

Examiner judgement of professional competence in exercise physiology

Author:

Naumann, Fiona

Publication Date:

2014

DOI:

<https://doi.org/10.26190/unsworks/18252>

License:

<https://creativecommons.org/licenses/by-nc-nd/3.0/au/>

Link to license to see what you are allowed to do with this resource.

Downloaded from <http://hdl.handle.net/1959.4/54610> in <https://unsworks.unsw.edu.au> on 2024-05-02

Examiner Judgement of Professional Competence in Exercise Physiology

Fiona Naumann

Z3344727

Thesis submitted to the University of New South Wales Australia,
for the degree of Masters of Philosophy in Higher Education



School of Education

UNSW Arts & Social Sciences

May, 2014

PLEASE TYPE

THE UNIVERSITY OF NEW SOUTH WALES
Thesis/Dissertation Sheet

Surname or Family name: NAUMANN

First name: Fiona

Other name/s: Leigh

Abbreviation for degree as given in the University calendar: MPhil

School: Education

Faculty: Arts & Social Sciences

Title: Examiner Judgement of Professional Competence in Exercise Physiology

Abstract 350 words maximum: (PLEASE TYPE)

Exercise Physiology courses have transitioned to competency based, forcing Universities to rethink curriculum design and assessment to ensure students are competent to practice. This thesis explored the reasoning behind clinical judgment, specifically, capturing the factors that contribute to assessor decision making about a students' competency. **Aims:** The aims were twofold. The first was to determine the sources of variance in the examination process. The second was to develop a conceptual framework of the factors impacting on examiner judgment and ratings of student performance. **Methodology:** Examiner judgement was explored from both a quantitative and qualitative perspective. 23 examiners viewed three video encounters of student performance on an OSCE. Once rated, analysis of variance was performed to determine where the variance in ratings was attributed. A semi-structured interview was conducted to draw out the examiners reasoning behind their ratings. **Results:** At a global level, analysis of variance indicated that the individual examiner had a minimal impact on the variance, with the majority of variance explained by the student performance on the task. A closer examination of the three domains indicated slightly different outcomes. The assessment of procedural skills, mimicked the global assessment ratings. For communication competency, examiners had different expectations about what they expected to observe. For technical competency, the examiners and the University contributed to the variance. Using qualitative analysis, five themes were identified to help explain the variance in examiner judgement. These were variable frames of reference, the use of varying inferences, the variable reaction to language, the varying use of the patient experience as a reference and varying levels of confidence in performing the assessment. **Discussion:** This study highlighted the variability of the process of observation, judgement and rating. Each examiner brings with them a unique set of characteristics, values and expectations, which contributed to each examiner viewing student performance from different lenses. Whilst general assessment determined good and bad performance, there were subtle differences which could be addressed via the development of standards-based assessment and the development of examiner capability. **Conclusion:** This research was the first thesis to explore the role of the examiner in the assessment of entry level exercise physiology competence that was grounded in a theoretical framework of the processes examiners use to make judgement on student competency. The examiners role is pivotal in determining graduate level student competency and ultimately the calibre of entry level exercise physiologists.

Declaration relating to disposition of project thesis/dissertation

I hereby grant to the University of New South Wales or its agents the right to archive and to make available my thesis or dissertation in whole or in part in the University libraries in all forms of media, now or here after known, subject to the provisions of the Copyright Act 1968. I retain all property rights, such as patent rights. I also retain the right to use in future works (such as articles or books) all or part of this thesis or dissertation.

I also authorise University Microfilms to use the 350 word abstract of my thesis in Dissertation Abstracts International (this is applicable to doctoral theses only).


Signature


Witness

22.05.2014

Date

The University recognises that there may be exceptional circumstances requiring restrictions on copying or conditions on use. Requests for restriction for a period of up to 2 years must be made in writing. Requests for a longer period of restriction may be considered in exceptional circumstances and require the approval of the Dean of Graduate Research.

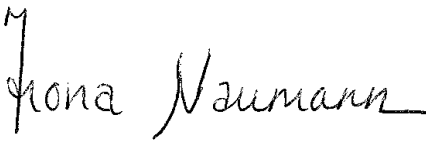
FOR OFFICE USE ONLY

Date of completion of requirements for Award:

THIS SHEET IS TO BE GLUED TO THE INSIDE FRONT COVER OF THE THESIS

ORIGINALITY STATEMENT

'I hereby declare that this submission is my own work and to the best of my knowledge it contains no materials previously published or written by another person, or substantial proportions of material which have been accepted for the award of any other degree or diploma at UNSW or any other educational institution, except where due acknowledgment is made in the thesis. Any contribution made to the research by others, with whom I have worked at UNSW or elsewhere, is explicitly acknowledged in the thesis. I also declare that the intellectual content of this thesis is the product of my own work, except to the extent that assistance from others in the project's design and conception or in style, presentation and linguistic expression is acknowledged.'

Signed 


Date 01.06.2014

COPYRIGHT STATEMENT

'I hereby grant the University of New South Wales or its agents the right to archive and to make available my thesis or dissertation in whole or part in the University libraries in all forms of media, now or here after known, subject to the provisions of the Copyright Act 1968. I retain all proprietary rights, such as patent rights. I also retain the right to use in future works (such as articles or books) all or part of this thesis or dissertation.

I also authorise University Microfilms to use the 350 word abstract of my thesis in Dissertation Abstract International (this is applicable to doctoral theses only).


I have either used no substantial portions of copyright material in my thesis or I have obtained permission to use copyright material; where permission has not been granted I have applied/will apply for a partial restriction of the digital copy of my thesis or dissertation.'

Signed 

Date 01.06.2015

AUTHENTICITY STATEMENT

'I certify that the Library deposit digital copy is a direct equivalent of the final officially approved version of my thesis. No emendation of content has occurred and if there are any minor variations in formatting, they are the result of the conversion to digital format.'

Signed 

Date 01.06.2015

Table of Contents

Originality Statement	4
List of Figures & Tables	5
Acknowledgments	6
Abstract	7

Chapter 1: Thesis Summary	8
---------------------------	---

- 1.1 Identification of the Issue*
- 1.2 Competence-based assessment*
- 1.3 Aims*
- 1.4 Approach*
- 1.5 Results*
- 1.6 Discussion*
- 1.7 Conclusion*

Chapter 2: Literature Review	17
------------------------------	----

- 2.1 Dimensions of Professional Competence*
- 2.2 Assessment*
- 2.3 Assessment of Clinical Competency*
- 2.4 Examiner Training Approaches*
- 2.5 Qualitative Analysis of Clinical Competency*

Chapter 3: Methodology	28
------------------------	----

- 3.1 Project Set Up*
- 3.2 Participants*
- 3.3 Qualitative Data Analysis*
- 3.4 Quantitative Data Analysis*

Chapter 4: Results	32
<i>4.1 Demographic Analysis</i>	
<i>4.2 Qualitative Analysis</i>	
<i>4.3 Quantitative Analysis</i>	
Chapter 5: Discussion	53
<i>5.1 Major Findings</i>	
<i>5.2 Implications from this Research</i>	
<i>5.3 Limitations</i>	
Chapter 6: Conclusions	62
Appendix A	63
<i>Appendix A: Information Sheet & Consent</i>	
Appendix B	67
<i>Appendix B: Station Task Descriptor</i>	
Appendix C	72
<i>Appendix C: Semi-Structured Interview Guide</i>	
References	73

List of Figures & Tables

Chapter 2

Figure 2.1 – Miller’s assessment levels.

16

Chapter 4

<i>Table 4.1: Demographics of Participants (n=23) of the study.</i>	27
<i>Table 4.2: Variance Estimates for the Rating of student performance</i>	44
<i>Table 4.3: Variance Estimates for the exercise physiology procedural skill competency rating.</i>	44
<i>Table 4.4: Variance Estimates for the exercise physiology procedural skill competency rating.</i>	45
<i>Table 4.5: Variance Estimates for the communication skill competency rating.</i>	45
<i>Table 4.6: Variance Estimates for the technical skill competency rating.</i>	45
<i>Table 4.7: Analysis of Variance between the Examiners from different Universities.</i>	46

Acknowledgements

To all of the academic, clinical educators and accredited exercise physiologists who participated in my research, none of this would have been possible without you. To the 3 students who volunteered to have their OSCE performance filmed and reviewed by examiners around the country. I would also like to thank the combined efforts of the Clinical Exercise Education Group (ExEd), for all of their support throughout the project. The assistance on this project and the field in general who have helped pave the way for future research and understanding in this emerging field.

To my primary supervisors, Professor Phil Jones, Professor Stephen Marshall and Associate Professor Boaz Shulruf, thank you so your guidance, support and insight over the past two years. You have helped me grow in this establishing field.

And finally to my family, Craig, Bianca and Lachie, who gave me time off from family duties, over many week-ends, in an attempt to get this research completed.

Abstract

Exercise Physiology courses have transitioned to competency based, forcing Universities to rethink curriculum design and assessment to ensure students are competent to practice. This thesis explored the reasoning behind clinical judgment, specifically, capturing the factors that contribute to assessor decision making about students' competency. **Aims:** The aims were twofold. The first was to determine the sources of variance in the examination process. The second was to document the factors impacting on examiner judgment and ratings of student performance. **Methodology:** Examiner judgement was explored from both a quantitative and qualitative perspective. 23 examiners viewed three video encounters of student performance on an OSCE. Once rated, analysis of variance was performed to determine where the variance in ratings was attributed. A semi-structured interview was conducted to draw out the examiners reasoning behind their ratings. **Results:** At a global level, analysis of variance indicated that the individual examiner had a minimal impact on the variance, with the majority of variance explained by the student performance on the task. A closer examination of the three domains indicated slightly different outcomes. The assessment of procedural skill, mimicked the global assessment ratings. For communication competency, examiners had different expectations about what they expected to observe. For technical competency, the examiners and the University they came from contributed to the majority of variance. Using qualitative analysis, five themes were identified to help explain the variance in examiner judgement. These were variable frames of reference, the use of varying inferences, variable reaction to language, varying use of the patient experience as a reference and varying levels of confidence in performing the assessment. **Discussion:** The major outcome from this study was that despite the documented differences in the examiner focus and judgement pathways, the ratings of student performance was relatively consistent amongst the examiners, with the differences in ratings largely attributed to the student's varied ability. This finding was unexpected, as the examiners had diverse backgrounds, roles and experience. The thought processes behind judgement were diverse and if the qualitative results had been used in isolation, may have led to the researchers drawing conclusions that the examined performances would have yielded widely different ratings. A strength of the current study was that by combining the use of quantitative and qualitative approaches, the researchers were able to gain a more comprehensive picture of the assessment rational and impact on the ratings. **Conclusion:** This research was the first to explore the role of the examiner in the assessment of entry level exercise physiology competence. The examiner role is pivotal in determining graduate level

student competency and ultimately the calibre of entry level exercise physiologists. This study highlighted the differences in the examiners judgement pathways but confirmed that the differences only had a small impact on the student ratings and that it was in fact the students' ability that accounted for the variance in rating.

Chapter 1: Thesis Summary

1.1 Identification of the Issue

It is well established that the cost associated with chronic disease are placing a substantial burden on the Australian healthcare system. It is also apparent that appropriate exercise interventions can assist in disease management and reduce these costs (ESSA, 2008). In 2006, Exercise Physiology was formally recognised by the Australian Government as an allied health profession, with the task of designing and delivering exercise, to both apparently healthy persons and to those with chronic and complex conditions, to better manage their health and well-being (ESSA, 2008). In response to the recognition as a new allied health profession, Exercise Physiology courses in Australia transitioned to a competency based degree, forcing Universities to rethink curriculum design and assessment practices. Clinical teaching and learning was subsequently required to play a significant role in Exercise Physiology courses, providing students with an opportunity to develop and refine skill competencies that will assist them prepare for professional practice.

In order to become accredited as an Exercise Physiologists, students are required to complete four years of University training that covers the core knowledge and skill competency to practice in the profession. An integral part of this training is to complete 500 hours of training in the field. It has never been questioned if the 500 hours ensures all graduates are competent to practice in the field. It remains to be seen whether the profession could be better served by assessing student competency and de-emphasising training compliance. In order to provide a greater emphasis on assessing competency, the profession needs to clearly articulate the essential competencies of practice, ensure competency can be assessed and that the students are aware of the assessment criteria and processes. However, to date there has been no exploration of the clinical competency assessment process in exercise physiology. We require supervisors in the field (largely AEPs), clinical academics and academics to assess students, with little regard to how well they are doing. As assessment is complex and such a vital part of ensuring professional standards are maintained, it is important that a critical evaluation of the assessment process be conducted. As the profession is in its infancy, there is a large scope for improving and consolidating the preparation of the next generation of Accredited Exercise Physiologists (AEPs).

In response to a need to better understand the clinical competency assessment process, this thesis explored the reasoning behind clinical judgment, specifically, capturing the factors that contribute to the assessor decision process with regard to a students' level of competency and a subsequent analysis of the source of variance examiners contribute to the clinical assessment process. Such an exploration will provide the exercise physiology profession with feedback on current practice in clinical assessment and provide possible directions for improving the assessment process and student training.

1.2 Competence-based assessment

Assessment is a critical component of the educational paradigm, whereby teaching, learning and assessment need to be aligned to optimise learning outcomes. It can be argued that assessment is one of the most important aspects of education, with the inherent feedback playing a role in shaping and motivating future learning (Westwood et al. 2013). Competence-based assessment is an integral part of determining clinical competence and readiness to practice in the medical and allied health professions. Competence-based assessment is a form of assessment derived from a specified set of outcomes and standards, which enable assessors to make an objective judgement of whether a student has achieved the standards and ultimately, ready to practice in the profession (Wolf, 2001). However, assessment of clinical competency is complex, particularly as it relies on human judgement to determine the outcome. For this reason, this thesis seeks to contribute to the research field by gaining a better understanding of the complexity of human judgement as it pertains to clinical competency assessment. This study is positioned in an emerging field as the exercise physiology profession seeks to formalise its assessment processes and promote a better understanding of examiner judgement and the implications of the assessment process for student teaching and learning.

One of the challenges underpinning clinical assessment is defining the term competence. By way of definition, clinical competence is a multifaceted and dynamic concept that includes core knowledge, understanding of knowledge, clinical skills, communication skills, interpersonal skills, problem solving, clinical judgement and technical skills relevant to the profession (Norman, 1985). However, what level of student performance is indicative of competent is a contentious issue, (Watson et al. 2002). Even more pressing is that the examiner assessment of student competency relies on an examiner making a professional

judgement. However, how the examiners make this judgement requires further exploration, particularly as to how the judgement process influences the ratings of student performance.

The Objective Structured Clinical Examination (OSCE) is a standardized method used for assessing clinical competence (Harden et al. 1975). The reliability and validity of OSCEs depend on the rigour of the examination development, the core competencies they assess, the assessment criteria developed and the examiners application of the criteria to formulate a judgment of performance. The rigour of the examination was explored previously (Naumann et al. 2014), concluding that the OSCE was a reliable and valid means to assess student competency in exercise physiology. However, one aspect of the examination process that was not explored in any depth was the role of the examiner or assessor in the examination process. As performance assessments rely on human judgement, these judgements are vulnerable to rater effects and warranted further investigation.

