



Libraries and Learning Services

University of Auckland Research Repository, ResearchSpace

Copyright Statement

The digital copy of this thesis is protected by the Copyright Act 1994 (New Zealand).

This thesis may be consulted by you, provided you comply with the provisions of the Act and the following conditions of use:

- Any use you make of these documents or images must be for research or private study purposes only, and you may not make them available to any other person.
- Authors control the copyright of their thesis. You will recognize the author's right to be identified as the author of this thesis, and due acknowledgement will be made to the author where appropriate.
- You will obtain the author's permission before publishing any material from their thesis.

General copyright and disclaimer

In addition to the above conditions, authors give their consent for the digital copy of their work to be used subject to the conditions specified on the [Library Thesis Consent Form](#) and [Deposit Licence](#).

Evaluating the Mini-Clinical Evaluation Exercise as an
assessment tool for Fourth Year Medical Students at the
University of Auckland.

Janak Rashme de Zoysa

A thesis submitted in partial fulfilment of the requirements for the degree of Masters of
Clinical Education. The University of Auckland, 2019.

Abstract.

Background.

In 2013 the University of Auckland medical programme was reinvigorated which expanded the use of the mini-clinical evaluation exercise (mini-CEX) assessment tool. In Year 4, the mini-CEX is used twice and along with nine other assessments, contributes to the Clinical and Communication Skills domain grade.

Aim.

There were two main aims. First, to explore how the mini-CEX was being used in Year 4 students, and secondly, to evaluate the contribution of the mini-CEX, and other assessments, to the grade of the Communication and Clinical Skills domain.

Methods.

Five years of de-identified data, from Year 4 students (2013 – 2017) was extracted. Demographic data, entry characteristics, training site, mini-CEX grade, Clinical and Communication Skills grade and the grades for other clinical assessments were obtained. Detailed information on the mini-CEX evaluations undertaken in 2013 and 2014 were obtained including assessor and case characteristics. These were subjected to various analysis using FACTOR and the Statistical Package for the Social Sciences.

Results.

A total of 1203 Year 4 students were included who completed 1196 General Medicine and 1198 Specialty Medicine mini-CEX. The majority of students received a pass or distinction; only 1.5% assessments (36/2398) receiving a fail or borderline mark. Although there were significantly higher mini-CEX scores awarded based on demographic, attachment and assessor characteristic, the effect of these influences were trivial, and essentially did not impact the grade.

All clinical assessments, including the mini-CEXs, correlated with the grade for the Clinical and Communication Skills domain ($r_s = .189$ to $.350$, $N = 1192$ to 1201 , $p < .05$). A Cronbach's alpha of 0.64 was calculated for the Clinical and Communication Skills domain grade. A hypothetically reliable construct, which retained all assessments, grouping these into three clusters, is described.

Conclusions.

The mini-CEX is effectively used and provides a naturalistic assessment of patients. The mini-CEX contributes to the Clinical and Communication Skills domain grade, however, consideration should be given to reviewing and revising the rubric for this assessment.

Acknowledgements.

I very much appreciate the support and guidance that was provided by Professor Jenny Weller, my primary supervisor for this thesis. I am also very grateful for the support and feedback provided by Professor Warwick Bagg and Associate Professor Elana Curtis who were co-supervisors of this thesis.

I would like to thank Dr Oleg Medvedev and Dr Ryan San Diego for their enthusiasm and support and providing advice in statistical analysis of this thesis.

Thanks goes to Ian Wood, in the Medical Programme Directorate at the Medical School, for providing me with the dataset and his support in cleaning up this data. I very much appreciate the support from Maggie Naidoo and Christine Ganly in the Department of Medicine at the University of Auckland for their help in completing the dataset.

I was helped by Mere Vercoe, Maggie Naidoo, Maria Vitas, Raewyn Wooderson, Irene Warren and Leonie Alley, the medical student coordinators at each clinical site, who advised me as to site specific details of the Year 4 students' rotations.

Special thanks to the Waitemata DHB library. Their efficiency and ability to obtain articles within a remarkably short turnaround time was amazing.

I appreciated the opportunity to discuss the assessment methods with Associate Professor Andy Wearn, Dr Nichola Wilson, Mrs Karen Hicks and Dr Kira Bacal.

I am grateful for the endless patience of my wife, Janeen. I have not spent as much time with her as I would like, nor as much as she deserves. And, of course, thanks to my children: Araliya, Lihini, Ishan and Rohan who constantly remind me what is important.

Table of Contents.

Title Page.....	i
Thesis Consent.....	ii
Abstract.....	iii
Acknowledgements.....	v
Table of Contents.....	vi
List of Figures.....	viii
List of Tables.....	ix
Abbreviations.....	x
Glossary.....	xii
Chapter 1 - Introduction.....	1
1.1 - The University of Auckland Medical School.....	2
1.2 - How is the mini-CEX used at the University of Auckland?.....	3
1.3 - Programmatic assessment and the assessment tools used in Year 4 students.....	6
1.4 - Workplace based assessment.....	7
1.5 - The student doctors' journey.....	8
1.6 - Purpose of this thesis.....	8
Chapter 2 - Use of the mini-CEX at the University of Auckland and medical schools in NZ and Australia.....	10
Chapter 3 - Narrative Literature Review of the mini-CEX.....	15
3.1 - The mini-CEX in Postgraduate Medical Education.....	16
3.2 - The mini-CEX in Undergraduate Medical Education.....	35
3.3 - Summary and Conclusions.....	43
Chapter 4 - The mini-CEX in Year 4 Medical Students at the University of Auckland.....	48
4.1 - Introduction.....	48
4.2 - Aim.....	49
4.3 - Ethical Statement.....	50
4.4 - Methods.....	50
4.5 - Results.....	52

4.6 - Discussion	63
Chapter 5 - Contribution of the mini-CEX and other assessments to the Clinical and Communications Skills domain	74
5.1 - Introduction.....	74
5.2 - Aim	76
5.3 - Methods	76
5.4 - Results	78
5.5 - Discussion	88
Chapter 6 - Discussion.....	97
6.1 - Introduction.....	97
6.2 - Use in the General Medicine and Specialty Medicine Rotations.....	98
6.3 - Use as a component of Programmatic Assessment.....	101
6.4 - Limitations	102
6.5 - Future Research	103
6.6 - Conclusions.....	104
References	105
Appendix 1	128
Appendix 2	135

List of Figures.

Figure 1 - Miller's Pyramid of Competence	7
Figure 2 - Flow chart of the literature selection process for the narrative review	16
Figure 3 - Number of mini-CEX assessments undertaken by assessors 2013 - 2014	54
Figure 4 - Histogram Plot of mini-CEX overall scores in General Medicine and Specialty Medicine	57
Figure 5 - Factorial Structure of the 11 item model of the Clinical and Communications skills domain grade	86
Figure 6 - Factorial Structure of the hypothesised 3 factor model of the Clinical and Communication Skills domain grade	87
Figure 7 - Possible rubric for marking of the hypothesised model of the Clinical and Communications Skills domain grade.....	92
Figure 8 - Rubric for marking of a possible revised model of the Clinical and Communications Skills domain grade	94
Figure 9 - Initial version of the University of Auckland mini-CEX form	128
Figure 10 - Current version of the University of Auckland mini-CEX form.....	129
Figure 11 - Page 1 of the Royal Australasian College of Physicians mini-CEX	130
Figure 12 - Page 2 of the Royal Australasian College of Physicians mini-CEX	131
Figure 13 - The University of Otago mini-CEX form.....	132
Figure 14 - The University of Wollongong's summative mini-CEX	133
Figure 15 - The University of Wollongong's formative mini-CEX	134

List of Tables.

Table 1 - Rubric to determine the final grade for the Clinical and Communication Skills Domain.....	4
Table 2 - - Medical Rotations at Year 4 University of Auckland Clinical Sites	5
Table 3 - Use of the mini-CEX in Australian medical schools.....	12
Table 4 - Demographic details of Year 4 medical students	53
Table 5 - Loadings on principal components	59
Table 6 - Marks achieve by Year 4 medical students for assessments contributing to the Clinical and Communication Skill domain grade	79
Table 7 - Correlation of clinical assessments contributing to the Clinical and Communication Skills domain grade	80
Table 8 - Item-total statistics for the Year 4 assessments	81
Table 9 - Standardised variance/covariance matrix of Year 4 Clinical and Communication Skills assessments	82
Table 10 - Explained variance based on eignenvvalues	83
Table 11 - Normalised varimax rotated loading matrix	84

Abbreviations.

The following abbreviations are used in the thesis.

AIC - Akaike's Information Criterion.

AMOS – Analysis of Movement Structures.

ANOVA - ANalysis Of VAriance.

BPT - Basic Physician Trainees.

CbD - Case-based Discussion.

CFI – Comparative Fit Index.

CSA - Clinical Skills Assessment.

DHB - District Health Board.

DOPS - Direct Observation of Procedural Skills.

EFA – Exploratory Factor Analysis.

EPA - Entrustable Professional Activities.

ERIC - Education Resources Information Center.

GFI – Goodness of Fit Index.

GPA - Grade Point Average.

G-theory - Generalisability theory.

IM – Internal Medicine.

IMG - International Medical Graduate.

KMO - Kaiser-Meyer-Olkin.

MANOVA – Multivariate analysis of variance.

MAPAS - Māori and Pacific Admission Scheme.

MBChB – Bachelor of Medicine Bachelor of Surgery.

Mini-CEX - Mini clinical evaluation exercise.

MSF - Multi-Source Feedback.

NZ – New Zealand.

OSCE - Observed Station Clinical Encounters.

PCA - Principal Component Analysis.

PDA - Personal Digital Assistant.

RACP – Royal Australasian College of Physicians.

RMO – Resident Medical Officer.

RMSEA - Root Mean Square Error of Approximation.

RRAS - Regional Rural Admission Scheme.

SEM - Structural Equations Modeling

SMO - Senior Medical Officer.

SPSS - Statistical Package for the Social Sciences.

UAHPEC - University of Auckland Human Participants Ethics Committee.

UMAT – Undergraduate Medicine and Health Sciences Admission Test.

UTAS - Undergraduate Targeted Admission Scheme.

UK - United Kingdom of Great Britain and Northern Ireland.

USA - United States of America.

WBA - Workplace Based Assessment.

Glossary.

AMOS – A statistical software package that allows simple models to be drawn displaying structural equations modeling. An add on to SPSS.

DHBs – District Health Boards. There are 20 DHBs which are responsible for funding or delivering healthcare in their district. However, some services may be funded or provided nationally by the New Zealand Ministry of Health.

ERIC - An authoritative database of indexed and full-text education literature and resources.

FACTOR – a free software programme developed at the Rovira i Virgili University, useful for polychoric correlation and exploratory factor analysis.

MBChB - The New Zealand medical degree.

RMOs – Junior doctors. These may be house officers or registrars. The equivalent of residents in the USA system.

SMOs – Senior doctors who have completed training. These are the equivalent of Faculty or Attending in the USA system.

SPSS – a commercial software package used widely for statistical analysis.

Whakapapa – a principle of Māori culture, which links an individual to their people and the land.

Whānau – a Māori language word for family and extended family. This includes emotional and spiritual components.

Chapter 1 - Introduction.

The mini-clinical evaluation exercise (mini-CEX) is a workplace based assessment (WBA) tool that has been widely adopted in medical education to assess clinical skills [Norcini, 1995]. The tool is easy to use, time-efficient and can evaluate trainees across a wide range of clinical settings. Importantly, it promotes immediate feedback and when used repeatedly has been shown to be reliable and valid in various clinical settings. The additional benefit of the mini-CEX is that trainees are evaluating patients in the workplace and thus have an assessment at the peak of Miller's "pyramid of competence" [Miller, 1990]. With the mini-CEX they are "showing how" they would interact with patients. Assessment drives learning [Newble, 1983; Van der Vleuten, 1996]; by assessing what a trainee does when interacting with a patient, we are examining real clinical practice and promoting clinical interactions rather than rote learning [Van der Vleuten, 2000].

My first experience with the mini-CEX was as an evaluation tool for the assessment of Basic Physician Trainees (BPTs) with the Royal Australasian College of Physicians (RACP) when it was introduced as a WBA in approximately 2008. Currently, the mini-CEX is used for formative assessment of BPTs who are required to complete eight mini-CEX over a 24 month period [The RACP, 2018]. The mini-CEX is not required for advanced trainees in Nephrology [The RACP, 2019]. Typically, I perform 4-6 mini-CEX assessments of registered medical officers (RMOs) annually at Waitematā District Health Board (DHB). In addition, the mini-CEX is used for the assessment of medical students from the University of Auckland. Since 2005 I have coordinated the Renal Medicine rotation, first at Auckland DHB and then at Waitematā DHB. In this role I assess medical students during their Specialty Medicine rotation and since 2013 have used the mini-CEX as part of the assessment of Year 4 medical students. Typically, I would perform ten Year 4 mini-CEX assessments annually during Specialty Medicine rotations. In addition, in the last two years, I have performed mini-CEX assessments of Year 4 students in their General Medicine rotation at Waitematā DHB. It was

in this context that I became interested in better understanding how the mini-CEX is used in Year 4 students from the University of Auckland.

1.1 - The University of Auckland Medical School.

The University of Auckland is one of two universities that offer medical degrees in New Zealand (NZ). The University of Auckland Bachelor of Medicine Bachelor of Surgery (MBChB) programme is a six year course with limited places available in the medical school. There has been a steady increase in the number of places being offered: in 2006 there were 148 places increasing to 282 in 2018; however, of note more than seven students apply for each place available. Students entering have a number of potential routes of admission with selection based on a combination of academic merit and personal qualities. Candidates may either be first year or graduate entrants. The candidate may enter via the general admission scheme, Regional Rural Admission Scheme (RRAS), Māori and Pacific Admission Scheme (MAPAS), Undergraduate Targeted Admission Scheme (UTAS) or be an International student. The majority of students enter via the general entry scheme with student selected from academic performance in their first year of study, a structured interview and the undergraduate medicine and health sciences admission test (UMAT), a standardised psychometric test. The RRAS is designed to train candidates from regional NZ in the hope that they will return to the areas which have a shortage of health care professionals. Candidates must have undertaken most or all of their undergraduate education in non-metropolitan NZ. The process of selection is no different to general entrant candidates, and includes year 1 academic grades, a structured interview and the UMAT, but candidates are ranked separately. MAPAS is part of the project to increase Māori and Pacific health care professionals to comprise 10% of health workforce. Candidates must be of Māori or Pacific whakapapa, and have a measure of prior academic performance and a MAPAS-specific interview and have separate ranking [Curtis, 2013]. The UTAS scheme allows entry for candidates from specific equity groups, including but not limited to a candidate with a disability or those who have been refugees.

1.2 - How is the mini-CEX used at the University of Auckland?

The first three years of the University of Auckland MBChB are pre-clinical; the fourth year introduces the student to clinical attachments. Currently, Year 4 students are based at one of six clinical sites: Waitematā, Auckland, South Auckland, Waikato, Rotorua and Tauranga. Rotations (also known as clinical attachments or clerkships) are undertaken in General Medicine, two specialty Medicine rotations, Geriatrics, General Practice, General Surgery, Musculoskeletal, Anaesthesiology and Emergency Medicine.

Five key domains are identified as important for the clinical practice of Auckland students which are: Applied Science for Medicine, Clinical and Communication Skills, Personal and Professional Skills, Hauora Māori and Population Health [University of Auckland Phase 2 (Year 4) Guidebook].

In 2013 the mini-CEX was introduced as an assessment in Year 4 General Medicine and Year 4 Specialty Medicine rotations. The mini-CEX and the clinical supervisors report contributes to the grade for the General Medicine rotation. The student does two specialty medicine runs and thus two clinical supervisors reports and one mini-CEX contribute to the grade for the Specialty Medicine rotation. In addition, the two mini-CEX contribute to the overall grade for Clinical and Communication Skills domain, along with two Musculoskeletal Objective Structured Clinical Encounters (OSCE) three Surgical Clinical Skills Assessments (CSAs), three Medicine CSAs, and one Drug & Alcohol assessment, which are undertaken through Year 4. Each of the eleven assessments is assigned a grade of distinction, pass, borderline performance or fail. The Clinical and Communication Skills domain overall grade may be a distinction, pass or fail. A student must obtain 5 distinctions with no more than one borderline performance and no grade which was a fail to achieve a distinction. A student will fail if they fail three or more of the eleven assessments; or they have two fails and two borderline performances; or one fail and three borderline performances; or four or more borderline performances. A pass grade is assigned for any combination which is not a distinction and not a fail [Table 1, adapted from University of Auckland Phase 2 (Year 4)

Guidebook]. Thus, the mini-CEX contributes as an assessment of the student at two time points, during the General Medicine and Specialty Medicine clinical rotations, and also as a longitudinal assessment for the year, as part of the clinical and communication skills domain.

Table 1 - Rubric to determine the final grade for the Clinical and Communication Skills Domain.

Grade	Requirement
Distinction	Distinction in five or more assessments of clinical skills and: No more than one borderline performance, and, No fail in clinical skills assessments.
Pass	Not meeting the criteria for fail or distinction.
Fail	Fail is assigned if any one of the following occurs: <ul style="list-style-type: none"> • Three or more fails of assessments of clinical skills assessments; or • Two fails and one borderline performances; or • One fail and three borderline performances; or • Four or more borderline performances.

Adapted from University of Auckland Phase 2 (Year 4) Guidebook.

