, M., & Wetherbee, E. (2014). How do we improve quality in clinical education? Examination of structures, processes, and outcomes. *Journal of Physical Therapy Education, 28*. https://journals.lww.com/jopte/Fulltext/2014/00001/How_Do_We_Improve_Quality_in_Clinical_Education_.4.aspx

Jochemsen-van der Leeuw, H. G. A. R., van Dijk, N., van Etten-Jamaludin, F. S., & Wieringa-de Waard, M. (2013). The attributes of the clinical trainer as a role model: a systematic review. *Academic Medicine, 88*(1), 26-34. https://doi.org/10.1097/ACM.0b013e318276d070

Jochemsen-van der Leeuw, H. G. A. R., van Dijk, N., & Wieringa-de Waard, M. (2014). Assessment of the clinical trainer as a role model: a Role Model Apperception Tool (RoMAT). *Academic Medicine, 89*(4), 671. https://doi.org/10.1097/ACM.0000000000000169

Johnson, C. E., Keating, J. L., Boud, D. J., Dalton, M., Kiegaldie, D., Hay, M., McGrath, B., McKenzie, W. A., Nair, K. B. R., & Nestel, D. (2016). Identifying educator behaviours for high quality verbal feedback in health professions education: literature review and expert refinement. *BMC Medical Education, 16*(1), 96. https://doi.org/10.1186/s12909-016-0613-5

Johnston, S., & Vaughan, B. (2020). 'We need one more hour solely based on anatomy… Give us anatomy!': Early-year learner perceptions of anatomy within an integrated & case-based learning osteopathy curriculum. *International Journal of Osteopathic Medicine, 36*, 49-54. https://doi.org/10.1016/j.ijosm.2020.02.001

Jones, A. (2001). Possible influences on clinical supervision. *Nursing Standard (through 2013), 16*(1), 38-42.

Kandiah, D. A. (2017). Perception of educational value in clinical rotations by medical students. *Advances in Medical Education and Practice, 8*, 149. https://doi.org/10.2147/AMEP.S129183

Kane, M. T. (1992). An argument-based approach to validity. *Psychological Bulletin, 112*(3), 527. https://doi.org/10.1037/0033-2909.112.3.527

Kane, M. T. (2013a). The argument-based approach to validation. *School Psychology Review, 42*(4), 448. https://doi.org/10.1080/02796015.2013.12087465

Kane, M. T. (2013b). Validating the interpretations and uses of test scores. *Journal of Educational Measurement, 50*(1), 1-73. https://doi.org/10.1111/jedm.12000

Kane, M. T. (2016). Explicating validity. *Assessment in Education: Principles, Policy & Practice, 23*(2), 198-211. https://doi.org/10.1080/0969594X.2015.1060192

Kelly, C. (2007). Student's perceptions of effective clinical teaching revisited. *Nurse Education Today, 27*(8), 885-892. https://doi.org/10.1016/j.nedt.2006.12.005

Kern, D. E., Branch Jr, W. T., Jackson, J. L., Brady, D. W., Feldman, M. D., Levinson, W., & Lipkin Jr, M. (2005). Teaching the psychosocial aspects of care in the clinical setting: practical recommendations. *Academic Medicine, 80*(1), 8-20. https://doi.org/10.1097/00001888-200501000-00006

Kilminster, S., & Jolly, B. (2001). Effective supervision in clinical practice settings: a literature review. *Medical Education, 34*(10), 827-840. https://doi.org/10.1046/j.1365-2923.2000.00758.x

Klassen, R. M., & Tze, V. M. C. (2014). Teachers' self-efficacy, personality, and teaching effectiveness: A meta-analysis. *Educational Research Review, 12*, 59-76. https://doi.org/10.1016/j.edurev.2014.06.001

Knight, A. W. (2018). How clinical instructor behavior affects student clinical engagement from a motivational perspective. *Journal of Nuclear Medicine Technology, 46*(2), 99-106. https://doi.org/10.2967/jnmt.118.209320

Knight, L. V., & Bligh, J. (2006). Physicians' perceptions of clinical teaching: A qualitative analysis in the context of change. *Advances in Health Sciences Education, 11*(3), 221-234. https://doi.org/10.1007/s10459-005-4035-0

Koh, L. C. (2008). Refocusing formative feedback to enhance learning in pre-registration nurse education. *Nurse Education in Practice, 8*(4), 223-230. https://doi.org/10.1016/j.nepr.2007.08.002

Kraut, A., Yarris, L. M., & Sargeant, J. (2015). Feedback: Cultivating a positive culture. *Journal of Graduate Medical Education, 7*(2), 262-264. https://doi.org/10.4300/JGME-D-15-00103.1

Ladha, M., Bharwani, A., McLaughlin, K., Stelfox, H. T., & Bass, A. (2017). The effect of white coats and gender on medical students' perceptions of physicians. *BMC Medical Education, 17*(1), 93. https://doi.org/10.1186/s12909-017-0932-1

Lalonde, F. (2013). Problem-based learning in osteopathic education. *International Journal of Osteopathic Medicine, 16*(4), 216-219. https://doi.org/10.1016/j.ijosm.2013.08.003

Lane, S. (2014). Validity evidence based on testing consequences. *Psicothema, 26*(1), 127-135. https://doi.org/10.7334/psicothema2013.258

Laurent, T., & Weidner, T. G. (2001). Clinical instructors' and student athletic trainers' perceptions of helpful clinical instructor characteristics. *Journal of Athletic Training, 36*(1), 58-61. http://www.ncbi.nlm.nih.gov/pmc/articles/PMC155403/pdf/attr_36_01_0058.pdf

LeBaron Wallace, T. (2011). An argument-based approach to validity in evaluation. *Evaluation, 17*(3), 233-246. https://doi.org/10.1177/1356389011410522

Lehmann, L. S., Brancati, F. L., Chen, M.-C., Roter, D., & Dobs, A. S. (1997). The effect of bedside case presentations on patients' perceptions of their medical care. *New England Journal of Medicine, 336*(16), 1150-1156. https://doi.org/10.1056/NEJM199704173361606

Liao, H. C., Yang, Y. M., Li, T. C., Cheng, J. F., & Huang, L. C. (2019). The effectiveness of a clinical reasoning teaching workshop on clinical teaching ability in nurse preceptors. *Journal of Nursing Management, 27*. https://doi.org/10.1111/jonm.12773

Lo, L., & Regehr, G. (2017). Medical students' understanding of directed questioning by their clinical preceptors. *Teaching and Learning in Medicine, 29*(1), 5-12. https://doi.org/10.1080/10401334.2016.1213169

Lombarts, K. M., Heineman, M. J., Scherpbier, A. J. J. A., & Arah, O. A. (2014). Effect of the learning climate of residency programs on faculty's teaching

performance as evaluated by residents. *PLoS ONE, 9*(1), e86512. https://doi.org/10.1371/journal.pone.0086512

Lombarts, K. M. J. M. H., Bucx, M. J. L., & Arah, O. A. (2009). Development of a system for the evaluation of the teaching qualities of anesthesiology faculty. *Anesthesiology: The Journal of the American Society of Anesthesiologists, 111*(4), 709-716. https://doi.org/10.1097/ALN.0b013e3181b76516

London, S. (2008). The assessment of clinical practice in osteopathic education: is there a need to define a gold standard? *International Journal of Osteopathic Medicine, 11*(4), 132-136. https://doi.org/10.1016/j.ijosm.2008.08.029

Luciani, E., Consorti, G., van Dun, P. L. S., Merdy, O., Lunghi, C., Petracca, M., Esteves, J. E., & Cerritelli, F. (2018). An overview of osteopathy graduates' perceived preparedness at transition from educational environment to clinic environment one year after graduation: a cross sectional study. *BMC Medical Education, 18*(1), 319. https://doi.org/10.1186/s12909-018-1429-2

Luciani, E., van Dun, P. L., Esteves, J. E., Lunghi, C., Petracca, M., Papa, L., Merdy, O., Jäkel, A., & Cerritelli, F. (2015). Learning environment, preparedness and satisfaction in osteopathy in europe: The PreSS study. *PLoS ONE, 10*(6), e0129904. https://doi.org/10.1371/journal.pone.0129904

Macias, C., Gold, P. B., Öngür, D., Cohen, B. M., & Panch, T. (2015). Are single-item global ratings useful for assessing health status? *Journal of Clinical Psychology in Medical Settings, 22*(4), 251-264. https://doi.org/10.1007/s10880-015-9436-5

Mann, K., Gordon, J., & MacLeod, A. (2009). Reflection and reflective practice in health professions education: a systematic review. *Advances in Health Sciences Education, 14*(4), 595-621. https://doi.org/10.1007/s10459-007-9090-2

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 and Communication Disorders, 45*(1), 1-14. https://doi.org/10.3109/13682820902745461

McArthur, L. (2016). *Developing clinical teacher's self-efficacy in Australian general practice* [Doctoral dissertion, University of Adelaide]. University of Adelaide. http://hdl.handle.net/2440/119792

McCallum, C. A., Mosher, P. D., Jacobson, P. J., Gallivan, S. P., & Giuffre, S. M. (2013). Quality in physical therapist clinical education: A systematic review. *Physical Therapy, 93*(10), 1298-1311. https://doi.org/10.2522/ptj.20120410

McEvoy, J. W., Shatzer, J. H., Desai, S. V., & Wright, S. M. (2019). Questioning style and pimping in clinical education: A quantitative score derived from a survey of internal medicine teaching faculty. *Teaching and Learning in Medicine, 31*(1), 53-64. https://doi.org/10.1080/10401334.2018.1481752

McSparron, J. I., Vanka, A., & Smith, C. C. (2019). Cognitive learning theory for clinical teaching. *The Clinical Teacher, 16*(2), 96-100. https://doi.org/10.1111/tct.12781

Mengel, F., Sauermann, J., & Zölitz, U. (2019). Gender bias in teaching evaluations. *Journal of the European Economic Association, 17*(2), 535-566. https://doi.org/10.1093/jeea/jvx057

Meyer, I. S., Louw, A., & Ernstzen, D. (2019). Perceptions of physiotherapy clinical educators' dual roles as mentors and assessors: Influence on teaching-learning relationships. *The South African Journal of Physiotherapy, 75*(1), 468-468. https://doi.org/10.4102/sajp.v75i1.468

Miller, G. E. (1990). The assessment of clinical skills/competence/performance. *Academic Medicine, 65*(9), S63-67. https://journals.lww.com/academicmedicine/Abstract/1990/09000/The_assessment_of_clinical.45.aspx%5D

Mintz, M., Southern, D. A., Ghali, W. A., & Ma, I. W. Y. (2015). Validation of the 25-Item Stanford Faculty Development Program Tool on Clinical Teaching Effectiveness. *Teaching and Learning in Medicine, 27*(2), 174-181. https://doi.org/10.1080/10401334.2015.1011645