In research by Kogan et al. (2011), the authors suggested the process of examining was not neat, predictable or straightforward and required thorough consideration, specifically what influences' the assessors' decision making process in rating student competency. Given the significance of the examination process, there has been no research in exercise physiology profession on the way examiners make judgement about the quality of student performance. We lack information on whether or not assessment is consistent across Universities or if there are varying levels of confidence in the examiners who are being tasked to perform the assessments. In such a new profession, these questions are important to provide a better understanding of the process of clinical judgement, with a view to ensure greater clarity, transparency and better standardised approaches to assessment. A review of such practices could also contribute to the development of professional capabilities as an assessor, all of whom could share a common understanding of the assessment criteria, standards and processes of clinical judgment. And finally, it could enhance students' preparation for the assessment process and better align teaching and learning to the clinical outcomes of practice.

1.3 Aims

The specific aims of the thesis were twofold. The first aim was to develop an understanding of the factors impacting on examiner judgment and rating of student performance. The second aim was to determine the sources of variance in the ratings of exercise physiology students performing a pre-determined clinical scenario (Exercise Physiology OSCE station).

The two research questions were:

1. What factors influence examiner judgment of professional competence of exercise physiology students?
2. Where does the variance exist when rating an exercise physiology students' performance on an OSCE station (clinical scenario)?

1.4 Approach

This project was underpinned by a social constructionist epistemology. Hence, knowledge and meaning were co-constructed rather than discovered, and the inter-subjectivity of the researcher and research participants was acknowledged. Within this approach, true objectivity was neither the aim of the research nor even considered possible; rather, the project aimed to achieve a deeper understanding of possible constructions of meaning within the data.

To address the aims, examiner judgement of exercise physiology student performance was explored from both a quantitative and qualitative perspective. The qualitative perspective explored the examiner thought process, providing the researchers with the reasons behind the judgement. Such an approach enabled researchers to explore the examiners internal logic and make better sense of their judgement. The quantitative approach then enabled the researchers to explore what factors contributed to the variance in the examiners ratings. The sources of variation included the examiners or raters, the student performance on task, and the three domains of competency: communication, procedural and technical competency. G studies were selected as a method as it enabled researchers to identify the source of variation. Ideally the examiners should have less of an impact on the ratings, with the student performance on the task being the major contributor to the student's rating by the examiner.

1.5 Results

Using qualitative analysis, five themes were identified to help explain the varied approaches to examiner judgement.

1. The first theme identified was that examiners used variable frames of reference. In order to make a judgment about the student's performance, the examiner needed a

standard to judge against. The standard was either themselves as a practitioner, themselves as an academic lecturing on the topic or the other students being examined.

2. A second theme was examiners made varying inferences whilst examining. Inferences were often made based on what the student did and what that behaviour meant, however these inferences were never confirmed. If the student was nervous, fumbled over their words or talked to themselves they were viewed as less professional and ultimately less competent. On a positive note, inferences were not made about behaviour not observed. If a student failed to complete all aspects of the assessment in the time allocated, the examiner did not assume they would have completed the tasks at a similar standard.
3. A third theme was the examiners had varying reactions to communication, specifically the student's language choice and the way the student and client interacted during the consultation. Some examiners noted that if the language was casual, it was perceived as unprofessional and the examiners perception of the student's ability dropped. Excellence in choice of words was also noted by some, attached to a belief that the patient was in more capable hands. In other instances, examiners failed to notice language choice. The interaction between the student and the patient was deemed very important by several examiners, including the ability of the student to listen to the client and respond accordingly.
4. A fourth theme that emerged was the examiners varying use of the patient outcome as a reference. Some examiners viewed the patient's response to the interaction and made judgement on whether they believed the patient had a positive or negative experience or encounter. The patient experience was not confirmed post assessment and could not be verified. Other examiners failed to make note of the patients' experience in the encounter.
5. A fifth theme that emerged was varying levels of confidence by the examiners in performing the assessment. Some examiners were confident in their ability to examine and the ratings they awarded. Other examiners were less than confident and after exploring their ratings with the researcher, wanted to adjust them.

Despite the varying reasons behind the ratings, Generalisability theory analysis indicated that the examiner made little contributed to the variance in ratings. When the examiners rating, the three domain ratings (communication, procedural & technical) and the student

performance were included in the analysis, it was in fact the student who was the biggest source of variance in the competency ratings. IN other words, if a student performed poorly, the examiners had good agreement that the performance was poor, despite the varying factors noted by the examiners to reach that judgment. At a global level, for communication and for the exercise physiology procedural skills, analysis of variance indicated that the individual examiner had a minimal impact on the variance in ratings (8-9%), with the majority of variance explained by the student performance on the station task (63-67%).

Unlike the other assessment domains, the examiners did contribute to the variance in the technical skill competency rating (24%), indicating the examiners interpretation of the technical aspects of the task contributed more to the variance in ratings than the student. A further analysis was conducted to compare the assessment results between examiners based on their home University. When comparing the means for global competency, exercise procedural and communication skills, there was no statistical difference between the examiners. However, there was a significant difference in how the examiners from different Universities judged technical skill.

1.6 Discussion

The major outcome from this study was that despite the documented differences in the examiner focus and judgement pathways, the ratings of the student performance was relatively consistent amongst the examiners, with the difference in ratings largely attributed to the student's varied ability. This finding was unexpected, as the 23 examiners in the study had diverse backgrounds, with a mixture of University academics, practicing AEPs and clinical educators, most of whom had little experience in examining clinical scenarios. The thought processes were diverse and if the qualitative results had been used in isolation, may have led to the research team drawing conclusions that the examined performances would have yielded widely different ratings. A strength of the study was that by combining the use of quantitative and qualitative approaches, the researchers were able to gain a more comprehensive picture of the assessment process and its outcomes.

Utilizing the general inductive approach, this study identified several factors influencing how examiners rate student competency during a standardised clinical scenario and highlighted the variability of the process of observation, judgement and rating. Each examiner brings with them a unique set of characteristics, values and expectations. This contributes to each

examiner viewing student performance from slightly different lenses. Despite the varying perspectives, the examiners as a cohort were still able to agree what constituted a good or poor performance. Differences in competency ratings were therefore, largely attributable to the student themselves. In other words, if a student performed poorly, most examiners agreed it was a poor performance.

One important anomaly was noted in the technical skill assessment, with the examiner having the highest contribution to the variance in the ratings. Further analysis also indicated that the University from where the examiner came from had a significant impact on how the student was rated. This indicated that the examiners from different Universities had varying views of what was acceptable technical competency and points to a need to better clarity as to what is an acceptable standard for the profession. Currently, exercise physiology lacks clear standards by which to judge student competency. Establishing competency frameworks and standards-based assessment for graduate exercise physiologists will be an important process to ensure all students are judged according to similar standards. These standards need to be clearly articulated to all examiners and students being assessed.

Another theme that arose from the qualitative research was that examiners varied in their confidence to rate. Once the ratings were explored in the interview with the researcher, some examiners wanted to modify their ratings and expressed value in providing mentors to assist them in being an examiner. This reflected the varying experience in assessing and the fact that the profession is still in its infancy and could potentially benefit from developing the examiner capabilities to assess student competency.

This thesis advances knowledge in the field of examiner judgement by exploring the factors that influenced examiner judgment of exercise physiology student competency and lays important foundations for future work in improved clarity of the clinical examination and assessment process. A greater understanding of these factors will assist the students understanding of the assessment criteria for exercise physiology clinical competence.

An awareness of the assessment process could also inform and enhance examiner preparation and the development of assessor capabilities. Currently the profession relies on practising AEPs and academic staff to both mentor and assess exercise physiology student competency, with little regard to the assessment capabilities of these staff. Development of the assessor capabilities is an important one for the profession, with an acknowledgment of the important

role the assessors play in determining the readiness of the next generation of AEPs to practice in the field. The ultimate implications are that as a profession, it needs to work towards a more standardised approach to clinical assessment, ensure students are aware of the assessment criteria and better development of assessor capabilities to ensure that competent exercise physiology graduates are ready to practice safely and competently in the profession.

1.7 Conclusion

This research was the first thesis to explore the role of the examiner in the assessment of entry level exercise physiology competence exploring the processes examiners use to make judgement on student competency. This research contributes to a shared understanding of the examination process from the examiners frame of reference and presents the foundations for future work in the development of a more transparent approach to clinical assessment in exercise physiology. The examiners role is pivotal in determining graduate level student competency and ultimately the calibre of entry level exercise physiologists.

Chapter 2: Literature Review

Exercise Physiology is the clinical modality dealing with the use of exercise in the treatment of patients with chronic disease and injuries. The current scope of practice of exercise physiology demands that graduates be trained in cardio-respiratory, neurological, muscular, metabolic, mental health and oncology fields (ESSA, Scope of Practice, 2008). The role of an exercise physiologist is to assess, prescribe and deliver exercise to assist clients manage complex injuries of chronic disease states. In order to prescribe exercise, the exercise physiologist must have a complex understanding of the injury or chronic disease and be able to take a comprehensive patient history to better understand the clients' condition. Exercise Physiology training and evaluation is now moving from the traditional knowledge-based focus to one of competence-based training and assessment. By assessing clinical competence it enables Universities and the professional body to determine readiness to practice, provide feedback and on-going motivation and direction for future learning and to judge the adequacy of student training. This chapter explores nature of competence-based assessment as it pertains to the exercise physiology profession.

2.1 Dimensions of Professional Competence

One of the challenges underpinning clinical assessment of competence is defining the term competence. By way of definition, clinical competence is a multifaceted and dynamic concept that includes core knowledge, understanding and application of knowledge, clinical skills, interpersonal skills, problem solving, clinical judgement and technical skills relevant to the profession (Norman, 1985). By way of definition, clinical competence is the degree to which an individual can use the knowledge, skills and judgement associated with the profession to perform effectively within the scope of professional practice (Kane, 1992). Epstein & Hundert (2002) proposed their own definition of professional competence to be the habitual and judicious use of communication, knowledge, technical skills, clinical reasoning, emotions, values and reflection of daily practice for the benefit of the individual and community being served.

In medical and the allied health care professions, competencies are used to define discipline specific standards and expectations of performance in practice. Competence-based assessment is also an important part of clinical training to ensure professional standards are met and upheld. In response to the establishment of Exercise Physiology as an allied health profession, these professional standards are in the developmental stage. Exercise Physiology

training and evaluation is now moving to training and assessment based on competencies such as clinical skills, attitudes, beliefs, management, communication and professionalism, but is very much in its infancy.

The set of competencies for exercise physiologists in Australia were first developed and published in 2008 (Selig et al. 2008). The key competencies of an Accredited Exercise Physiologists include professional behaviour, procedural skills, technical skills and communication skills. Specifically, the procedural role of an exercise physiologist is to conduct a client consultation, documenting a detailed patient history in a logical and detailed manner and use this information to prescribe and deliver exercise to assist clients manage complex injuries of chronic disease states. In order to prescribe exercise, the exercise physiologist must have a complex understanding of the injury or chronic disease and be able to recognise the implications of the patient history for exercise prescription. They also need to be able to review medical reports and perform a baseline health and fitness assessment and utilise the results to inform exercise prescription. Once a program is written, the exercise physiologists must work closely with the client to deliver and monitor the exercise program, ensuring the client maintains a level of compliance and motivation towards the exercise program. This competency extends beyond an exercise prescription and captures the important role that education and facilitating behaviour change have in the overall success of any clinical exercise physiology intervention (Selig et al. 2008).

Ongoing assessments are then fed back to the patient and GP, outlining the clients' response to the exercise program. Throughout the encounter, exercise physiologists are required to demonstrate professional integrity and ensure the well-being for all clients. An exercise physiologist is also required to demonstrate technical competence through an ability to correctly use all pieces of equipment for health and fitness assessments and exercise delivery. This competency also takes into account that the clinical exercise physiologist employs a wide range of assessment tools and procedures and performs their role as part of a wider healthcare team (Selig et al. 2008). Finally an exercise physiologist must demonstrate communication competence including verbal and non-verbal communication, an ability to develop excellent patient rapport, demonstrate empathy, an ability to facilitate informed client decision making and effective communication working within a team of professionals. This competency is about using effective communication styles and techniques to allow the

clinical exercise physiologist to successfully deal with a diverse range of individuals and groups across the scope of clinical exercise physiology practice (ESSA Scope of Practice, 2008).

The current scope of practice of exercise physiology demands that graduates be able to apply these skill competencies in areas such as cardio-respiratory, neurological, muscular, metabolic, mental health and oncology (ESSA, Scope of Practice, 2008). The 2008 scope of practice of an Accredited Exercise Physiologist is currently under review, with an updated scope of practice due for release in 2015. Assessment of the exercise physiology professional, procedural, technical and communication skills is essential for professional continuance.

2.2 Assessment

Assessment is a process by which information is observed against a known objective or outcome (Hodges & McNaughton, 2009). It is a judgement or appraisal of someone's ability and allows the examiner to make a decision about the learners' current level of knowledge and behavioural skills (Westwood et al. 2013). Well-designed assessment is the key to engaging students in active and productive learning. To promote effective learning, assessment should be educational, formative, aligned with the teaching outcomes, provide quality feedback to students, allowing them to build on their knowledge and skill set (Wass et al. 2001) and be a true measure of student academic achievement (Sadler, 2010). Assessment also needs to be reliable, valid, transparent, promotes relevance, motivate students to engage in a deep learning approach, encourage students to think critically, and have a positive effect on their learning (Struyven et al. 2005). On the opposite spectrum, poor assessment is brought about when students sense a lack of control, feel the effort required does not match the perceived grade or reward, perform irrelevant tasks, and are stimulated to engage in poor learning practices such as rote and surface learning (Struyven et al. 2005).

Assessment can provide either formative or summative evaluations, depending on the timing and purpose of assessment (Boud, 2000). Formative assessment provides feedback to students and can be instrumental in guiding the learner's participation in the educational process (Boud, 2000). Summative assessment, on the other hand, provides an overall grade or rating, signifying the overall achievement of the student (Boud, 2000). Assessment must take into account what is assessed, how it is being assessed, the usefulness of the assessment in fostering future learning, inspire confidence in the learner, foster habits of self-reflection, self-

remediation and drive institutional self-assessment and curricular change (Epstein & Hundert, 2002).

A fundamental change in the view of assessment is represented by the notion of 'assessment as a tool for learning' (Dochy & McDowell, 1997). In the past, assessment was primarily seen as a means to determine grades; to find out to what extent students had reached the intended objectives. Today, there is a realisation that the potential benefits of assessing are much wider and impinge on all stages of the learning process. Research findings suggest that students' perceptions about assessment, has considerable influences on students' approaches to learning (Struyven et al. 2005).

2.3 Assessment of Clinical Competency

Within the scope of assessment, lies the assessment of clinical competence. Competency is defined as an observable ability of a health professional, integrating multiple components such as knowledge, skills, values and attitudes (Frank et al. 2010). Since competencies are observable, they can be measured and assessed to ensure their acquisition. Assessment of clinical competency is based on the notion that the competence required to practice within the profession, must be assessed prior to becoming accredited. Such assessment enables the institutions and profession to make decisions about whether a student graduate has attained a minimum level of competency and is ready to practice safely within the profession. Currently, accreditation in Exercise Physiology is linked to compliance of hours of training and a competency report from the clinical supervisor, charged with mentoring the student on clinical placement (Selig et al. 2008).