Renal Medicine, is one of several possible Specialty Medicine rotations that students may undertake, however, this is not offered at all sites. In fact, how Internal Medicine and Specialty Medicine rotations are delivered is quite different across the Year 4 clinical sites. Rotorua Hospital and Tauranga Hospital are smaller teaching hospitals. Waikato, South Auckland, Waitematā and Auckland are larger teaching hospitals. In Rotorua, General Medicine and Specialty Medicine are integrated with each medical team having a specialty in which they have an interest. In Tauranga, students will do General Medicine with some, but not all teams, having an interest in Specialty Medicine areas, and all students will have specialty rotations in Cardiology and Oncology. In the larger hospitals there is a greater demarcation between General and Specialty Medicine and a wider range of Specialty Medicine rotations are offered (Table 2). In general, there are one to two students per team per rotation although the Cardiology rotation in Waitematā DHB will take up to three students.

Table 2 - Medical Rotations at Year 4 University of Auckland Clinical Sites.

Teaching Site	Hospital Beds	General medical Teams (Students per team)	Specialty Medicine Rotations (Students per team)
Waitematā	953	Ten General Medicine teams (one student per team with students rotating between up to three teams).	Cardiology (1-3) Diabetes/Endocrinology (1-2) Gastroenterology/Hepatology (1-2) Respiratory (1-2) Infectious Disease (1)
Auckland	1124	Twelve General Medicine teams (1).	Respiratory (1-2) Cardiology (1-2) Diabetes/Endocrinology (1-2) Neurology (1-2) Oncology (1-2) Renal (1-2) Gastroenterology (1-2) Infectious Disease (1-2)
South Auckland	745	Twelve General Medicine teams (1).	Cardiology Ward (1) Cardiology Coronary Care Unit (1) Endocrinology (1) Gastroenterology (1-2) Haematology (1) Renal Medicine (1-2) Respiratory Green team (1) Respiratory Blue team (1) Rheumatology (1) Stroke (1)
Waikato	673	Six General Medicine teams (1-2).	Oncology (1) Renal (1) Gastroenterology (1) Endocrinology (1) Respiratory (1) Cardiology (1-2) Haematology (1) Neurology (1)
Tauranga	360	General Medicine teams with an interest in: Respiratory (1) Infectious Diseases (1) Endocrinology (1) General Medicine (1)	Cardiology (1-2) Oncology (1-2)
Rotorua	233	General Medicine teams with an interest in: Respiratory (2) Cardiology (2) Diabetes/Endocrinology (2) Gastroenterology (2)	

Personal Comms (Mere Vercoe, Maggie Naidoo, Maria Vitas, Raewyn Wooderson, Irene Warren and Leonie Alley. Medical Student Co-ordinators at Waitematā (MV), Auckland (MN), South Auckland (MV), Waikato (RW), Tauranga (IW), and Rotorua (LA)).

The mini-CEX is also used in Year 5 during rotations in Paediatrics, Psychiatry and General Practice and in Year 6 during rotations in General Practice, Obstetrics and Gynecology, Paediatrics and Surgery.

1.3 - Programmatic assessment and the assessment tools used in Year 4 students.

In 2013 the University of Auckland MBChB programme was reinvigorated and adopted a model of Programmatic Assessment [van der Vleuten, 2012]. Programmatic assessment takes a holistic approach to assessment; high stakes decisions such as the final grade, are based on an aggregation of multiple data points, at multiple time points across multiple interactions. This assumes that by combining multiple assessment methods that the validity and reliability of the overall assessment will be improved. In other words, the intention of programmatic assessment is that if one assessment tool fails to effectively evaluate an aspect of the student's learning, then another method will evaluate this aspect. In the case of Year 4 medical student the grade for Clinical and Communications Skills is assigned using two WBAs (the mini-CEXs), two OSCE stations (two musculoskeletal), one standardised patient assessment (Drug and Alcohol) and six clinical skills assessments.

In contrast, prior to 2013, the clinical supervisors report and two short cases were undertaken during the medicine rotation. The student could fail one short case and fail the rotation, leading to the student having to repeat the medicine rotation. The specialty medicine rotation was assessed from the Clinical Supervisors report and written case reports. The student would also need to take an end-of-year "barrier" assessment, at the end of Year 5, consisting of medical and surgical short cases occurring over one day. Thus, we see each mini-CEX used currently, as a single low-stakes assessment, contributing a data point both as a single time point and to longitudinal assessment, for the Year 4 medical student.

1.4 - Workplace based assessment.

An additional benefit of the mini-CEX is that the assessment is performed in the workplace. Student assessments such as the OSCE or standardised patient simulation assess the students' ability to perform a task under controlled conditions, and assess the student's competence. For postgraduate doctors there has been a move away from standardised assessments as they may not correlate with their performance in clinical practice [Rethans, 2002]. WBA with tools such as Direct Observation of Procedural Skills (DOPS), Case-based discussion (CbD) and the mini-CEX are now used widely to evaluate trainees during everyday practice. It is recognised that assessing trainees in the workplace gives greater insight into how the person behaves in a real-world setting and is a better measure of their performance. Introduction of WBA for student doctors offers the opportunity to assess what a student "does" rather than the student "showing how". For the student the mini-CEX, OSCE and assessment of the standardised patient are all near to the peak of Miller's pyramid of competency; they are "showing how" they would interact with patient rather than traditional assessments of multiple choice questions or short answers where they demonstrate "knows" or "knows how" [Figure 1 Miller, 1990].

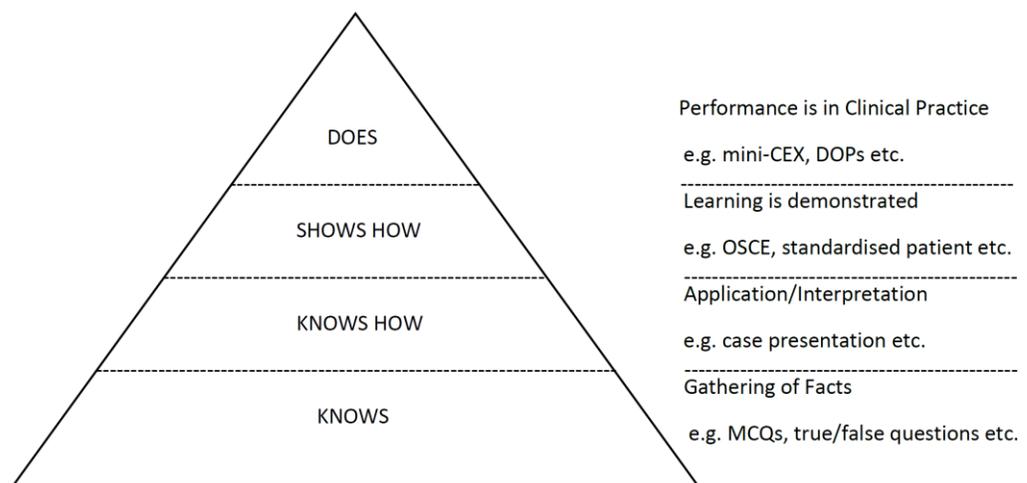


Figure 1 - Miller's Pyramid of Competency

Adapted from Miller, GE. 1990. The assessment of clinical skills/competence/performance. *Academic Medicine*, 65: s63-s67.

1.5 - The student doctors' journey.

The transition from lecture-based learning to clinical interactions is a major step in the student doctors' journey. The Dreyfuss model describes skill acquisition and professional development where learners progress through a series of stages: novice, advanced beginner, competent, proficient and expert [Dreyfus, 1980]. Using this model Year 4 Auckland medical students are novices, who should be "memorizing and following rules" in how they interact, evaluate and treat patients. To develop, the medical student should have monitoring, "either by self-observation or instructional feedback, so as to bring his behavior more and more completely into conformity with the rule" [*sic*] [Dreyfus, 1980]. The mini-CEX offers a valuable opportunity to do this. However, the risk with using the mini-CEX in this fashion is that the student's focus will be on obtaining a good grade and this will detract from feedback, reflection and the learning opportunity. This concern was raised in a study from the Netherlands by teaching faculty [Daelmans 2016]. In contrast the RACP require BPT, who could be viewed as being at the competent stage in the Dreyfuss model, to complete eight mini-CEX over a 24 month period and use the mini-CEX as a formative assessment [The RACP, 2018]. This allows the BPT to focus on feedback and skill acquisition rather than on a grade. So here we have a potential conflict: a strength of programmatic assessment is that it recognises assessment can be used for learning. Using the mini-CEX purely as a formative assessment would embody this approach.

1.6 - Purpose of this thesis.

An important component of programmatic assessment is the recognition that the programme of assessments can influence learning, as does the medical curriculum itself, and thus should be reviewed. The main purpose of this thesis is to investigate how mini-CEX assessments contribute in the context of programmatic assessment. To do this I will analyse how they function as an assessment tool at the University of Auckland medical programme and how they contribute to the grade for the Clinical and Communication Skills domain.

For this dissertation Chapter one provides an overview of the mini-CEX and describes its use in the context of the University of Auckland MBChB programme. I was interested in how the mini-CEX was used in undergraduate medical education in other centres and Chapter two will summarise an informal survey of how the mini-CEX is used in medical schools in NZ and Australia. To better understand the use of the mini-CEX as an educational tool being used for medical students in the context of a programmatic assessment I undertook a narrative literature review. Chapter three summarises the published evidence of the use of the mini-CEX, first looking at its development and use in the post-graduate setting and secondly the use of the mini-CEX in undergraduate medical programmes. In order to determine how the mini-CEX was being used as an assessment tool, deidentified data on students in the Year 4 of the University of Auckland MBChB programme between 2013 and 2017 was obtained. More detailed information about assessor, the patient characteristics and the assessment for 2013 and 2014 was obtained. In Chapter four descriptive statistics and measurement characteristics of the Year 4 mini-CEX are outlined. Differences between the General Medicine and Specialty Medicine rotations and a focus on how whether there is any difference in grade for different groups of medical students are assessed. In Chapter five data on the mini-CEX and its association on the grade for the Clinical and Communication Skills domain are described. Finally, Chapter six will summarise the findings from this research.

Chapter 2 - Use of the mini-CEX at the University of Auckland and medical schools in NZ and Australia.

At the University of Auckland the mini-CEX is used twice in Year 4 and forms part of the final grade of the Clinical and Communication Skills domain grade. The mini-CEX is used five times in Year 5 during rotations in Paediatrics, Psychiatry and General Practice and five times in Year 6 during rotations in General Practice, Obstetrics and Gynecology, Paediatrics and Surgery. In Year 5 Paediatrics a single mini-CEX is done and contributes to the overall grade for the rotation. In Year 5 Psychiatry and Year 5 General Practice the student must undertake a formative mini-CEX and then must pass a summative mini-CEX with the summative mini-CEX contributing to the grade for the rotation. In Year 6 Paediatrics and Surgery a single mini-CEX is done and contributes to the overall grade for the rotation. In Year 6 Obstetrics and Gynecology a single mini-CEX is done and contributes to the overall grade for the rotation. Students are encouraged but not required to undertake formative mini-CEX. In Year 6 General Practice the student must undertake a formative mini-CEX and then must pass a summative mini-CEX with the summative mini-CEX contributing to the grade for the rotation. In each academic year, the mini-CEX grade contributes to both the attachment grade and the longitudinal Clinical and Communication Skills domain grade.

A standard form is used for all mini-CEX assessed at the University of Auckland. The overall clinical competence of the student is determined as well as four subcompetencies: history taking/communication, physical examination, clinical judgement/reasoning and humanistic qualities/professionalism. These are assessed using a four point ordinal scale (Appendix 1). The original post-graduate tool developed by the American Board of Internal Medicine used a nine-point scale which fell within four bands which described the trainees performance: unsatisfactory, marginal, satisfactory and superior [Norcini, 1995]. When adopting the tool for undergraduate use it was determined, that a nine-point scale would not be accurate for Auckland medical students and thus a four point ordinal scale was adapted to use in the undergraduate setting (Personal Comms A/Prof Andrew Wearn). The four point scale is

used for other clinical assessments in the University of Auckland undergraduate programme. A space to record the time taken observing the student is provided and a space for the assessor to sign the form is provided. Instructions that describe how the tool should be used and a rubric for marking are included as part of the mini-CEX form (Appendix 1).

The RACP mini-CEX assesses the overall clinical competence of the BPT but, in contrast, assesses six subcompetencies and uses a nine point interval scale (Appendix 1). The complexity of the case is assessed. Space for recording the time the trainee is observed and the amount of time spent providing feedback to the trainee is provided. Both the assessor and the trainee must sign the form. A description of the specific skills a BPT must demonstrate and a rubric to assess these skills is provided with more space provided on the form for recordings. Eight formative assessments are required over 24 months.

So it can be seen that there are some similarities but also differences between how an undergraduate student and a postgraduate trainee are assessed. I was interested in understanding how other local medical programmes use the mini-CEX and thus undertook an informal survey by contacting the medical program directorate or equivalent at the University of Otago and medical school in Australia.

The University of Otago, a six year MBChB programme, uses the mini-CEX but only in Paediatric rotations in Year 5 and Year 6. There it is used at a formative assessment, and only in Paediatric rotations, twice in year 5, and four times in year 6 (Prof Tim Wilkinson and Dr Tony Walls, Personal communications, September 2018). The form assesses two domains: the quality of the clinical skills and the interaction with the child and uses a nine-point interval scale which falls within three broad descriptors where 1, 2, 3 are unsatisfactory, 4, 5, 6 are satisfactory and 7, 8, 9 are superior and examples of what behaviours would correspond to these scores. A space for both the assessor and the trainee to sign the form is provided but no documentation of the time spent with the student is recorded (Appendix 1).

The mini-CEX is used in the majority of Australian medical schools and their use of the assessment tool is summarized in Table 3.

Table 3 - Use of the mini-CEX in Australian medical schools.

Medical School	Course Duration (Years)	Use of the Mini-CEX	Number of Assessments	Formative or Summative	Subcompetencies	Scale Length
Australian National University Medical School.	4	Yes	16 in Year 3 6 in Year 4	Formative	No response	No response
The University of Sydney School of Medicine	4	No				
University of New South Wales Faculty of Medicine	6	Yes	4 in Year 5 4 in Year 6	Formative	5	4
Western Sydney University School of Medicine.	5	Yes	4 in Year 3 5 in Year 4 9 in Year 5	Summative	5	5
University of Wollongong Graduate School of Medicine.	4	Yes	35 in Phase 2 4 in Phase 3 12 in Phase 4	Formative and summative	0	Formative 0 Summative 5
University of Newcastle School of Medicine & Public Health. University of New England School of Rural Medicine.	5	Yes	2 in Year 4 6 in Year 5	Formative	6	9
University of Notre Dame School of Medicine, Sydney	4	Yes	1 per week Year 3 1 per week Year 4	Formative	6	4
Curtin Medical School	4	Planned*	To be confirmed	Formative and Summative	To be confirmed	To be confirmed
University of Western Australia School of Medicine	4	Modified*	9	Summative	6	8

University of Notre Dame School of Medicine, Fremantle	4	Yes	18 in Year 3 18 in Year 4	Summative	7	6
Deakin University School of Medicine	4 [‡]	Planned	In each clinical rotation of Year 3 and Year 4.	To be confirmed	To be confirmed	To be confirmed
Monash University Faculty of Medicine, Nursing & Health Sciences	4 -5 [^]	Modified [#]	8 in Year 3	Formative and summative	3	8
Melbourne Medical School	4	Yes	10 in Year 3	Formative and Summative	8	6
University of Tasmania School of Medicine	5	Yes	10 in Year 4 10 in Year 5	Summative	6	9
Flinders University School of Medicine	4	Yes	16 in Year 3 6 in Year 4	Formative	6	8
University of Adelaide Medical School & Adelaide Rural Clinical School	6	Yes	9 in Year 4 9 in Year 5 3 in Year 6	Formative and summative	7	9
University of Queensland School of Medicine	4	Yes	16 in year 3 14 in Year 4	Formative and Summative	4	4
James Cook University School of Medicine & Dentistry	6	Yes	12 in Year 5 8 in Year 6	Summative	7	2
Griffith University School of Medicine	4	Yes	12 in Year 3 6 in Year 4	Summative	6	9
Bond University Faculty of Health Sciences & Medicine	5	Yes	12 in Year 5	Summative	6	9

* University of Western Australia use a modified mini-CEX called the structured clinical assessment.

[^] Monash University have a 4 year graduate or 5 year undergraduate programme.

[#] Monash University used a modified mini-CEX called the mini-case record.

[‡] Curtin Medical School is establishing its programme and plan to use the mini-CEX but have not confirmed the specific requirements.

[‡] Deakin University School of Medicine were in the process of revising their programme to include the mini-CEX, however the exact use is yet to be confirmed.

Thus, in NZ and Australia, there is a wide range of use of the mini-CEX in undergraduate medicine. One medical school, the University of Sydney, does not use the tool. Other

medical schools use the tool for formative assessment only (e.g. Otago), summative assessment only (e.g. Tasmania) or both (e.g. Queensland). Of note the University of Notre Dame has medical schools both in Freemantle and Sydney but uses slightly different versions of the mini-CEX.

Interestingly, the University of Wollongong has different marking sheets for formative and summative assessments. In the formative assessment no grade is assigned. This embraces the concept of assessment for learning. The focus is on feedback with space for the assessor to comment on strengths, weaknesses and an action plan for the student. In summative assessment the tool provides an overall grade and has moved to an entrustability scale (Appendix 1).

The original post-graduate scale developed by the American Board of Internal Medicine used a nine-point scale which fell into four bands: unsatisfactory, marginal, satisfactory and superior. Quite understandably, when adapting the mini-CEX to the undergraduate setting the University of Auckland's mini-CEX has simplified the scale and reduced the range of sub-competencies being assessed. Of note, as well as being part of the assessment for the clinical rotation, the mini-CEX at the University of Auckland is used as a component of programmatic assessment.