Moore, K., Bacon, R., Bevitt, T., Bialocerkowski, A., Ciccone, N., Haworth, N., Horstmanshof, L., Milne, N., Naumann, F., & Sanderson, B. (2018). Report: The university health clinic: Definition, educational practices and outcomes. *Focus on Health Professional Education: A Multi-Disciplinary Journal, 19*(2), 1. https://doi.org/10.11157/fohpe.v19i2.213

Moore, K., & Field, B. J. (2017). Enhancing clinical education in the private practice setting: A case study in osteopathy. *International Journal of Osteopathic Medicine, 23*, 42-50. https://doi.org/10.1016/j.ijosm.2016.06.003

Moore, K., & Vaughan, B. (2015). Assessment of Australian osteopathic learners' clinical competence during workplace learning. *International Journal of Osteopathic Medicine, 19*, 50-60. https://doi.org/10.1016/j.ijosm.2015.06.004

Nation, J. G., Carmichael, E., Fidler, H., & Violato, C. (2011). The development of an instrument to assess clinical teaching with linkage to CanMEDS roles: a psychometric analysis. *Medical Teacher, 33*, e290-e296. https://doi.org/10.3109/0142159X.2011.565825

Notzer, N., & Abramovitz, R. (2008). Can brief workshops improve clinical instruction? *Medical Education, 42*(2), 152-156. https://doi.org/10.1111/j.1365-2923.2007.02947.x

Olmos-Vega, F., Dolmans, D., Donkers, J., & Stalmeijer, R. E. (2015). Understanding how residents' preferences for supervisory methods change throughout residency training: a mixed-methods study. *BMC Medical Education, 15*(1), 177. https://doi.org/10.1186/s12909-015-0462-7

Orrock, P., Grace, S., Vaughan, B., & Coutts, R. (2014). Developing a viva exam to assess clinical reasoning in pre-registration osteopathy students. *BMC Medical Education, 14*(1), 1. https://doi.org/10.1186/1472-6920-14-193

Orsini, C., Evans, P., & Jerez, O. (2015). How to encourage intrinsic motivation in the clinical teaching environment?: a systematic review from the self-determination theory. *Journal of Educational Evaluation for Health Professions, 12*. https://doi.org/10.3352/jeehp.2015.12.8

Osteopathy Board of Australia. (2018). *Statistics*. Retrieved June 2, 2020, from https://www.osteopathyboard.gov.au/About/Statistics.aspx

Osteopathy Board of Australia. (2019). *Capabilities for Osteopathic Practice*. Retrieved July 4th, 2019, from https://www.osteopathyboard.gov.au/Codes-Guidelines/Capabilities-for-osteopathic-practice.aspx

Pallant, J. F., & Tennant, A. (2007). An introduction to the Rasch measurement model: an example using the Hospital Anxiety and Depression Scale (HADS). *British Journal of Clinical Psychology, 46*(1), 1-18. https://doi.org/10.1348/014466506x96931

Parsell, G., & Bligh, J. (2001). Recent perspectives on clinical teaching. *Medical Education, 35*(4), 409-414. https://doi.org/10.1046/j.1365-2923.2001.00900.x

Pascoe, J. M., Nixon, J., & Lang, V. J. (2015). Maximizing teaching on the wards: review and application of the One-Minute Preceptor and SNAPPS models. *Journal of Hospital Medicine, 10*(2), 125-130. https://doi.org/10.1002/jhm.2302

Passi, V., Johnson, S., Peile, E., Wright, S., Hafferty, F., & Johnson, N. (2013). Doctor role modelling in medical education: BEME Guide No. 27. *Medical Teacher, 35*(9), e1422-e1436. https://doi.org/10.3109/0142159X.2013.806982

Pelgrim, E. A. M., Kramer, A. W. M., Mokkink, H. G. A., & Van der Vleuten, C. P. M. (2012). Quality of written narrative feedback and reflection in a modified mini-clinical evaluation exercise: an observational study. *BMC Medical Education, 12*(1), 97. https://doi.org/10.1186/1472-6920-12-97

Perry, C., Henderson, A., & Grealish, L. (2018). The behaviours of nurses that increase student accountability for learning in clinical practice: An integrative review. *Nurse Education Today, 65*, 177-186. https://doi.org/10.1016/j.nedt.2018.02.029

Phillips, K. F., Mathew, L., Aktan, N., & Catano, B. (2017). Clinical education and student satisfaction: An integrative literature review. *International Journal of Nursing Sciences, 4*(2), 205-213. https://doi.org/https://doi.org/10.1016/j.ijnss.2017.03.004

Pintrich, P. R., Smith, D. A. F., Garcia, T., & McKeachie, W. J. (1993). Reliability and predictive validity of the Motivated Strategies for Learning Questionnaire (MSLQ). *Educational and Psychological Measurement, 53*(3), 801-813. https://doi.org/10.1177/0013164493053003024

Polyzois, I., McLouglin, J., Kelly, A., & Claffey, N. (2010). Clinical teaching in restorative dentistry and the variation between students' and supervisors' perceptions of its effectiveness. *European Journal of Dental Education, 14*(2), 92-98. https://doi.org/10.1111/j.1600-0579.2009.00597.x

Prideaux, D., Alexander, H., Bower, A., Dacre, J., Haist, S., Jolly, B., Norcini, J., Roberts, T., Rothman, A., & Rowe, R. (2000). Clinical teaching: maintaining an educational role for doctors in the new health care environment. *Medical Education, 34*(10), 820-826. https://doi.org/10.1046/j.1365-2923.2000.00756.x

Ragsdale, J. W., Berry, A., Gibson, J. W., Herber-Valdez, C. R., Germain, L. J., & Engle, D. L. (2020). Evaluating the effectiveness of undergraduate clinical education programs. *Medical Education Online, 25*(1), 1757883. https://doi.org/10.1080/10872981.2020.1757883

Raj, K. S. (2016). Well-being in residency: a systematic review. *Journal of Graduate Medical Education, 8*(5), 674-684. https://doi.org/10.4300/JGME-D-15-00764.1

Ramani, S. (2003). Twelve tips to improve bedside teaching. *Medical Teacher, 25*(2), 112-115. https://doi.org/10.1080/0142159031000092463

Ramani, S., & Leinster, S. (2008). AMEE Guide no. 34: teaching in the clinical environment. *Medical Teacher, 30*(4), 347-364. https://doi.org/10.1080/01421590802061613

Ramani, S., & Orlander, J. D. (2013). Human dimensions in bedside teaching: focus group discussions of teachers and learners. *Teaching and Learning in Medicine, 25*(4), 312-318. https://doi.org/10.1080/10401334.2013.827979

Ramsey, P. G., Gillmore, G. M., & Irby, D. M. (1988). Evaluating clinical teaching in the medicine clerkship. *Journal of General Internal Medicine, 3*(4), 351-355. https://doi.org/10.1007/BF02595793

Read, E. K., Bell, C., Rhind, S., & Hecker, K. G. (2015). The use of global rating scales for OSCEs in veterinary medicine. *PLoS ONE, 10*(3), e0121000. https://doi.org/10.1371/journal.pone.0121000

Recker-Hughes, C., Wetherbee, E., Buccieri, K. M., FitzpatrickTimmerberg, J., & Stolfi, A. M. (2014). Essential characteristics of quality clinical education experiences: Standards to facilitate student learning. *Journal of Physical Therapy Education, 28*, 48-55. https://journals.lww.com/jopte/Fulltext/2014/00001/Essential_Characteristics_of_Quality_Clinical.9.aspx

Regehr, G., & Eva, K. (2006). Self-assessment, self-direction, and the self-regulating professional. *Clinical Orthopaedics and Related Research, 449*, 34-38. https://doi.org/10.1097/01.blo.0000224027.85732.b2

Rich, M. E., & Lamiman, K. (2019). Five concepts for collaborative clinical teaching. *The Clinical Teacher*. https://doi.org/10.1111/tct.12802

Rieck, T., Callahan, J. L., & Watkins Jr, C. E. (2015). Clinical supervision: An exploration of possible mechanisms of action. *Training and Education in Professional Psychology, 9*(2), 187. https://doi.org/10.1037/tep0000080

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 and the Health Professions, 39*(1), 33-48. https://doi.org/10.1177/0163278713511456

Roff, S., McAleer, S., & Skinner, A. (2005). Development and validation of an instrument to measure the postgraduate clinical learning and teaching environment for hospital-based junior doctors in the UK. *Medical Teacher, 27*(4), 326-331. https://doi.org/10.1080/01421590500150874

Rotenstein, L. S., Ramos, M. A., Torre, M., Segal, J. B., Peluso, M. J., Guille, C., Sen, S., & Mata, D. A. (2016). Prevalence of depression, depressive symptoms, and suicidal ideation among medical students: a systematic review and meta-analysis. *JAMA, 316*(21), 2214-2236. https://doi.org/10.1001/jama.2016.17324

Rusticus, S. A., Wilson, D., Casiro, O., & Lovato, C. (2019). Evaluating the quality of health professions learning environments: development and validation of the Health Education Learning Environment Survey (HELES). *Evaluation and the Health Professions, 43*(3), 162-168. https://doi.org/10.1177/0163278719834339

Said, M., Jochemsen-van der Leeuw, R. H. G. A., Spek, B., Brand, P. L. P., & van Dijk, N. (2019). Role modelling in the training of hospital-based medical specialists: a validation study of the Role Model Apperception Tool (RoMAT). *Perspectives on Medical Education, 8*(4), 237-245. https://doi.org/10.1007/s40037-019-00527-6

Sandhu, G., Robinson, A. B., Magas, C. P., Purkiss, J. A., & Reddy, R. M. (2016). Teaching across the continuum: variations in rankings and valued teaching components between surgery residents and medical students. *The American Journal of Surgery, 212*(5), 1005-1010. https://doi.org/10.1016/j.amjsurg.2016.03.006

Scheepers, R. A., Arah, O. A., Heineman, M. J., & Lombarts, K. M. (2016). How personality traits affect clinician-supervisors' work engagement and

subsequently their teaching performance in residency training. *Medical Teacher, 38*(11), 1105-1111. https://doi.org/10.3109/0142159X.2016.1170774

Scheepers, R. A., Lombarts, K. M., van Aken, M. A., Heineman, M. J., & Arah, O. A. (2014). Personality traits affect teaching performance of attending physicians: Results of a multi-center observational study. *PLoS ONE, 9*(5), e98107. https://doi.org/10.1371/journal.pone.0098107

Schlair, S., Dyche, L., & Milan, F. (2017). Longitudinal faculty development program to promote effective observation and feedback skills in direct clinical observation. *MedEdPORTAL, 13*. https://doi.org/10.15766/mep_2374-8265.10648

Schopper, H., Rosenbaum, M., & Axelson, R. (2016). 'I wish someone watched me interview:' medical student insight into observation and feedback as a method for teaching communication skills during the clinical years. *BMC Medical Education, 16*(1), 286. https://doi.org/10.1186/s12909-016-0813-z