A framework for the assessment of clinical competency has been described by Miller (1990), which distinguished four levels at which the learner can be assessed. The different layer's in Miller's model represent a developmental sequence of stages of clinical competence development. Knowledge of what is required to carry out professional functions effectively is at the base of the framework and action to perform in day to day practice is at the top. All levels are needed and have their own impact on clinical competence. The first level is what the student knows (basic facts), traditionally assessed through written tests and essays. The second level is what the student knows how to do (applied knowledge), within the clinical context. The third level includes the student showing how to do something, typically in the format of a performance based assessment. It is also the level that is referred to as

competency-based assessment. And finally, the forth level or summit of the pyramid is evidenced by what the student does in practice (actual performance). Since the adoption of performance-based frameworks, emphasis has shifted towards what a student can do rather than what they know. The OSCE is one modality that is well placed to enable students to demonstrate what they can do, represented by the Miller's shows how level.

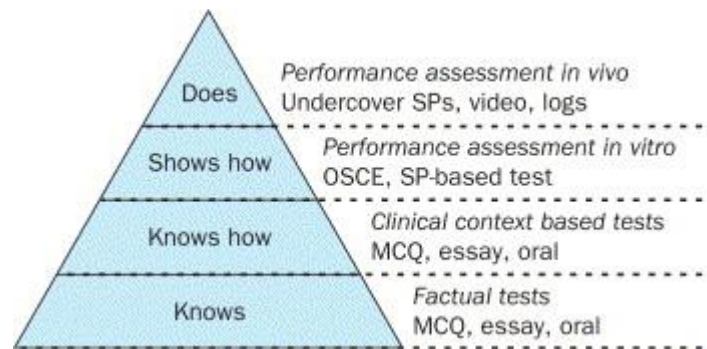


Figure 2.1: Miller's assessment levels.

The goal of Exercise Physiology programmes nationally, is to ensure that their students on graduation will be competent beginning practitioners in the discipline of exercise physiology, able to practice across multiple contexts. Tests of clinical competence, which allow decisions to be made about exercise physiology qualification and fitness to practise, must be designed with respect to key issues including development, validity, reliability, and standard setting, as well as clarity about their formative or summative function (Wass et al. 2001). The distinguishing features of work-based learning and assessment situations are that they are inherently variable, unpredictable, high-risk learning events that are not replicable. This varying environment, combined with varying experience of the supervisors to mentors and assess, presents challenges for the quality assurance of the assessment process (Cooper, Orrell & Bowden; 2003, Hodges, 2011; Yorke, 2011). As a consequence, there is a need to simulate the practice environment as much as possible during assessment, to assess student's skills in an authentic manner, allowing the examiner to view what they can do (Miller's Shows How level).

Competence-based assessment need to be derived from a specified set of outcomes or criteria and standards, enabling assessors to make an objective judgement on whether a student has achieved or not achieved the outcome (Wolf, 2001). Often competence-based assessment includes pre-defined criteria that enable the examiner to form a judgement. Criteria are attributes that are useful as levers for making judgments and are essential when attempting to explain particular judgments (Sadler, 2005). In addition are standards, which are fixed

reference levels of attainment. In order to have high quality assessment of skill competency, both criteria and standards need to be clearly defined and applied consistently (Sadler, 2005).

An extensive review of the competency-based literature identified four overarching themes: a focus on curricular outcomes that ensure all graduates demonstrate competence in all domains of their intended practice, an emphasis on abilities, a de-emphasis on time-based training, with a shift in favour of developing the learners' abilities and the promotion of learner-centeredness (Frank et al. 2010).

One potential method for assessing clinical competence in the work-simulated setting is the Objective Structured Clinical Examination (OSCE). The OSCE was first described by Harden et al. (1975) as a means to assess the clinical skills of final year medical students and has been used in the assessment of student clinical competence for over three decades. The OSCE aims to enhance the validity of clinical assessment by simulating realistic clinical scenarios to reflect real-life professional tasks (Harden, 1975). Typically, an OSCE consists of a series of stations in which students are required to perform a range of clinical tasks to demonstrate competency in relevant skills. During the OSCE, students rotate through the various stations spending a predetermined time at each station to complete the described task. Prior to developing an OSCE, it becomes necessary to identify the key competencies of the profession and develop OSCE stations that present a standardized portrayal of typical clinical scenarios. Once the stations are developed, establishing reliability of the assessment is critical. The underlying premise is that standardised assessment procedures ensure objectivity and maximise reliability (Bartfay et al. 2004). Since its development, the OSCE has gained acceptance as a benchmark for assessing clinical competence (Bartfay et al. 2004) and has been adopted by several health professions including Radiography (Marshall and Harris, 2000), Nursing (McKnight et al., 1987; Alinier, 2006), Physiotherapy (Nayer, 1999) and Dentistry (Brown et al. 1999). In order for the expansion of OSCEs use within other allied health professions it is important that each discipline explore the validity and reliability of this strategy, to their own satisfaction. The validity and reliability of the OSCEs use in Exercise Physiology has been recently published (Naumann et al. 2013).

The reliability and validity of OSCEs depend on four important factors:

- 1) the rigour of the examination development,
- 2) the competencies they assess,

- 3) the assessment criteria developed and
- 4) the examiners application of the criteria to formulate a judgment of performance.

Just as key competencies to be examined require careful attention, so do the issues surrounding assessment validity and reliability. Validity focusses on whether a test actually succeeds in testing the competencies it is designed to test (Downing, 2004). Reliability on the other hand is a measure of the reproducibility or consistency of a test over time, and is affected by many factors including station design, candidate nerves, expectations attached to the standards and examiner judgement (Downing, 2004). For a reliable measure of clinical skills, performance has to be sampled across a range of patient problems. This is the basic principle underlying the development of objective structured clinical examinations (OSCEs).

2.3 Assessor Judgment

Assessment of exercise physiology trainee clinical competence remains central to clinical skill assessment. However, the accuracy of the assessment process needs careful attention to ensure accurate and reliable assessment outcomes. Inter-rater agreement is determined by having different raters evaluate the same performance. Under these circumstances, differences in ratings reflect differences in rater focus (aspects of performance observed), standards (stringency and leniency), expectations of acceptable clinical performance levels and application of the assessment criteria (Williams et al 2009). Clinical assessment often suffers from poor accuracy, poor reliability and rating error, with the raters themselves being a source of error (William et al. 2003; Downing, 2005; Roch et al. 2012). Assessors often give learners the benefit of the doubt and fail to fail incompetent performance, unless there is very clear evidence of unsafe practice (Watson et al. 2002).

In an ideal situation, an examiners ability to assess a students' performance would be calibrated against an accepted gold standard. Unfortunately, such a gauge does not exist and makes the task of examining, a difficult one. Clinical assessment needs to be supported through the provision of a rich source of information to guide the assessors' judgment (Jones, 2000). The goal is to ensure examiners are consistent in how they distinguish between various scores, emphasizing the precision of the scoring process (Hefti & Preshaw, 2012). The primary issue in ensuring reliability in examiners becomes how to support quality judgment in assessment and ensure a certain level of inter-rater agreement.

Quality judgement cannot occur in the absence of clearly defined competencies and without clearly defined threshold standards for performance. At the same time, it is important to allow for the expression of professional judgment and not over specify the competencies by resorting to checklists which trivialize and neglect the holistic nature of the competency (Norman et al. 1991; Jones, 2000). Using global judgment that focus on the global competency have been found to improve reliability and validity in ratings based performance in health education (Cox, 2000; Epstein & Hundert, 2002). Global judgement is often grounded in tacit knowledge and professional experience. Tacit knowledge refers to the expertise people carry around with them in their various academic and professional fields (Sadler, 2005). Tacit knowledge commonly exists in an unarticulated form but can be shared amongst experts, transmitted from expert to novice and importantly, be used in making judgments about student competence. Whilst using tacit knowledge alone would prove challenging when providing transparent and consistent assessments, they need to be acknowledged in the contribution to examiner judgment.

An important factor contributing to higher rater agreement is when raters collaborated in developing the rating scales and had a shared definition of competent clinical performance (Williams et al. 2009). Frame of reference training programs have also been demonstrated to enhance rater agreement. These programs typically provide raters with samples of behaviours representing each level of the rating scale (Williams et al. 2009). At a minimum, all raters need to be familiar with the rating criteria at the beginning of the examination and be able to apply them across the full range of performance standards.

The reality, however, is that not all assessors examine in a consistent manner. Having an awareness of examiner bias and differences, may be one step towards reducing error. Bias occurs when examiners disagree in their interpretation of the scoring system, have different standards that relate to leniency or stringency, demonstrate a central tendency error, contrast errors, demonstrate a halo effect or have deficiencies in their own competencies that may potentially compromise their ability to detect errors in the student performances (Hefti & Preshaw, 2012). The halo effect describes the error that occurs when the rater attaches excessive significance to a single factor of performance. This leads to similar ratings on the other performance elements, regardless of how they actually performed. This can occur in both a positive or negative bias (Downing, 2004). Central tendency error describes the systematic reluctance to rate at the low extreme or the high extreme of the scale, so that the

majority of the scores cluster in the middle, regardless of performance (Hefti & Preshaw, 2012). Contrast error occurs when a rater allows an extremely weak or strong student to become the benchmark by which the other student performances are evaluated against (Hefti & Preshaw, 2012). A final factor which may impact the ratings is whether the examinee is known to the examiner and whether an emotional response is evoked that could impact judgment (Albanese, 2000).

Where assessments involve rating a student performance on a clinical task, human judgement is central to the process. As assessments often have significant outcomes for students, understanding the decision making process behind judgement is important for the quality of assessment. As examiner clinical judgement is a complex process, this thesis plans to investigate examiner judgement of exercise physiology performance from a qualitative and quantitative perspective.

2.4 Examiner Training Approaches

Given the important role of examiners in the evaluation of clinical competencies, Universities and the professions must train examiners in this important task, if consistency, reliability, validity and transparency are to be achieved. An assessment of clinical competency is only as good as the individual judging it. Research in medical education has demonstrated that various approaches to training examiners, can improve the examination process (Holmboe, 2004). The first of these approaches include behavioural observation training, focussed on improving the detection, perception and recall of actual performance (Heinemann & Wesley, 1983). A second approach to improving examiner performance is performance dimension training, designed to teach and familiarise examiners with the appropriate performance dimensions (Holmboe, 2004). Performance dimension training incorporates a review of the definitions and criteria for each dimension of competency. Once examiners are familiar with the criteria, the next step is to get the examiners to interact with the definitions and criteria and reach a high level of consensus. A third approach to improving examiner rating is frame of reference training (Holmboe, 2008). Frames of reference training, focusses on the application of the performance criteria to distinguish levels of performance. Holmboe (2008) described four levels of performance, ranging from unsatisfactory, borderline, satisfactory, through to superior. It was recommended to establish the minimum criteria required for a satisfactory performance, which provide an anchoring point from which the other levels of performance can be determined.

2.5 Qualitative Analysis of Clinical Competency

Qualitative research acknowledges the contextual nature of inquiry and focuses on interpretation of phenomena in their natural settings to make sense in terms of the meanings people bring to these settings (Denzin & Lincoln, 1984; Glesne & Peshkin, 1992). Its intellectual roots extend well over a hundred years to the birth of social science itself (Tesch, 1990), and it is utilized in contemporary social research in the fields of education, sociology, nursing, social work (Denzin & Lincoln, 1984). Qualitative research enables researchers to “get close” to participants, to “penetrate their internal logic and interpret their subjective understanding of reality” (Shaw, 1999, p. 60). It is integrated with desire to gain a fresh slant on phenomena that are difficult to convey with quantitative methods (Strauss and Corbin, 1990).

The qualitative researcher begins by observing phenomena and continues to find patterns in the form of themes, categories, concepts and typologies that emerge. Tentative hypothesis are introduced and additional information are collected to explain the phenomenon. The purpose of understanding is not necessarily to predict what might occur, but rather to understand in depth the characteristics of the situation and the meaning brought by participants and what is happening to them at the moment (Patton, 1996). The great contribution of qualitative research is the culturally specific and contextually rich data it produces.

There is a vast array of literature documenting the various assumptions and procedures associated with analysing qualitative data. These approaches include grounded theory (Strauss & Corbin, 1990), phenomenology (van Manen, 1990), narrative analysis (Leibicj, 1998) and discourse analysis (Potter & Wetherall, 1994). To explore the process of clinical judgement, a general inductive approach can be utilized. The general inductive approach involves a detailed set of procedures for analysing and reporting qualitative data (Thomas, 2006). Researchers using the general inductive approach typically limit their theory building to the presentation and description of the most important categories (Cresswell, 1998). The intended outcome of the process is to create a small number of summary categories, ideally between three and eight categories, that in the evaluator’s view capture the key aspects of the themes identified and are assessed to be the most important themes given the evaluation objectives. Inductive coding that finishes up with more than eight major themes, some of the categories may need combining, or the evaluator must make a decision about which themes or categories are most

important. The findings are often presented as a description of the most frequent, dominant and significant themes inherent from the raw data.

The three purposes underlying the use of the general inductive analysis approach is to:

- i) Condense extensive and raw data into a summary format
- ii) Establish clear links between the research question and the summary findings
- iii) Develop a model about the experiences and processes that emerge from the data.

This approach will be used in the current thesis to analyse examiners thoughts and judgements of exercise physiology student performances when determining competency. A central assumption to this approach is that the examiners are central to the process and are active information processors. The rater's information processing is influenced by their understanding of effective performance as well as other factors within the social context of the assessment process (Govaerts et al. 2007).

Chapter 3: Methodology

3.1 Project Set Up

Prior to assessing student performance on the OSCE, a station was developed from a series of pilot testing (Naumann et al. 2014). The mock station involved an initial 2 minutes for the student to read the station task descriptor and prepare their approach. This was followed by a 7 minute student-patient interaction on a task. The station required students to conduct a patient history to gather information on the patients' cardiovascular disease history, their physical activity history and determine if it was safe for the participant to participate in a cardio-respiratory fitness assessment. The student was then required to set up the patient for the test and monitor the patient on the test for the first minute. The station task represented several key exercise physiology competencies including patient interviewing skills and patient monitoring skills during an assessment. The station incorporated clear task descriptors for the student and the examiners and had good alignment of the assessment criteria to the task. The station also involved a standardised patient, who was the same for all students and was instructed to follow a detailed script. Three exercise physiology students were invited to perform the mock station, as outlined by the task descriptor. The student participants were fourth year exercise physiology students who had just completed their final course OSCE the week before. The students' performance on the mock station was captured on film and prepared for assessment by the examiners. Examiners were subsequently recruited between May 2013 and January, 2014. All participants provided informed consent. Ethical approval was provided by the UNSW Human Ethics Committee, Biomedical HREA Panel D (HC13148).

3.2 Participants

The participants comprised 23 examiners who included University clinical academics, practicum coordinators, exercise physiology academics and practicing accredited exercise physiologists. The participants were recruited from 9 Universities across Australia and had a range of clinical skills and examination experience. Demographic details were recorded for each examiner, detailing their career position, number of years of experience in the role and number of OSCEs examined. The goal was to recruit examiners from a variety of career backgrounds and years of experience to determine if these factors influence the process of clinical judgement and an ability to articulate their clinical judgement. Examiners were recruited through the Exercise Physiology Clinical Education Group (ExEd) and Universities

within Australia. An email was sent by a third party to the potential examiners, inviting them to be an examiner on a filmed, mock OSCE station. Participants willing to participate in the study were subsequently contacted by the researcher and sent an information and consent form, detailing the project. An appointment was made to view the student performance, complete the assessment and be interviewed by the researcher.