Chapter 3 - Narrative Literature Review of the mini-CEX.

To better understand the mini-CEX and the use of the mini-CEX as an assessment tool, I undertook a narrative literature review [Ferrari, 2015; Siddaway, 2019]. A systematic review offers a comprehensive, standardised and reproducible overview of the literature, and will provide both a qualitative and quantitative analysis around a specific topic. However, a protocol describing a planned systematic review of the mini-CEX has been published [Mortaz Hejri, 2017]. A narrative review, although less rigorous than a systematic review, in terms of approach can be used to review and summarise key aspects of the literature highlighting the historical development and important components of the mini-CEX. Reviewing the mini-CEX with this method has allowed me to consider the mini-CEX from both the lens of a postgraduate and undergraduate assessor.

For this narrative review I will provide a historical account of the development of the mini-CEX, describe studies that have evaluated validity, and reliability, and review the sources of variance and strategies that have been evaluated to limit this variance, the role of feedback, the effect of the mini-CEX on trainees, the use of the mini-CEX in different disciplines and finally other assessment tools that have been developed based on the mini-CEX, all in the postgraduate setting. Next, I will discuss the use of the mini-CEX in the undergraduate medical setting covering feasibility, validity, reliability, feedback, and electronic formatting of the mini-CEX.

To comprehensively review the literature articles on the mini-CEX were obtained from searched databases which include PubMed, Education Resources Information Centre (ERIC), the Cochrane library, and PsycINFO. Search terms used include mCEX, mini-CEX and mini-Clinical Evaluation Exercise. In addition, the reference lists and citations from articles that were obtained from these searches were also reviewed and included if appropriate. No restriction was placed on study design. The search was censored *ab initio* to the 1st of

August 2017. Articles that were not in English, without a translation readily available, and non-contributory articles were not included (Figure 2).

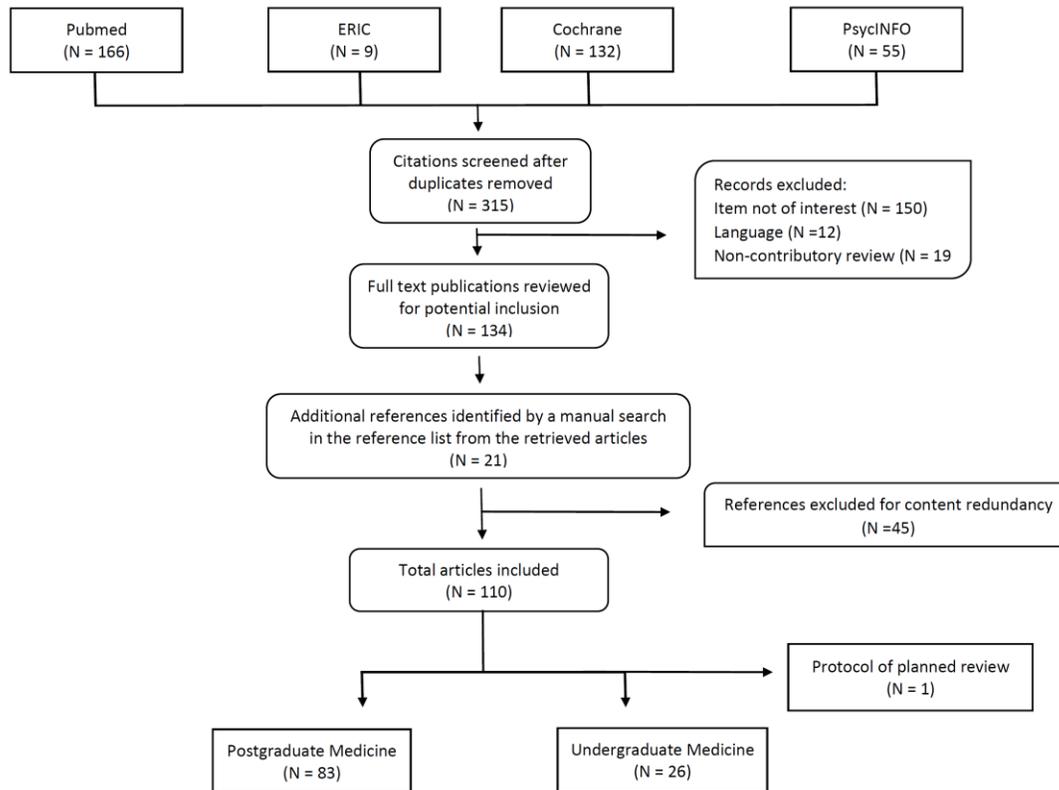


Figure 2 – Flow chart of the literature selection process for the narrative review.

3.1 - The mini-CEX in Postgraduate Medical Education.

Initial descriptions and development.

The mini-CEX was created by the American Board of Internal Medicine to assess the clinical skills of medical residents [Norcini, 1995]. Previously, Internal Medicine (IM) residents were observed by a single faculty member performing a thorough history and physical examination. The resident would then discuss their finding and plan management; in essence perform a “long case”. The long case was noted to have several potential limitations: 1.) a limited number of assessors; 2.) a single patient interaction; and 3.) a limited format [Norcini, 1995]. The mini-CEX was developed to address these issues while

effectively assessing the clinical skills that residents, the equivalent of a registered medical officers (RMO), require in patient interactions with a faculty member, the equivalent of a senior medical officer (SMO). Faculty would observe a medical resident undertake a focused history and examination, then discuss an impression and treatment plan, before finally providing educational feedback. Multiple patient-resident interactions with multiple faculty would occur.

The first description of the tool [Norcini, 1995] was undertaken in Pennsylvania, USA. In the preliminary study 389 mini-CEX encounters in 88 residents were included. Each resident undertook 2 – 10 assessments with 78 residents (89%) in their first year of training, 9 (10%) in their second year, and 1 (1%) in their third year. Assessments were performed by attending physicians, who completed 1 - 26 evaluations. A wide range of conditions were assessed including common IM problems and also conditions in other specialties such as Psychiatry, Gynecology, and Ophthalmology. The time taken ranged widely: from 5-100 minutes, with significantly longer assessments for inpatient visits, compared to outpatients, and those that occurred in the Emergency Department. It was noted that both residents and evaluators were significantly more satisfied after longer encounters. The assessment was undertaken in overall clinical competence and four components of competence: history taking, physical examination, clinical judgement and humanistic qualities. A nine-point scale was used to assess encounters, with scores of 1, 2 and 3 ranked as unsatisfactory, 4 marginal, 5 and 6 ranked as satisfactory and 7, 8 and 9 ranked as superior. A significant correlation between the components of competence (range: .65 – .81, $p < .01$) and between the components and overall competence (range: .61 - .68, $p < .01$) was demonstrated. Residents did better the longer they were given (Pearson product moment correlations of .10 - .16 for duration of evaluation and performance ratings, $p \leq .05$). The reproducibility, the likelihood that the same resident would receive the same score if tested again, was determined using Generalisability theory (G-theory) and between 12-14 encounters were required to produce satisfactory reproducibility. It was noted that second years residents'

overall competency score was significantly better than first years' score (6.98 v 6.45 respectively, $p < .05$).

When focusing on those undertaking the evaluation [Norcini, 1997] 64 assessors were included. The scores for overall clinical competence ranged from 5.5 to 8.0 (mean 6.6 +/- 0.6). There was correlation among the components of competence (history, examination, clinical judgment, and humanism). There was not a large difference seen between evaluators in the components of competence. Little difference between setting, inpatient versus outpatient, was noted. Examiners were generally satisfied with the tool, especially when more time was taken for the assessment.

A larger study was performed using non-systematically selected IM training programmes from USA which assessed data from 1228 mini-CEX encounters involving 421 residents and 316 assessors [Norcini, 2003]. A nine-point scale was used with the resident rated on overall competence and seven sub-competencies: interviewing, physical examination, professionalism, clinical judgment, counseling, organization, and efficiency. Additional data on the complexity of the patient's problem on a 3-point scale (low, moderate, and high), gender of the patient, the type of visit (new or return), the setting (ambulatory, inpatient, emergency department, or other), the time spent observing the encounter, and providing feedback was obtained. The examiner also noted whether the focus of the encounter was data gathering, diagnostic, therapeutic, or focused on counseling. Interestingly, the scores varied significantly, with residents receiving lower scores for *less* complicated patients (low complexity 6.60 +/- 1.02, moderate complexity 6.66 +/- 1.04, high-complexity 6.94 +/- 1.07; $p < .01$). The total score and the scores for each sub-competency achieved by residents significantly improved through the academic year (total quarter 1 6.40 +/- 1.05, total quarter 2 6.54 +/- 1.03, total quarter 3 6.81 +/- 1.01, total quarter 4 6.97 +/- 0.99; $p < .01$). The more mini-CEXs that were assessed the smaller the confidence intervals around overall score. Ten encounters produced a tight confidence interval with only minor gain from more encounters. Examiners were in general satisfied with the tool.

These initial reports by Norcini *et al* are important as they establish the role of the mini-CEX as a formative assessment tool developed to encourage education rather than evaluation. The first paper demonstrates that the mini-CEX could be used to assess residents effectively, and that the tool was acceptable to those involved in the assessment. G-theory was used to demonstrate that multiple uses of the tool were needed for reproducibility. And of course exposing the trainee to multiple clinical scenarios was another key aim of developing the tool. A key finding was that residents' scores improved with time which supported the construct validity of the tool. The second paper focused on the assessors and demonstrated that they were satisfied with the mini-CEX, supporting face validity. It was also determined that the mini-CEX was broadly comparable over different clinical settings. The third paper replicated the previous studies but in a different and larger population. It demonstrated the feasibility of the tool as well as construct and face validity. G-theory was again used to show that multiple assessments were required to reliably assess residents. The benefit of multiple assessments with the mini-CEX is that the trainee has greater patient and examiner diversity. The disadvantages are more taxing administration requirements and a failure to evaluate the resident performing a complete history and physical examination. These pioneering papers established the mini-CEX as a valuable and effective assessment tool.

“A good assessment should be objective, reliable and valid.”

Mirjana Knorr & Dietrich Klusmann, *Medical Education* 2015, 49: 550-552

Validity of the mini-CEX.

Validity, the ability of an assessment to measure what it is supposed to measure, is a vital aspect of any assessment tool [Schuwirth, 2011a]. Classically, this is divided into three components: criterion, content and construct validity [Cronbach, 1955].

Criterion validity.

Criterion validity, the extent to which the grade relates to the trainee's clinical abilities, may be further separated into either concurrent validity, where the comparison between the measure and outcome are determined at the same time or predictive validity where the measure is compared to an outcome at a later time. The criterion validity of the mini-CEX has been evaluated in a number of studies.

The mini-CEX has been compared to other evaluation tools. The Clinical Skills Assessment (CSA) is used by the Educational Commission for Foreign Medical Graduates to determine if International Medical Graduates (IMGs) can gather and interpret clinical patient data and communicate effectively in English at a level comparable to students graduating from United States medical schools. This was found to correlate closely to the mini-CEX [Boulet, 2002]. The American Board of IM have monthly evaluation forms which evaluate overall clinical competence and specific subcompetencies. This was compared to the mini-CEX and found to closely correlate [Durning, 2002]. The Royal College of Physicians and Surgeons of Canada have a Comprehensive Examination in Internal Medicine (RCPSC IM). Trainees preparing for this exam had standard mini-CEX performed and their scores correlated with outcomes of the RCPSC IM exam [Hatala, 2006]. In the UK, IM training is conducted by three Royal Colleges of Physicians. A study of trainees who completed the mini-CEX and other WBAs demonstrated significantly correlation with the score achieved in the mini-CEX and other WBAs including CbD, the mini-peer assessment tool and DOPS [Davies, 2009].

Audit of clinical notes is used for continuous professional development of medical professionals. A study from USA, performed a chart review of patients seen by IM residents and determined how frequently six basic health preventative measures: bowel cancer screening, breast cancer screening, lipid screening, pneumonia vaccination, screening for tobacco use, and counselling to stop smoking were performed [Willett, 2009]. There was a weak correlation between quit smoking advice and the overall mini-CEX score of trainees and no correlation with other health preventative measures. The Clinical Skills Assessment

(CSA) is used as summative assessment tool for General Practice trainees in the UK. A study of General Practice trainees who sat the CSA did not show that the number of mini-CEX performed during training was a predictor of CSA outcome [Shaw, 2014].

Thus, when considering criterion validity we see mixed results in the published literature. The paper by Boulet *et al* showed that the mini-CEX displays concurrent validity. In a small number of trainees the mini-CEX appeared to discriminate between candidates who met and did not meet the required standard. The study also highlighted differences in the stringency of examiners. The paper by Durning *et al* demonstrates concurrent validity. However, as the monthly evaluation forms and the mini-CEX were completed by the same assessors, it is perhaps unsurprising that a close correlation was found. The report by Hatala *et al* demonstrated predictive validity and supports the view that the mini-CEX provides a reliable and valid assessment of clinical competence with a clear demarcation between passing and failing candidates. The paper by Davies *et al* demonstrated concurrent validity from correlation with other work-based assessments tool, although the correlation with DOPS was weak. The papers by Willets *et al* were less supportive of concurrent validity with only a weak correlation seen in mini-CEX scores and health preventative measure. The paper by Shaw *et al* also did not show the number of mini-CEX could predict the likelihood of passing an exit exam. However, this paper looked only at the number of mini-CEX that were performed rather than the scores that were achieved by the candidates and thus the outcome of this paper should be viewed cautiously. Taken together these studies demonstrate both concurrent and predictive validity and support the criterion validity of the mini-CEX.

Content validity

Content validity is the degree to which an assessment measures all components of a construct. This was determined in the early studies by Norcini *et al* [Norcini, 1995; Norcini, 2003], where a close correlation between the components of competence (range: .65 – .81,

$p < .01$) and the components and overall competence (range: .61 - .68, $p < .01$) were demonstrated.

Construct validity.

Construct validity, the ability of an evaluation tool to differentiate between levels of performance, is the other component of any assessment tool and has been evaluated for the mini-CEX.

Videotaped encounters of standardised patients and standardised residents displaying unsatisfactory, satisfactory, and superior clinical skills were evaluated using the mini-CEX and found that scores significantly increased with each higher level of performance of the standardised resident [Holmboe, 2003].

Studies from Argentina [Alves de Lima, 2007], UK [Wilkinson, 2008; Davies, 2009] and Taiwan [Liao, 2013] have all demonstrated that the mini-CEX grades were significantly better as residents' seniority increased. A retrospective analysis of trainees who used WBAs including the mini-CEX, Cbd, DOPs and the mini-peer assessment tool found that the mean mini-CEX scores were lower for trainees who were identified as "in difficulty" by their educational supervisors [Mitchell, 2011].

Using the multitrait-multimethod matrix significant convergence of scores between the mini-CEX and a number of other assessments were demonstrated in IMGs entering into the Canadian medical programme [Baig, 2010].

These studies consistently demonstrate the construct validity of the mini-CEX by the ability of the tool to discriminate between junior doctors with satisfactory or unsatisfactory performances and by demonstrating that as junior doctors gain experience then mini-CEX scores improve.

Kane's Validity Framework

A more sophisticated approach has been developed where the process of validation involves development of a structured argument to evaluate the tool. Kane's validity framework, which focusses on four key inferences, is commonly used to conceptualise this process [Cook, 2015]. A validity argument for the mini-CEX using Kane's validity framework has been constructed [Hawkins, 2010]. Evidence was sought to support an argument for validity in the components of scoring, generalisation, extrapolation and interpretation. This paper highlighted that in the constructs of generalization and extrapolation there was evidence to support the validity of the mini-CEX. There were some arguments to support the construct of interpretation/decision. However, in the construct of scoring, there were noted to be some limitations in evidence to support the use of the tool. (which remain unaddressed).

Historic studies have highlighted variability in the clinical skills of junior doctors. A key role of the mini-CEX is to determine the level of competence of medical practitioners. These studies demonstrate validity and support that a framework to construct a validity argument exists to support the use of the mini-CEX.

Reliability of the mini-CEX

Reliability refers to the consistency on an assessment when it is repeated; a reliable tool will achieve the same result when repeated under similar conditions [Downing, 2004]. The concept of reliability also encompasses the generalisability of an assessment and the amount of random error that occurs with the measurement. Error that occurs may be due to the assessment but also may be due other factors such as the assessors administering the test so called rater error or to case itself. If the assessment has multiple components that are being assessed then internal consistency may also be determined. So, for the mini-CEX, we can determine reliability, looking at the error from the assessment tool, the assessors and the cases and we can look at the internal consistency of the different components. G-theory analysis or intraclass correlation coefficient may be used to estimate

the variance that may contribute to error. G-theory analysis is viewed as a more elegant analysis and is used to determine G-coefficient which informs the consistency of the rating, where 0 is not at all reliable and 1 is perfectly reliable. Estimates of the components of variability may also be determined using this method where 0 is not contributing any variance and 1 accounting for all the variance. Typically a G-coefficient of 0.8 is expected for an assessment to be considered reliable for a high stakes decision, such as progression in training, although an even greater cutoff may be desired in very high stakes examinations [Downing, 2004].

The initial work on the mini-CEX determined that between 12-14 evaluations were needed to achieve satisfactory reliability, a G-coefficient of > 0.8 [Norcini, 1995]. Other studies of IM trainees have determined that 8 [Jackson, 2010] or 10 assessments [Norcini, 2003; Alves de Lima, 2007] are needed to achieve satisfactory reliability.