Schultz, K. W., Kirby, J., Delva, D., Godwin, M., Verma, S., Birtwhistle, R., Knapper, C., & Seguin, R. (2004). Medical students' and residents' preferred site characteristics and preceptor behaviours for learning in the ambulatory setting: a cross-sectional survey. *BMC Medical Education, 4*(1), 12. https://doi.org/10.1186/1472-6920-4-12

Schum, T. R., Koss, R., Yindra, K., & Nelson, D. B. (1993). Students' and residents' ratings of teaching effectiveness in a department of paediatrics. *Teaching and Learning in Medicine, 5*(3), 128-132. https://doi.org/10.1080/10401339309539606

Seabrook, M. A. (2003). Medical teachers' concerns about the clinical teaching context. *Medical Education, 37*(3), 213-222. https://doi.org/10.1046/j.1365-2923.2003.01437.x

Shayne, P., Heilpern, K., Ander, D., Palmer-Smith, V., & Emory University Department of Emergency Medicine Education Committee. (2002). Protected clinical teaching time and a bedside clinical evaluation instrument in an emergency medicine training program. *Academic Emergency Medicine, 9*(11), 1342-1349. https://doi.org/10.1197/aemj.9.11.1342

Shellenberger, S., & Mahan, J. (1982). A factor analytic study of teaching in off-campus general practice clerkships. *Medical Education, 16*, 151-155. https://doi.org/10.1111/j.1365-2923.1982.tb01076.x

Silber, C., Novielli, K., Paskin, D., Brigham, T., Kairys, J., Kane, G., & Veloski, J. (2006). Use of critical incidents to develop a rating form for resident evaluation of faculty teaching. *Medical Education, 40*, 1201-1208. https://doi.org/10.1111/j.1365-2929.2006.02631.x

Smedley, A., Morey, P., & Race, P. (2010). Enhancing the knowledge, attitudes, and skills of preceptors: An Australian perspective. *The Journal of Continuing Education in Nursing, 41*(10), 451-461. https://doi.org/10.3928/00220124-20100601-08

Smith, J. R., & Lane, I. F. (2015). Making the most of five minutes: the clinical teaching moment. *Journal of Veterinary Medical Education, 42*(3), 271-280. https://doi.org/10.3138/jvme.0115-004R

Snell, L., Tallett, S., Haist, S., Hays, R., Norcini, J., Prince, K., Rothman, A., & Rowe, R. (2000). A review of the evaluation of clinical teaching: new perspectives and challenges. *Medical Education, 34*(10), 862-870. https://doi.org/10.1046/j.1365-2923.2000.00754.x

Solarte, I., & Könings, K. D. (2017). Discrepancies between perceptions of students and deans regarding the consequences of restricting students' use of electronic medical records on quality of medical education. *BMC Medical Education, 17*(1), 55. https://doi.org/10.1186/s12909-017-0887-2

Solomon, D. J., Speer, A. J., Rosebraugh, C. J., & DiPette, D. J. (1997). The reliability of medical student ratings of clinical teaching. *Evaluation and the Health Professions, 20*(3), 343-352. https://doi.org/10.1177/016327879702000306

Spencer, D. C., Choi, D., English, C., & Girard, D. (2012). The effects of electronic health record implementation on medical student educators. *Teaching and Learning in Medicine, 24*(2), 106-110. https://doi.org/10.1080/10401334.2012.664513

Spencer, J. (2003). Learning and teaching in the clinical environment. *British Medical Journal, 326*(7389), 591-594. https://doi.org/10.1136/bmj.326.7389.591

Spooren, P., Brockx, B., & Mortelmans, D. (2013). On the validity of student evaluation of teaching: The state of the art. *Review of Educational Research, 83*(4), 598-642. https://doi.org/10.3102/0034654313496870

Spooren, P., & Christiaens, W. (2017). I liked your course because I believe in (the power of) student evaluations of teaching (SET). Students' perceptions of a teaching evaluation process and their relationships with SET scores. *Studies in Educational Evaluation, 54*, 43-49. https://doi.org/10.1016/j.stueduc.2016.12.003

Stachura, K., Garven, F., & Reed, M. (2000). Quality assurance: Measuring the quality of clinical education provision. *Physiotherapy, 86*(3), 117-126. https://doi.org/10.1016/S0031-9406(05)61154-6

Stalmeijer, R. E., Dolmans, D., Snellen-Balendong, H., van Santen-Hoeufft, M., Wolfhagen, H., & Scherpbier, A. (2013). Clinical teaching based on principles of cognitive apprenticeship: views of experienced clinical teachers. *Academic Medicine, 88*(6), 861-865. https://doi.org/10.1097/ACM.0b013e31828fff12

Stalmeijer, R. E., Dolmans, D., Wolfhagen, H., Muijtjens, A. M., & Scherpbier, A. (2008). The development of an instrument for evaluating clinical teachers: involving stakeholders to determine content validity *Medical Teacher, 30*, e272-e277. https://doi.org/10.1080/01421590802258904

Stalmeijer, R. E., Dolmans, D., Wolfhagen, H., Muijtjens, A. M., & Scherpbier, A. (2010a). The Maastricht Clinical Teaching Questionnaire (MTCQ) as a valid and reliable instrument for the evaluation of clinical teachers. *Academic Medicine, 85*(11), 1732-1738. https://doi.org/10.1097/ACM.0b013e3181f554d6

Stalmeijer, R. E., Dolmans, D., Wolfhagen, I., & Scherpbier, A. (2009). Cognitive apprenticeship in clinical practice: can it stimulate learning in the opinion of students? *Advances in Health Sciences Education, 14*(4), 535-546. https://doi.org/10.1007/s10459-008-9136-0

Stalmeijer, R. E., Dolmans, D. H. J. M., Wolfhagen, I. H. A. P., Peters, W. G., van Coppenolle, L., & Scherpbier, A. J. J. A. (2010b). Combined student ratings and self-assessment provide useful feedback for clinical teachers. *Advances in Health Sciences Education, 15*(3), 315-328. https://doi.org/10.1007/s10459-009-9199-6

Stark, P., & Freishtat, R. (2014). An evaluation of course evaluations. *ScienceOpen Research*. https://doi.org/10.14293/S2199-1006.1.SOR-EDU.AOFRQA.v1

Steiner, I. P., Yoon, P. W., Kelly, K. D., Diner, B. M., Donoff, M. G., Mackey, D. S., & Rowe, B. H. (2003). Resident evaluation of clinical teachers based on teachers' certification. *Academic Emergency Medicine, 10*(7), 731-737. https://doi.org/10.1111/j.1553-2712.2003.tb00067.x

Steinert, Y., Basi, M., & Nugus, P. (2017). How physicians teach in the clinical setting: The embedded roles of teaching and clinical care. *Medical Teacher, 39*(12), 1238-1244. https://doi.org/10.1080/0142159X.2017.1360473

Steinert, Y., Mann, K., Anderson, B., Barnett, B. M., Centeno, A., Naismith, L., Prideaux, D., Spencer, J., Tullo, E., & Viggiano, T. (2016). A systematic review of faculty development initiatives designed to enhance teaching effectiveness: A 10-year update: BEME Guide No. 40. *Medical Teacher, 38*(8), 769-786. https://doi.org/10.1080/0142159X.2016.1181851

Stoddard, H. A., & O'Dell, D. V. (2016). Would Socrates have actually used the "Socratic Method" for clinical teaching? *Journal of General Internal Medicine, 31*(9), 1092-1096. https://doi.org/10.1007/s11606-016-3722-2

Strand, P., Edgren, G., Borna, P., Lindgren, S., Wichmann-Hansen, G., & Stalmeijer, R. E. (2015). Conceptions of how a learning or teaching curriculum, workplace culture and agency of individuals shape medical student learning and supervisory practices in the clinical workplace. *Advances in Health Sciences Education, 20*(2), 531-557. https://doi.org/10.1007/s10459-014-9546-0

Strand, P., Sjöborg, K., Stalmeijer, R., Wichmann-Hansen, G., Jakobsson, U., & Edgren, G. (2013). Development and psychometric evaluation of the Undergraduate Clinical Education Environment Measure (UCEEM). *Medical Teacher, 35*(12), 1014-1026. http://informahealthcare.com/doi/abs/10.3109/0142159X.2013.835389

Streiner, D. L., Norman, G. R., & Cairney, J. (2014). *Health measurement scales: a practical guide to their development and use* (5th ed.). Oxford University Press.

Stritter, F. T., Hain, J. D., & Grimes, D. A. (1975). Clinical teaching reexamined. *Journal of Medical Education, 50*(9), 876-882. https://doi.org/10.1097/00001888-197509000-00005

Sulis, I., Porcu, M., & Capursi, V. (2019). On the use of student evaluation of teaching: A longitudinal analysis combining measurement issues and implications of the exercise. *Social Indicators Research, 142*(3), 1305-1331. https://doi.org/10.1007/s11205-018-1946-8

Summers, J. A. (2017). Developing competencies in the novice nurse educator: An integrative review. *Teaching and Learning in Nursing, 12*(4), 263-276. https://doi.org/10.1016/j.teln.2017.05.001

Sutkin, G., Wagner, E., Harris, I., & Schiffer, R. (2008). What makes a good clinical teacher in medicine? A review of the literature. *Academic Medicine, 83*(5), 452-466. https://doi.org/10.1097/ACM.0b013e31816bee61

Tavares, W., Brydges, R., Myre, P., Prpic, J., Turner, L., Yelle, R., & Huiskamp, M. (2018). Applying Kane's validity framework to a simulation based assessment of clinical competence. *Advances in Health Sciences Education, 23*(2), 323-338. https://doi.org/10.1007/s10459-017-9800-3

ten Cate, O., Snell, L., Mann, K., & Vermunt, J. (2004). Orienting teaching toward the learning process. *Academic Medicine, 79*(3), 219-228. https://doi.org/10.1097/00001888-200403000-00005

Tung, Y.-C., & Chang, G.-M. (2009). Patient satisfaction with and recommendation of a primary care provider: associations of perceived quality and patient education. *International Journal for Quality in Health Care, 21*(3), 206-213. https://doi.org/10.1093/intqhc/mzp006

Valencia, E. (2020). Acquiescence, instructor's gender bias and validity of student evaluation of teaching. *Assessment & Evaluation in Higher Education, 45*(4), 483-495. https://doi.org/10.1080/02602938.2019.1666085

Valiee, S., Moridi, G., Khaledi, S., & Garibi, F. (2016). Nursing students' perspectives on clinical instructors' effective teaching strategies: A descriptive study. *Nurse Education in Practice, 16*(1), 258-262. https://doi.org/10.1016/j.nepr.2015.09.009

Van den Berg, J. W., Mastenbroek, N. J. J. M., Scheepers, R. A., & Jaarsma, A. D. C. (2017). Work engagement in health professions education. *Medical Teacher, 39*(11), 1110-1118. https://doi.org/10.1080/0142159X.2017.1359522