3.3 Qualitative Data Analysis

3.3.1 Training

Prior to examining student performance, a standardized training and assessor alignment session was held for each of the 23 examiners. Leading up to the session, each examiner was emailed the station descriptor, enabling them to become familiar with the station tasks and the assessment criteria. The examiners read through the station task, instructions to student and the notes for examiners, which detailed the standards required to be demonstrated by the student for each assessment domain. Importantly the examiner alignment session reviewed the station task descriptor, the assessment criteria, application of the criteria and how to make global assessments on the three domains of communication, procedural competency and technical proficiency. The performance criteria included a detailed checklist and global rating scale for 3 key exercise physiology domains: communication skills, exercise physiology procedural skills and technical proficiency with equipment. Examiners were required to apply the criterion-based scoring to assess the students' level of mastery which included a 4 point scale: F for incorrect (score 3), P- for partially correct (score 5) and P for correct (score 7) and P+ for excellence in skill execution (score 9) which have been validated previously (Naumann et al. 2014). This scoring relied on the professional judgement of the clinical examiners. Filmed performances were shown in a random order and all examiners were blinded to the scores of the other examiners. The unique contribution of this study was that it investigated and compared assessor rating of the same student clinical performance.

3.3.2 Semi-Structured Interview

At completion of the assessment, each examiner was interviewed by the principal investigator to explore the key factors which contributed to their judgment of the students' performance. The semi-structured format allowed flexibility in sequencing and wording of interview questions and allowed for additional probing to clarify certain participant responses. The

method of semi-structured interviews was adopted as the research tool for the current study. Interviews were selected as the methodological tool as this approach can yield useful information relevant to the project goals that would not be available through questionnaires or other forms of questioning (Brenner, 1985; Fontana & Frey, 1994).

The questions in the semi-structured interview were aimed at understanding the examiners' paradigms and what they deem as critical aspects of clinical performance and ultimately competency. The semi-structured interview was piloted in 4 examiners, prior to commencing the data collection phase, to ensure the questions and approach captured the relevant information. Pilot interviews indicated that it was necessary to press examiners to be explicit about how they came to their rating. The interviews began with a broad generalisation about each student's score and progressed to more specific statements and explanations of what they saw in the student performance and what counted as evidence of attainment of the performance criteria. Probing questions were needed to explore the examiners' expectations, what aspects of performance examiners considered important and if there were any biases or influences they were aware of that contributed to their own interpretation of the rating standards and application of the standards to the final rating.

The semi-structured interview took approximately 20 to 30 minutes to complete. Identifying information about the examiners was removed once the interviews were transcribed. The questions commenced with a documentation of global competence, followed by documentation of the reasoning behind each rating. The interview then focussed on the domains of competency, specifically communication skills, technical competency and exercise physiology/procedural competency, determining the aspects of performance that were important to the examiner (Appendix A). As researchers, we were particularly interested in determining the strength and influences of the examiner's methodological paradigm on the process of examination. One could hypothesise that the examiner paradigm would be one of the major differentiating factors in the outcome. Examiners were also asked to discuss their confidence to perform the examination and provide a rating for the three students.

All interviews were taped, transcribed and analysed using a general inductive approach (Thomas, 2006). The data was reviewed and meaning created from the raw data through the development of summary themes. This analysis involved multiple readings and interpretation

of the raw data. This coded data was used to develop categories and a framework which captures the key aspects of examiners judgement which play an important role in determining the final student rating. Each time a new code or theme emerged, the transcripts were reread and refined. The process continued until no new themes were identified.

3.4 Quantitative Data Collection & Analysis

Descriptive statistics summarised the examiner demographics with respect to age, gender, role and experience. Examiner agreement was assessed by having the 23 examiners evaluate the same student performances. Under these circumstances, differences in ratings reflect differences in focus and rigor. Twenty three examiners rated the clinical performances using structured forms designed to direct attention to specific dimensions of performance. A combination of a skill competency check-list and global competency rating were recorded from all examiners. Once the performance has been rated by the 23 examiners, Generalizability (G) studies were performed to estimate the sources of variance in the student scores rated by the different examiners. Generalizability studies were selected for this study as it enabled the identification and estimation of the effects of potential sources of measurement error. A factorial ANOVA was conducted to acknowledge that multiple factors contributed to the variance in observations and enabled partition of the variance (Shavelson & Webb, 1991). The sources of variation included the examiners or raters and the student performance of the station task. These facets are potential sources of error and the purpose of generalizability theory was to quantify the amount of error caused by each facet and interaction of facets.

Chapter 4: Results

4.1 Demographic Analysis

Exercise physiology academics, clinical practicum coordinators and accredited exercise physiologists from 9 Universities across Australia were invited to join the study. Of the 23 examiners who agreed to participate in the study, all completed the examination task and follow up interview. 18 of the 23 examiners were female, with 5 males. The range of roles varied between being an exercise physiology academic, a clinical coordinator/educator or a practicing Accredited Exercise Physiologist (AEP), with some examiners having multiple roles.

<i>Characteristic</i>	<i>Mean (std) Range</i>	
Male, n (%)	5 (21.7%)	
Female, n (%)	18 (78.3%)	
Role		
Academic	16	
Clinical Practicum Coordinator/Educator	13	
Accredited Exercise Physiologist	16	
Years in the role		
Academic	5.39 (± 6.4)	(1 - 20)
Clinical Practicum Coordinator/Educator	2.13 (± 2.8)	(1 - 10)
Accredited Exercise Physiologist	3.54 (± 3.45)	(1-10)
Number of OSCEs examined	1.17 (± 2.91)	(0-9)

Table 4.1: Demographics of Participants (n=23) of the study.

Not many of the examiners had participated in an OSCE prior to the study, with an average of 1.17 OSCEs examined across the cohort, reflecting the newness of this type of clinical assessment to the profession. The experience of the examiners in their various roles was also limited, at 5.4 years as an academic, 2.2 years as a clinical coordinator/educator and only 3.5 years as a practitioner, again reflecting the newness of the profession.

4.2 Qualitative Analysis

Five themes were identified from the semi-structured interviews which documented the variance in examiner judgement. These were: 1) The use of varying frames of reference by the examiners to assess student performance; 2) Varying use of inferences made by the examiner on aspects of the student performance and an interpretation of what the aspect of performance meant; 3) Examiners varying reaction to communication, specifically the student's language choice and the interaction between the patient and student; 4) Examiner variance in using the patient outcomes and perceived patient reactions as a point of reference; 5) Varying levels of confidence expressed by the examiners, especially as it related to performing the assessment, having the confidence to fail a student or confidence in the rating they made.

Theme 1: Varied Frames of Reference

The first theme identified was the use of varied frames of reference used by the examiners. In order to make a judgment about the student's performance, the examiner needed a standard to judge against. In most instances it was either them self as a practitioner, themselves as an academic lecturing on the topic, themselves as a practicum coordinator or against the other students being examined. Examples of using themselves as a practitioner as a frame of reference included:

<i>Frames of Reference – self</i>
<i>I perceived the silence as awkward. It was a time where I would ask a question of the patients to determine how they are doing on the test [Ex1].</i>
<i>Interviewer: So you used yourself as a reference for the performance?</i>
<i>I suppose I did. When the silence came, I immediately thought of questions in my head and things I would be doing at that moment in time. When they weren't picked up on or done, I would mark the student down [Ex1].</i>
<i>I imagined my response in that situation and I kept comparing them to me [Ex 22].</i>

Although one examiner acknowledged what he would have done as a practitioner, he did acknowledge that he was an experienced clinician and potentially it was an unfair comparison.

Frames of Reference - self

Both candidate B and C didn't really get into the name of the drug and what does the doctor say it does da da da da..... Now if someone was being sent to me, and I would be interested to getting other AEPs take on this, if someone is cardiac, I would also be asking do you have angina meds with you, if something should happen during the test. And if he does, you know where they are (the medication) when he exercises [Ex7].

Interviewer: So did you actually look at that, thinking what would I do?

Yes, I would have been asking a lot more clinical questions, then again, I am very experienced and have dealt with these kind of clients for many years before I came here, to Australia. So it is a bit of an unfair comparison [Ex7].

Comparisons in performance standards were made between the 3 student candidates, with the other student performers providing the frame of reference for some examiners.

Frames of Reference – other candidates

Although I rated A and B the same, I felt A was the slightly stronger EP, as time management can be mastered in the field through training and practice [Ex2].

I felt I needed to use the other performances as a calibration. For me candidate C was clearly lacking and set the benchmark of what wasn't acceptable. I then returned to candidate A and B and thought they had strengths in different areas. (Candidate) A was really good on the interpersonal skills and was well structured in terms of patient history, whereas candidate B was better on the patient monitoring during the exercise [Ex5].

I thought she was going down a slippery slope but she ended up with some good information. He (the patient) said I've got lots of work pressures, I'm really stressed, eight out of ten. It gave information that the other two (candidates) didn't get which I think for an exercise program, this information was really important [Ex6].

I liked the way she spoke slowly and explained herself. I think she was able to put him more at ease than the two boys (candidate A & C), just the way she spoke to him [Ex10].

I got so uptight with student A, that by the time student B came along, it was a relief. She didn't bombard the patient and had some warmth in the exchange [Ex21].

One comparison between the students was related to the communication style between the exercise physiology candidate and the patient. Most examiners stressed the importance of not dominating the consultation and giving time to the patient to tell their story.

Frames of Reference – other candidates
<i>Candidate B got the same amount of information in a shorter time, whereas the other guy (Candidate A) talked A LOT. It came across as either he was nervous or a hyped up person. He didn't even stop to come up for air [Ex5].</i>
<i>Student A virtually had verbal diarrhoea, he said a lot, and once cut the patient off while he was trying to get the patient back on track, whereas the girl (candidate B) had a slower pace and was able to have a conversation with the client [Ex6].</i>
<i>He (student A) spoke really fast. And he asked several questions in a row and didn't break them up. Because he spoke fast he didn't leave much open space for the client to tell him things. I think he would do really well if given feedback around him communication and the importance of listening to a client. Candidate B on the other hand..... [Ex9].</i>
<i>She (candidate B) had the communication with the client, which in the scenario of what they were doing, was probably most important at this point in time. She actually listened to the client. OK. She listened, she probed, she just seemed a little more engaged with the client, whereas the others (Candidate A and C) were pathetic [Ex7].</i>

A third frame of reference that emerged from the interviews was themselves as an academic. The examiner stated that in the course they teach at their University, cardiac history and patient management was taught in a certain way. When students missed taught elements, they were marked down.

Frames of Reference – lecturer
<i>Just before I answer that can I just clarify something, because this is how we teach this at our University, if someone exercised a client without having there angina medication with them, we would fail them. And it wasn't really clarified by any of them as to what medication he was on and what it was for [Ex6].</i>
<i>My expectations of what I wanted to see was linked to what I have taught as an academic [Ex20].</i>

What this data indicated was the examiners were not approaching the assessment with the same frames of reference and that the process was both complex and highly variable. It was also acknowledged that all examiners had different experience, expertise in the area and expectations of what they wanted to see in the student performance and that this was likely to influence the rating process.

Theme 2: Inferences made by the examiner on aspects of the student performance and what the performance meant.

The second factor contributing the examiner ratings was the use of inferences by some examiners, in order to interpret and fix meaning to the student behaviour on the task. Inferences were often made based on what the student did and what that behaviour meant. If the student was nervous, fumbled over their words or talked to themselves they were viewed as less professional and less competent.

<i>Inferences</i>
<i>For the third candidate, I felt he wasn't quite confident. He tended to talk to himself a lot and it sounded like he was doubting himself and wanted reassurance from the patient that the consultation was going well [Ex1].</i>

Similarly, if a student was all over the place in their patient consultation, it was assumed the student lacked practice in the area of competency and not rated highly.

<i>Inferences</i>
<i>For me, when the tasks are out of sequence and the student jumped forward and back, it strikes me that random thoughts are popping into the students head and I suppose not practiced [Ex2].</i>
<i>I just got the sense this guy (candidate C) was incompetent. He was just coming across as disorganised, not efficient. I actual got the feeling that the patient would have sensed this guy was incompetent. Asking, am I in the right place here? [Ex4].</i>

Inferences were also made about behaviour not observed. If a student failed to complete all aspects of the assessment in the time allocated, some examiners assumed that in a real

consultation the student may have got there, whereas others did not assume they would have completed the tasks at a similar standard.

Inferences

I gave him (candidate A) a P-, he did most bit but there were a few elements missing there, hence the minus. And I don't think there were any elements that were missing that could have caused harm to the patient. I kind of got a feeling if we had filmed for a bit longer we might have seen him monitor the patient during the assessment and check how he was going [Ex12].

In contrast several examiners were clear about not grading performance not observed in candidates who failed to manage their time well and complete all aspects of the station.

Inferences

Candidate C did poorly on this aspect as they did not take much of a patient history. Therefore, I was unable to assess the flow of the consultation. It was very difficult to judge an incomplete performance as it was not clear if it would have been done badly as it was not done at all [Ex1].

Theme 3: Examiners variable reaction to communication and the students' interaction with the patient.

A third theme identified from the semi-structured interviews was the examiners variable reaction to communication, including choice of language and the interaction between the client and student exercise physiologist. There were an overwhelming number of comments by the examiners regarding the candidates' verbal communication skills, including the word selection and the appropriateness of the use of such language. If the language was casual it was perceived as unprofessional and the examiners perception of the student's ability dropped.

Language & Patient Interaction

"Cool" It was his terminology that was short, single words and lead me to believe the patient was in less capable hands [Ex1.]

<i>A standout thing for me was his language was way too casual. When I heard the introductory phrase “morning mate” I cringed a little inside. The response to the patients’ condition being fair enough and cool was very poor. At this point I really had to bite my tongue as this [angina] was definitely not cool or fair enough [Ex3].</i>
<i>When the students used very casual greetings and single word responses to information provided - like cool, yep - this didn’t sit well with me. Mate is a word they should use with their friends in a social setting, not a paying client who is seeking professional guidance to manage their conditions [Ex2].</i>
<i>Look sometimes it is a hard thing for students to know how to be friendly, whilst being professional. And I think those type of scenarios are good for them to get that feedback, that they have perhaps gone too far one way, with “good job mate”, when you have only just met the person [Ex6].</i>
<i>Well he (candidate A) was trying hard to level with the gentleman, but he was a bit blokey with him [Ex8].</i>
<i>I don’t really like the use of mate, it is contentious. He did it repeatedly and I’d want to talk to him about it [Ex9].</i>
<i>In terms of Candidate A, he did things very well, his communication was very clear and flowed well, although it was the rapidness of his speech and the use of mate. You can’t assume everyone is your mate [Ex14].</i> <i>Interviewer: So when you heard the term mate, what was your reaction? (Visual cringe.....) I not sure how to put it in words.... I suppose a bit unprofessional [Ex14].</i>

Single syllable responses to patient answers such as yep and cool was also negatively perceived by several examiners. The student was perceived as being too focussed on working through a series of questions and completing the medical history, as opposed to focussing on the patient in front of them.