A single study has been reported that looked at the internal structure of the mini-CEX and using factor analysis demonstrated that although an assessor utilizing the mini-CEX rates competence in multiple domains (e.g. interviewing, physical examination, humanistic qualities/professionalism, clinical judgment, counseling, and organization/efficiency) as well as an overall, global impression that these subcompetencies are tightly linked to the overall score [Cook, 2010]. Thus, we may be only measuring overall clinical competence rather than multiple domains.

The reliability of the mini-CEX has been further evaluated using more sophisticated analyses and four studies are worth considering in more detail. An USA study of 24 assessors used the mini-CEX to retrospective assess 10 trainees who each completed 6 cases for an alternate exam used G-theory to determine the components of variance [Margolis, 2006]. Most of the variance was due to the rater (variance 0.47), then the examinee (variance 0.46) with the least effect due to the case (variance 0.04). Importantly, the study reported that the number of raters was more important than the number of patient interaction in

improving reliability: it was determined the G-coefficient for assessment of overall clinical competence based on one examiner assessing each examinee on ten occasions was $g = 0.39$ while the reproducibility for an assessment where ten raters each rated an examinee on one case would be $g = 0.83$. A UK study of 128 IM trainees who undertook 658 mini-CEX assessed by between 1 and 21 assessors used G-theory to evaluate reliability and found that only 5 assessments were needed to achieve a reliable measure ($g = 0.80$) [Wilkinson, 2008]. It was estimated that trainee-to-trainee variation was 0.13, assessor-to-assessor variation 0.19, and encounter-to-encounter variation 0.24. A study of Anaesthesia trainees found that the mini-CEX, which assessed overall care as well as nine other domains specific to Anaesthesia, was unreliable and lenient [Weller, 2009a]. G-theory was used to analyse 297 assessments on 38 trainees by 58 assessors and found most of the variance related to the case specificity (variance .41), then assessor stringency (variance .40), assessor subjectivity (variance .15) with only a small component related to the trainees ability (variance .04). This paper determined that over 100 assessments were needed for a G-coefficient of > 0.8 . Finally, an interesting study utilized a controlled set-up to estimate the reproducibility of the mini-CEX [Alves de Lima, 2013]. Three standardised patients, were seen by 21 Cardiology residents with encounters then videotaped. Three experienced assessors scored the videotaped encounters using a nine-point scale mini-CEX. The value of increasing the number of assessors was again highlighted: 9 mini-CEX were needed to achieve a G-coefficient of > 0.8 with one assessor ($g = 0.81$), 5 mini-CEX with two assessors ($g = 0.83$) and 3 mini-CEX if there were three assessors ($g = 0.81$). The major sources of error related to general error, and variability (stringency/leniency) of the examiners with the least due to the ability of the candidate.

These papers highlight that the reliability of the mini-CEX increases with the number of cases, that increasing the number of assessors is more effective than just increasing the number of cases and that the assessor is noted to have more impact on variance than the ability of the candidate. The number of cases needed to achieve a robust generalisability was notably different between IM trainees and Anaesthesia trainees.

Factors contributing to the variance between raters.

We see the assessor has an effect on the reliability and thus the validity of the mini-CEX. A number of studies have further evaluated factors that contribute to variance between raters.

An interesting study highlighted that the clinical skills of faculty may be associated with how they rate trainees. Forty-four IM faculty performed 8 standardised patient encounters themselves and then rated videotaped resident encounters using the mini-CEX (Kogan, 2010). There did not appear to be any correlation between the rating and demographic factors such as age, gender, clinical experience or preceptor experience. However, it was noted that faculty who demonstrated more complete history taking during their assessments of standardised patient were more stringent in their ratings in interviewing and organization domains, and faculty who demonstrated more complete examination skills were more stringent with interviewing, examination and organization. Thus the implication is that the better the clinical skills of the assessor the greater their stringency of assessment perhaps due to the assessors expectation of the trainee.

Social judgements are where an assessor makes inferences about a trainee and the reasons why they do something. Work by Gingerich *et al* highlights that raters may make social judgements: causal explanations for observed behaviours, and this may influence and explain the variance between raters scores. In their first study, using videotaped resident encounters, it was shown that IM Physicians when asked to make social judgements gave different explanations for the behavior of residents. Using latent partition analysis, a method where judgements of behaviours are grouped into common themes, to analyse these causal explanations it was shown these explanations could be grouped into themes, and these judgements could account for some of the variance seen between raters [Gingerich, 2014]. In a second paper, using Q methodology, which focuses on describing viewpoints that describe a population, it was shown that different assessors have distinct points of view about components of a trainee's performance. These were not always

consistent between assessors with different perspectives shown to emphasize a strength or weakness of different components of a trainee's performance. Each perspective was consistent with the mini-CEX rating assigned by the rater [Gingerich, 2017].

In a series of papers by Yeates *et al* the impact of bias, caused by comparing the performance of recently assessed trainees rather than an absolute standard was demonstrated to contribute to variance between raters. It was hypothesized that two opposite effects may be seen: anchoring bias or contrast bias. Anchoring bias will lead to a recent experience of a good performance increasing scores compared to a recently reviewed poor performance. Contrast bias will lead to a recent experience of good performance and may tend to decrease subsequent scores compared to a recently reviewed poor performance. In the first study videos of scripted performances with simulated patients using real junior doctors was developed and scored using a modified mini-CEX with seven domains. This demonstrated that contrast bias occurred and accounted for variance between raters [Yeates, 2012]. A qualitative study demonstrated three themes that contributed to variance: assessors weighed aspects of the performance differently, criteria for marking varied between assessors, and when integrating components of the assessment different emphasis led to a global impression [Yeates, 2013]. Finally, using qualitative analysis, it was demonstrated that even a single poor or single good performance by a trainee could lead to contrast bias [Yeates, 2015].

These papers give valuable insight into factors that may influence raters' judgement and contribute to variance. Kogan *et al* highlight that the examiners ability may influence their stringency, Gingerich *et al* highlights that assessors make social judgements and Yeates *et al* demonstrate that the evaluation of previous trainees may impact on assessors scores.

Improving reliable assessment of medical trainees.

A number of studies have been performed to determine whether training of assessors can improve assessment of trainees but have shown mixed results.

A controlled study using a cluster randomisation design was performed to determine if a multi-faceted method of faculty development could improve the accuracy of the assessment [Holmboe, 2004a]. A combination of mini-lectures and evaluation exercises were used to train assessor. Although these interventions were rated as outstanding and led to participants being more confident with administering the mini-CEX, this did not translate into significantly different scores to a control group. Another study to train raters in the mini-CEX used a workshop that covered the need for resident assessment, potential errors and biases, discussed the domains of the mini-CEX, and had facilitated discussion of videotaped interaction of residents and standardised patients but did not significantly improve reliability of raters [Cook, 2009a]. The effect of training faculty on the mini-CEX with a two hour workshop found that for assessors that attended had a significant improvement in cognitive knowledge but did not address the effect on trainee assessment [Liao, 2013]. Finally a qualitative study assessing two training methods: performance dimension training, where raters are made familiar with the dimensions on which the tool is assessing and a modified approach to frame of reference training, where raters are made familiar with the dimensions and also the appropriate standard for each dimension that is being assessed, was used to train assessors in the mini-CEX. Although participants perceived benefit from both methods a number of challenges were identified in incorporating these training methods including both individual and institutional issues [Kogan, 2015].

So, a little surprisingly, we do not see that training assessors has clear benefits in terms of improving reliability. It does, however, have benefits in terms of improving the confidence of assessors and their understanding of the tool. Any workshop should take into consideration local information to overcome challenges in implementing the assessment.

Structure of the mini-CEX.

Whether the structure of the mini-CEX can be altered to improve reliability has also been explored.

The mini-CEX was developed with a nine-point rating scale. Scales with fewer points may be more efficient and potentially easier to use [Preston, 2000]. However, simulation studies in non-medical education have shown that scales with fewer points may be less reliable than those with more points [Nishisato, 1970; Jenkins, 1977]. It was hypothesized that raters may be unable to make reasonable distinctions between categories on a nine-point scale, and that a five-point scale for rating the usual mini-CEX domains may be preferable [Cook, 2009]. To evaluate this faculty members rated videotaped encounters using the mini-CEX with both a five-point and nine-point scale. Scores using a nine-point scale were significantly more accurate in classifying competence for overall rating than a five-point scale and the variance between raters was not significantly different.

Entrustable Professional Activity.

The concept of entrustable professional activity (EPA) allows assessors to decide on the competency based on the degree of supervision required by trainees. This concept was applied to the mini-CEX where it was noted that assessors will have different interpretations of what constitutes different levels of performance and hypothesized that a scale that linked performance to entrustability: “Do I trust this trainee with this task?”, would be more effective than the standard scale used for the mini-CEX [Crossley, 2011]. In the study of IM trainees the standard scale for the mini-CEX was used, and a second scale added, in which pre-determined training level anchors were accompanied by behavioural descriptors aligned to the constructs of developing clinical sophistication and independence. It was shown that assessors discriminated more widely between high- and low-performing trainees using the new scale than they did using the conventional scale and, when a trainee saw several assessors, those assessors scored the trainee more similarly. In addition, it was noted that when using the construct-aligned scale that only 4 mini-CEX assessments ($g = .82$) were needed to achieve a satisfactory reliability compared to 10 mini-CEX with the standard scale ($g = .80$).

A pilot and feasibility study to determine if EPAs could be used to assess IM residents identified that trainees and assessors found this scale useful for learning and feedback but concluded that significant barriers to participation made using EPAs challenging to implement [Hauer, 2013].

The impact of a scale using both a conventional scoring scale and a novel scale based on the need for direct or more distant supervision using EPA has been evaluated in Anaesthesia trainees [Weller, 2014]. Reliability of assessments significantly improved with the new scoring system based on whether direct or distant supervision was required. This work led to a 9-point scale with three entrustability categories being used for NZ and Australia Anaesthesia training: “Trainee needs assessor in the theatre suite”; “Trainee needs assessor in the hospital”; and “Trainee could manage this case independently and does not require direct supervision”. Using an online database from the Australia and New Zealand College of Anaesthesia it was shown that to achieve a G-coefficient of > 0.8 only 24 assessments were needed compared to previous work which demonstrated that over 100 assessments were needed [Weller, 2017].

These studies highlight that the scale initially developed for the mini-CEX may be less than ideal. A move to a scale which uses descriptors of entrustability is feasible and deliverable in Anaesthesia training. Using an entrustability scale will improve the reliability of the mini-CEX as an assessment tool and reduce the number of assessments for a trainee.

Combining the mini-CEX with other Workplace Based Assessments.

Many postgraduate programmes use several WBA tools when making a decision on whether to accredit a trainee. For example an UK study highlighted that the mini-CEX, MSF, and DOPS are used to assess the performance of IM specialties and that these were reliable and feasible tools [Wilkinson, 2008]. Two papers are of note when considering this and these will be discussed in detail.

The separate and composite reliability of the mini-CEX, MSF and DOPS were evaluated in 953 residents across multiple specialties using G-theory with 7467 mini-CEX, 4765 DOPS and 547 MSF utilised for analysis [Moonen-Van Loon, 2013]. For reliable assessment of the mini-CEX 8 assessments were required (2 mini-CEX $g = .52$, 4 mini-CEX $g = .68$, for 6 mini-CEX $g = .76$ and for 8 mini-CEX $g = 0.81$). Nine DOPs and nine MSF were needed to exceed a G-coefficient of 0.8. When combined, the number of assessments needed to achieve satisfactory reliability was less (e.g. 7 mini-CEXs, 8 DOPS, and 1 MSF ($g = 0.80$) or 6 DOPS, 5 mini-CEX and 2 MSF ($g = 0.81$)). The study included trainees that could be first, second, third or fourth year residents. When focusing on first year residents only 6 mini-CEX ($g = 0.82$), 6 DOPS and 6 MSF were needed to achieve satisfactory reliability when used individually and again this number could be reduced when combining assessments (e.g. 5 mini-CEX, 6 DOPS and 1 MSF ($g = 0.81$)).

The reliability of WBA in an Obstetrics and Gynecology training program has also been described [Homer, 2013]. A combination of 187 mini-CEX in Obstetrics, 142 mini-CEX in Gynecology, 239 CbD in Obstetrics, 147 CbD in Gynecology and 2409 Objective Structured Assessment of Technical Skills in nine procedures were used to evaluate trainees assessments. Multi-level modelling and G-theory were used to determine the reliability of these tools. Interestingly, this demonstrated a wide variation in the number of assessments needed to achieve reliability for specific topics i.e. 4 assessments for the mini-CEX in Gynecology were needed for a $G > 0.8$ and more than 10 for the mini-CEX in Obstetrics or 4 CdB in Gynecology were needed to achieve a satisfactory G-coefficient and 10 CbD in Obstetrics. So we can see that the number of WBA to achieve a reliable assessment of trainees varies for different assessments. And it may be that the number of assessment needed for each tool may vary based on the subject.

Importantly, the paper by Moonen-Van Loon *et al* demonstrates that assessment tools can be combined to reduce the overall number of tests needed for reliable assessment although a little surprisingly, less assessments were needed in first year trainees. These studies

highlight that it is possible to make summative decisions using a combination of assessment tools and supports the use of the mini-CEX as a component of Programmatic Assessment [Van der Vleuten, 2012], but when considering how they are combined some consideration may be needed.

“The most powerful single moderator that enhances achievement is feedback.”

John Hattie. University of Auckland. 1999. Third Inaugural Professorial Lecture.

Feedback, a key benefit of the mini-CEX.

Feedback enhances learning and a strength of the mini-CEX is that feedback occurs immediately and can be targeted to specific areas. The ability of the mini-CEX to offer the trainee feedback was noted as an important component of the tool in the initial paper [Norcini, 1995] and has been evaluated in more detail by a number of studies.

A prospective observational cohort study from USA that analysed audio-recordings of mini-CEX sessions to determine the nature of feedback showed that 90% of sessions included recommendations with learner reaction facilitated by assessors in 61% of assessments and self-assessment from the trainee requested in 34% of sessions [Holmboe, 2004b].

Two early studies focused on the role of feedback when using the mini-CEX in Anaesthesia trainees [Weller, 2009a; Weller, 2009b]. The initial studies showed a positive effect of feedback and perceived a positive educational impact. In the written feedback fields of the Mini-CEX form, 95% of specialists wrote comments under ‘things that the trainee did well’, 70% recorded comments in ‘areas for improvement’, and 60% wrote down an ‘agreed action’. Trainees felt the mini-CEX facilitated feedback. A later study in done in Anaesthesia trainees, after the entrustability scale was adopted, identified that a strength of the mini-CEX was that it facilitated feedback [Castanelli, 2016].

The theme, of the value of feedback, has been explored in other studies. The majority of Nepalese doctors utilising WBAs noted that the immediate and specific feedback made the mini-CEX useful [Butterworth, 2010]. However, two UK study have shown than although there is potential benefit of the mini-CEX to provide feedback that trainees were not well versed in its use, that the mini-CEX was often used incorrectly with direct observation of the trainee not occurring and feedback not occurring or being poor [Jackson, 2010; Weston, 2014].

The quality of written feedback has also been evaluated. A Dutch study reviewed written feedback in the mini-CEX and identified that although there was frequent reporting of feedback there was less frequent reporting of reflection and rare reporting of action plans [Pelgrim, 2012]. A Canadian showed that written feedback identified specific tasks for the trainee in over half of assessments but less frequently gaps in the trainees' performance (4%) and action plans (14%) to address areas to improve [Gauthier, 2015]. A third study from Taiwan identified that positive feedback occurred frequently (85.3%), however suggestions for development occurred less frequently (54.8%) and an agreed action plan for the trainees was also less common (29.5%) [Chang, 2017].

When considering why feedback is provided it is likely that assessor will also account for how the feedback will impact on the trainee. A study from USA determined how feedback was delivered and found that there was variability of techniques. The delivery of feedback was tempered by what faculty thought the goal of the feedback was and how they felt about providing feedback. It was highlighted that there was a balance between providing both positive and negative feedback, the faculty considered the perception of the resident's insight, receptivity, skill and potential. The faculty member and resident relationship and context of providing feedback were also consider important [Kogan, 2012].

These studies demonstrate that the mini-CEX is a valuable tool to provide feedback to trainees. The feedback component is perceived as a positive educational benefit by both

assessors and trainees. How feedback is delivered is quite variable. Assessors consider how the feedback will impact on the trainee. The quality of written feedback could be improved and active involvement of the learner could occur more frequently.

Effect of the mini-CEX on those being assessed.

The effect of the mini-CEX on trainees has been evaluated and has been shown to be a stressful process in Cardiology [Alves de Lima, 2005], Internal Medicine [Malhotra, 2008], Dermatology [Cohen, 2009] and Anaesthesia trainees [Castenelli, 2016].

So the papers on feedback indicate that trainees and assessors recognise that the assessment process and that feedback is of benefit, however, these studies highlight that the mini-CEX is likely to produce anxiety which may impact on performance.

The Mini-CEX for assessment in medical disciplines.

The initial development and descriptions of the mini-CEX focused on its use in the assessment of IM trainees and the majority of studies are in the setting of IM with some in Anaesthesia. The mini-CEX has been used in other medical specialties including: Primary Care Physicians [Sidhu, 2009], Surgical trainees [Phillips, 2013; Joshi, 2017] although an online survey in the UK suggested the mini-CEX was one of the least frequently used work-placed assessment tools [Phillips, 2015], Paediatric trainees [Singh, 2010; Goel, 2015; Khalil, 2017], Obstetrics and Gynecology trainees [Chandra, 2013; Homer, 2013], Ophthalmology trainees [Kapoor, 2010] and IMGs [Nair, 2008].