Van den Berg, J. W., Verberg, C. P. M., Scherpbier, A. J. J. A., Jaarsma, A. D. C., Arah, O. A., & Lombarts, K. M. (2018). Faculty's work engagement in patient care: impact on job crafting of the teacher tasks. *BMC Medical Education, 18*(1), 312. https://doi.org/10.1186/s12909-018-1411-z

Van der Hem-Stokroos, H., van der Vleuten, C., Daelmans, H., Haarman, H., & Scherpbier, A. (2005). Reliability of the Clinical Teaching Effectiveness Instrument. *Medical Education, 39*, 904-910. https://doi.org/10.1111/j.1365-2929.2005.02245.x

Van der Leeuw, R., Lombarts, K., Heineman, M. J., & Arah, O. (2011). Systematic evaluation of the teaching qualities of obstetrics and gynecology faculty: reliability and validity of the SETO tools. *PLoS ONE, 6*(5), e19142. https://doi.org/10.1371/journal.pone.0019142

Van Gelderen, I., Matthew, S. M., Hendry, G. D., & Taylor, R. (2018). Variation in clinical placement supervisors' conceptions of and approaches to supervision in a veterinary internship programme. *Higher Education Research & Development, 37*(2), 418-432. https://doi.org/10.1080/07294360.2017.1355894

Van Lierop, M., de Jonge, L., Metsemakers, J., & Dolmans, D. (2018). Peer group reflection on student ratings stimulates clinical teachers to generate plans to improve their teaching. *Medical Teacher, 40*(3), 302-309. https://doi.org/10.1080/0142159X.2017.1406903

Varpio, L., Day, K., Elliot-Miller, P., King, J. W., Kuziemsky, C., Parush, A., Roffey, T., & Rashotte, J. (2015). The impact of adopting EHRs: how losing connectivity affects clinical reasoning. *Medical Education, 49*(5), 476-486. https://doi.org/10.1111/medu.12665

Vaughan, B. (2016). Confirmatory factor analysis of the Study Process Questionnaire in an Australian osteopathy student population. *International Journal of Osteopathic Medicine, 20*, 62-67. https://doi.org/10.1016/j.ijosm.2016.03.001

Vaughan, B. (2018). A Rasch analysis of the Revised Study Process Questionnaire in an Australian osteopathy student cohort. *Studies in Educational Evaluation, 56*, 144-153. https://doi.org/10.1016/j.stueduc.2017.12.003

Vaughan, B., Carter, A., Macfarlane, C., & Morrison, T. (2014a). The DREEM, part 1: measurement of the educational environment in an osteopathy teaching program. *BMC Medical Education, 14*(1), 99. https://doi.org/10.1186/1472-6920-14-99

Vaughan, B., Fleischmann, M., Fitzgerald, K., Grace, S., McLaughlin, P., Jolly, B., & Trumble, S. (2020). Profile of an allied health clinical supervision workforce: results from a nationally representative Australian practice-based research network. *Health Professions Education, 6*(3), 376-385. https://doi.org/10.1016/j.hpe.2020.04.008

Vaughan, B., Florentine, P., & Carter, A. (2013). Introducing a portfolio assessment in a pre-professional osteopathy program. *International Journal of Osteopathic Medicine, 17*(2), 129-134. https://doi.org/10.1016/j.ijosm.2013.06.003

Vaughan, B., Macfarlane, C., & Florentine, P. (2014b). Clinical education in the osteopathy program at Victoria University. *International Journal of Osteopathic Medicine, 17*(3), 199-205. https://doi.org/10.1016/j.ijosm.2013.10.010

Vaughan, B., & Moore, K. (2015a). The mini Clinical Evaluation Exercise (mini-CEX) in a pre-registration osteopathy program: Exploring aspects of its validity. *International Journal of Osteopathic Medicine, 19*, 61-72. https://doi.org/10.1016/j.ijosm.2015.07.002

Vaughan, B., & Moore, K. (2015b). The Osteopathic Clinical Practice Assessment–a pilot study to develop a new workplace-based assessment tool. *International Journal of Osteopathic Medicine, 18*(4), 287-296. https://doi.org/10.1016/j.ijosm.2015.05.002

Wald, H. S., George, P., Reis, S. P., & Taylor, J. S. (2014). Electronic health record training in undergraduate medical education: bridging theory to practice with curricula for empowering patient-and relationship-centered care in the computerized setting. *Academic Medicine, 89*(3), 380-386. https://doi.org/10.1097/ACM.0000000000000131

Walker, S., Rossi, D., Anastasi, J., Gray-Ganter, G., & Tennent, R. (2016). Indicators of undergraduate nursing students' satisfaction with their learning journey: An integrative review. *Nurse Education Today, 43*, 40-48. https://doi.org/10.1016/j.nedt.2016.04.011

Wallace, J. S., & Infante, T. D. (2008). Outcomes assessment of dental hygiene clinical teaching workshops. *Journal of Dental Education, 72*(10), 1169. http://www.jdentaled.org/content/72/10/1169.abstract

Warboys, I., Mok, W. Y., & Frith, K. H. (2014). Electronic medical records in clinical teaching. *Nurse Educator, 39*(6), 298-301. https://doi.org/10.1097/NNE.0000000000000072

Wasson, L. T., Cusmano, A., Meli, L., Louh, I., Falzon, L., Hampsey, M., Young, G., Shaffer, J., & Davidson, K. W. (2016). Association between learning environment interventions and medical student well-being: a systematic review. *JAMA, 316*(21), 2237-2252. https://doi.org/10.1001/jama.2016.17573

Wear, D., Kokinova, M., Keck-McNulty, C., & Aultman, J. (2005). Pimping: Perspectives of 4th year medical students. *Teaching and Learning in Medicine, 17*(2), 184-191. https://doi.org/10.1207/s15328015tlm1702_14

Wearne, S., Dornan, T., Teunissen, P. W., & Skinner, T. (2012). General practitioners as supervisors in postgraduate clinical education: an integrative review. *Medical Education, 46*(12), 1161-1173. https://doi.org/10.1111/j.1365-2923.2012.04348.x

Wheeler, N. L., & Gill, S. S. (2010). Recognising excellence in medical education: a student-led award scheme. *The Clinical Teacher, 7*(3), 153-156. https://doi.org/10.1111/j.1743-498X.2009.00342.x

Wijbenga, M. H., Bovend'Eerdt, T. J. H., & Driessen, E. W. (2019). Physiotherapy students' experiences with clinical reasoning during clinical placements: A qualitative study. *Health Professions Education, 5*(2), 126-135. https://doi.org/10.1016/j.hpe.2018.05.003

Williams, B. C., Litzelman, D. K., Babbott, S. F., Lubitz, R. M., & Hofer, T. P. (2002). Validation of a global measure of faculty's clinical teaching performance. *Academic Medicine, 77*(2), 177-180. https://doi.org/10.1097/00001888-200202000-00020

Wolpaw, T. M., Wolpaw, D. R., & Papp, K. K. (2003). SNAPPS: a learner-centered model for outpatient education. *Academic Medicine, 78*(9), 893-898. https://doi.org/10.1097/00001888-200309000-00010

Woolliscroft, J. O., Van Harrison, R., & Anderson, M. B. (2002). Faculty views of reimbursement changes and clinical training: a survey of award-winning clinical teachers. *Teaching and Learning in Medicine, 14*(2), 77-86. https://doi.org/10.1207/S15328015TLM1402_03

Wormley, M. E., Romney, W., & Greer, A. E. (2017). Development of the Clinical Teaching Effectiveness Questionnaire in the United States. *Journal of Educational Evaluation for Health Professions, 14*. https://doi.org/10.3352/jeehp.2017.14.14

Yazigi, A., Nasr, M., Sleilaty, G., & Nemr, E. (2006). Clinical teachers as role models: perceptions of interns and residents in a Lebanese medical school. *Medical Education, 40*(7), 654-661. https://doi.org/10.1111/j.1365-2929.2006.02502.x

Young, M. E., Cruess, S. R., Cruess, R. L., & Steinert, Y. (2014). The Professionalism Assessment of Clinical Teachers (PACT): the reliability and validity of a novel tool to evaluate professional and clinical teaching behaviors. *Advances in Health Sciences Education, 19*(1), 99-113. https://doi.org/10.1007/s10459-013-9466-4

Zhao, J., & Gallant, D. J. (2012). Student evaluation of instruction in higher education: Exploring issues of validity and reliability. *Assessment & Evaluation in Higher Education, 37*(2), 227-235. https://doi.org/10.1080/02602938.2010.523819

Zibrowski, E. M., Myers, K., Norman, G., & Goldszmidt, M. A. (2011). Relying on others' reliability: challenges in clinical teaching assessment. *Teaching and Learning in Medicine, 23*(1), 21-27. https://doi.org/10.1080/10401334.2011.536754

Zinbarg, R. E., Revelle, W., Yovel, I., & Li, W. (2005). Cronbach's α, Revelle's β, and McDonald's ω H: Their relations with each other and two alternative conceptualizations of reliability. *Psychometrika, 70*(1), 123-133. https://doi.org/10.1007/s11336-003-0974-7

Zoghi, M., Brown, T., Williams, B., Roller, L., Jaberzadeh, S., Palermo, C., McKenna, L., Wright, C., Baird, M., & Schneider-Kolsky, M. (2010). Learning style preferences of Australian health science students. *Journal of Allied Health, 39*(2), 95-103. https://www.ingentaconnect.com/content/asahp/jah/2010/00000039/00000002/art00008

Zuberi, R. W., Bordage, G., & Norman, G. R. (2007). Validation of the SETOC instrument - student evaluation of teaching in outpatient clinic. *Advances in Health Sciences Education, 12*, 55-69. https://doi.org/10.1007/s10459-005-2328-y

APPENDICES

Osteopathy Clinical Teaching Questionnaire

Clinical Educator name: _____

Please indicate your gender: ☐ Male ☐ Female

OCTQ
OSTEOPATHY CLINICAL TEACHING QUESTIONNAIRE

Using the following scale, please rate your Clinical Educator on the statements below:

5 – Strongly agree
4 – Agree
3 – Neither agree nor disagree
2 – Disagree
1 – Strongly disagree

This Clinical Educator…	**Rating**
1. Maintained a positive attitude towards me	5 4 3 2 1
2. Demonstrated humanistic attitudes in relating to patients (integrity, compassion and respect)	5 4 3 2 1
3. Showed genuine concern for my professional well-being	5 4 3 2 1
4. Has good communication skills	5 4 3 2 1
5. Is open to student questions and alternative approaches to patient management	5 4 3 2 1
6. Adjusted teaching to my needs (experience, competence, interest)	5 4 3 2 1
7. Promoted reflection on clinical practice	5 4 3 2 1
8. Emphasises a problem-solving approach rather than solutions	5 4 3 2 1
9. Asked questions to enhance my learning	5 4 3 2 1
10. Stimulates me to learn independently	5 4 3 2 1
11. Offered me suggestions for improvement when required	5 4 3 2 1
12. Demonstrated osteopathic, clinical examination and rehabilitation knowledge and skill(s)	5 4 3 2 1

Please rate your Clinical Educator on the following statements:

I would do more clinics with this Clinical Educator	5 4 3 2 1
I would recommend other students to work with this Clinical Educator	5 4 3 2 1

Rate the overall effectiveness of this Clinical Educator as an educator/supervisor:

☐ Poor ☐ Fair ☐ Good ☐ Very good ☐ Excellent

Please turn over to provide any comments on the Clinical Educator you have rated above.