<i>Language & Patient Interaction</i>
<i>The patient was giving up his life story along with very personal information and to have “yep and cool” thrown back at them was not good. I felt they could have used the answers and information to perhaps probe a little more or reassure the patient the information was noted and understood [Ex4].</i>

He also had a word he kept using. I can't remember his, but the girls' was "fantastic". Everything was fantastic [Ex8].

Interviewer: So terminology was key?

"Excellent", "Fantastic" and he's just told you he has angina and it is not fantastic [Ex7].

By using single word responses to the medical information provided by the patient, the perceived ability of the student diminished. The examiners were quick to point out that a single word response to a patient providing medical history was indicative of a student simply going through a checklist. They believed that this sort of information is very personal and it is an ideal time for a client and Exercise Physiologist to build rapport, trust and demonstrate empathy.

Language & Patient Interaction

The problem with this is that the student is only at the cognitive stage of learning, so they are process driven and less likely to pick up what is in front of them. They are thinking about the next step (or question) as opposed to the information the client is giving them [Ex5].

Several examiners also stressed the importance of demonstrating a patient-centred approach within the patient consultation to ensure the patient was provided with an opportunity to tell their story. It was stated as an important clinical competency, especially in relation to building rapport with the client.

Language & Patient Interaction

She left more space for the client to say his story. She actually asked fewer questions but actually left more room for her client to give her more information. And it shows me that she is more advanced in the development of her interviewing skills. That she is not just going through a checklist to get the information but that she is engaging in a conversation with another human being to get that information [Ex9].

Additionally, the use of technical exercise physiology terminology without a suitable explanation was not viewed favourably by several examiners.

<i>Language & Patient Interaction</i>
<i>And the type of language they used. If they do use a technical term, it is not that you can't use it but you need to explain it to the patient, especially if it is something that the patient will hear regularly, such as RPE. So if you are going to use technical terms you need to explain what it is and how it fits in with what you are doing and why is it important for them to know [Ex6].</i>
<i>He indicated we are only doing a sub-max test and we will extrapolate the results. And to me that would have meant nothing to the client and missed the important point of this is why we are doing the test [Ex7].</i>

Incorrect use of terminology was also highlighted by examiners as being a negative performance indicator.

<i>Language & Patient Interaction</i>
<i>Oh and one term that was really confusing was cardiac strength. In the third one (Candidate C) we will be having a look at your cardiac strength, so I knew what he meant but..... It was kind of putting two terms together that don't mean anything together and could be confusing for a patient.... If I heard it correctly [Ex6].</i>

Excellence in choice of words was also noted by several examiners, attached to a belief that the patient was in more capable hands.

<i>Language & Patient Interaction</i>
<i>Candidate A was also clear with instructions on what to do, had an excellent choice of language and was not confusing for the patient [Ex1].</i>

In others instances, examiners failed to notice language or register language to be a contributing aspect to their performance ratings. Discussion focussed on the consultation flow, listening to the client and responding based on the information provided.

Language & Patient Interaction

For me, it was the ability to get the detailed information from the patient, without having to ask so many questions. Not repeating yourself, and being able to respond to the patient and go with the information the client is giving you. So you kind of flow in with what they have said but not miss anything critical. As opposed to going just with what you had on a bit of paper [Ex6].

Several examiners also raised the issue of the candidates not listening to the client as being a real negative.

Language & Patient Interaction

And the patient with candidate C, clearly stated he didn't really like the bike, but I am going to ignore what you just said and put you on the bike anyway for the protocol, instead of something like the treadmill, that the guy might find a little more to his liking [Ex7].

The client stated it (the bike test) was a little unnatural, and then I thought.... he said, oh. OK. You're going to do it anyway. So that was a bit strange. So he did check with the patient that the mode was suitable, but he didn't really listen to what the patient had to say [Ex10].

Some examiners also highlighted the important skill of being able to listen to the client as being an important competency.

Language & Patient Interaction

So with candidate B, there were a few things with her, which she wasn't picking up on. She asked the guy how cardiac outpatient was, and he said "oh, it was a bit repetitive, but I got through, blah blah" and the response, excellent. I'm like, there is just a little bit of detached communication there, ok, or inappropriate communication [Ex7].

A component missed by all candidates was empathy. It was raised as a more advanced skill competency but an essential one for an entry level exercise physiologist.

Patient Interaction

But just like the others, not great empathy. I think if you have a patient with angina that has been through cardiac treatment and rehabilitation, it will be a big emotional thing and you should at least acknowledge that [Ex10].

The challenge of assessing empathy was also raised by the examiners and the alignment of the curriculum to assessment criteria was raised.

Patient Interaction

Probably a final point was that I found it quite difficult to assess empathy. I wasn't sure if I was looking for something really obvious or subtle [Ex14].

When the patient told candidate B they have angina, her response phase was "Fantastic". Now I know she didn't mean it like that, but her approach was from a checklist, not from an approach of empathy whereby you acknowledge how this news may have been difficult for the patient., I am wondering from a curriculum point of view, whether we teach it well or provide any opportunities to students to practice it. Maybe we need more time on the course to develop empathy, because we are expecting the students to demonstrate it to us, without really teaching it well.

Theme 4: Using the patient outcomes as a reference

A fourth theme that emerged was using the patient outcome as a reference. Some examiners viewed the patient's response to the interaction and made judgement on whether they believed the patient had a positive or negative outcome or encounter. Examples of negative outcomes from the patient included a reaction to the questions asked of him.

Patient Outcomes

I felt the patient was more defensive with his answer, particularly in the exercise history. Perhaps it was the way the student asked the questions that were kind of putting pressure on the patient [Ex1].

He never really checked whether the patient actually understood what he was saying. At one point he asked him, is that clear? And he responded "Yer ... sort of". But then he didn't really stop and explain it to him [Ex10].

<p><i>Subject number C, just doesn't have it. Um, clumsy, wasn't actually looking for client understanding. If I was the client and I came in, I wouldn't be very confident with him running my test as he didn't seem to know what he was doing [Ex7].</i></p>
<p><i>And the third person I gave a fail to. He didn't tick many boxes and had lots of things missing, such as getting information of diagnosis, didn't conduct a medical history or current physical activity history. I think it is important to know that if you are going to actually conduct an exercise test. So I reckon this guy needs to fail as he is putting the patient at risk [Ex12].</i></p> <p><i>Interviewer: So patient safety was key to the fail?</i></p> <p><i>If you are going to kill someone or put them at significant risk, that is an automatic fail in my mind. But other things to support this were, I just got the sense and I think that the patient would have thought that this guy was incompetent. He was just coming across as disorganised, not efficient. I actual got the feeling that the patient would have sensed this guy was incompetent. Sensing am I in the right place here, have I got Mr dodgy brother? Perhaps I do [Ex12].</i></p>
<p><i>Interviewer: And you also mentioned his rapidness.</i></p> <p><i>Yep, right, moving on, ok, um and it just means the patient won't want to open up and speak to the EP [Ex14].</i></p> <p><i>I: So you looked at it from the point of view from the patient.</i></p> <p><i>Well it is funning. When I am listening to the consultation, I feel like I am the patient and I felt tense. I just felt uncomfortable [Ex14].</i></p>

In contrast, positive patient encounters included:

<i>Patient Outcome</i>
<i>The patient looked more relaxed and comfortable with this candidate and was more receptive to answering the consultation questions [Ex1].</i>
<i>The information obtained from the client and information explained to the client was clear, concise and at the patient level. I really got a sense that the patient understood [Ex2].</i>
<i>I thought she (candidate B) did really well. She had all of the components required for the consult and she responded to the clients answers. She was also having a</i>

conversation with the patient, instead of going through a checklist. It helped to put the patient at ease [Ex12].

The patient experience was not followed up post assessment. Other examiners failed to note the patients' role in the encounter.

Theme 5: Examiners Varied Levels of Confidence and Certainty with their Ratings of Student Competency.

A fifth theme that emerged was varying levels of examiner confidence to participate in the OSCE assessment and make a rating with certainty. Many of the less experienced examiners, as determined by years in the profession, found the role of examining stressful and were not very confident in their ratings. Once they articulated their findings they wanted to make changes to them.

Confidence

I found the examination process difficult and now that I have spoken about my ratings, would want to change them. Given the opportunity, I would want to be mentored by an experienced examiner first, before being put on a station by myself [Ex22].

The examination process was a little stressful and I don't feel confident with the ratings I gave [Ex21].

Some examiners found the task of examining straight forward and were happy with the process and their assessment.

Confidence

I found it easy to use and easy to be consistent and to justify my positions [Ex9].

Other examiners varied in their confidence to fail a student. Although the student may not have been up to the required standard of competence, some examiners would search for positive aspects of the encounter, giving the student the benefit of the doubt.

<i>Confidence to Fail</i>
<i>I look at my ratings and now realise I made allowances for the students' performance [Ex19].</i>
<i>Interviewer: In your opinion, are each of the students observed in the video ready to practice as a graduate entry exercise physiologist?</i> <i>Candidates A and B are definitely ready to practice. However, C I was a bit concerned about going out to practice. He came across as he didn't quite know what he was doing and that he had potential to make errors and put a client at risk. I feel with guidance and more time in the field, he would get there [Ex1].</i>
<i>It is hard to fail students, but I think we need to, for the good of the profession. Under the current system, these students could get through a placement [Ex6].</i>

This candidate was however, awarded a borderline pass by this examiner. Similarly, examiner 6 wanted to fail the student but looked for reasons to pass the student.

<i>Confidence</i>
<i>I sought of had him sitting at a P-....but</i> <i>C was probably umm.... He was another one where I would probably He kind of totally missed the medical history in the consultation but did the exercise monitoring well. And although he got him started on the bike, he didn't really know anything about him. So that for me would kind of tip him into a fail. Like, I would want him to do it again [Ex6].</i>
<i>Well, that is where I was like P-, should I fail him on it, because he didn't show it or do I assume that because he had done the other stuff, so maybe you would scrape through on that one [Ex8].</i>
<i>Now I debated around this one. Because I was going to fail..... on the basis of at the end of the day, if you don't have enough information before you commence, it is just not safe. But then I looked at everything else, assessing the other skills or components, risk assessment and execution and he was OK on the execution, so that's how I ended up in the middle, with a P- [Ex14].</i> <i>Interviewer: But is risk assessment vital?</i> <i>In a clinical scenario yes.....but, yes there was a definite reluctance to fail because..... I</i>

knew him..... and he was such a great student in the program [Ex14].

Whilst the reluctance to fail a student was acknowledged, the examiner also made it clear that they would be willing to fail a student and that it can be a good learning opportunity for the student, as long as adequate feedback was provided by the examiners.

Confidence to Fail

I hate failing people, I really do and I always try to, you know, well yes you've failed but here is an opportunity to really learn something from it. So take a few steps back and these are the bits you are missing. I would never say no, you are not going to be a good clinician. Everyone has good points and something to offer. It is just a matter of trying to build up those deficiencies, to be a good clinician [Ex7].

Several examiners also raised the fact that they would be happy to pass a candidate, if they could be guaranteed he would receive further mentoring and training on the job.

Confidence to Fail

I would be happy to recommend him to a graduate internship, although I don't think there are many of these around [Ex2].

Candidate A and B all hit the main points, and ahhh they are a classic case of yes they can go out and practice but you would want them to go out and practice in a mentored environment...but....these roles don't exist [Ex5].

The mentoring role was subsequently discussed in more detail in reference to each of the candidates' readiness to practice in the exercise physiology profession.

Readiness to Practice

I think under the current system we have a light switch approach that as long as you come from an accredited institution and have ticked the box in hours, you are ready to practice.

This exam indicated candidate A and B could improve substantially and under mentoring could be so much more effective..... but probably won't be afforded the

mentoring.

So yes, if the mentoring is available, A and B are ready to practice but potentially we need to ensure a minimum standard to ensure they are safe to practice.

Even candidate C has the potential to get there if they are trained on the safety aspects, even though they are a bit rough, if they had some good mentoring, they could be in the ball park [Ex5].

Others were quick to fail a student without any questions asked, if the student provided incorrect information or compromised the patients safety.

Confidence to Fail

Candidate C failed as he had no history, identified no risk factors and failed to take an exercise history. My major issue was that not enough information was gained prior to testing to ensure that the test was safe to run. He didn't ask about his medications and are the meds working. This would all impact on patient safety and possibly put the patient at risk [Ex3].

A and C are definitely not ready to practice. At the end of the day it comes down to patient safety. If the patient is not in safe hands I would certainly recommend further training [Ex3].

C would fail outright and pulled back and hit with a baseball bat and questioned what they are doing. They have violated a fundamental principle of patient safety [Ex5].

And candidate C, there was just there nothing there, just jumped straight in.... so that was an easy one. They didn't know what is going on at all with the client and basically they perceived their role was to do an exercise test regardless of whether the patient had stable angina, if the person was 8 months pregnant, they were going to do an exercise test no matter what....and it was going to be sub-max and it was going to be on a bike [Ex5].

Interviewer: And is danger is a straight out fail?

Straight out fail. Every red flag, I mean the sirens would be going off. To me that is a mistake that a personal trainer would make. Straight to the exercise, no introduction, no investigation. And that's what sets us apart from someone who graduates with a cert 3, cert 4. There was nothing about his medications. He had no exploration of what

other conditions he had. Oh, I am glad you mentioned that. Big red flag [Ex7].

I feel comfortable failing this candidate as he didn't do so many of the things listed on the checklist and there were just too many components missing. And these were the critical components. I was sort of Ok that he missed some components which weren't so critical. But he missed the critical one's that could put the patient in danger [Ex12].

Interviewer: And I suppose he can miss one or two, but what is basic competency?

That's what we need to determine. What are the critical components that the student needs to demonstrate. It is only the one's that put the patient at risk. Perhaps that's a conversation we need to have as a profession [Ex12].

And finally some examiners who made a rating but then explored the rating in the interview, wanted to change their rating based on further discussion. By discussing their judgment process and determining how they came to their rating, allowed a detailed examination of their own process and rationalisation. In most cases, examiners didn't suggest large changes but generally moved the student down a rating.

Confidence

Ah you know what, I think I would actually move that to a pass now that I have spoken about it [Ex10].

Interviewer: And what about candidate 3?

Well this was a hard one. He went straight into the stress test, so he didn't actually ask the client anything. I don't know if he was instructed not to do the interview? [Ex7].

Interviewer: He had the same task as the others.

Really. Oh. See, that is where I would have failed him straight out. He knows nothing about this client. He has not developed a rapport. I thought... what I got from this was that oh, he must be starting from a different stage in the interview. No, he has absolutely no idea what is going on with this client. What this client has or...yer no, that was dangerous [Ex7].

To me she was a pass +, she was smooth and efficient and really focussed on the patient, but now that we start to look at it, I am wondering if I have been too generous there [Ex12].

So now I look at this, it should have been a fail [Ex14].

Perhaps this process will assist examiners clarify how they assess student performance and get clearer standards in their minds about what is acceptable clinical competency.

In contrast, some examiners felt confident with the process and the ratings they made.

<i>Confidence to Fail</i>
<i>I found it easy to use and easy to be consistent and to justify my positions [Ex8].</i>
<i>I think the credit criteria provided to us were sound and provide a good understanding of how to grade a student who has performed really well. I think it was easy to distinguish between a student that was really strong verses a student that is just good enough. However, it would still be good to benchmark across the profession [Ex9].</i>

Final Comments on how the examiners perceived the importance of the OSCE in assessing entry level graduate competency and a need to identify minimum standard of competency for Accreditation.