The Mini-CEX in the wider health care community.

The use of the mini-CEX as an assessment tool for health care professionals has also been described in social workers [Bond, 2017], chiropractors [Paravicini, 2015], nurses [Milner, 2014], ayurvedics [Sadanand, 2013], midwives [Sweet, 2013], veterinarians [Weijs, 2015; Bok, 2016], dentists [Behere, 2014] and osteopaths [Moore, 2016; Vaughan, 2016] and highlights its benefits as an education tool.

Assessment tools based on the Mini-CEX.

A variety of assessment tools have been developed based on the mini-CEX. In general these tools share common characteristics. They are: 1.) workplace based assessments, 2.) time-efficient, 3.) identify specific behaviors, 4.) assess these behaviors using a Likert scale and 5.) provide immediate feedback to the person being assessed based on the evaluation. The development of these tools highlights the utility of the mini-CEX in medical education.

These can be divided into tools that are specialty specific or focused on specific behaviours and clinical scenarios. Examples of specialty specific tools include the Clinical Evaluation Exercise for Emergency Medicine [Shayne, 2002], the Ophthalmic Clinical Evaluation Exercise [Golnik, 2004; Golnik, 2005], the Palliative Care Clinical Evaluation Exercise [Han, 2005] and Neurology Clinical Skills Examination [Schuh, 2009; London 2013]. Adaptations of the mini-CEX to focus professional behaviours include the Professionalism Mini-Evaluation Exercise [Cruess, 2006; Tsugawa, 2009; Tsugawa, 2011; Bajwa, 2017], the Teamwork Mini-Clinical Evaluation Exercise [Olupeliyawa, 2014; Olupeliyawa, 2014] and the Hand-Off Clinical Evaluation Exercise [Farnan, 2009; Horwitz, 2013; Arora, 2014].

3.2 The mini-CEX in Undergraduate Medical Education.

First description of the mini-clinical evaluation exercise in medical students.

The initial description of the mini-CEX in undergraduate assessment was in a core IM clerkship at the University of California, USA [Hauer, 2000]. The mini-CEX used a nine-point scale with evaluation of overall clinical competence and the student's history, physical examination, clinical judgment, and humanism with a single assessment for each student using only patients in hospital. It was recognized that the reproducibility of a single encounter was limited and thus the assessment did not contribute to the students' grade. The benefits noted were that faculty could observe and provide feedback to students and potentially more easily identify students who would require remediation.

Feasibility, validity and reliability of the mini-CEX in undergraduate programmes.

The mini-CEX was used in a six-week inpatient and three-week outpatient medicine clerkship in July 2001 at the University of Pennsylvania, USA [Kogan, 2002]. Students were each asked to complete nine mini-CEX of which three had to be assessed by faculty, three by residents and three by outpatient Physicians. A nine-point scale assessing overall clinical competence and six sub-competencies (interviewing, physical examination, professionalism, clinical judgment, counseling, and organization) was used. In the study 32 students performed 232 mini-CEX (mean of 7.3, range 2-9 mini-CEX) with a completion rate of just over 80%, establishing the feasibility of performing multiple assessments.

The University of Pennsylvania, USA published more details of use of the mini-CEX in an IM clerkship [Kogan, 2003]. A nine-point scale assessing overall clinical competence and six sub-competencies as described above was used. Students were each asked to complete nine mini-CEX and 162 medical students who had 1297 mini-CEX were included. A mean of 7.9 mini-CEX (range 2-10) were done per student with a completion rate of 89%. Residents spent significantly longer than Physicians assessing students (residents mean time 22minutes, outpatient faculty mean time 19 minutes, inpatient faculty mean time 17 minutes; $p < .01$). The mean mini-CEX scores were lowest in the first quartile of the academic year and significantly increased through the year (first block mean = 7.4 +/- .6, second block 7.7 +/- .6, third block 7.8 +/- .6, fourth block 7.9 +/- .6; $p < .01$). Each student also undertook subject exams, case write up and had summative assessments by clinical staff for the rotation. The mini-CEX score was compared to these other assessments and a significant correlation was demonstrated. Reliability increased with the number of assessors ($g = 0.62$ for four raters, $g = 0.7$ for six raters and $g = 0.77$ for eight raters).

The mini-CEX was adapted for use at the University of Southampton, UK and introduced into the undergraduate curriculum in 2004 [Hill, 2007]. Final year students were asked to undertake three mini-CEX in clinical attachments of Child Health, Medicine, Mental Health, Obstetrics and Gynecology, and Surgery and this replaced a long case assessment. In total

2340 assessments were completed by 156 students in the 2004/2005 year. A nine-point scale was used in the first year of implementation (1-3 unsatisfactory, 4-6 satisfactory, 7-9 superior), and then changed to a six-point scale (1 – 2: below expectations; 3: borderline; 4: meets expectations; 5-6 above expectations) following feedback from students and assessors and it was demonstrated that it was feasible to rapidly introduce this extent of assessment.

The validity and reliability of the mini-CEX at the University of Southampton, UK was described in more detail [Hill, 2009]. There were 3499 mini-CEX encounters included in 326 students by 813 examiners who were either consultants and specialist registrars, across five specialties with a nine-point scale used in the first year of implementation, and then changed to a six-point scale as described above. A high degree of internal consistency was noted. The mean scores between attachments varied significantly and it was noted that consultants rated students more stringently than specialist registrars (consultant mean 4.35 +/- .62, specialist registrar 4.50 +/- $p < .01$). Cases which were more complex were scored highest (low complexity mean score 4.17 +/- .64, medium complexity mean score 4.40 +/- .59, high complexity mean score 4.50 +/- .62; $p < .01$). G-theory was used to determine that examiner stringency contributed to variation on the score (29%). The aptitude of the students for the attachment was determined to contribute 13% to score variation. Of note the reliability of the scores increased with the number of encounters. Within attachments for two mini-CEX $g = 0.31$, for 11 encounters $g = 0.71$.

In a further study from the University of Pennsylvania, USA, students were asked to complete 8 formative mini-CEX during an twelve-week Internal and Family Medicine rotation [Ney, 2009]. The mini-CEX used a nine-point scale that assessed overall clinical competence and six sub-competencies (interviewing, physical examination, professionalism, clinical judgment, and counseling). In addition, students had to complete a Medicine Standardised Patient exam at the end of their Medicine rotation and a further end-of-year, Multidisciplinary Standardised Patient exam. Satisfactory reliability was achieved with 8

assessments ($g = 0.90$) and there was correlation between overall mini-CEX grade and the Medicine Standardised Patient and the Multidisciplinary Standardised Patient exam.

The mini-CEX, was used as part of a suite of assessments in a Psychiatry clerkship at Kaohsiung Medical University, Taiwan with 196 medical students completing a 4 week clerkship with evaluation comprising of a mini-CEX, a multiple choice question test, and a standardised patient evaluation [Wang, 2011]. There was a weak but significant positive relationship between the multiple-choice examination and standardised patient score and the score of the overall clinical competence in the mini-CEX.

A study from the Rural Clinical School of Western Australia, Australia evaluated the mini-CEX over a three year study period and across multiple disciplines [Playford, 2013]. Of note a modified 8 point scale was used where 0, 3 and 4.5 were a substandard interaction, 5.5 a clear pass, 6.5, 7.5, 8.5 were scores for increasing competence and 10 for a flawless interaction with 5686 formative and summative mini-CEX assessments included in the study. A progressive and significant increase in the average score over the course of the academic year was demonstrated (February mean score 6.74, June mean score 7.54, October mean score 7.72; $p < .01$). Junior doctors awarded higher grades than General Practitioners which were higher than Specialists (Junior medical officer mean score 8.22 +/- .10, Registered medical officer mean score 7.71 +/- .08, General Practitioner mean score 7.47 +/- .03, Specialist mean score 7.08 +/- .05; $f = 146.6$; $p < .01$). Students with more complex cases received higher marks than students with the least complicated cases ($f = 33.3$; $p < .01$).

Two types of clinical assessments, the OSCE and mini-CEX, were compared in a study from the University of Bern, Switzerland [Rogausch, 2015]. In this setting the mini-CEX was modified to have a 10 point scale (1 great need of improvement, 10 little need of improvement) and evaluated overall assessment and subcompetencies of history, physical exam, counselling, clinical reasoning, organisation, and professionalism. At the end of year

three of a six year curriculum all medical student at the University of Bern take part in a mandatory 9 station OSCE. In the study Year 4 students performed 1773 mini-CEX assessed by 512 trainers. Feasibility was good with 92% of students completing the required 11 mini-CEX for the year. OSCE scores only showed a weak correlation with the overall score of the mini-CEX ($r = .26$) and the mean domain scores ($r = .27$). Interestingly, mini-CEX scores did not improve through the course of the year. When grouping the students into tertiles for the OSCE, the mini-CEX scores for students from the low-performing OSCE group were lower but not significantly so compared to the high-performing OSCE group ($r = -0.13$, 95% confidence interval -0.33 to 0.078). The mini-CEX scores for overall clinical performance for students from the medium-performing OSCE group were significantly lower compared to the high-performing OSCE group ($r = -0.18$, 95% confidence interval -0.33 to -0.038 ; $p < .05$). The seniority of the assessor was evaluated and it was determined that Residents gave significantly higher mini-CEX scores than Heads of Departments ($r = .55$, 95% confidence interval $.26$ to $.84$; $p < .01$) but there was no significant difference between Residents and Senior Physician, nor Senior Physicians and Heads of Department. The clinics where training occurred were arbitrarily classified, based on the number of student who completed a clerkship annually, as small (less than 15 students/year), medium (16-30 students/year), and large (> 30 students/year). Smaller clinics assigned statistically higher marks than large clinics ($r = -0.25$, 95% confidence interval -0.48 to -0.019 ; $p < .05$), with no difference seen between small and medium and medium and large clinics.

The mini-CEX was introduced into third year clinical rotations at the Rutgers Robert Wood Johnson Medical School, USA [Kim, 2016]. The mini-CEX form was modified in three ways to promote direct learning. First, the “not observed” notation was removed and it was a requirement for medical interviewing and physical examination to be observed. Secondly, specific clinical systems were required to be assessed e.g. the gastrointestinal system in the Surgical rotation etc. Thirdly, an area was provided for the student to confirm that specific suggestions for development were made. Multiple assessments were required during each clerkship with a minimum number by faculty and additional assessments by residents.

There was a high completion rate of forms with 92% of the required mini-CEX submitted. The rate of specific feedback varied between specialties: Surgery 70%, IM 88%. No other changes in the curriculum were made but of note the pass rate for the end of year observed clinical skills examination rate significantly improved after the mini-CEX requirement was introduced (where 118 of 121 student versus 100 of 114 students passed, $p < .01$).

A high feasibility for using the mini-CEX to assess medical student is consistently demonstrated by these studies [Kogan, 2003; Hill 2007; Hill 2009; Playford, 2013; Rogausch, 2015; Kim, 2016]. Assessments can be performed both by junior and senior medical staff, however, it is consistently noted that grade assigned by junior medical officers are lenient or perhaps the scores awarded by senior medical officers are more stringent [Hill, 2009; Playford, 2013; Rogausch, 2015]. The tool can be used in multiple disciplines and across multiple settings both in hospital and in the community [Hill, 2007; Ney, 2009; Playford, 2013; Rogausch, 2015; Kim, 2016].

Construct validity for the mini-CEX is supported with an improvement in grades through the academic year [Koogan, 2003; Playford, 2013]. Concurrent validity is suggested by the correlation with other clinical assessments [Koogan, 2003; Ney, 2009; Wang, 2011]. However, it should be noted that one study did question both the construct and concurrent validity with variable correlation between clinical performance in the OSCE and of mini-CEX scores and lack of improvement in scores through the clinical year [Rogausch, 2015]. The mini-CEX was shown to have predictive validity by showing that performance on the mini-CEX has a small, but modest, correlation with overall future clinical skills performance [Ney, 2009].

Reliability was achieved with eight to eleven mini-CEX suggesting that in the undergraduate setting a similar number of assessments are needed to that seen in the postgraduate setting [Kogan, 2003; Hill, 2007]. In the undergraduate setting, we see similar factors to those seen in the postgraduate setting, contributing to variance. Internal factors such as the students'

aptitude for the attachment, and external variables such as examiner status; the complexity of the case, the attachment specialty and the size of the institution which trains the students contribute to variance [Hill, 2007; Rogausch, 2015].

Provision of feedback to medical students during the mini-CEX.

Engaging the student with feedback has long been recognized as a powerful tool to stimulate learning [Ramaprasad, 1983, Sadler, 1989]. The role of feedback for the mini-CEX in undergraduate education has been evaluated by a number of studies.

Several papers have identified that students and assessors who use the mini-CEX perceive that it provides a forum for and increases feedback [Dewi, 2010; Pernar, 2011; Suhoyo, 2014, Sudarso, 2016, Yanting, 2016].

Evaluation of the written feedback in the mini-CEX of undergraduates has been determined. A paper from the University of Aberdeen, UK found that in 22.7% of cases the positive aspects were not identified in 28.2% of cases no suggestions for development were identified and in 49.7% of cases no plan of action was made. [Fernando, 2006]. A qualitative study from Monash University, Australia that explored feedback found that 20% of cases there was no comments related to performance. Of the 80% that did have feedback the majority of feedback affirmed the students' performance with only 5.2% having feedback that provided directed goals to improve clinical performance [Harvey, 2013]. A third study from the University of Bern, Switzerland found that although learning needs were identified in 65% of cases these were linked to a learning goal in only 38% of cases [Montagne, 2014].

Thus we see that both students and assessors view a strength of the mini-CEX as its ability to provide appropriate and timely feedback. However, the literature also highlights that there is a gap in the written feedback provided. Of course, the written feedback does not account for verbal feedback that hopefully the assessor will have provided. Feedback that is

immediate is more effective than feedback that is delayed. However, the opportunity to reflect and consider documented feedback is also useful. The lack of this may represent a missed opportunity to enhance the learning for the student.

Electronic format for the mini-CEX.

It was postulated that the uptake of the mini-CEX may not be extensive because it is cumbersome as a paper-based tool [Torre, 2007]. Use of mobile technology to perform WBAs offers potential benefits in freedom and flexibility of the learning environment. However, it may or may not be easy to access in an environment where Information Technology access may be challenging.

Third year medical students were asked to perform 2 mini-CEX during an IM clerkship using personal digital assistants (PDAs) at the Medical College of Wisconsin, USA [Torre, 2007]. Feasibility was demonstrated with 100% of required assessments completed. The PDA was used both inpatients and outpatients and satisfaction with the PDA was high. A randomised trial to compare PDAs to paper forms for the mini-CXE was designed in McMaster University, Canada but could not be completed as students were unwilling to use the device [Norman, 2008]. A report from the University of Leeds, UK of a study that used PDAs to record the mini-CEX in final year medical students undergoing a remediation program had a high rate of use with only minor problems reported [Fuller, 2009]. A further study from the Medical College of Wisconsin, USA comparing paper-based and personal digital assistants for the mini-CEX and the reliability of both formats were similar [Torre, 2011]. Interestingly, the time taken to perform both the observation and the feedback time were significantly shorter using the paper form. Both formats were acceptable, however, faculty were significantly more satisfied using the paper form, residents more significantly satisfied using the digital form and students had no preference.

Thus, variable results exist in the literature about the benefits of the use of personal digital assistants. The studies by Torre *et al* and Fuller *et al* showed that a personal digital

assistants-based mini-CEX was feasible whereas the study by Norman *et al* suggested otherwise. Paper based forms are used at the University of Auckland.

Assessors view of the mini-CEX in Undergraduate education.

A study from the Vrije University Medical Center School of Medical Sciences, Netherlands explored faculty perceptions in performing workplace based assessments to better understand challenges in providing feedback [Daelmans 2016]. Fifty five members of the faculty undertook active discussions with the study authors. A number of issues were highlighted including that the clinical clerkships were too short to allow adequate time for observing the student, that the student choose the time to undertake the mini-CEX, the student focused on the grade received rather than the feedback, the form had too little space for narrative feedback, there was a conflict between the grading of the student which comments on the quality of teaching; and that by failing a student this created more work.

Training of Assessors.

A few studies have assessed how to train undergraduate assessors. A study from the University of Auckland, NZ used an online format and videotaped scripted performances to train General Practitioners in summative assessment with the mini-CEX. Using intraclass correlation good reliability was demonstrated [Eggleton, 2016]. A paper from the University of Cork, Ireland described using two bespoke video recordings and a 2 hour workshop to train General Practitioner assessors. The majority of participants felt the workshop was useful but it did not alter variance [Walsh, 2017].

3.3 Summary and Conclusions.

The mini-CEX is a widely used evaluation tool created to assess postgraduate trainees in the workplace. It was originally developed with a nine-point rating scale with three levels of performance: unsatisfactory (1-3), satisfactory (4-6) and highly satisfactory (7-9). In the postgraduate setting the tool provides an overall assessment of the trainee as well as a

number of subcompetencies. Initially the tool evaluated four specific aspects of a trainee's performance but this was subsequently expanded and now includes areas such as history taking, physical examination, professional qualities, counselling, clinical judgement and organization. Scores are assigned to each of these aspects of performance however studies have shown these are closely related and it has been observed that these represent a global measure of competence.

A number of studies have evaluated validity of the mini-CEX and have demonstrated criterion, content and construct and support a validity argument for the use of the mini-CEX as an assessment tool.