What have you enjoyed about working with this Clinical Educator?

Are there any areas where this Clinical Educator could improve or develop?

APPENDICES

Chapter 2

	Medline	CINAHL
Search term	**Articles found**	**Articles found**
Clinical teacher NOT nursing	Butani et al. [1] Boerboom et al. [2] Arah et al. [3] Boerebach et al. [4] Makama & Ameh [5] Al-Qahtani [6] Bell et al. [7] Butvidas [8] Boerboom et al. [9] Stenfors-Hays et al. [10] Schwartz [11] Bannister et al. [12]	Al Haqwi et al. [13] Arah et al. [3] Schwartz [11]
Medical teacher NOT nursing AND instrument	Archer et al. [14] Riquelme [15] Fluit et al. [16] Boerboom [2] Kamran et al. [17] Nation et al. [18] Boor et al. [19] Boerboom et al. [9]	
Medical teacher NOT nursing AND validity	Archer et al. [14] van Es et al. [20] Fluit et al. [16] Boerboom et al. [2] Yu et al. [21] Nation et al. [18] Boerboom et al. [9]	
Clinical teaching NOT nursing AND instrument	Schonrock-Adema et al. [22] Archer et al. [14] Fluit et al. [16] Egbe & Baker [23] Boerboom et al. [2] Nation et al. [18] Boerboom et al. [9] Zibrowski et al. [24] Stalmeijer et al. [25]	Conigliaro & Stratton [26] Zibrowski et al. [24]
Clinical teaching NOT nursing AND evaluation AND instrument	Schonrock-Adema et al. [22] Archer et al. [14] Fluit et al. [16] Egbe & Baker [23] Boerboom et al. [2] Nation et al. [18] Boerboom et al. [9] Zibrowski et al. [24] Stalmeijer et al. [25]	Conigliaro & Stratton [26] Zibrowski et al. [24]

Medical education AND clinical teacher	Arah et al. [3] Makama & Ameh [5] Schwartz [11]	Al Haqwi et al. [13] Arah et al. [3]
Medical education AND effectiveness AND instrument	Schonrock-Adema et al. [22] Archer et al. [14] Zibrowski et al. [24]	XiaoJing [27] Zibrowski et al. [24]
Medical education AND teaching AND instrument AND evaluation	Archer et al. [14] Fluit et al. [16] Egbe & Baker [23] Arah et al. [3] Nation et al. [18] Zibrowski et al. [24] Stalmeijer et al. [25]	Conigliaro & Stratton [26] XiaoJing [27] Zibrowski et al. [24]
Medical education AND teaching AND validity AND instrument	Schonrock-Adema et al. [22] Archer et al. [14] Fluit et al. [16] Egbe & Baker [23] Arah et al. [3] Zibrowski et al. [24] Stalmeijer et al. [25]	Conigliaro & Stratton [26] XiaoJing [27]
Instrument AND clinical teacher	Bergjan & Hertel [28] Henriksen et al. [29] Bos et al. [30] Boerboom et al. [2] Boerboom et al. [9] Johansson et al. [31]	Johansson et al. [31] Henriksen et al. [29] Bos et al. [30]
Instrument AND validity AND clinical teacher	Henriksen et al. [29] Boerboom et al. [2] Boerboom et al. [9] Johansson et al. [31]	Johansson et al. [31] Henriksen et al. [29]

Additional File 1. Articles identified and selected from the systematic search of the literature from 1st March 2010 to 1st January 2013.

REFERENCES

1. Butani L, Paterniti DA, Tancredi DJ, Li S-TT: **Attributes of residents as teachers and role models-a mixed methods study of stakeholders.** *Med Teach* 2012, **35:**e1052-e1059.
2. Boerboom TB, Mainhard T, Dolmans DH, Scherpbier AJ, Van Beukelen P, Jaarsma AC: **Evaluating clinical teachers with the Maastricht clinical teaching questionnaire: How much'teacher'is in student ratings?** *Med Teach* 2012, **34:**320-326.
3. Arah OA, Heineman MJ, Lombarts KM: **Factors influencing residents' evaluations of clinical faculty member teaching qualities and role model status.** *Med Educ* 2012, **46:**381-389.
4. Boerebach BC, Lombarts KM, Keijzer C, Heineman MJ, Arah OA: **The teacher, the physician and the person: how faculty's teaching performance influences their role modelling.** *PLoS ONE* 2012, **7:**e32089.
5. Makama J, Ameh E: **Quality of teaching provided by surgical residents: an evaluation of the perception of medical students.** *Nigerian journal of medicine: journal of the National Association of Resident Doctors of Nigeria* 2010, **20:**341-344.
6. Al-Qahtani MF: **What constitutes a good clinical teacher?** *The Journal Of The Egyptian Public Health Association* 2011, **86:**27-32.
7. Bell MA, Wales PS, Torbeck LJ, Kunzer JM, Thurston VC, Brokaw JJ: **Do personality differences between teachers and learners impact students' evaluations of a surgery clerkship?** *J Surg Educ* 2011, **68:**190-193.
8. Butvidas LD, Anderson CI, Balogh D, Basson MD: **Disparities between resident and attending surgeon perceptions of intraoperative teaching.** *The American Journal of Surgery* 2011, **201:**385-389.
9. Boerboom T, Dolmans D, Jaarsma A, Muijtjens A, Van Beukelen P, Scherpbier A: **Exploring the validity and reliability of a questionnaire for evaluating veterinary clinical teachers' supervisory skills during clinical rotations.** *Med Teach* 2011, **33:**e84-e91.
10. Stenfors-Hayes T, Hult H, Dahlgren LO: **What does it mean to be a good teacher and clinical supervisor in medical education?** *Adv Health Sci Educ* 2011, **16:**197-210.
11. Schwartz AJ: **Resident/fellow evaluation of clinical teaching: an essential ingredient of effective teacher development and educational planning.** *Anesthesiology* 2010, **113:**516-517.
12. Bannister SL, Raszka WV, Maloney CG: **What makes a great clinical teacher in pediatrics? Lessons learned from the literature.** *Pediatrics* 2010, **125:**863-865.
13. AlHaqwi AI, van der Molen HT, Schmidt H, Magzoub M: **Determinants of effective clinical learning: a student and teacher perspective in Saudi Arabia.** *Education for Health* 2010, **23:**369.
14. Archer J, Swanwick T, Smith D, O'Keeffe C, Cater N: **Developing a multisource feedback tool for postgraduate medical educational supervisors.** *Med Teach* 2013, **35:**145-154.
15. Riquelme A, Padilla O, Herrera C, Olivos T, Román JA, Sarfatis A, Solís N, Pizarro M, Torres P, Roff S: **Development of ACLEEM questionnaire, an instrument measuring residents' educational environment in postgraduate ambulatory setting.** *Med Teach* 2013, **35:**e861-e866.
16. Fluit C, Bolhuis S, Grol R, Ham M, Feskens R, Laan R, Wensing M: **Evaluation and feedback for effective clinical teaching in postgraduate medical education: Validation of an assessment instrument incorporating the CanMEDS roles.** *Med Teach* 2012, **34:**893-901.
17. Kamran A, Zibaei M, Mirkaimi K, Shahnazi H: **Designing and evaluation of the teaching quality assessment form from the point of view of the Lorestan University of Medical Sciences students-2010.** *Journal of education and health promotion* 2012, **1**.

18. Nation JG, Carmichael E, Fidler H, Violato C: **The development of an instrument to assess clinical teaching with linkage to CanMEDS roles: a psychometric analysis.** *Med Teach* 2011, **33:**e290-e296.
19. Boor K, Van Der Vleuten C, Teunissen P, Scherpbier A, Scheele F: **Development and analysis of D-RECT, an instrument measuring residents' learning climate.** *Med Teach* 2011, **33:**820-827.
20. van Es JM, Schrijver CJ, Oberink RH, Visser MR: **Two-dimensional structure of the MAAS-Global rating list for consultation skills of doctors.** *Med Teach* 2012, **34:**e794-e799.
21. Yu T-C, Wheeler BR, Hill AG: **Clinical supervisor evaluations during general surgery clerkships.** *Med Teach* 2011, **33:**e479-e484.
22. Schönrock-Adema J, Boendermaker PM, Remmelts P: **Opportunities for the CTEI: disentangling frequency and quality in evaluating teaching behaviours.** *Perspectives on medical education* 2012, **1:**172-179.
23. Egbe M, Baker P: **Development of a multisource feedback instrument for clinical supervisors in postgraduate medical training.** *Clin Med* 2012, **12:**239-243.
24. Zibrowski EM, Myers K, Norman G, Goldszmidt MA: **Relying on others' reliability: challenges in clinical teaching assessment.** *Teach Learn Med* 2011, **23:**21-27.
25. Stalmeijer RE, Dolmans D, Wolfhagen H, Muijtjens AM, Scherpbier A: **The Maastricht Clinical Teaching Questionnaire (MTCQ) as a valid and reliable instrument for the evaluation of clinical teachers.** *Acad Med* 2010, **85:**1732-1738.
26. Conigliaro RL, Stratton TD: **Assessing the quality of clinical teaching: a preliminary study.** *Med Educ* 2010, **44:**379-386.
27. XiaoJing H, Zhu D, Zheng M: **Clinical Nursing Faculty Competence Inventory–development and psychometric testing.** *J Adv Nurs* 2011, **67:**1109-1117.
28. Bergjan M, Hertel F: **Evaluating students' perception of their clinical placements—Testing the clinical learning environment and supervision and nurse teacher scale (CLES+ T scale) in Germany.** *Nurse Educ Today* 2013, **33:**1393-1398.
29. Henriksen N, Normann HK, Skaalvik MW: **Development and testing of the Norwegian version of the Clinical Learning Environment, Supervision and Nurse Teacher (CLES+ T) evaluation scale.** *Int J Nurs Educ Scholarsh* 2012, **9**.
30. Bos E, Alinaghizadeh H, Saarikoski M, Kaila P: **Validating the 'clinical learning environment, supervision and nurse teacher'CLES+ T instrument in primary healthcare settings using confirmatory factor analysis.** *J Clin Nurs* 2012, **21:**1785-1788.
31. Johansson UB, Kaila P, Ahlner‐Elmqvist M, Leksell J, Isoaho H, Saarikoski M: **Clinical learning environment, supervision and nurse teacher evaluation scale: psychometric evaluation of the Swedish version.** *J Adv Nurs* 2010, **66:**2085-2093.