Through the process of examining student competency, many examiners commented on the need to determine minimum levels of competency that are transparent across the board, to ensure we are producing graduates that are safe to practice independently in the profession.

<i>Confidence</i>
<i>Because at the end of the day, with the accreditation, we (the Universities) are the last stop. And I think as a profession we need to agree on what is minimum standard. And even though I am a hard ass, it is all about certain competencies that we are happy with. So if they are not up to speed with the clinical skills, they don't even sit the final exam and can't graduate [Ex7].</i>

Examiners also discussed the need to determine standards of competency and the shared responsibility for this task lying with both the Universities who are training the students and the Professional Association who are providing Accreditation to Practice.

<i>Confidence</i>
<i>And I agree, I think ESSA (the professional association) need to get on board about minimum competencies. I mean you can't have everyone at the same standard, but at</i>

the same time, you do need to have a clear perspective on minimum competency [Ex7].

4.3 Quantitative Analysis

There were 3 possible sources of variance to the ratings of performance, the examiner, the student performing the scenario and the domains by which the students were being assessed against. The analysis of variance indicated that the individual examiner had a minimal impact on the variance in ratings (6.5%), with the majority of variance explained by the student performance on the station task (27.1%) (table 4.2).

Variance Estimates for the Rating of Student Performance			
Component	Estimate	%Variance	
Variance (Examiner)	.258	6.5%	
Variance (Student)	1.078	27.1%	
Variance (Domain)	.260	6.5%	
Variance (Student* Examiner)	.399	10.0%	
Variance (Examiner* Domain)	.203	5.1%	
Variance (Student* Domain)	.461	11.6%	
Variance (Examiner*Student*Unknown)	1.32	33.2%	Unknown source of error

Table 4.2: Variance Estimates for the Rating of student performance

Examiners were asked to rate the global or overall performance of the three students, incorporating the three assessment domains. The analysis of variance indicated that the examiners contributed to a low proportion of the variance in ratings (9.0%), with the majority of the variance being attributable to the students' performance on the station task (62.8%) (table 4.3).

Variance Estimates for Global Performance		
Component	Estimate	%Variance
Variance (Examiner)	.368	9.0%
Variance (Student)	2.56	62.8%
Variance (Examiner*Student*Unknown)	1.14	28.0%

Table 4.3: Variance Estimates for the exercise physiology procedural skill competency rating.

A closer examination of the three domains indicated slightly different outcomes for each of the three domains. The assessment of exercise physiology procedural skills, mimicked those of

the global assessment ratings. The examiners again contributed to a low proportion of the variance in ratings (7.6%), with the majority of the variance being attributable to the students' performance on the station task (67.1%) (table 4.4).

Variance Estimates for exercise physiology procedural skill		
Component	Variance	%Variance
Variance (Examiner)	.392	7.6%
Variance (Student)	3.454	67.1%
Variance (Examiner*Student*Unknown)	1.299	25.2%

Table 4.4: Variance Estimates for the exercise physiology procedural skill competency rating.

For the communication skill assessment, the examiner continued to have a low contribution to the variance (8.7%). The greatest source of variation was attributable to the interaction between the examiner, the student and other unknown factors (57.2%) (table 4.5). The student themselves contributed to 34% of the variance in ratings.

Variance Estimates for communication skill		
Component	Estimate	%Variance
Variance (Examiner)	0.258	8.7%
Variance (Student)	1.017	34.1%
Variance (Examiner*Student*Unknown)	1.70	57.2%

Table 4.5: Variance Estimates for the communication skill competency rating.

For the technical skill competency assessment, the examiner had the largest contribution to the variance in ratings, when compared to the other skill competency domains. The examiners' impact on the variance of ratings was 24.2%. The greatest contribution to the variance in the technical competency rating was attributable to the interaction between the examiner and the student (70.9%) (Table 4.6).

Variance Estimates for technical skill		
Component	Estimate	%Variance
Variance (Examiner)	0.732	24.1%
Variance (Student)	0.150	4.9%
Variance (Examiner*Student*Unknown)	2.154	70.9%

Table 4.6: Variance Estimates for the technical skill competency rating.

A final analysis was conducted on the results, comparing the assessment results between the 9 Universities. When comparing the means for global competency, exercise procedural and communication skills, there was no statistical difference between the Universities. However, there was a significant difference in how the Universities judges technical skill (Table 4.7).

ANOVA between Universities	
Component	Significance
Global	0.475
Exercise Physiology Procedural	0.741
Communication	0.669
Technical	0.004

Table 4.7: Analysis of Variance between the Examiners from different Universities.

Chapter 5: Discussion

Competency-based assessment is an integral part of determining clinical competency and readiness of trainees to practice in the profession.

The major outcome from this study was that despite the documented differences in the examiner focus and judgement pathways, the ratings of the student performance was relatively consistent amongst the examiners, with the difference in ratings largely attributed to the student's varied ability. This finding was unexpected, as the examiners in the study had diverse backgrounds and roles, with a mixture of University academics, practicing AEPs and clinical educators, most of whom had little experience in examining clinical scenarios. The thought processes were diverse and if the qualitative results had been used in isolation, may have led to the research team drawing conclusions that the examined performances would have yielded widely different ratings. A strength of the study was that by combining the use of quantitative and qualitative approaches, the researchers were able to gain greater insight of the assessment process and the implications on the rating outcomes.

The first aim was to explore and understand the factors that contribute to assessor judgement and concluded that there were varying factors and that judgement was attributed to multiple themes which integrate to produce the assessors' ratings. The first theme identified was the use of varying frames of references by the examiners to judge the student performance. In support of our findings, previous research focusing on clinical performance rating suggested that evaluators use different frames of reference to judge performance. These varying frames of reference are not always considered or questioned prior to recruiting for the role of examiner, so it not always clear what level of impact the examiner has or is expecting to observe in the student performance (Kogan et al. 2010; Vukanovick-Criley et al. 2006). Several studies in the medical education literature have focussed on assessor training, particularly in the realm of frames of reference training, showing limited or no effect (Holmboe et al. 2004; Cook et al. 2009) through to moderate to large effects (Woehr 1994; Schleicher et al. 2002). The subsequent quantitative analysis in the current study confirmed that despite the varying frames of reference, the examiners were still able to judge consistently, what was considered a good performance and what was considered a poor performance and that the varying frames of references used by examiners should be noted but also embraced..

A second theme identified was the varying use of inferences by some examiners. The inferences made were often around what the observed student behaviour meant. The use of inferences was prominent by some examiners and noticeably absent in others. Some examiners made note of an action, interpreted those actions and affixed meaning to the action. These interpretations went unchallenged and were never clarified by the student. This can become problematic for the student if the inferences are not clarified, confirmed or articulated clearly to the student and could present problems when providing feedback to the student on their performance. Examiners need to be alerted to the use or potential use of inferences and could be incorporated into assessor capability training within the exercise physiology profession.

A third theme identified in the findings was the examiners' varied reaction to communication and the language used by the students. In the current OSCE, if the language was casual it was perceived by the examiners as unprofessional and indicated a lack of connection with the patient and the examiners' perception of the student's ability dropped. Excellence in choice of words was also noted, attached to a belief that the patient was in capable hands. In other instances, examiners failed to notice language. The different aspects of communication noted by the examiners in this study reflect the different perspectives by which examiners make judgement which has been eloquently referred to as examiner idiosyncrasies (Govaerts et al. 2007). From an educational perspective these differences can be a positive for student learning if the multiple perspectives of feedback can be triangulated for the student.

The interaction between the student and the patient was deemed very important by several examiners', including the ability of the student to listen to the client and respond accordingly, establish a rapport with the client and demonstrate empathy towards the patient and their situation. Several examiners raised the issue that there were very few instances where the students were competent in demonstrating empathy towards the patient and their situation. It was a required component of the checklist, but was not rated as being demonstrated by any examiner. An examination of assessment outcomes provides valuable feedback on the curriculum content and clinical training and in this instance has served to highlight a deficiency in the teaching and learning process (Holmboe et al. 2010). The outcome was that the teaching of empathy needs to be more explicit and opportunities for students to observe and develop their skills of competency.

A fourth theme identified was the varying reactions to the exercise physiology student-client interaction and varying use of viewing the consultation outcome from the patient's perspective. Some examiners viewed the patient's response to the interaction and made judgement on whether they believed the patient had a positive or negative outcome or encounter. Such inferences were not confirmed by the patient and could be problematic if the assumptions were incorrect. Producing ratings for student performance requires examiners to use data, such as student behaviour, and affix meaning to them. If patient experience is to be used in making a rating, it needs to be clarified with the patient. Working in partnership with patients is a cornerstone of modern medical practice, with research focused on the important role of an individual patient and identifying insights they can provide for the learner, not only into the role of 'culture' in health and illness, but also into areas such as health and illness beliefs and behaviour, how problems present, the importance of eliciting the patient's ideas and concerns and expectations, as opposed to focussing solely on their medical conditions. Patient contact can facilitate effective exploration of these areas (Stewart et al. 1995).

The final theme identified was the varying levels of confidence in their role as an examiner and varying levels of confidence in the decisions they made. Emerging from the interviews was a reluctance to fail a student by some examiners, especially those examiners who knew the candidates as being from their own University. Assessors often give the students the benefit of the doubt and 'fail to fail' incompetent students, unless there is clear evidence of unsafe practice (Lankshear, 1990; Watson et al. 2002; Dudek et al. 2005). Clinical supervisors at McMaster's University reported an unwillingness to record negative evaluations as the biggest problem with evaluation in the clinical setting (Cohen et al. 1990). This research was supported by the research by the Association of American Medical Colleges, where again the unwillingness to record a negative evaluation was rated highly (74.5%) across 10 medical schools (Tonesk & Buchanan, 1987). The reasons provided for the resistance to fail poorly performing students were lack of supporting evidence or documentation and the threats regarding the appeals process and the time associated with such an appeal and lack of time for student remediation (Dudek et al. 2005). This presents particular problems for the profession as patient outcomes may be compromised by poor practice when allowed to pass through the system (Lankshear, 1990). An examiners judgement, without explicit criteria and training has the potential to be biased in either direction, with the judgement biased even further if the assessor gets to know the student (Howard, 1990). Examiners confidence to fail a student who is not yet competent may be enhanced within the exercise physiology profession if framed

within the context of agreed standards. This may improve the effectiveness of assessment by relocating it from assessment about the person to one of assessment of competency within a standards framework. Benefits for the profession could also arise from an opportunity for the less experienced examiners to practice and receive feedback and to work alongside more experienced examiners.

As exercise physiology is a new profession, there appears to be a real need to develop the capabilities of examiners or assessors. The profession asks practicing AEPs to mentor and assess students but offers very little training. Although the current examination process offered guidance on the assessment criteria, many of the less experienced examiners indicated they were still lacked confidence in performing the assessment and would have benefited from having more experienced examiners as mentors. Additional support suggested in the literature includes having opportunities to develop the criteria as a collective group of examiners, reviewing virtual student performances, having opportunities to openly discuss what they were viewing and attach meaning to the performance (Jones, 2000; Holmboe, 2004). Finally, as a profession, policy development needs to occur that considers the importance of training and assessing the next generation of AEPs. Greater recognition of the role needs to occur and perhaps provide recognition to those who do train to be an assessor. Similarly, Universities curriculums could also consider training student AEP in the role of mentor and assessor. Fourth year students could be taught mentoring and assessing skills with their first year counterparts.

The varied themes identified that contribute to making a judgement illustrates the complex nature of examining and that a simple score on a rating scale is such a small part of the overall process. This single score also fails to provide the student with any feedback as to how the score was derived. To maximize feedback to the students, meaningful interpretations of ratings requires narrative comments that provide insight into the raters reasons. Such feedback is also invaluable in guiding the student in their competence development. For these reasons it is recommended the assessors be encouraged to provide as much narrative feedback to expand on their rating and provide the student with rich feedback to guide their on-going competency development.

As the examiners application of the criteria to formulate a judgment of performance is an essential element of a valid and reliable OSCE, it was important to determine if the difference in reasons for determining ratings was impacting on the rating scores. On a global level they did not. Quantitative analysis revealed that it was the student that was the major contributor

to the variance in the ratings, not the examiner. At the global level and for the exercise procedural competency ratings, the examiner had a minimal impact on the variance in ratings, with the majority of variance explained by the student performance on the station task. For the communication competency assessment, the examiner had a slighter greater contribution to the variance, indicating the examiners had slightly different expectations about what they expected to observe in communication competency. The different expectations were picked up through the interviews, where examiners reported varying reaction to language, the students choice of words (single syllable words such a cool, fine, yep), the casualness of the terminology (use of the term mate), the pace of the conversation (described as verbal diarrhoea), the dominance of the student in the client-student interaction and the ability of the student to listen to the client and respond appropriately. Some examiners mentioned these aspects, others made little comment. A suggestion is that it would be advantageous as a profession to determine what is appropriate at the graduate entry level, and ensure that effective communication and patient-client interactions are taught within the curriculum and examined consistently.

For the technical skill competency assessment, the examiner and what University the examiner came from had a larger impact on the variance of ratings when compared to the student. These results suggested that the examiners used different standards and criteria to judge clinical performance. Despite not being evident in the technical domain of the current study, it is possible to achieve good inter-rater agreement. MacRae and colleagues (1995), compared physician ratings of 120 videotaped medical student clinical encounters across 4 cases using an eight-item rating scale. The inter-rater reliability coefficients ranged from 0.65 to 0.93, demonstrating that good inter-rater agreement is possible when judging clinical performance. The authors postulated that factors contributing to the high level of agreement were that the examiners collaborated when developing the rating scales and had a shared definition of good clinical performance. This mutual development of the standards and definitions did not occur in the current study and would be recommended in future examination set ups. A shared definition also does not exist within the assessment realm of the profession. A major challenge to measuring clinical competency will be determining the level of performance that is indicative of competent and at what level a student can be deemed incompetent. Once determined, examiner training must aim to create a shared level of understanding of the standards of competency.

A recommendation from the study would be to explore the use of standards based assessment, whereby the profession is able to clearly articulate acceptable performance standards of its graduate entry across all domains of performance. Examiner training programs have been successfully implemented in medicine, which trains examiners with regard to performance standards and performance dimensionality (Williams et al. 2009). These programs typically provide examiners with samples of behaviour representing each level or standard of performance, practice at examining and the provision of feedback to the examiners. A suggestion from this research would be to clearly define and articulate the standards to the examiners, potentially enhancing the rigour of the assessment. It would also be advantageous to articulate such standards to the students to allow for greater transparency for assessment and a better framework to provide feedback to the student on their performance. By developing assessor capability, this would also possible assist increasing examiner confidence in being able to perform the assessment task.

5.2 Implications from this Research

A greater understanding of the factors that influence the assessment process was completed to assist the exercise physiology profession work towards a competent approach to the assessment of clinical competence. A concern was raised as to the varying backgrounds of our clinical examiners and whether this was impacting on the assessment process. This study confirmed that they did not and that it was the student's actual performance on task that contributed the greatest variance in ratings. Although this research used a station from the OSCE exam as a surrogate marker of student performance, it is important to note that no single measure of assessment should be utilized to determine student competency. At our University, assessment of clinical competency is determined by a suite of assessment modalities which include the Work-Integrated-Learning supervisor reports, VIVAs, OSCEs and practical exams. If a student fails a single component of assessment, all other assessment items are reviewed in order to make a global decision about a student's readiness to graduate and practice in the profession.