Multiple encounters are needed for a reliable assessment of the postgraduate trainee, with the number needed to achieve satisfactory reliability variable between specialties. Sources of variance in the assessment relate to the examiner and the case seen as well as the ability of the trainee. Variance related to examiner factors have been explored in some detail with a number of studies, but not all, showing that more experienced examiners are more stringent in their ratings. Some studies have shown that stringency of the examiner relates to their own competence and thus possibly the expectation of the examiner. Examiners give weight to different aspects of the mini-CEX and this also may lead to differences in ratings. The previous performance of the trainee may introduce contrast bias and influence the mark awarded to subsequent trainees. How this would impact on the mini-CEX, where assessments are not generally conducted sequentially and occur more ad-hoc is unclear. It has been shown that examiners make social judgements about trainees and that this may also influence rating, however, these studies were conducting in a somewhat artificial manner involving videotaped performances and whether they would have the same effect when an assessor is observing a trainee is unclear. Interestingly, training of examiners has not consistently been shown to reduce variance. Variance related to case factors has also been explored; cases that took longer to complete and cases that were more complex were

associated with higher scores. Adopting a scale that uses descriptors of entrustability may also reduce variance.

A strength of the tool is that immediate feedback is provided and this is viewed positively by both trainees and assessors. Studies have consistently indicated that the feedback given to trainees is insufficient, however, it has been shown that examiners are cautious when delivering feedback and consider the impact that the feedback may have on trainees. The mini-CEX is stressful and generates anxiety for trainees. Written feedback offers an opportunity for the trainee to reflect and consider their performance but is not fully utilised with gaps in learning and action plans not well completed.

The work by Moonen-van Loon *et al*, highlights that the mini-CEX may be combined with other workplace based assessments to evaluate the postgraduate trainee and that by combining assessments it is possible to reduce the number of assessments needed to effectively assess the trainee.

A number of studies have demonstrated that the mini-CEX is feasible to implement as an assessment tool for undergraduate medical students. As a workplace based assessment the mini-CEX provides the student with a more authentic assessment than an OSCE station or evaluation of a simulated standardised patient. As we increasingly utilise assessment for learning, it can be seen that by using the mini-CEX as an assessment tool we encourage the student to develop and integrate clinical skills with a focus on patient evaluation.

In the postgraduate setting, only a single study has evaluated scale length, and did not demonstrate any difference between a nine-point and five-point scale [Cook, 2009b]. The initial reports of the mini-CEX in the undergraduate setting used a nine-point assessment scale with an overall assessment of the students' performance and a number of subcompetencies. Subsequent studies have used both shortened [Hill, 2009] and expanded

scales [Rogausch, 2015]. This suggests that the scale length may not be important in the undergraduate setting.

There is less support for a validity argument for the use of the mini-CEX as an undergraduate assessment tool with conflicting results in the literature. The work of Rogausch *et al* questions whether the inference which links the assessment score with the interpretation of the score occurs. This is regarded as the most important step using Kane's validity framework. [Cook, 2015].

Reliability was achieved with eight to eleven assessments. Variance is again influenced by the examiners, case and the students' ability. Training of examiners did not reduce variance. External variables such as the attachment specialty and the size of the institution also were found to contribute to variance. Examiners do not appear to assess different aspects of the students' performance separately; only one factor accounted for variance in the different aspects assessed in the mini-CEX.

The mini-CEX is valued for promoting prompt and targeted feedback and provides an appropriate segue for the supervisor to provide feedback on how the student can develop. However, we still see gaps in the written feedback provided to students.

In the University of Auckland MBChB context two mini-CEX are being assessed in Year 4. These assessments occur at different clinical sites, with medicine and specialty medicine taught and practiced in slightly different ways across the sites. The literature suggests that a greater number of assessments are needed. I was interested to determine if, with only two evaluations, any particular group of students would be disadvantaged or if there was any difference in mini-CEX scores between the clinical sites and this is evaluated in Chapter 4.

The work by Moonen-van Loon *et al*, highlights that as a component of postgraduate assessment the number of evaluations can be reduced when combined with other

assessments. In Year 4 the two mini-CEX are used as part of a programme of assessment, combined with OSCE stations and the alcohol and drug assessment to evaluate clinical and communication skills. However it is unclear if we have the correct “cocktail of assessments”. In Chapter 5 I will evaluate the mini-CEX in the context of programmatic assessment.

Chapter 4 - The mini-CEX in Year 4 Medical Students at the University of Auckland.

4.1 - Introduction.

The literature suggests that eight or more mini-CEX assessments are needed for the tool to be reliable. However, in the University of Auckland it is being used twice, in Year 4, as students are starting their clinical training. A number of studies have shown that variance in grade is influenced by the rater and by the case [Margolis, 2006; Wilkinson, 2008; Weller, 2009a; Hill, 2009; Alves de Lima, 2013]. The impact of the difficulty of an individual case or the stringency of an individual rater is minimised by increasing the number of cases and assessors [Margolis, 2006].

In Year 4 the students are assessed twice, in their General Medicine rotation and in their Specialty Medicine rotation. I was interested in determining how the mini-CEX was being used in the University of Auckland and whether there was any difference between the mini-CEX in the two clinical rotations.

Places in medical schools are limited. Students entering the University of Auckland medical school have a number of potential routes of admission with selection based on a combination of academic merit and personal qualities. Candidates may either be first year or graduate entrants. The first year or graduate candidate may enter via the general admission scheme, Regional Rural Admission Scheme (RRAS), Māori and Pacific Admission Scheme (MAPAS), Undergraduate Targeted Admission Scheme (UTAS) or be an International student. I was interested to determine, whether, any particular group may be significantly disadvantaged by this.

Students in Year 4 undertake attachments at six different clinical sites. Rotorua Hospital (233 beds) and Tauranga Hospital (360 beds) are smaller teaching sites than Waikato, South Auckland, Waitematā, and Auckland (673 beds, 745 beds, 953 beds, and 1124 beds

respectively) [Ministry of Health, 2018]. As previously highlighted the way the Specialty Medicine and General Medicine are delivered is different between these hospitals. In a larger hospital, such as Waitematā, there is a dedicated renal service, where Year 4 students are exposed to inpatients with acute kidney injury, chronic kidney disease and end stage kidney disease and attend clinics for patients who have received a renal transplant. In Tauranga there is a visiting renal service and patients who need in-hospital care are transferred to Waikato Hospital, hence no renal rotation is offered.

In the postgraduate setting the original paper describing the mini-CEX was undertaken across 5 residency programmes in Pennsylvania. Detailed descriptions of these programmes were not provided, however, it was noted that in terms of mini-CEX scores there were no significant differences between programmes [Norcini, 1997]. A larger study included 21 postgraduate programmes, and described that there were a range of examiners, residents and mini-CEX encounters but did not evaluate if there were any differences between the programmes [Norcini, 2003]. In the undergraduate setting, work from the University of Bern identified that smaller sites awarded statistically higher marks to students than larger sites. I was thus interested to determine if the differences in scores awarded for the mini-CEX may be apparent between University of Auckland clinical training sites.

4.2 - Aim.

The overall aim of this chapter was to explore in detail how the mini-CEX was being used in Year 4 students at the University of Auckland. The specific research questions to address to evaluate this were:

- 1.) To understand how the mini-CEX is being used in Year 4 students at the University of Auckland.
- 2.) To determine if there were any differences in scoring for the mini-CEX in General Medicine and Specialty Medicine rotations.
- 3.) To evaluate the measurement characteristics of the mini-CEX.

4.) To determine any variation in grade achieved in the mini-CEX based on demographic variables, entry scheme into the medical school or the clinical teaching site.

4.3 - Ethical Statement.

Ethical approval for this component of the thesis was granted by the University of Auckland Human Participants Ethics Committee (Ref 020974).

4.4 - Methods.

Deidentified data from Year 4 medical students enrolled between 2013 and 2017 (inclusive) in the University of Auckland MBChB programme was extracted from the Medical Programme Directorate and Department of Medicine databases.

Demographic data obtained included the year that the student undertook Year 4 (2013, 2014, 2015, 2016, 2017), gender (female, male), entry route to medical school (General, RRAS, MAPAS, UTAS, International), graduate status (undergraduate, graduate) and clinical teaching site where the student spent the year (Waitematā, Auckland, South Auckland, Waikato, Rotorua, Tauranga).

Review of individual mini-CEX forms from students in the 2013 and 2014 cohort was undertaken as scores for the subcompetencies of History taking/Communication Skills, Physical Examination Skills, Clinical Judgement/Reasoning and Humanistic qualities/Professionalism were not recorded in the Medical Programme Directorate database for 2013 and 2014 but was available for 2015, 2016 and 2017. Not all the forms were available and some had incomplete data, however, where available the information on subcompetencies was recorded. In addition, data from the 2013/2014 cohort was obtained about the assessor, their seniority (SMO, RMO), the time the assessor spent with the student, the setting of the assessment (inpatient, outpatient, acute), the age, gender and main problem of the patient, and the case complexity.

The marks for the General Medicine and Specialty Medicine mini-CEX were recorded including the overall grade and scores for the subcompetencies of History taking/Communication Skills, Physical Examination Skills, Clinical Judgement/Reasoning and Humanistic qualities/Professionalism from the Medical Programme Directorate database.

Statistical Analysis.

Data was analysed using Statistical Package for the Social Sciences (SPSS) version 25 (IBM Corp, Armonk, NY, USA).

Descriptive statistics including mean, mode and standard deviation were determined for the mini-CEX grades.

A paired t-test was used to compare the age of the patient seen, complexity of cases and duration of the mini-CEX in General Medicine and Specialty Medicine. A one-way analysis of variance (ANOVA) was used to compare different complexity with a p value of $< .05$ considered significant. To compare the overall score in the General Medicine and Specialty Medicine mini-CEX Pearson's χ^2 test and regression analysis was performed.

Principal component analysis using varimax rotation was undertaken to evaluate dimensionality of the mini-CEX. The Kaiser-Meyer-Olkin (KMO) measure of sampling adequacy and Bartlett's test of sphericity were determined to evaluate the suitability of the data for structure detection with a KMO of greater $.70$ and a Bartlett's test of sphericity less than $.05$ considered appropriate [Field, 2013].

The Mann-Whitney U test and ANOVA was used to compare scores for different student groups, with a p value of $< .05$ considered significant. The effect size was calculated using a cutoff of 0.1 for a small effect, 0.3 for a medium size and 0.5 for a large size.

Finally, multivariate analysis of variance (MANOVA) using the General Medicine and Specialty medicine overall mini-CEX scores as dependent variables and cohort, gender, graduate status, entry scheme, and clinical site as independent variables was conducted with a Wilks' lambda of $< .05$ considered significant.

4.5 - Results.

4.5.1 - Study Population.

A total of 1203 Year 4 students were included in the analysis. There were more female (634) than males students (569) enrolled in the programme. The majority of students, 723, were admitted via the general admissions scheme, 222 were admitted under the MAPAS, 165 students were admitted under the RRAS, 91 were International medical students and 2 admitted under the UTAS. There were more undergraduates (911) than postgraduate students (292) (Table 4). The majority of students had clinical rotations in the greater Auckland metropolitan region (Table 4). Of note, in 2013, the Waikato clinical site had students who undertook some clinical attachments in Rotorua. However, the students that undertook mini-CEX assessments in Rotorua could not be identified and information from these students were analysed as part of the Waikato cohort. From 2014 onwards Year 4 students were able to spend the entire year in Rotorua. In 2015 Year 4 students started in training in Tauranga.

Table 4 - Demographic details of Year 4 medical students.

Demographics/Years	2013	2014	2015	2016	2017	Total
Number of Year 4 Students	202	236	234	251	280	1203
Gender Females	105	124	111	143	151	634
Admission Scheme						
General	128	136	147	140	172	723
MAPAS*	38	51	40	51	42	222
RRAS^	20	30	31	40	44	165
International	16	19	16	20	20	91
UTAS#	0	0	0	0	2	2
Graduate Status						
Undergraduate	161	174	186	187	203	911
Graduate	41	62	48	64	77	292
Clinical Site						
Waitematā	49	54	45	46	54	248
Auckland	76	78	59	71	73	357
South Auckland	58	57	61	64	72	312
Waikato	19 [‡]	36	35	35	46	171
Tauranga	0	0	22	23	24	69
Rotorua	0	11	12	12	11	46

*MAPAS - Māori and Pacific Admission Scheme (MAPAS)

^RRAS - Rural regional admission scheme

#UTAS - Undergraduate targeted admissions scheme

‡ Some Year 4 students from Waikato did clinical rotations in medicine at Rotorua in 2013

4.5.2 - The mini-CEX in Year 4 Medical Students at the University of Auckland.

In total 2394 mini-CEX were completed by 1203 Year 4 students between 2013 and 2017: 1196 in general medicine and 1198 in specialty medicine.

Of the 1196 mini-CEX that were completed during the General Medicine rotation 1 student failed the mini-CEX, 14 received borderline performance, 900 passed the mini-CEX and 281 received a distinction. Of the 1198 mini-CEX were completed during Specialty Medicine rotations 3 students failed the mini-CEX, 18 received borderline performance, 911 passed the mini-CEX and 266 received a distinction.

4.5.3 - Research Question 1: The mini-CEX in Year 4 students at the University of Auckland.

In total, 438 students undertook the Year 4 course in 2013 and 2014. The mini-CEX forms were that were available from 2013 and 2014 were individually reviewed.

Assessors

The assessor was documented in 380 mini-CEX forms in General Medicine and 389 mini-CEX forms in Specialty Medicine. The 2013 and 2014 Year 4 students were assessed by 254 people: 25 people performed mini-CEX assessments in both General and Specialty Medicine, 106 in only General Medicine and 123 in only Specialty Medicine. Over the two year period, individual assessors undertook between 1 and 21 mini-CEX (range 1-18 mini-CEX in General Medicine 1-18, and range 1-21 mini-CEX in Specialty Medicine), with 90 assessors performing 1 assessment, 59 assessors performing 2 assessments, 27 assessors performing 3 assessments and 28 assessors performing 4 assessments in the two year period (Figure 3). Thus, we see the majority of assessors are performing less than 2 assessments per year.

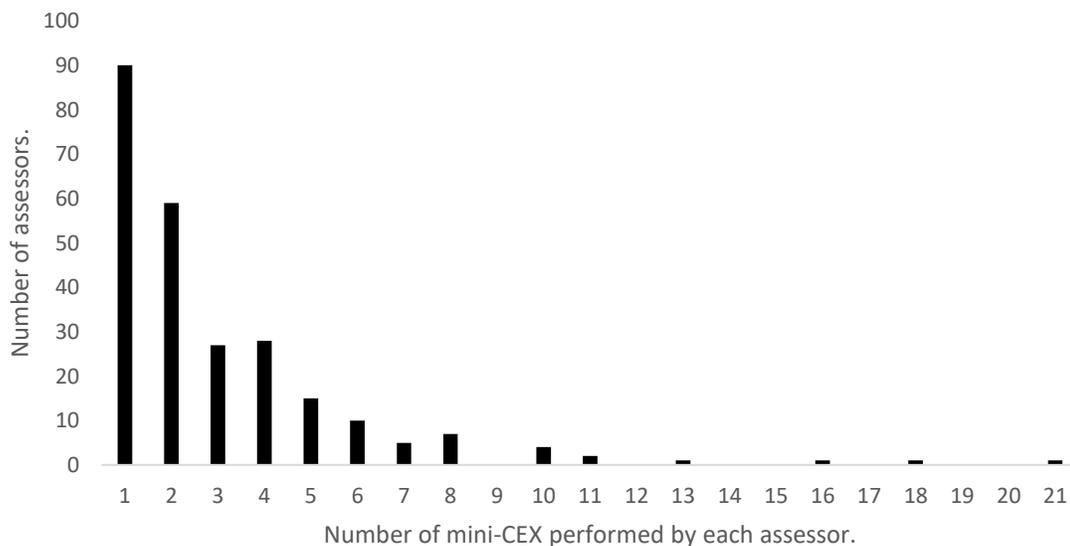


Figure 3 - Number of mini-CEX assessments undertaken by assessors 2013-2014.

Regarding the seniority of the assessor: 148 assessments were performed by RMOs and 621 by SMOs. Of these 327 SMO and 53 RMO performed the General Medicine mini-CEX and 294 SMOs and 95 RMO performed the Specialty Medicine mini-CEX. Using the paired t-test it was more likely for students to be assessed by RMOs in the Specialty Medicine mini-CEX than the General Medicine mini-CEX ($t(338) = -3.96, p < .01$).

Duration

The duration of observing the student ranged from 3 to 120 minutes. The mean duration for all assessments was 23.5 minutes ($n = 657$, median 20.0 minutes, range 3 - 120 minutes, interquartile range 15 - 30 minutes). Using a paired t-test it was determined that there was no significant difference in the time spent observing students in the General Medicine mini-CEX and the Specialty Medicine mini-CEX ($t(328) = 1.96$, $p = .05$). However, it can be seen that there is a wide range in the time assessors spend with the students.

Patients

The majority of mini-CEX assessments, 84%, were on inpatients. A further 10% were undertaken on outpatients and 6% were acute assessments, undertaken in the Emergency Department (644 patients inpatient, 77 outpatient and 45 acute assessments). The patients ranged in age from 15-97 years, mean 63.5 years. Male patients were 58.2% ($n = 431$) and 41.8% ($n = 303$) were female. Using a paired t-test it was determined that patients participating in the General Medicine mini-CEX were significantly older than those seen in the Specialty Medicine mini-CEX (66.2 years \pm 18.0 versus 60.2 \pm 17.6 years, $t(292) = 4.50$, $p < .01$). For 310 students the gender of the patient was recorded for both assessments. Using the paired t-test there was no significant difference in the General Medicine mini-CEX or Specialty Medicine mini-CEX ($t(310) = -0.31$, $p = .76$).