APPENDICES

Chapter 3

Osteopathy Clinical Teaching Questionnaire

Using the following scale, please rate your Clinical Educator on the statements below:

5 – Strongly agree
4 – Agree
3 – Neither agree nor disagree
2 – Disagree
1 – Strongly disagree

Please indicate the gender of the Clinical Educator who is being rated:
☐ Male ☐ Female

Please indicate your gender:
☐ Male ☐ Female

This Clinical Educator…	**Rating**
Treated me with respect	5 4 3 2 1
Maintained a positive attitude towards me	5 4 3 2 1
Fostered an environment of respect in which I felt comfortable participating	5 4 3 2 1
Established a good learning environment (approachable, focused, nonthreatening, professional and enthusiastic)	5 4 3 2 1
Demonstrated humanistic attitudes in relating to patients (integrity, compassion and respect)	5 4 3 2 1
Was approachable for discussion	5 4 3 2 1
Showed genuine concern for my professional well-being	5 4 3 2 1
Had reasonable expectations of students	5 4 3 2 1
Has good communication skills	5 4 3 2 1
Is open to student questions and alternative approaches to patient management	5 4 3 2 1
Gave me the opportunity to offer opinions on patient problems or treatment	5 4 3 2 1
Adjusted teaching to my needs (experience, competence, interest)	5 4 3 2 1
Is an effective clinical teacher	5 4 3 2 1
Encouraged me to think	5 4 3 2 1
Promoted reflection on clinical practice	5 4 3 2 1
Emphasises a problem-solving approach rather than solutions	5 4 3 2 1
Asked questions that promote learning (clarifies, probes, reflective questions etc.)	5 4 3 2 1
Asked questions to enhance my learning	5 4 3 2 1
Encouraged questions and active participation	5 4 3 2 1
Stimulates me to learn independently	5 4 3 2 1
Gave timely feedback to me	5 4 3 2 1
Gave me regular, useful feedback about my knowledge and performance	5 4 3 2 1
Offered me suggestions for improvement when required	5 4 3 2 1

Please turn over the page

Osteopathy Clinical Teaching Questionnaire

This Clinical Educator…	**Rating**
Identified areas needing improvement	5 4 3 2 1
Identified my strengths	5 4 3 2 1
Explained to me why I was correct or incorrect	5 4 3 2 1
Promoted keeping of medical records in a way that is thorough, legible, efficient and organised	5 4 3 2 1
Encouraged me to assume responsibility for patient care	5 4 3 2 1
Demonstrates knowledge of current medical and manual therapy literature	5 4 3 2 1
Demonstrated osteopathic, clinical examination and rehabilitation knowledge and skill(s)	5 4 3 2 1

Please rate your Clinical Educator on the following statements:

I would do more clinics with this Clinical Educator	5 4 3 2 1
I would recommend other students to work with this Clinical Educator	5 4 3 2 1

Rate the overall effectiveness of this Clinical Educator as an educator/supervisor:

☐ Poor ☐ Fair ☐ Good ☐ Very good ☐ Excellent

What have you enjoyed about working with this Clinical Educator?

Are there any areas where this Clinical Educator could improve or develop?

Item	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30
1	1																													
2	0.647	1																												
3	0.538	0.542	1																											
4	0.185	0.262	0.390	1																										
5	0.081	0.123	0.108	0.175	1																									
6	0.272	0.351	0.348	0.390	0.036	1																								
7	0.108	0.131	0.116	0.119	0.174	0.333	1																							
8	0.133	0.185	0.153	0.223	0.068	0.110	0.169	1																						
9	0.185	0.108	0.251	0.218	0.181	0.092	0.021	0.257	1																					
10	0.110	0.104	0.124	0.171	0.151	0.084	-0.067	0.148	0.218	1																				
11	0.074	0.104	0.081	0.061	0.044	0.062	-0.037	0.162	0.075	0.264	1																			
12	0.127	0.108	0.056	0.066	-0.043	0.015	-0.051	0.105	0.164	0.153	0.302	1																		
13	0.054	0.013	0.091	0.133	0.012	0.026	-0.112	-0.076	0.201	0.111	0.076	0.228	1																	
14	-0.178	-0.192	-0.196	-0.14	-0.128	-0.136	-0.161	-0.195	-0.121	-0.131	-0.085	-0.088	0.202	1																
15	-0.225	-0.174	-0.198	-0.099	-0.081	-0.052	0.080	-0.071	-0.171	-0.105	-0.111	-0.042	-0.030	0.224	1															
16	-0.171	-0.203	-0.186	-0.101	-0.099	-0.206	-0.051	-0.080	-0.034	-0.007	0.013	-0.064	0.079	0.295	0.277	1														
17	-0.131	-0.188	-0.199	-0.164	-0.094	-0.126	-0.124	-0.142	-0.113	-0.077	0.056	0.067	0.096	0.314	0.159	0.202	1													
18	-0.168	-0.168	-0.184	-0.134	-0.084	-0.172	-0.149	-0.19	-0.146	-0.084	-0.057	-0.079	0.082	0.332	0.092	0.206	0.558	1												
19	-0.074	-0.061	-0.108	-0.007	-0.042	-0.042	-0.001	-0.126	-0.116	-0.060	-0.036	-0.056	0.039	0.118	0.031	0.045	0.214	0.376	1											
20	-0.166	-0.069	-0.134	-0.087	0.059	-0.136	0.080	-0.114	-0.09	-0.157	-0.081	-0.044	-0.092	0.006	-0.006	0.042	-0.103	0.049	0.063	1										
21	-0.091	-0.097	-0.082	-0.206	-0.141	-0.134	-0.066	-0.132	-0.113	-0.203	-0.139	-0.165	-0.241	-0.205	-0.011	-0.041	-0.120	-0.141	-0.053	0.013	1									
22	-0.095	-0.186	-0.219	-0.235	-0.248	-0.098	-0.132	-0.151	-0.093	-0.182	-0.156	-0.166	-0.227	-0.150	-0.027	-0.099	-0.112	-0.116	-0.004	-0.020	0.541	1								
23	-0.086	-0.139	-0.158	-0.188	-0.169	-0.184	-0.124	-0.144	-0.112	-0.197	-0.178	-0.12	-0.220	-0.066	-0.042	-0.062	-0.095	-0.097	-0.166	0.007	0.178	0.408	1							
24	-0.296	-0.349	-0.297	-0.306	-0.208	-0.257	-0.157	-0.261	-0.254	-0.276	-0.299	-0.332	-0.216	0.035	-0.061	-0.053	-0.050	0.001	-0.124	-0.004	0.151	0.219	0.279	1						
25	-0.018	-0.085	-0.091	-0.147	-0.144	-0.095	-0.124	-0.056	-0.23	-0.109	-0.216	-0.122	-0.133	-0.219	-0.180	-0.250	-0.224	-0.216	-0.137	-0.081	0.082	0.112	0.125	0.331	1					
26	-0.141	-0.154	-0.102	-0.100	-0.148	-0.074	-0.153	-0.115	-0.141	-0.133	-0.163	-0.066	-0.076	-0.137	-0.141	-0.193	-0.110	-0.034	-0.053	-0.108	0.056	0.037	0.180	0.240	0.213	1				
27	-0.260	-0.255	-0.261	-0.233	-0.093	-0.184	-0.082	-0.091	-0.203	-0.198	-0.090	-0.096	-0.233	0.040	-0.043	-0.059	-0.082	-0.109	-0.163	0.031	0.017	0.055	-0.029	0.155	0.070	0.042	1			
28	-0.205	-0.169	-0.173	-0.247	0.049	-0.142	-0.016	-0.163	-0.175	-0.155	-0.114	-0.118	-0.155	0.123	-0.025	0.061	-0.086	-0.072	-0.088	0.070	0.093	-0.048	0.030	0.064	-0.052	0.044	0.255	1		
29	-0.177	-0.109	-0.128	-0.059	0.039	-0.156	-0.130	-0.125	-0.041	0.017	-0.053	-0.169	0.000	0.104	0.020	0.012	0.073	0.018	-0.006	-0.048	-0.138	-0.143	-0.099	0.056	-0.095	-0.081	-0.021	0.109	1	
30	-0.096	-0.121	-0.076	-0.044	0.008	-0.102	-0.138	-0.103	-0.033	0.015	0.058	-0.096	0.119	0.024	-0.098	-0.079	0.108	0.052	0.089	-0.054	-0.135	-0.156	-0.045	-0.007	-0.088	-0.040	-0.020	-0.026	0.396	1

Note: residual correlations >0.20 are highlighted.

Iteration 1

Step	Description	Chi-square	PSI	Items	Person	Outcome
1	Initial analysis	407.42, p<0.0001	0.910	2.34	2.03	Item fit issues for 1, 3, 4, 17, 19, 24, 25, 27 Disordered thresholds for items 1, 9, 27, 30 DIF for institution (items 14, 27, 28), clinical educator gender (item 14), student gender (item 19) 122 misfitting persons Local dependence multiple items
2	Rescore item 27	316.97, p<0.0001	0.928	2.13	1.11	Item fit issues for 17, 19, 24, 25, 27 Disordered thresholds for items 1, 30 DIF for institution (items 14, 27), clinical educator gender (item 14), student gender (item 19) 4 misfitting persons Local dependence multiple items
3	Rescore item 30	316.87, p<0.0001	0.928	2.13	1.11	Item fit issues for 17, 19, 24, 25, 27 Disordered thresholds for items 1, 14 DIF for institution (items 14, 27), clinical educator gender (item 14), student gender (item 19) 4 misfitting persons Local dependence multiple items
4	Delete item 27 (fit SD 6.497)	237.22, p<0.0001	0.926	1.83	1.07	Item fit issues for 19, 24, 25 Disordered thresholds for item 1 DIF for institution (items 14, 28), clinical educator gender (item 14), student gender (item 19) 3 misfitting persons Local dependence multiple items
5	Delete item 24 (fit SD 3.985)	220.17, p<0.0001	0.924	1.83	1.02	Item fit issues for 19, 25 Disordered thresholds for item 1 DIF for institution (items 14, 28), clinical educator gender (item 14) 3 misfitting persons Local dependence multiple items