Competence-based assessment is a form of assessment derived from a specified set of outcomes and standards, which enable assessors to make an objective judgement of whether a student has achieved the standards and ultimately, ready to practice in the profession (Wolf, 2001). It is an integral part of determining readiness to practice in all medical and allied health

professions and could play a more prominent role in the exercise physiology profession. In order to integrate standards based assessment into the profession, it will be necessary for the profession to determine what the minimal level of acceptable competency a student must demonstrate in order to gain entry into the profession.

An awareness of the assessment process can also inform and enhance the development of assessor capabilities. Currently the profession relies on AEPs to both mentor and provide assessment reports on exercise physiology student competency, with little regard to the assessment capabilities of these mentors. Development of the assessor capabilities is an important one for the profession, with an acknowledgment of the important role the assessors play in determining the readiness of the next generation of AEPs to practice in the field. The ultimate implications are as a profession to work towards a more standardised approach to clinical assessment and better development of assessor capabilities to ensure that competent exercise physiology graduates are ready to practice safely and competently in the profession.

The first recommendation from this research would be that exercise physiology needs to focus on the development of clearly defined set of skill competencies and take on a standards-based approach to assessment. These standards should be generic across the profession and institutions. To achieve better clarity and alignment, the professional members and institutions need to collaborate when developing the assessment items and would benefit from being involved in a national approach to assessment, with the outcome being that the examination process clearly identifies the standards of performance of a graduate-entry level, exercise physiologists. This could result in the development of a shared definition of good clinical performance. This outcome has the potential to strengthen professional and community confidence in the safety and competence of new EP graduates, as well as to improve the quality of education, training and assessment this field. Having robust processes of assessment, with a strong evidence base for their outcomes, will likely assist in the recognition of the profession and contribute to being registrable under the national regulation authority, AHPRA.

Although the examiners didn't have a big influence on the performance ratings, the research did suggest that the profession could benefit from mentoring the less experienced examiners, designed to improve examiner capability and examiner confidence. Several more inexperienced examiners commented on their lack of confidence in the rating they made, especially after exploring the ratings with the researcher. The less confident examiners

suggested that an examiner training package, including samples of performance standards, practice sessions and providing the examiner with feedback would be beneficial. It was also deemed valuable to provide the less experience examiners with an examination mentor, believing they would be more confident to make a rating and importantly, fail a student if they have not reached the required performance level. A second recommendation from this research is to develop the professional capabilities of assessors, who share a common understanding of the assessment criteria and assessment standards for clinical competency.

Currently, exercise physiology accredits students into the profession based on a time-based paradigm, with a tendency to recognize the successful completion of 500 hours on placement, as opposed to abilities acquired. A review of the competency-based literature suggests that a focus of teaching, learning and assessment needs to ensure all graduates can demonstrate competency in all domains of their intended profession, with an emphasis on competencies and a de-emphasis on time-based training). Clinical assessment and quality judgement cannot occur in the absence of clearly defined competencies and standards of performance. Exercise Physiology may be better served moving to a competency-based model, requiring the successful demonstration of attainment of competency as part of the accreditation process. Competency thresholds must be clearly defined and understood by the examiner and student, in order to determine whether the student is competent and ready to enter the profession.

5.3 Limitations

There are several limitations to this research. Study participants were only examined on a single station, which was not representative of an OSCE. A typical OSCE examines students across many stations in an attempt to sample a broad range of competencies relevant to the profession. The views depicted in the current study are therefore, relevant to the station task examined. This study also only examined one type of assessment modality, that of the OSCE. To gain a broader picture of examiner judgment it is recommended to explore student capability through a suite of assessment types both within the University setting and within the Work-Integrated Learning environment. Finally, this study only captured the views of 23 examiners and is limited to the conclusions of these participants.

Chapter 6: Conclusions

The study concluded that despite the fact that the examiners had different backgrounds and reasons for formulating their ratings', the overall ratings were consistent and that it was the students varying levels of ability that contributed most to the variance in ratings. This finding was reassuring, as the examiners in the study had diverse backgrounds and experience in examining clinical scenarios and assessing clinical competency. When a University administers an Objective Structured Clinical Exam, it requires a large number of examiners. In order to staff the exam, Universities often need to look outside the academic staff and call on volunteer AEPs and other clinical staff associated with the University to assist in the assessment. The fact that such diversity in examiners did not have a major influence the assessment outcomes was reassuring for the future recruitment of potential examiners. The overall outcomes from the assessment outcomes also suggests that as a novice profession we are progressing well in clinical competency assessment, however, further exploration indicated that we could work towards consolidating examination standards and the examiner capabilities.

Ensuring that exercise physiology students are sufficiently competent to practice in the profession is of critical importance to the profession and is an essential aspect of curricula in higher education. We have attempted to explore factors that contribute to the examination process within a competency-based degree. This thesis advances knowledge in the field of examiner judgement by exploring the factors that influenced examiner judgment of exercise physiology student competency and lays important foundations for future work in understanding of the clinical examination process and the examiner capability to conduct the assessment. This improved understanding of the examination process could also be feedback into teaching and learning practices.

Currently, little consideration has gone into performance standards within the exercise physiology profession, with each University taking an isolated approach. What is needed is a common understanding of performance standards amongst all Universities and the professional body for the purpose of quality assurance and safety nationwide. Knowledge of expected competencies in exercise physiology and elucidation of essential standards of competency for practice could be a valuable contribution to the profession. The profession may also benefit from the development of examiner expertise by developing a training package for the examiners to become more expert users of performance assessment, and

working towards better standardisation within the examiner group and the profession as a whole.

Appendix A: Information Sheet & Consent



UNSW Medicine, School of Medical Sciences,

Clinical Examiner Participant Information Statement

Examiner Reliability and Judgment in an Objective Structured Clinical Examination (OSCE)

1. Introduction

Dr Fiona Naumann from the University of New South Wales is working collaboratively with the academic staff from the Faculty of Medicine and Learning and Teaching Unit to assess examiner reliability and the judgment process for assessing clinical competence during Objective Structured Clinical Exams.

This research project has been developed to measure the inter examiner reliability when grading student performance during the OSCE style of assessment, designed to appraise students clinical skills. This research is undertaken with the UNSW's Research Code of Conduct and the Australian Code for the Responsible Conduct of Research in mind.

2. Aim of the research

The aim of this research is to determine examiner reliability in an objective structured clinical examination (OSCE) in the exercise physiology course at UNSW. The accuracy of OSCE assessment is related to the reliability of the scoring sheet and to the reliability of the examiner. Consistency of examiner judgment of student performance against the scoring criteria is vital to success. Therefore, the aim of the research is to test the consistency of different examiners to grade students' performance so that future iterations of the OSCE exam are equitable. Additionally, the research also seeks to document the factors which influence examiners judgment of clinical performance. An improved understanding of these factors could inform assessor training of key attributes that are important for clinical competence in the profession. Similarly, explicit statements about the factors assessors notice could enhance the potential educational impact because it can inform the preparation of students whilst training at clinical placements and preparing to perform the examination.

3. What does the research involve?

The research involves inviting 20 clinical examiners to assess 3 student performances, on one clinical task or OSCE station. The OSCE station will involve a patient who will follow a script. The filmed performances will be approved by the student and then viewed by the clinical examiners. The assessments will be pooled to compare examiner reliability for assessment of the exercise physiology objective structured clinical examinations (OSCEs). Once the students' performance is viewed and assessed, the footage will be returned to Dr Fiona Naumann. As the student's performance is a mock or simulated OSCE, the examiners results will not contribute to the students' formal assessment.

At the completion of the mock assessment, the examiner will be interviewed to determine the factors that contributed to them making the various clinical judgments of student

performance. At the completion of the study, it is planned to store the student performances on file for up to 7 years, to be used as a training tool for future examiners and exercise physiology students.

4. How long will I be involved?

Following written consent you will act as an examiner for 3 students on a mock OSCE station. This should take approximately 30 minutes. Following the examination process, you are required to participate in a semi-structured interview that explores the important factors which contributed to you making your clinical judgment of the students' performance. This interview will take a further 30 minutes. You will be asked to agree not to discuss the student's performance during mock examinations outside the bounds of the research project. This will involve approximately one hour of your time.

5. Can I withdraw from the research?

This research is undertaken in compliance with the National Statement on Ethical Conduct in Research Guidelines, Section 2. Participation in this research is completely voluntary - you are not under any obligation to consent and - if you do consent - consent may be withdrawn at any time without affecting your relationship with The University of New South Wales. Once you have been filmed you have the right to withdraw your consent *prior to* the review of their performance by the team of clinical educators for the purpose of moderating the assessment process.

6. Will anyone else know the results?

The findings of this research will be submitted for publication. The video footage of the mock OSCE performance will be used as a training tool to train future clinical examiners and exercise physiology students. A disclaimer will be made that performances are mock and do not necessarily reflect the real student ability.

7. Will the study benefit me?

This study will not benefit particular students or clinical examiners but in the long term the information gained from the research is designed to enhance the rigor of the OSCE assessment process in Exercise Physiology undergraduate education.

8. What if I require further information?

The study has received approval from the HREA Biomedical Panel, at the University of New South Wales, Approval No. HC 12104. The University of New South Wales requires that all participants are informed that if they have any concerns regarding the manner in which this research is conducted they may forward them to the Ethics Secretariat, UNSW, Tel 9385 4234, Fax 9385 6648, email ethics.sec@unsw.edu.au

9. Where can examiners get help if I feel distressed during or after examining?

If you become distressed while examining you may stop and withdraw from the study at any time. Dr Fiona Naumann will be conducting the semi-structured interview, so please indicate your willingness to stop and the session will be terminated. Below are the contact details of the researchers involved who can assist you with further information.

Name	Contact details
Dr Fiona Naumann	f.naumann@unsw.edu.au
Prof Philip Jones	philip.jones@unsw.edu.au Ph: 02 9385 3375

Signed

Dr Fiona Naumann



Exercise Physiology Clinical Examiner Consent Form

Examiner Reliability and Judgment in an Objective Structured Clinical Examination

(HC 12104)

I am an Exercise Physiology Clinical Examiner and I consent to:

☐ The investigators observing my work as I assess students clinical skills during a *mock* OSCE Assessment.

☐ Completing an interview regarding what factors contributed to my clinical judgment scores of the student performances on the OSCE.

I understand:

☐ My name will not be mentioned

☐ No details of my health status will be noted

☐ The research project will be conducted as outlined in the Participant Information Statement, a copy of which I have retained

☐ I will have the opportunity to have questions answered to my satisfaction at any time during this research project

☐ The researcher intends to publish the results and the report will include my contribution but will not identify me in anyway.

Signature of participant

Please PRINT name

Date

Signature of investigator

Please PRINT name

Date

Ethics Approval

The study has received approval from the HRE Biomedical Panel, of the University of New South Wales, Approval No. HC 12104. The University of New South Wales requires that all participants are informed that if they have any concerns regarding the manner in which this research is conducted they may forward them to the researcher(s) f.naumann@unsw.edu.au or to the Ethics Secretariat, UNSW, Tel 9385 4234, Fax 9385 6648, email ethics.sec@unsw.edu.au

Appendix B: Station Task Descriptor

Station 1: Assessment of Cardiovascular Fitness in Chronic Disease

Description of Station:

A 44 year old male presents with poor fitness having recently been diagnosed with ischaemic heart disease and commenced on medication for his angina. The patient attended the Prince of Wales Hospital Cardiac Rehabilitation Outpatient Program. The patient has been referred to an exercise clinic to improve his baseline fitness.

This station involves taking a brief cardiovascular disease history, documentation of co-morbidities, reviewing past physical activity levels, selecting an appropriate test of cardiovascular fitness and explaining the test to the client (including start and finish points). Finally, the student needs to set the patient up for testing, including adjustments of the cycle and commence the first minute of the test. The student is required to monitor the patient closely during the minute. Finally, towards the end of the first minute, the student is required to take an exercise heart rate reading.

The patient would be a male volunteer surrogate, aged 30 to 40 years.

Station Focus: Exercise Physiology Procedural Skills.

Instructions to Student:

The patient to whom you will be introduced has presented to your exercise clinic. He has attended the Prince of Wales Hospital Outpatient Cardiac Rehabilitation Program. He is interested in commencing an exercise program and needs to have his cardio-respiratory fitness assessed.

You will be expected to:

- greet the patient,
- take a brief cardiac medical history
- assess the patient for other co-morbidities
- take a thorough physical activity history.

After completing the history you will be asked to:

- select an appropriate cardio-respiratory fitness assessment, that can be performed on a stationary cycle
- explain the procedure to the patient
- set the patient up ready to commence the test
- conduct the first minute of the exercise protocol, monitoring the patient closely
- take an exercise heart rate towards the end of the minute.

Instructions to Examiner:

Aim(s): Assess the student's ability to communicate effectively with a patient, take a history relevant to cardiovascular disease, select an appropriate cardio-respiratory assessment, explain the test to the patient (including the test end point), set the patient up for testing and monitor the patient during the first minute of the test. An exercise heart rate should be recorded towards the end of this minute.

Description of station:

- The student is expected to introduce themselves to the patient and put the patient at ease.
- The student must then take a focused history of the patient's presenting complaint.
- Line of questioning needs to be in a logical sequence and efficiently details to extract all relevant information.
- In the **medical history** the student should elicit the following features:
 - Dates associated with diagnosis, treatment.
 - Patient's participation in the cardiac outpatient program
 - Patient's co-morbidities (hypertension, diabetes and hyperlipidemia)
 - The student's questioning should be able to recognise clinical features which would make exercise testing in this man contra-indicated eg unstable angina, heart failure, low BP?
- In documenting the CVD history, the student should be able to:
 - Recognize that the angina was fairly mild, brought on through exertion and is now stable.
 - The patient is now on daily aspirin and nicorandil
- *Nicorandil (brand name Ikorel) is a newer medication for the treatment of stable angina. It works by opening the coronary arteries and relaxing muscle in the blood vessels and so reduces angina.*
- The student also needs to document the patient's **physical activity history**, both past and current.
- Next, the student should use the patient history to select an appropriate cardiovascular fitness test that can be performed on a cycle.
- In selecting a suitable exercise test, the student is best to select:
 - A sub-maximal test and a cycle protocol – YMCA, Astrand
- The student is then required to **explain the purpose and procedure** of the test to the patient. This needs to be clear and concise, in a manner that can be understood by the participant.
- The student is required to **set the patient up for testing**, attaching a heart rate monitor.
- The student should correctly set the bike up for the patient (seat approx hip height, with a slight bend in the knees and stable hips) and attach a Monark heart rate monitor, wetting the electrodes and assisting the patient to hook the heart rate band together.
 - Finally, the student should instruct the patient to commence the first workload, closely **monitoring the patient** and taking an exercise heart rate towards the end of this minute. In monitoring the patient, the student is best to:
 - Refer to the RPE scale to register how the patient is feeling
 - Record the first workload exercise heart rate

GRADING

Criteria 1: Communication - Description

Introduction to patient, purpose of meeting, rapport, questioning technique, active listening to patient, explanation of assessment, task or encounter, patient education.