The patients' clinical problem was categorised (by JdZ) by the main system recorded: 28.8% were Respiratory, 27.7% were Cardiology, 14.5% were Neurology, 7.9% were Infectious Disease, 7.2% were Gastroenterology, 3.2% were Oncology, 2.8% were Rheumatology, 2.4% were Renal, 2.4% were Dermatological and 0.8% Endocrinology. Thus, the students were exposed to wide range of clinical problems.

The complexity of the patients' problem was recorded for all of 2013, however, this information stopped being recorded part way through 2014. The complexity of the case was available for 210 students in the General Medicine mini-CEX, 236 students in the Specialty

Medicine mini-CEX with 166 students having the complexity recorded for both assessments. In total 4.9% of cases were low in complexity (22 patients), 74.4% were medium in complexity (332 patients) and 20.6% were high in complexity (92 patients). Using a paired t-test there was no significant difference in the complexity of patients seen in the General Medicine and Specialty Medicine mini-CEX ($t(166) = .44, p = .66$). Using ANOVA there was no difference in the score assigned to the General Medicine mini-CEX based on the case complexity (3.17 +/- .41 low, 3.19 +/- .42, medium, 3.24 +/- .44, high; $F(2, 207) = .28, MSE = .05, p = .76$), nor the Specialty Medicine mini-CEX (3.00 +/- 0 low, 3.24 +/- .44, medium, 3.29 +/- .46, high; $F(2, 233) = 2.86, MSE = .54, p = .06$).

Thus the students saw few easy cases, reflective of the case mix of patients currently seen in hospital. They were asked to assess patients with a wide range of clinical complaints. They were a little older in General Medicine but otherwise similar.

4.5.4 - Research Question 2: Comparing the Mini-CEX in General Medicine and Specialty Medicine Rotations.

All 2,394 mini-CEX forms from 2013 to 2017 (inclusive) were used to calculate descriptive statistics for the mini-CEX test scores. A four-point ordinal scale was used for the grades assigned to overall clinical competence where 1 = fail, 2 = borderline performance, 3 = pass and 4 = distinction.

Of the 1196 mini-CEX that were completed during the General Medicine rotation by 1203 Year 4 students, one student failed the mini-CEX, 14 received a borderline performance, 900 passed and 281 received a distinction. Using the four-point nominal scale the mean score was 3.22 +/- .45 with a mode of 3 (Figure 4). Of the 1198 mini-CEX that were completed during Specialty Medicine rotations, three students failed, 18 received borderline performance, 911 passed and 266 received a distinction. Using the four-point scale the mean score was 3.20 +/- .45 with a mode of 3 (Figure 4). Thus for both the General

Medicine and Specialty Medicine mini-CEX most students received a pass with very students failing over the entire 5 year period.

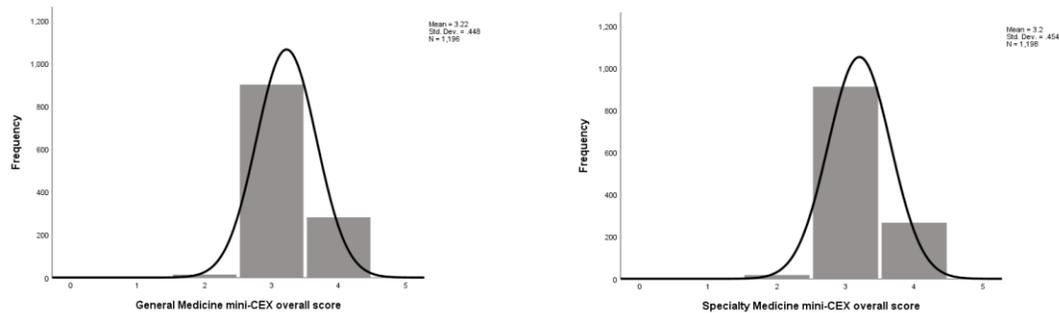


Figure 4 - Histogram plot of mini-CEX overall scores in General Medicine and Specialty Medicine.

The difference in the means was .02. Using the χ^2 test there was a significantly higher score in General Medicine than in Specialty Medicine ($\chi^2 (2, n = 1195) = 25.06 p < .01$). The effect size was found to have weak or non-significant effect ($\psi = .14$; $\psi_c = .08$). Essentially the same mark was awarded in General Medicine and Specialty Medicine.

A regression analysis was conducted with the General Medicine mini-CEX as the outcome variable and the Specialty Medicine mini-CEX as the predictor. The results demonstrated that the Specialty Medicine mini-CEX significantly predicted General Medicine mini-CEX score with a small effect size ($R^2 = 0.004$, standardised $\beta = .07$, $p = .02$). Thus, the students' score in one assessment did not always align with those in the second assessment.

4.5.5 - Research Question 3: Measurement characteristics of the mini-CEX.

In addition to the overall score, four subcompetencies were assessed: History taking/communications skills; Physical examination skills; Clinical judgement/reasoning; and Humanistic qualities/professionalism. Again a four-point nominal scale was used for the

grades assigned to these subcompetencies where 1 = major deficiencies, 2 = some reservations, 3 = good and 4 = excellent.

Principal component analysis (PCA) was conducted to evaluate dimensionality for the subcompetencies of both the General Medicine and Specialty Medicine. KMO = .71 and significant Bartlett's test of sphericity ($\chi^2 (28) = 1328.14, p < .01$) confirmed the sample adequacy for PCA. For the General Medicine mini-CEX this revealed that a two-factor solution explained 68.7% of variance. The overall communalities ranged from .56 to .99. History taking/communications skills, Physical examination skills and Clinical judgement/reasoning load on the first component explaining 43.6% of variance while the only Humanistic qualities/professionalism loaded on the second component adding 25.1% of the total variance (Table 5). For the Specialty Medicine mini-CEX this showed that a single factor solution explained 61.0% of variance with the overall communalities ranging between .56 and .65 confirming unidimensionality of the test (Table 5). When combining the General Medicine and Specialty Medicine mini-CEX, PCA showed a single factor solution explained 66.7% of variance (Table 5) with the overall communalities ranging between .47 and .76 confirming unidimensionality of the test with KMO = .80 and significant Bartlett's test of sphericity ($\chi^2 (6) = 3372.74, p < .01$) confirmed the sample adequacy. This suggests that although we are assessing four items, that these closely correlate, and we may not be able to discriminate between these components of the students' ability.

Table 5 - Loadings on principal components.

Mini-CEX subcompetencies/Components	1	2
<i>Mini-CEX General Medicine</i>		
General medicine mini-CEX History Taking /Communication Skills domain	0.78	-0.05
General medicine mini-CEX Physical examination skills	0.75	-0.05
General medicine mini-CEX Clinical Judgement/Reasoning	0.76	0.09
General medicine mini-CEX Humanistic Qualities/Professionalism	0.01	0.99
<i>Mini-CEX Specialty Medicine</i>		
Specialty medicine mini-CEX Humanistic Qualities/Professionalism	0.81	
Specialty medicine mini-CEX History taking/Communication	0.79	
Specialty medicine mini-CEX Clinical Judgement/Reasoning	0.77	
Specialty medicine mini-CEX Physical examination skills	0.75	
<i>Mini-CEX Combined</i>		
Combined mini-CEX History taking/Communication	0.47	
Combined mini-CEX Physical examination skills	0.71	
Combined mini-CEX Clinical Judgement/Reasoning	0.74	
Combined mini-CEX Humanistic Qualities/Professionalism	0.76	

The model used Varimax rotation with Kaiser Normalization.

4.5.6 - Research Question 4: Variation in the mini-CEX based on student characteristics or training site.

To better understand the scores awarded, the overall score for mini-CEX in General Medicine and Specialty Medicine were determined for each of these variables: assessor, gender, graduate status, cohort, entry scheme and clinical site.

Assessor

The mean score for SMOs who assessed the mini-CEX in General Medicine was 3.20 +/- .42 and the mean score for RMOs was 3.38 +/- .49. Using the Mann-Whitney U test the overall score in the General Medicine mini-CEX was significantly higher from RMOs than SMOs

($U(380) = 7148, Z = -2.76, p < .01$). The mean score for SMOs who assessed the mini-CEX in the Specialty Medicine was 3.19 +/- .41 and the mean score for RMOs was 3.34 +/- .48. Using the Mann-Whitney U test the overall Specialty Medicine mini-CEX score was significantly higher from RMOs than SMOs ($U(389) = 11984.5, Z = -2.83, p < .01$). When combining the data for both the General Medicine and the Specialty Medicine mini-CEX the mean overall score awarded by SMOs was 3.20 +/- .41, the mean score awarded by RMOs was 3.35 +/- .48. Using the Mann-Whitney U test the overall score was significantly higher from RMOs than SMOs ($U(769) = 38940, Z = -3.92, p < .01$). However, the effect size was found to be negligible ($r < .01$) suggesting that the student was not advantaged or disadvantaged by the seniority of the assessor.

Gender

Using the Mann-Whitney U test the overall score for female students (mean 3.23 +/- .44) did not differ significantly from male students (mean 3.21 +/- .46) in the General Medicine mini-CEX ($U(1195) = 1753000, Z = -.66, p = .51$). However, the overall score for female students (mean 3.24 +/- .45) was significantly greater than male students (mean 3.16 +/- .45) in the Specialty Medicine mini-CEX ($U(1197) = 16654.5, Z = -2.78, p < .01$). The effect size was found to be negligible ($r = .08$). When combining the scores for both the General Medicine and the Specialty Medicine mini-CEX the mean score for female students (mean 3.24 +/- .45) was significantly higher than for male students (mean 3.19 +/- .46) however, the effect size was negligible ($U(2394) = 683670, Z = -2.43, p = .02, r < .01$). Overall, the gender of the student did not have an important impact on their grade.

Graduate status

Using the Mann-Whitney U test the overall score for students who entered medical school as undergraduates (mean 3.23 +/- .46) was the same as graduates (mean 3.20 +/- .42) in the General Medicine mini-CEX ($U(1195) = 127731, Z = -.87, p = .38$). However, the overall score for graduates (mean 3.26 +/- .46) was significantly greater than undergraduates (mean 3.18 +/- .45) in the Specialty Medicine mini-CEX ($U(1197) = 123106.5, Z = -.23, p = .02$). The effect

size was found to be negligible ($r = .07$). When combining the scores for both the General Medicine and the Specialty Medicine mini-CEX the mean score for undergraduates (mean 3.21 +/- .45) was the same for graduate students (mean 3.23 +/- .44) ($U(2394) = 514976$, $Z = -1.02$, $p = .31$).

Cohort (year at medical school)

Using a one way ANOVA no difference in mean overall score assigned to each cohort was seen in the General Medicine mini-CEX ($F(4,1191) = 1.69$, $MSE = .34$, $p = .15$) nor the Specialty Medicine mini-CEX scores ($F(4,1193) = .84$, $MSE = .17$, $p = .50$). When combining the scores of the General Medicine and Specialty Medicine mini-CEX there was no difference in scores assigned to each cohort ($F(4, 2394) = .483$, $MSE = .10$, $p = .75$).

Entry scheme

As there were only two students who entered by the UTAS, for this component of the analysis, their scores were merged with students in the General entry scheme.

Using a one-way ANOVA no significant difference in the Specialty Medicine mini-CEX overall score was found in students entering by different schemes ($F(3,1194) = 2.41$, $MSE = .50$, $p = .07$). A significant difference was seen between students entering by different schemes in the general medicine mini-CEX overall score ($F(3,1192) = 5.85$, $MSE = 1.16$, $p < .01$). A Tukey's honestly significance difference post-hoc test showed the mean score achieved by International students (3.09 +/- .44, $p < .01$) and MAPAS students (3.15 +/- .45, $p = .02$) were significantly lower compared to General/UTAS students (3.25 +/- .45). The mean score achieved by International students were also significantly lower than RRAS students ($p = .04$). There was no significant difference between RRAS students and MAPAS students score ($p = .20$). There was no significant difference between General/UTAS and RRAS students (3.24 +/- .46, $p = .99$). There was no significant difference in mean scores between MAPAS students and International students ($p = .64$). However, the effect size, where there were differences seen, were all insignificant ($\eta^2 = .01$). When combining the scores of the General

Medicine and Specialty Medicine mini-CEX a significant difference was seen between students entering by different schemes in the general medicine mini-CEX overall score ($F(3,2393) = 7.74, MSE = 1.56, p < .01$). A Tukey's honestly significance difference post-hoc test showed the mean score achieved by International students ($3.10 \pm .42, p < .01$) and MAPAS students ($3.15 \pm .43, p < .01$) were significantly lower compared to General/UTAS students ($3.24 \pm .46$). The mean score achieved by International students ($p < .01$) and MAPAS students ($p < .04$) were significantly lower compared to RRAS students ($3.24 \pm .46$). There was no significant difference between General/UTAS and RRAS students ($p = .99$). There was no significant difference in mean scores between MAPAS students and International students ($p = .61$). Of note, the effect size these differences were insignificant ($\eta^2 < .01$). So we see that General entrant students and RRAS students achieve similar marks, however MAPAS and International students achieve lower grades. In practice, as the effect size is trivial, this did not impact on the overall grade awarded to MAPAS and International students.

Clinical site (Geographical site of training)

Using a one-way ANOVA no difference in mean overall score in the General medicine mini-CEX assigned to students was seen in different clinical sites ($F(5,1190) = 1.203, MSE = .24, p = .31$). A significant difference was seen in the mean overall score in Specialty Medicine mini-CEX assigned to students in different clinical sites ($F(5,1192) = 3.124, MSE = .64, p < .01$). However, a Tukey's honestly significance difference post-hoc test did not show there was a significant difference between clinical sites ($p = 0.06 - 1$). The number of students at each site varied ($n = 65 - 356$). Post-hoc test using Games-Howell suggested significantly higher scores awarded to students at the Waitematā clinical site ($3.25 \pm .45, n = 248$), than at Tauranga ($3.07 \pm .31, n = 69, p < .01$), and South Auckland ($3.24 \pm .47, n = 308$) and Tauranga ($p < .01$). No differences between other sites were noted. Again the effect size was negligible ($\eta^2 = .01$). When combining the scores of the General Medicine and Specialty Medicine mini-CEX a significant difference was seen in the mean overall score in Specialty Medicine mini-CEX assigned to students in different clinical sites ($F(5,2393) = 2.264, MSE =$

.46, $p = .046$). However, post-hoc tests using a Tukey's honestly significance ($p = 0.20 - 1$) and Games-Howell ($p = .14 - 1$) did not suggest a significant difference.

Group Comparisons

To reduce the probability of Type 1 error all group comparisons were conducted using a single omnibus MANOVA test. Dependent variables were General Medicine, Specialty medicine and overall mini-CEX scores and independent variables included assessor, cohort, gender, graduate status, entry scheme and clinical site (3 x 6 design). This did not show a statistically significant effect for the General Medicine mini-CEX score, the Specialty Medicine mini-CEX score or the combined General Medicine and Specialty Medicine mini-CEX score ($F(127) = 1.22, p = .10$). Thus, overall when assessing all variables, no difference in mini-CEX grade is seen attributable to assessor, gender, graduate status, cohort, entry scheme or clinical site.

4.6 - Discussion.

In this chapter we see how the mini-CEX is being used in Year 4 medical students in the MBChB University of Auckland programme. A wide range of assessors use the mini-CEX, examining students using patients with a broad spectrum of clinical problems. Of note there was a large variation on the time spent examining the student, with one assessor completing their evaluation within 3 minutes and another taking 2 hours. The recommended duration for an assessment is 15-20 minutes observing the student with a further 5 minutes for feedback (Appendix 1). This suggests that developing a training resource for the assessors may be of benefit. Reassuringly, there was little difference between the mini-CEX in the General Medicine and Specialty Medicine run, irrespective of the seniority of the assessor. The vast majority of students passed or achieved a distinction with very few students having a borderline pass or fail for the assessment. If interpreting feedback then the Course Supervisor or Board of Examiners should be cautious in interpreting the individual grades assigned to the subcompetencies in the mini-CEX. Although there is information on the four subcompetencies, it is challenging to discriminate

between these components of the student's performance. Reassuringly, there does not seem to be a significant difference between the mini-CEX scores for candidates, irrespective of their entry characteristics or clinical training site.

The feasibility of the mini-CEX as an assessment tool is demonstrated with more than 99% of Year 4 students able to complete two mini-CEX assessments during the study period. This is unsurprising; the initial descriptions of the feasibility of using the mini-CEX in an undergraduate programme, where nine mini-CEX were requested over a nine week period, demonstrated over 80% of mini-CEX were completed [Kogan, 2002; Kogan, 2003]. Subsequent undergraduate studies demonstrated that more than 90% of required mini-CEX were completed [Hill 2007; Hill 2009; Playford, 2013; Rogausch, 2015; Kim, 2016]. Although new as an assessment tool for Year 4 medical students in the Auckland MBChB program most assessors are familiar with the mini-CEX as it has been used to assess Physician trainees for a number of years in the postgraduate setting and of course, the number of assessments is much fewer than reported in other programmes.

Overall, we see a very high pass rate for the mini-CEX. More than 98.4% of assessments were a pass or distinction with 1.5% assessments (36/2398) receiving a fail or borderline mark over the 5-year study period. It would have been interesting to see if a fail in the mini-CEX was predictive of the final overall year grade for Year 4 students, but that beyond the scope of this study. Of note, the student has two specialty medicine rotations, and can undertake a mini-CEX in each of these rotations, and choose which assessment to record. In addition, if the student receives a fail or borderline mark they are offered the opportunity to retake the assessment, with the better of the assessments recorded. Also, the structure of the mini-CEX is such that the student is offered immediate, and specific feedback. This allows the student the opportunity to immediately reflect, and hopefully correct any areas of concern. Unfortunately, I was unable to determine how many students had earlier assessments which were a fail or borderline mark.