6	Delete item 25 (fit SD 4.904)	186.89, p=0.002	0.916	1.74	1.01	Item fit issues for 19, 26 Disordered thresholds for item 1 DIF for institution (item 14), clinical educator gender (item 14) 7 misfitting persons Local dependence multiple items
7	Delete 7 misfitting persons	186.21, p=0.002	0.922	1.73	1.01	Item fit issues for 19, 26 Disordered thresholds for items 1, 3 DIF for institution (item 14), clinical educator gender (item 14), student gender (item 19) No misfitting persons Local dependence multiple items
8	Delete item 14 (due to DIF)	184.17, p=0.001	0.920	1.75	1.02	Item fit issues for 19, 26 Disordered thresholds for items 1, 3 DIF for institution (item 28), student gender (item 19) 2 misfitting persons Local dependence multiple items
9	Delete item 19 (due to DIF)	167.77, p=0.006	0.915	1.70	1.00	Item fit issues for 3, 26 Disordered thresholds for items 1, 3 No DIF 1 misfitting person Local dependence multiple items
10	Delete item 3 (fit SD -2.566)	170.11, p=0.001	0.912	1.59	0.99	Item fit issue for 26 Disordered thresholds for item 1 DIF for institute (item 28) 1 misfitting person Local dependence multiple items
11	Delete item 26 (fit SD 2.974)	138.57, p=0.066	0.903	1.57	0.97	Item fit issues Disordered thresholds for item 1 No DIF 5 misfitting persons Local dependence multiple items

12	Delete 5 misfitting persons	137.08, p=0.078	0.907	1.56	0.95	No item fit issues Disordered thresholds for item 1 No DIF No misfitting persons Local dependence multiple items
13	Delete item 1 (unable to resolve disordering)	139.43, p=0.030	0.904	1.50	0.95	No item fit issues No disordered thresholds No DIF 2 misfitting persons Local dependence multiple items
14	Delete 2 misfitting persons	135.76, p=0.048	0.904	1.52	0.93	No item fit issues No disordered thresholds DIF for institute (item 28) No misfitting persons Local dependence multiple items
15	Delete item 28 (due to DIF)	133.23, p=0.030	0.904	1.58	0.92	No item fit issues No disordered thresholds No DIF 4 misfitting persons Local dependence multiple items
16	Delete 4 misfitting persons	129.61, p=0.051	0.901	1.55	0.90	No item fit issues No disordered thresholds No DIF No misfitting persons Local dependence multiple items
17	Delete item 17 (after subtest of items 17/18 due to r=0.550)	124.98, p=0.046	0.894	1.50	0.81	Item fit issues for 4, 9 No disordered thresholds No DIF No misfitting persons Local dependence multiple items
18	Delete item 21 (after subtest of	134.01, p=0.005	0.888	1.62	0.85	Item fit issues for 4, 9 No disordered thresholds

	items 21/22 due to r=0.540)					No DIF 2 misfitting persons Local dependence multiple items
19	Delete 2 misfitting persons	134.01, p=0.005	0.890	1.62	0.85	Item fit issues for 4, 9 No disordered thresholds No DIF 2 misfitting persons Local dependence multiple items
20	Delete item 9 (fit SD -2.718)	111.62, p=0.061	0.881	1.49	0.84	Item fit issues for 4 No disordered thresholds No DIF 1 misfitting person Local dependence multiple items
21	Delete misfitting person	111.62, p=0.061	0.882	1.49	0.84	Item fit issues for 4 No disordered thresholds No DIF No misfitting persons Local dependence multiple items
22	Delete item 4 (fit SD -2.697)	102.94, p=0.090	0.874	1.35	0.84	No item fit issues No disordered thresholds No DIF No misfitting persons Local dependence multiple items
23	Delete item 6 (after subtest of items 6/7 r=0.325, and 2/6 r=0.276)	97.44, p=0.089	0.868	1.35	0.83	No item fit issues No disordered thresholds No DIF 2 misfitting persons Local dependence multiple items
24	Delete 2 misfitting persons	97.19, p=0.092	0.868	1.34	0.82	No item fit issues No disordered thresholds No DIF No misfitting persons

						Local dependence multiple items
25	Delete item 22 (after subtest of items 22/23 r=0.410)	88.535, p=0.135	0.854	1.50	0.85	No item fit issues No disordered thresholds No DIF 7 misfitting persons Local dependence multiple items
26	Delete 7 misfitting persons	86.89, p=0.163	0.843	1.44	0.80	No item fit issues No disordered thresholds DIF for institution (item 13) No misfitting persons Local dependence multiple items
27	Delete item 13 (due to DIF)	91.17, p=0.045	0.830	1.29	0.83	No item fit issues No disordered thresholds No DIF 1 misfitting persons Local dependence multiple items
28	Delete item 12 (after subtest of items 11/12 r=0.220)	66.47, p=0.425	0.809	1.20	0.82	No item fit issues No disordered thresholds No DIF 1 misfitting persons Local dependence multiple items
29	Delete item 29 (after subtest of items 29/30 r=0.338)	66.06, p=0.275	0.800	1.20	0.83	No item fit issues No disordered thresholds No DIF 3 misfitting persons No local dependence
30	Delete 3 misfitting persons	66.06, p=0.275	0.802	1.20	0.83	p-value range 0.061 Lower CI 0.031-0.038 Upper CI 0.091-0.101

Iteration 2

Step	Description	Chi-square	PSI	Items	Person	Outcome
1	Initial analysis	407.42, p<0.0001	0.910	2.34	2.03	Item fit issues for 1, 3, 4, 17, 19, 24, 25, 27 Disordered thresholds for items 1, 9, 27, 30 DIF for institution (items 14, 27, 28), clinical educator gender (item 14), student gender (item 19) 122 misfitting persons Local dependence multiple items
2	Delete item 27 (fit SD 6.138)	237.36, p<0.0001	0.926	1.83	1.07	Item fit issues for 19, 24, 25 Disordered thresholds for items 1, 30 DIF for institution (items 14, 28), clinical educator gender (item 14), student gender (item 19) 3 misfitting persons Local dependence multiple items
3	Delete item 19 (due to DIF and fit SD -3.229)	236.59, p<0.0001	0.922	1.74	1.05	Item fit issues for 24, 25 Disordered thresholds for items 1, 14, 30 DIF for institution (items 14, 28), clinical educator gender (item 14) 3 misfitting persons Local dependence multiple items
4	Delete 3 misfitting persons	239.02, p<0.0001	0.921	1.68	1.03	Item fit issues for 24, 25 Disordered thresholds for items 1, 21, 30 DIF for institution (items 14, 28), clinical educator gender (item 14) No misfitting persons Local dependence multiple items
5	Rescore item 21	239.02, p<0.0001	0.921	1.68	1.03	Item fit issues for 24, 25 Disordered thresholds for items 1, 30 DIF for institution (items 14, 28), clinical educator gender (item 14) No misfitting persons Local dependence multiple items
6	Rescore item 30	239.02, p<0.0001	0.921	1.68	1.03	Item fit issues for 24, 25

						Disordered thresholds for item 1 DIF for institution (items 14, 26), clinical educator gender (item 14) No misfitting persons Local dependence multiple items
7	Delete item 1 (unable to resolve disordering)	218.47, p<0.0001	0.919	1.63	1.03	Item fit issues for 24, 25 No disordered thresholds DIF for institution (items 14, 28), clinical educator gender (item 14) 1 misfitting person Local dependence multiple items
8	Delete item 24 (fit SD 3.720)	204.69, p<0.0001	0.916	1.62	0.98	Item fit issue for 25 No disordered thresholds DIF for institution (items 14), clinical educator gender (item 14) No misfitting persons Local dependence multiple items
9	Delete item 25 (fit SD 4.164)	155.73, p=0.032	0.908	1.56	0.98	Item fit issue for 26 No disordered thresholds DIF for institution (items 14), clinical educator gender (item 14) 5 misfitting persons Local dependence multiple items
10	Delete 5 misfitting persons	155.73, p=0.032	0.913	1.56	0.98	Item fit issue for 26 No disordered thresholds DIF for institution (items 14), clinical educator gender (item 14) No misfitting persons Local dependence multiple items
11	Delete item 26	148.91, p=0.037	0.905	1.53	0.95	No item fit issues No disordered thresholds DIF for institution (items 14), clinical educator gender (item 14) 6 misfitting persons Local dependence multiple items
12	Delete 6 misfitting persons	139.61, p=0.106	0.909	1.54	0.93	No item fit issues No disordered thresholds DIF for institution (items 14), clinical educator gender (item 14)

						No misfitting persons Local dependence multiple items
13	Delete item 14	133.91, p=0.109	0.906	1.56	0.93	No item fit issues No disordered thresholds No DIF No misfitting persons Local dependence multiple items
14	Delete item 17 (after subtest 17/18, r=0.567)	129.22, p=0.101	0.900	1.55	0.91	Item fit issue for 4 No disordered thresholds DIF for institution (item 13) No misfitting persons Local dependence multiple items
15	Delete item 3 (after subtest 2/3, r=0.487)	117.81, p=0.185	0.896	1.42	0.91	No item fit issues No disordered thresholds DIF for institution (item 13) No misfitting persons Local dependence multiple items
16	Delete item 13	106.72, p=0.304	0.890	1.31	0.91	No item fit issues No disordered thresholds No DIF No misfitting persons Local dependence multiple items
17	Delete item 6 (after subtest 2/6 r=0.325 & 4/6 r=0.343	103.15, p=0.266	0.886	1.23	0.90	No item fit issues No disordered thresholds No DIF No misfitting persons Local dependence multiple items
18	Delete item 11 (after subtest 10/11 r=0.202 & 11/12 r=0.251)	97.61, p=0.273	0.880	1.14	0.90	No item fit issues No disordered thresholds DIF for institution (item 28) No misfitting persons Local dependence multiple items

19	Delete item 29 (after subtest 29/30 r=0.374)	87.78, p=0.396	0.876	1.14	0.88	No item fit issues No disordered thresholds DIF for institution (item 28) 2 misfitting persons Local dependence multiple items
20	Delete 2 persons	87.78, p=0.396	0.878	1.14	0.88	No item fit issues No disordered thresholds DIF for institution (item 28) No misfitting persons Local dependence multiple items
21	Delete item 28	88.62, p=0.238	0.873	1.15	0.87	No item fit issues No disordered thresholds No DIF 2 misfitting persons Local dependence multiple items
22	Delete 2 persons	88.62, p=0.238	0.874	1.15	0.87	No item fit issues No disordered thresholds No DIF No misfitting persons Local dependence multiple items
23	Delete item 21 (after subtest 21/22 r=0.511)	96.36, p=0.04	0.864	1.33	0.84	No item fit issues No disordered thresholds No DIF 2 misfitting persons Local dependence multiple items
24	Delete 2 persons	96.36, p=0.04	0.866	1.33	0.84	No item fit issues No disordered thresholds No DIF No misfitting persons Local dependence multiple items
25	Delete item 8 (after subtest 8/9	73.60, p=0.360	0.859	1.17	0.85	No item fit issues No disordered thresholds

	r=0.202)					No DIF 1 misfitting person Local dependence multiple items
26	Delete item 4 (after subtest 2/4 r=0.207)	62.40, p=0.568	0.846	1.05	0.85	No item fit issues No disordered thresholds No DIF 3 misfitting persons Local dependence items 22 & 23
27	Delete 3 persons	62.63, p=0.560	0.844	1.05	0.84	No item fit issues No disordered thresholds No DIF No misfitting persons Local dependence items 22 & 23
28	Delete item 22 (after subtest 22/23 r=0.379)	70.78, p=0.160	0.826	1.24	0.85	No item fit issues No disordered thresholds No DIF 8 misfitting persons No local dependence
29	Delete 8 persons	65.26, p=0.298	0.827	1.18	0.82	p-value range 0.02 Lower CI 0.002-0.008 Upper CI 0.037-0.049