P+

- Warmly introduces themselves to the patient and determine who the patient is and reason for visit
- Excellent rapport with the patient
- Used a mix of open and closed questions appropriately
- Listened actively to the patient and responds appropriately
- Able to explain clearly and concisely to the patient the purpose of the assessment, task or encounter

P

- Introduce themselves to the patient and determine who the patient is and reason for visit
- Very good rapport with the patient
- Mostly used open and closed questions appropriately
- Listened to the patient
- Able to explain clearly to the patient the purpose of the assessment, task or encounter

P-

- Casual introduction to the patient and didn't clearly determine who the patient is and reason for visit
- Good rapport with the patient
- Used mainly closed questions or questions that casual
- Listened to the patient sometimes
- Able to explain to the patient the purpose of the assessment, task or encounter

F

- No introduction to the patient and didn't clearly determine who the patient is and reason for visit
- Poor rapport with the patient
- Used poor questions
- Interrupted the patient
- Did not listen to the patient
- Failed to explain clearly to the patient the purpose of the assessment, task or encounter

Criteria 2: Exercise Physiology Procedural Skills

Client history, well executed assessment, task or encounter, with attention to patient safety, comprehensive assessment, task or encounter, providing feedback throughout the execution, patient monitoring during encounter.

P+

- Able to perform a complete and relevant history of the patients' condition, which was well organised and efficient
- Performed a well executed assessment, task or encounter, with careful attention paid to patient safety.
- Able to perform a comprehensive assessment, task or encounter, providing effective feedback throughout the execution
- Effective monitoring of the patient during the assessment, task or encounter

P

- Able to perform a relevant history, which is organised and logical.
- Performed a good assessment, task or encounter, with careful attention paid to patient safety.
- Basic monitoring of the patient during the assessment, task or encounter

P-

- Able to perform a basic history, which is somewhat organised.
- Performed a basic assessment, task or encounter, with some attention paid to patient safety.
- Limited monitoring of the patient during the assessment, task or encounter

F

- Inadequate, disorganised history, illogical sequence.
- Performed a poorly executed assessment, task or encounter, with no or minimal attention paid to patient safety.
- Limited or no monitoring of the patient during the assessment, task or encounter

Criteria 3: Exercise Physiology Technical Skills

Correct set up for testing, correct use of equipment, correct monitoring of patient during assessment, accurate results obtained.

P+

- Efficient and correct set up of the patient for testing
- Efficient and correct use of equipment
- Efficient and correct monitoring of a patient during an assessment
- Efficient and accurate result obtained

P

- Correct set up for testing
- Correct use of equipment
- Correct monitoring of a patient during an assessment
- Accurate result obtained

P-

- Poor set up of the patient for testing
- Poor use of equipment
- Limited monitoring of a patient during an assessment
- Poor result obtained

F

- Incorrect set up the patient for testing
- Incorrect use of equipment
- No monitoring of a patient during an assessment
- Inaccurate result obtained

Information for Patient :

Your story:

- You are a male aged 35 years.
- You are receiving medical treatment for mild angina.
- Your treatment involved medication to control the angina which now includes a daily aspirin tablet and a daily Ikorel tablet.

Nicorandil

- *Nicorandil (brand name Ikorel) is a newer medication for the treatment of stable angina. It works by opening the coronary arteries and relaxing muscle in the blood vessels and so reduces angina.*
- You have just completed participation in an 8 week outpatient cardiac rehabilitation program at the Prince of Wales Hospital. This involved lifestyle education and an exercise program.
- You finished the program 2 weeks ago and are keen to build on your exercise regime and improve your fitness
- Your past health was poor and included limited exercise.
- You have hypertension and hyperlipidemia
- Fortunately, you do not have diabetes.
- You would like to keep active.

Appendix C: Semi-Structured Interview Guide

Global Questions on Student Performance

1. On a global level, how did you rate the three student performances?

What was your rationale behind the ratings?

Domain Specific Questions on Student Performance

2. In the communication domain what aspects of the students' performance were key to you making your rating?
3. In the technical domain what aspects of the students' performance were key to you making your rating?
4. In the exercise physiology procedural domain what aspects of the students' performance were key to you making your rating?

Exercise Physiology Competency Question

5. In your opinion, are each of the students observed in the video ready to practice as exercise physiologist? What was your rationale for these decisions?

References

- Albanese M.A. (2000) Challenges in using rater judgments in medical education. *J Eval Med Educ.* 6:305-319.
- Bartfay, W., Rombough, R., Howse, E., Leblanc, R. (2004) The OSCE approach in nursing education. *Canadian Nurse* 100: 18–23.
- Boud, D. (1995). Assessment and learning: contradictory or complementary? In P. Knight (Ed.) *Assessment for Learning in Higher Education*. London: Kogan Page, 35-48.
- Brenner, M. (1985) Intensive interviewing, in: M. Brenner, J. Brown & D. Canter (Eds) *Research Interview: uses and approaches* (London, Academic Press).
- Cohen G.S., Henry N.L., Dodd P.E. (1990) A self-study of clinical evaluation in the McMaster clerkship. *Med Teacher.* 12:62;265-272.
- Cooper, L., Orrell, J., & Bowden, M. (2010). *Work integrated learning: A guide to effective practice*. Routledge: USA.
- Cox, K. (2000) Examining and recording clinical performance: a critique and some recommendations. *Education for Health*, 13:45-52.
- Creswell, J.W. (1998). *Qualitative inquiry and research design: Choosing among five traditions*. Thousand Oaks, CA: Sage.
- Denzin, N.K. and Lincoln, Y.S. (1984), *Handbook of Qualitative Research*, Newbury Park: Sage Publications.
- Dochy, F. and McDowell, L. (1997). 'Assessment as a tool for learning', *Studies in Educational Evaluation* 23(4), 279-298.
- Downing, S.M. (2004) Reliability: on the reproducibility of assessment data. *Medical Education.* 38:9;1006-1012.
- Downing S.M. (2005) Threats to validity of clinical teaching assessments: What about rater error? *Medical Education.* 39;350-355.
- Dudek N.L, Marks M,B, Reghr G. (2005) Failure to Fail: The Perspectives of Clinical Supervisors. *Academic Medicine*, 80:10;S84-S87.
- Epstein, R. M., Hundert, E.M. (2002) Defining and assessing professional competence. *J of American Medical Association.* 287;226-235.
- ESSA Exercise Physiology Scope of Practice (2012) http://www.essa.org.au/wp/wp-content/uploads/AEP_scope_of_practice.pdf.
- Flick, U. (2002) Qualitative research — State of the art. *Social Science Information*, 41:1;5–24.

Fontana A. & Frey., J. (1994) Interviewing: the art of science, in: N. DENZIN & Y. LINCOLN (Eds) *Handbook of Qualitative Research*, pp. 361–376 (Thousand Oaks, CA, Sage).

Frank J.R., Snell L.S., Ten Cate O., Holmboe E.S., Carraccio C., Swing S.R., Harris P., Glasgow N.J., Campbell C., Dath D., Harden R., Iobst W., Long D.M., Mungroo R., Richardson D.L., Sherbino J., Silver I., Taber S., Talbot M., Harris K.A. (2010) Competency-based medical education: theory to practice. *Medical Teacher* 32:638-645.

Glesne, C. and Peshkin, A. (1992), *Becoming Qualitative Researchers An Introduction*, Longman, New York.

Covaerts, M., Van der Vleuten C, Schuwirth L, Muijtijens A. (2007) Broadening perspectives on clinical performance assessment: Rethinking the nature of in-training assessment. *Advances in Health Sciences Education*. 12:239-260.

Harden, R.M., Stevenson, M., Downie, W.W., Wilson, G.M., (1975). Assessment of clinical competency using objective structured examination. *British Medical Journal* 1: 447–451.

Hefti A., Preshaw P.M. (2012) Examiner alignment and assessment in clinical periodontal research. *Periodontology*. 59:41-60.

Heinemain R.L., Wesley K.N. (1983) The effect of delay in rater and the amount of information observed on performance rating accuracy. *Acad Management J*. 26:677-686.

Hodges D., McNaughton N. (2009) Who should be an OSCE Examiner. *Academic Psychiatry*. 33:4;282-284.

Hodges, D. (2011) The assessment of student learning in cooperative and work-integrated education. In R. K. Coll & K. Zegwaard (Eds.), *International handbook for cooperative and work-integrated education: International perspectives of theory, research and practice* (2nd ed., pp. 53–62). Hamilton, New Zealand: Waikato Print.

Holmboe E.S., Hawkins R.E., Huot S.J. (2004) Effect of training in direct observation of medical residents' clinical competence: A randomized trial. *Am College of Physicians*. 140:874-881.

Holmboe E.S. (2008) Assessment of the practicing physician: Challenges and opportunities. *J Continuing Education Health Professional*. 28:S4-S10.

Holmboe E.S., Sherbino J, Long D.M., Swing S.R., Frank J.R. (2010) The role of assessment in competency-based medical education. *Medical Teacher* 32:676-682.

Homer M, Pell G. (2009) The impact of the inclusion of simulated patient ratings on the reliability of OSCE assessments under the borderline regression method. *Med Teach* 31:5;420–425.

Howard, E. (1990) Measurement of student performance. In: Waltz, C., Strickland, O. (Eds) *Measurement of Nursing Outcomes*. 3. Springer, New York.

Johnston J.L., Lundy G., McCullough M., Gormley G.J. (2013) The view from over there: reframing the OSCE through the experience of standardised patient raters. *Medical Education*. 47:9;899-909.

Jones A. (2000) The place of judgement in competency-based assessment. *J of Vocational Education and Training*. 51:145-160.

Kane M.T. (1992) The assessment of professional competence. *Evaluation and Health Professionals*. 15:163-182.

Kogan JR, Hess BJ, Conforti LN, Holmboe ES. (2010) What drives faculty ratings of residents' clinical competence. The impact of faculty's own clinical skills. *Academic Medicine*. 85:25-28.

Kogan, J.R., Conforti L., Bernabeo E., Iobst W., Holmboe E. (2011) Opening the black box of clinical skills assessment via observation: a conceptual model. *Medical Education*. 45:1048-1060.

Lankshear, A. (1990) Failure to fail: the teacher's dilemma. *Nursing Standard* 4 20:35-37.

Lieblich A., Tuval-Mashiach, R., and Zilber, T. (1998). Narrative research: Reading, analysis and interpretation. Thousand Oaks, CA: Sage.

MacRae H.M., Vu N.V., Graham B., Word-Sims M., Colliver J.A., Robbs R.S., (1995) Comparing checklists and databases with physicians ratings as measures of students' history and physical-examination skills. *Academic Medicine*. 70:3;1-6.

Mazor K.M., Zanetti M.L., Alper E.J., Hatem D., Barrett SV, Meterko V, Gammon W, Pugnaire MP. (2007) Assessing professionalism in the context of an objective structured clinical examination: an in-depth study of the rating process. *Med Educ* 41:4;331–40.

Miller, G. (1990). The assessment of clinical skill, competence, performance. *Academic Medicine* 65:9;S63-S67.

Naumann F.L., Morre K., Jones P.J. (2012) Developing an Objective Structured Clinical Examination (OSCE) to Assess Clinical Competence. *ACEN Conference Proceedings*. Geelong.

Naumann, F.L., Moore K, Mildon S, Jones P. (2014) Developing an Objective Structured Clinical Examination to Assess Work-Integrated Learning in Exercise Physiology. *Asia Pacific Journal of Cooperative learning*. (In Press)

Norman G.R. (1985) Assessing Clinical Competence. New York: Springer; p330-341.

Patton, J. (1996). Analysis of thinking and research about qualitative methods. New Jersey: Lawrence Erlbaum.

Potter, J., Wetherell, M. (1994). Analyzing discourse, in Bryman, A., Burgess, B., (Eds), *Analyzing Qualitative Data*, London; Routledge.

Roch S.G, Woehr D.J, Mishra, V., Kieszczyńska, U. (2012) Rater training revisited: an updated meta-analytic review of frame of reference training. *Journal of Occupational and Organisational Psychology*. 85:2;370-395.

Sadler R. (2005) Interpretations of criteria-based assessment and grading in higher education. *Assessment and Evaluation in Higher Education*. 30:2;175-194.

Sadler, D .R. (2010). Fidelity as a precondition for integrity in grading academic achievement. *Assessment and Evaluation in Higher Education*, 35(6), 727-743.

Selig, S., Torode, M., Coombes, J., Groeller, H., Spinks, W., Leicht, A., LeRossignol, P., McDonald, M., Otago, L., Pascoe, D., Raymond, J. (2008) *Meeting the challenges of clinical exercise science and practice : a collaborative university-industry approach*. PANDORA electronic collection. Australian Learning and Teaching Council, Strawberry Hills, N.S.W.

Shavelson R.J., Webb N.J. (1991) Generalizability theory: A primer. Newbury Park. CA. Sage.

Shaw, E. (1999) A guide to the qualitative research process: evidence from a small firm study. *Qualitative Market Research*, 2;2:59–70.

Stewart M, Brown JB, Weston WW, McWhinney IR, McWilliam CI, Freeman TR. Patient-Centred Medicine. Transforming the Clinical Method. Thousand Oaks, CA: Sage Publications; 1995.

Strauss, A. and Corbin, J. (1990), *Basics of Qualitative Research. Grounded Theory Procedures and Techniques*, Sage Publications, Newbury Park, USA.

Struyven K, Dochy F, Janssens S. (2005) Students' Perceptions about evaluation and assessment in higher education: a review. 30:4;325-341.

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

Tesch, R. (1990), *Qualitative Research: Analysis Types and Software Tools*. The Falmer Press, London, UK.

Thomas, D. (2006) A General Inductive Approach for Analyzing Qualitative Evaluation Data. *American Journal of Evaluation*. 27:237-246.

Tonesk X, Buchannan R.G. (1987) An AAMC pilot study by 10 medical schools of clinical evaluation of students. *J Med Education*. 62:707-718.

van Manen, M. (1990). *Researching lived experience: Human science for an action sensitive pedagogy*. Abany, NY: State University of New York Press.

Vukanovick-Criley J.M., Criley S., Warde C.M., Boker J.R., Guevara-Matheus L., Churchill W.H., Nelson W.P. (2006) Competency in cardiac examination skills in medical students, trainees, physicians and faculty: a multi-centre study. *Arch Internal Medicine*. 166:6:610-616.

Wass V., Van der Vleuten, C., Shatzer, J., Jones, R. (2001) Assessment of clinical competence. *The Lancet*, 357:9260;945-949.

Watson, R., Stimpson, A., Topping, A., Porock, D. (2002) Clinical competence in nursing: a systematic review of the literature. *Journal of Advanced Nursing*. 39;5:421-431.

Westward O, Griffin A., Hay F. (2013) How to Assess Students and Trainees in Medicine and Health. Wiley Publishers. USA.

Williams R.G., Klamen D.A., McGaghie W.C. (2003) Cognitive, Social and Environmental Sources of Bias in Clinical Performance Ratings. *Teaching & Learning in Medicine: An International Journal*. 15:4;270-292.

Woehr, D.J. (1994) Rater training for performance appraisal: A quantitative review. *J of Occupational and Organisational Psychology*. 67:189-205.

Wolf, A. (2001) Competency Based Assessment. Chapter 25, in Competence in the Learning Society, John Raven and John Stephenson (eds), published Peter Lang, New York, 453-466.

Yorke, M. (2011). Work-engaged learning: towards a paradigm shift in assessment. *Quality in Higher Education*. 17:1;117-130.