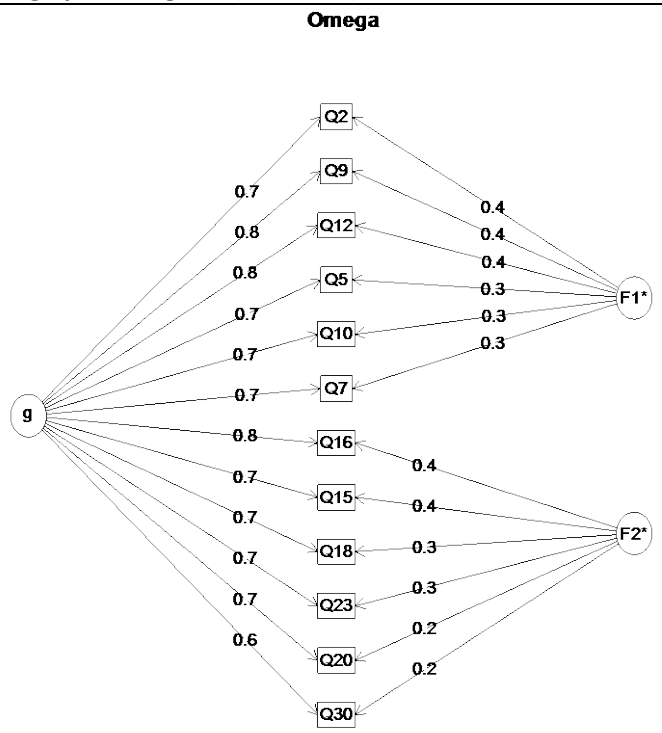
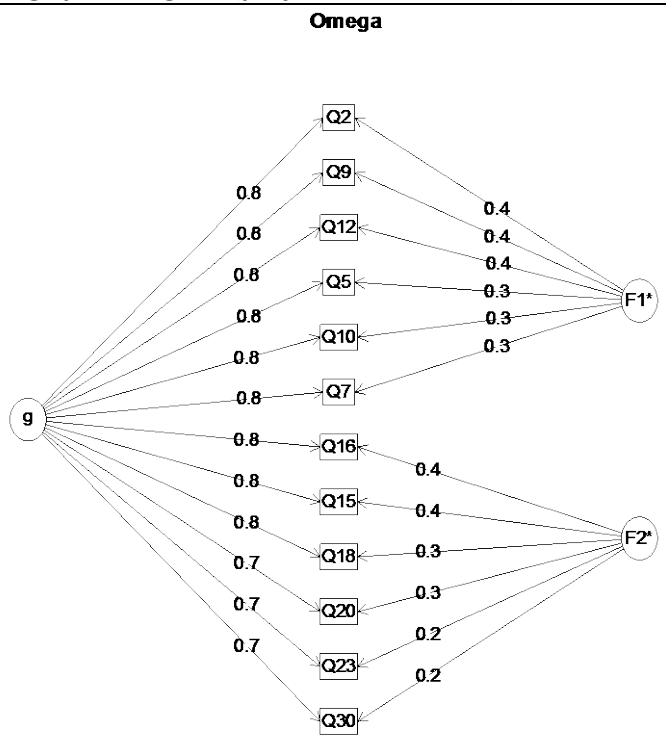
Item	Factor*	Description
1. Treated me with respect	LE	Disordered thresholds unable to be resolved through rescoring
2. Maintained a positive attitude towards me	LE	Retained
3. Fostered an environment of respect in which I felt comfortable participating	LE	Removed due to local dependence with item 2
4. Established a good learning environment (approachable, focused, nonthreatening, professional and enthusiastic)	LE	Removed due to local dependence with item 2
5. Demonstrated humanistic attitudes in relating to patients (integrity, compassion and respect)	LE	Retained
6. Was approachable for discussion	LE	Removed due to local dependence with item 2
7. Showed genuine concern for my professional well-being	LE	Retained
8. Had reasonable expectations of students	LE	Removed due to local dependence with item 9
9. Has good communication skills	LE	Retained

10. Is open to student questions and alternative approaches to patient management	LE	Retained
11. Gave me the opportunity to offer opinions on patient problems or treatment	LE	Removed due to local dependence with items 10 and 12
12. Adjusted teaching to my needs (experience, competence, interest)	LE	Retained
13. Is an effective clinical teacher	LE	Demonstrated DIF for university
14. Encouraged me to think	RP	Demonstrated DIF for both university & clinical educator gender
15. Promoted reflection on clinical practice	RP	Retained
16. Emphasises a problem-solving approach rather than solutions	RP	Retained
17. Asked questions that promote learning (clarifies, probes, reflective questions etc.)	RP	Removed due to local dependence with item 18
18. Asked questions to enhance my learning	RP	Retained
19. Encouraged questions and active participation	RP	Fit residual SD = -3.229

20. Stimulates me to learn independently	RP	Retained
21. Gave timely feedback to me	FB	Removed due to local dependence with item 22
22. Gave me regular, useful feedback about my knowledge and performance	FB	Removed due to local dependence with item 23
23. Offered me suggestions for improvement when required	FB	Retained
24. Identified areas needing improvement	FB	Fit residual SD = 3.720, χ^2=36.96, p<0.001
25. Identified my strengths	FB	Fit residual SD = 4.164, χ^2=36.56, p<0.001
26. Explained to me why I was correct or incorrect	FB	Fit residual SD = 3.179
27. Promoted keeping of medical records in a way that is thorough, legible, efficient and organised	PM	Fit residual SD = 6.938, χ^2=85.27, p<0.001
28. Encouraged me to assume responsibility for patient care	PM	Demonstrated DIF for university
29. Demonstrates knowledge of current medical and manual therapy literature	MD	Removed due to local dependence with item 30

30. Demonstrated osteopathic, clinical examination and rehabilitation knowledge and skill(s)	MD	Retained but rescored

*Factors identified in the paper by Vaughan (2015): LE – learning environment; RP – reflective practice; FB – feedback; PM – patient management; MD – modelling.

Omega path diagram (raw data)	Omega path diagram (polychoric correlation)
	

2. Maintained a positive attitude towards me
5. Demonstrated humanistic attitudes in relating to patients (integrity, compassion and respect)
7. Showed genuine concern for my professional well-being
9. Has good communication skills
10. Is open to student questions and alternative approaches to patient management
12. Adjusted teaching to my needs (experience, competence, interest)
15. Promoted reflection on clinical practice
16. Emphasises a problem-solving approach rather than solutions
18. Asked questions to enhance my learning
20. Stimulates me to learn independently
23. Offered me suggestions for improvement when required
30. Demonstrated osteopathic, clinical examination and rehabilitation knowledge and skill(s)

Scoring structure for the 12-item Osteopathy Clinical Teaching Questionnaire.

Item	Strongly disagree	Disagree	Neither agree nor disagree	Agree	Strongly agree
1. Maintained a positive attitude towards me	1	2	3	4	5
2. Demonstrated humanistic attitudes in relating to patients (integrity, compassion and respect)	1	2	3	4	5
3. Showed genuine concern for my professional well-being	1	2	3	4	5
4. Has good communication skills	1	2	3	4	5
5. Is open to student questions and alternative approaches to patient management	1	2	3	4	5
6. Adjusted teaching to my needs (experience, competence, interest)	1	2	3	4	5
7. Promoted reflection on clinical practice	1	2	3	4	5
8. Emphasises a problem-solving approach rather than solutions	1	2	3	4	5
9. Asked questions to enhance my learning	1	2	3	4	5
10. Stimulates me to learn independently	1	2	3	4	5
11. Offered me suggestions for improvement when required	1	2	3	4	5
12. Demonstrated osteopathic, clinical examination and rehabilitation knowledge and skill(s)	1	1	2	3	4

Raw score to Rasch score conversion for the 12-item Osteopathy Clinical Teaching Questionnaire.

Total score	Rasch score	Strata
12	0	1
13	1	1
14	1	1
15	2	1
16	3	1
17	4	1
18	6	1
19	8	2
20	10	2
21	13	2
22	15	2
23	18	2
24	21	2
25	24	3
26	27	3
27	30	3
28	33	3
29	36	3
30	39	3
31	42	3
32	45	3
33	48	4

Total score	Rasch score	Strata
34	51	4
35	53	4
36	56	4
37	59	4
38	61	4
39	64	4
40	67	4
41	69	4
42	72	4
43	74	4
44	76	4
45	78	4
46	80	4
47	82	4
48	84	4
49	86	4
50	88	4
51	90	4
52	91	4
53	93	4
54	94	4
55	96	4

Total score	Rasch score	Strata
56	97	4
57	98	4
58	99	4
59	100	4

APPENDICES

Chapter 4

Scoring structure for the 12-item Osteopathy Clinical Teaching Questionnaire.

Item	Strongly disagree	Disagree	Neither agree nor disagree	Agree	Strongly agree
1. Maintained a positive attitude towards me	1	2	3	4	5
2. Demonstrated humanistic attitudes in relating to patients (integrity, compassion and respect)	1	2	3	4	5
3. Showed genuine concern for my professional well-being	1	2	3	4	5
4. Has good communication skills	1	2	3	4	5
5. Is open to student questions and alternative approaches to patient management	1	2	3	4	5
6. Adjusted teaching to my needs (experience, competence, interest)	1	2	3	4	5
7. Promoted reflection on clinical practice	1	2	3	4	5
8. Emphasises a problem-solving approach rather than solutions	1	2	3	4	5
9. Asked questions to enhance my learning	1	2	3	4	5
10. Stimulates me to learn independently	1	2	3	4	5
11. Offered me suggestions for improvement when required	1	2	3	4	5
12. Demonstrated osteopathic, clinical examination and rehabilitation knowledge and skill(s)	1	1	2	3	4

Raw score to Rasch score conversion for the 12-item Osteopathy Clinical Teaching Questionnaire.

Total score	Rasch score	Strata
12	0	1
13	1	1
14	1	1
15	2	1
16	3	1
17	4	1
18	6	1
19	8	2
20	10	2
21	13	2
22	15	2
23	18	2
24	21	2
25	24	3
26	27	3
27	30	3
28	33	3
29	36	3
30	39	3
31	42	3
32	45	3
33	48	4

Total score	Rasch score	Strata
34	51	4
35	53	4
36	56	4
37	59	4
38	61	4
39	64	4
40	67	4
41	69	4
42	72	4
43	74	4
44	76	4
45	78	4
46	80	4
47	82	4
48	84	4
49	86	4
50	88	4
51	90	4
52	91	4
53	93	4
54	94	4
55	96	4

Total score	Rasch score	Strata
56	97	4
57	98	4
58	99	4
59	100	4