What is apparent is the large number of assessors who evaluate students. This also occurs in other undergraduate programmes. For example, at the University of Southampton 326 students were evaluated by 813 examiners [Hill, 2009], and at the University of Bern 116 students were evaluated 512 assessors [Rogausch, 2015]. In the Rural Clinical School of Western Australia “Hundreds of doctors” were noted to complete 5306 mini-CEX in 215 students [Playford, 2013]. Each student, at these institutions, undertook more mini-CEX evaluations annually than at the University of Auckland, with students asked to undertake 15, 11 and 21 mini-CEX evaluations respectively. However, it is unclear from these papers how many assessments are undertaken by individual examiners. In the postgraduate setting the initial description of the mini-CEX included 97 examiners who undertook 1-26 assessments [Norcini, 1997]. Of note, at the University of Auckland, the majority of assessors undertook very few assessments: 58.7% of assessors (149/254) undertook only 1 or 2 assessments, with a further 21.6% (55/254) undertaking between 3 and 4 assessments over a two year period. This reflects the deployment of students across a variety of clinical campuses and that the assessment is undertaken in the workplace. In addition, we see a wide range of time spent performing the assessment. The mini-CEX instructions suggest that the entire assessment should take 15-20 minutes (Appendix 1). In our study it was found that the time spent on an assessment ranged from 3 to 120 minutes. In my experience, it would be challenging to evaluate a student in less than 15 minutes, to do so within 3 minutes suggests that an inadequate assessment was undertaken, and this was viewed as a “tick-box exercise”. Conversely, if an assessment is taking over one hour, then the student is likely to be undertaking a “long-case” [Norcini, 2002]. The long case is, of course, a valuable assessment and authentic, but also has issues around reliability due to case and examiner factors [Wass, 2004]. At the University of Southampton no prior training was required to undertake a mini-CEX, however, notes on how to use the mini-CEX and an instructional video were made available to assessors [Hill, 2009]. At the Rural Clinical School of Western Australia no training was undertaken [Playford, 2013]. At the University of Bern an interactive workshop, written material and course videos with student-patient encounters are offered for training [Rogausch, 2015]. At the University Auckland, there is

instructional material available with the mini-CEX score sheet (Appendix 1) but no other training material is available. The mini-CEX is used to assess Physician trainees, so most RMOs and SMOs will be familiar with the format. As previously noted, it has not been consistently found that training in the use of the mini-CEX is of benefit and to mandate this may be 1) viewed as an administrative burden and 2) used as a reason for potential assessors not to perform the assessment. Nevertheless, developing and improving access to training material for the mini-CEX, specifically highlighting the focus of an assessment and how the marking rubric is used, may be of beneficial.

The seniority of the assessor has been noted to significantly affect the grade assigned to the student [Kogan, 2003; Hill, 2009; Playford, 2013; Rogausch, 2015]. This was again demonstrated in the University of Auckland context, with RMOs awarding a significantly higher score than SMOs. We see more assessments completed by RMOs in Specialty Medicine rotations. It may be that this occurs as there are more advanced trainees who undertake teaching and assessment in the specialty medicine rotations; General Medicine RMOs tends to rotate more frequently than Specialty Medicine trainees. It may be that Physicians supervising General Medicine rotations are more diligent in undertaking assessment than Physicians supervising Specialty Medicine rotations, or this may be due to random error, as this sampling was taken from only two of the five years of data. RMOs will rotate between clinical training sites and will have less opportunity to develop teaching skills than SMOs. However, for Year 4 students, at this stage of their development the focus is on acquisition of basic examination skills and clinical problems so assessment and feedback from RMOs seems reasonable (or to have face validity). Interestingly, at the Waitemata clinical campus, there has been a move for all General Medicine mini-CEX to be performed by SMOs. Importantly, it should be noted that whether an RMO or SMO assessed the student had a negligible effect on the grade. Essentially, the most common mark was a “Pass” for students.

The complexity of the case has been shown to influence the mark award to postgraduate trainees [Norcini, 2003] and undergraduate students [Hill, 2009, Playford, 2013]. The fact that more complex cases are associated with better scores suggests that assessors may account for, and overcompensate for more complex patients when marking. This may be especially true for Year 4 students; and an assessor is unlikely to fail a student when they are seeing a complex patient with a rare or complicated illness. When the mini-CEX was first introduced to the University of Auckland programme information about the complexity of the case was documented. However, part way through 2014 the assessment form was revised and this information was no longer recorded. In our cohort, we did not find any significant difference in the students' mini-CEX grade based on the complexity of the case. This may be due to the reduced range in grades, the small sample size, leading to underestimating any difference and that by reducing the scale that students received a pass or distinction grade thus preventing possible overcompensation of grade.

As demonstrated in other studies [Norcini, 1995; Norcini, 2003] a wide range of problems are assessed during the mini-CEX. In our cohort more male than female patients were seen and this was no different between General Medicine and Specialty Medicine rotations. Slightly older patients were seen in General Medicine assessments, and anecdotally this mimics the profile of patients seen in clinical practice. Patient characteristics, such as the age or gender have not particularly been described in the literature. The Respiratory (27.7%) and Cardiology (14.5%) make up the majority of hospital admissions and were the two main clinical problems assessed. These both involve examination of the thorax, and it may be that the student or assessor approached more men to be evaluated to maintain the patients' dignity. A similar amount of time was spent with students during assessment of the General Medicine and Specialty Medicine mini-CEX. We did see a significantly higher score awarded to students in the General Medicine mini-CEX compared to the Specialty Medicine mini-CEX (3.22 versus 3.20), however the effect size was again negligible.

It was shown, using regression analysis, that the mark awarded in the Specialty Medicine mini-CEX predicted the mark in the General Medicine mini-CEX. As these are medical students starting their clinical attachments this is perhaps unsurprising. We see that a wide range of clinical problems are assessed and although the student may be outstanding in one area they may be weak in others. As the students' knowledge and experience expands it would be expected that they develop more consistent abilities in all areas. In addition, it should be recognised that some variance is likely to be due to features other than candidates' clinical abilities, such as the assessor and case characteristics [Hill, 2009; Rogausch, 2015].

Dimensionality of the internal structure of the mini-CEX considers how the tool's scores can be interpreted and used. The University of Auckland MBChB mini-CEX provides a global evaluation of the student's performance and also provides a grade for the student in areas of History taking/Communication Skills, Physical Examination Skills, Clinical Judgement/Reasoning and Humanistic qualities/Professionalism. Previous studies of the mini-CEX have shown that assessors do not assess different subcompetencies separately. It has been shown that mini-CEX subcompetency scores are highly correlated [Norcini, 2003, Kogan, 2003, Cook, 2009b] and that error also correlated across subcompetencies [Margolis, 2006]. Factor analysis can be done to evaluate dimensionality in more detail. Only one study in postgraduate trainees used factor analysis to evaluate dimensionality of the internal structure of the mini-CEX and demonstrated a single dimension [Cook, 2010]. Two studies evaluated dimensionality in undergraduate students, both using a mini-CEX with six subcompetencies (history taking, physical examination, communications/counselling, clinical judgement, professionalism and organisation and efficiency). The first study from the University of Southampton found a single dimension [Hill, 2009]. However, this study averaged item scores from multiple observations prior to analysis which may have reduced the discrimination of items and reduced dimension score. A more recent study from the University of Bern in Year 4 medical students, which used a ten-point mini-CEX scale showed the six domains closely correlated and only a single factor

emerged [Berendonk, 2018]. I used factor analysis and found that a single factor explained the Specialty Medicine domain scores, but that a two-factor solution explained the General Medicine scores with the first component being History taking/communications skills, Physical examination skills and Clinical judgement/reasoning load and the second factor being Humanistic qualities/professionalism. Although research in the mini-CEX has consistently shown a single factor explaining dimensionality, other research evaluating clinical performance has demonstrated two factors usually with a technical component, i.e. clinical examination, and a professional dimension i.e. empathy and interpersonal skills. Using a 12-item rating form medical students were shown to have variance accounted for by a two-factor solution (clinical skills and humanism) [Wilkinson, 2003]. Some studies of postgraduate students have also suggested a two-factor solution [Nasca, 2002; Silber, 2004; Hojat, 2007; Greenburg, 2007]. The Auckland MBChB mini-CEX assesses four domains, fewer than many other mini-CEX. The two-factor solution seen in General Medicine has a first cluster that evaluates the students' clinical abilities and a second that evaluates professionalism. An explanation would be that Year 4 students are developing clinical abilities and also developing professional standards separately. However, when combining both the General Medicine and the Specialty Medicine mini-CEX data only a single factor explains variance and this suggests that the two-factor solution seen when looking at the General Medicine scores may be a false positive error. Of note significantly fewer assessments were undertaken by RMOs in General Medicine than in Specialty Medicine who are known to be less stringent markers and this may also have impacted on the assessment. As an assessor, it is important to recognise that only a single factor emerges and this highlights that it is challenging to discriminate between components of the student's performance and that it is likely that a "halo effect" occurs [Thorndike, 1920].

I was also interested in determining if there was any variation in grade achieved in the mini-CEX, based on demographic variables of the candidate, entry scheme into the medical school or clinical teaching site. The initial work describing the mini-CEX did not show any difference in mini-CEX scores between training programmes [Norcini, 1997]. However, in a

review article of assessment methods in medical education, it was highlighted that with the mini-CEX variance may occur due to differences in patients, assessors and sites [Norcini, 2007].

We see that at the University of Auckland more females are offered places in medical school than males and that females received an overall higher score than males for the mini-CEX. There was no difference in the mini-CEX scores awarded to undergraduates or graduates, nor did the scores differ between each of the years of this study.

Students who entered via the General Entry or Rural and Regional Admission Scheme received similar scores which were significantly higher than students who entered via the Māori and Pacific Admission Scheme or International students. Ideally, I would have been interested to link these to the UMAT scores or the grade point average (GPA) of the Year 4 students when they enter medical school. However, I did not have access to this information. The entry GPA reflects prior academic achievement. Higher GPAs have been shown to be associated with student retention in medical school, and their performance during medical school [Fergusson, 2002]. In New Zealand, the UMAT has also been shown to be predictive of medical student achievement, but less so than the grade point average [Poole, 2012]. The GPA that students achieve in Year 2 and 3 of medical school at the University of Auckland MBChB programme has been shown to be significantly higher for General admission students than RRAS students and both categories significantly higher than MAPAS students (average Year 2–3 GPA General scheme = 6.35 +/- 1.52; RRAS = 5.82 +/- 1.65, $p < 0.01$; MAPAS = 4.33 +/- 1.56, $p < .01$) [Curtis, 2017]. No information on International student GPA is reported in the literature. In this study no difference was seen in the Specialty Medicine mini-CEX scores for students entering by different schemes. For the General Medicine mini-CEX the scores achieved by MAPAS students were lower than students entering by the General scheme but no different to RRAS students. When combining the Specialty Medicine and General Medicine mini-CEX it was demonstrated that MAPAS and International students had lower scores than General or RRAS students. Given

the difference reported in Year 2-3 GPAs, which reflects the academic achievement immediately prior to Year 4, it is likely that the difference in scores is reflective of academic ability. It is important to note that dedicated academic and pastoral support is available for MAPAS students, including culturally safe learning spaces, specific online tutorials and provisions of targeted tutorials [Curtis, 2013]. However, when considering the impact of this in terms of grade, the effect of these statistically significant differences is minimal.

The paper from the University of Bern looked at the effect of size of the clinical training site on the mini-CEX scores and demonstrated that small sites tend to award lower scores than large training sites [Rogausch, 2015]. Depending on the number of students at each site, clinics were divided into small (< 15 students/year), medium (16-30 students/year) or large training sites (> 30 students/year). Using these somewhat arbitrary parameters for the site size, this translates to Rotorua being a small, Tauranga medium and Waitematā, Auckland, South Auckland and Waikato being large. Although possible differences in scores between clinical campuses were raised, the post-hoc analysis did not show a difference in scores between the sites.

To consider the potential variability between entry characteristics and the site the candidates undertake their training an omnibus test was undertaken. Reassuringly, there was no difference seen in scores awarded, with tested combinations.

There are, of course, limitations to this section of the study. As highlighted (in Chapter 2) the majority of medical schools in Australia and NZ use the mini-CEX, however, as this data is from a single centre, its conclusions can only be applied to the University of Auckland MBChB programme. Detailed data on the assessor, patient and complexity was extracted for only two years, and this data was incomplete, in part because some of the forms were unavailable, or incomplete. However, this provided a sample of over 300 cases which was determined to be large enough to provide informative results. If a fail or borderline pass is awarded for the mini-CEX then the student will always be offered an opportunity to resit

the assessment. I was unable to determine how many students had an earlier assessment to that which was recorded, which was a fail or borderline mark. The internal consistency of the different components of the mini-CEX was evaluated using principal components analysis. Initially, a generalisability analysis was also planned to evaluate the reliability of the mini-CEX. The grade awarded to students in one assessment is only weakly predictive of the grade that is awarded in the other assessment. Thus, it would have been useful to estimate variance that may be contributing when the mini-CEX is used. However, as there were a large number of assessors, many of whom would only have evaluated the student in only one case, a generalisability analysis was not possible. When considering whether the entry scheme of the student had an impact on the mini-CEX grades the scores achieved by the two UTAS students were merged with those of entering by the General scheme. This is a small number and is unlikely to have impacted the analysis, but it could be argued that their scores should have been excluded from the analysis.

To better understand the mini-CEX grade awarded it would be interesting to have the students' entry GPA and UMAT. A regression analysis could be undertaken to provide a clearer understanding of the impact of prior academic achievement on the mini-CEX score. Another area of future research could be to identify students that had a fail or a borderline pass in the mini-CEX and determine if this predicted their final Year 4 grade, not just the grade for the Clinical and Communications skills domain. Annually, only 3-5 students fail Year 4. It would be interesting to determine, if the mini-CEX grade is a predictor for poor overall performance and allow early intervention to support selected students. Of interest, is that the mini-CEX in the University of Auckland MBChB programme uses a four point ordinal scale. Rasch analysis [Rasch, 1960] of the mini-CEX could be undertaken to determine the extent to which the responses on the scale model match the descriptors of the scale and thus the construct being assessed. This information could be used to modify the scale and improve the construct [Medvedev, 2017].

In conclusion, this chapter informs how the mini-CEX is being used by Year 4 students at the University of Auckland, where it is used as a component of Medicine rotations. The mini-CEX is also a component of programmatic assessment and this discussed further in Chapter 5. We see some differences between patient and assessor characteristics between the General Medicine and Specialty Medicine mini-CEX, however these are minor. Increased training resources may help assessors in using the mini-CEX. Reassuringly, although we do see significant differences in the overall grades in the mini-CEX awarded, based candidates' factors, pragmatically these are not important due to an insignificant effect.

Chapter 5 - Contribution of the mini-CEX to the Clinical and Communication Skills domain.

5.1 - Introduction.

Historically, an end-of-year high stakes clinical examination was used, in the University of Auckland MBChB programme, to decide whether students could progress with training. This approach to learning has been viewed as having a number of disadvantages. A summative end-of-year examination may promote shallow or rote learning [Cilliers, 2012], discourage the learners from using feedback [Harrison, 2016], and lacks qualitative feedback which could guide the student in their future development [van der Veleuten, 2012]. By utilising a programmatic approach we have a number of datapoints that ideally promote assessment for learning [van der Veleuten, 2012]. These datapoints should be evaluated at multiple time points through the year and then will have the potential to allow the student to address areas of weakness and to develop and consolidate skills. In addition, if feedback is consistently provided with these assessments throughout the year, the final, cumulative result is not surprising to the student. A move to programmatic assessment, with multiple assessments though the year, was introduced for the University of Auckland MBChB course in 2013.

Clinical and Communication Skills are one of five key domains that have been identified as important for clinical practice [University of Auckland Phase 2 (Year 4) Guidebook]. The ability to communicate sensitively with patients and their whānau, to clearly elicit and provide information, and to examine patients and interpret information are all core skills for a doctor.

To evaluate the Clinical and Communication Skills of the Year 4 student, a total of 11 assessments are used and determine the grade for this domain: the mini-CEX assessed in the General Medicine rotation, the mini-CEX in the Specialty Medicine rotation, two OSCEs in the Musculoskeletal rotation, the communication component of the Drug and Alcohol assessment, three Medicine station CSAs and three Surgical station CSAs. For each of these

individual assessments the student receives a grade of distinction, pass, borderline performance or fail and students receive a final grade of distinction, pass or fail for Clinical and Communication Skills domain based on the combination of these assessments (Table 2).

The Drug and Alcohol assessment is a simulated patient assessment, with an actor taking the role of a patient. The Year 4 student demonstrates their ability to carry out a brief intervention of screening for drug and/or alcohol (ab)use. The interaction is videotaped, and the student undertakes self-reflection. All students undertake the assessment at either the Tamaki campus, in Auckland or at the Waikato clinical site and are marked by a single assessor (Personal Communications Karen Hicks).

The Musculoskeletal rotation incorporates both Orthopaedic and Rheumatological components. The students undertake the musculoskeletal assessment at a single site: Middlemore Hospital. They have four OSCE stations where they are assessed by two assessors at each station, i.e. eight assessors, over half-a-day. Two OSCE stations contribute to the mark for Musculoskeletal OSCE 1 and another two stations to the mark for Musculoskeletal OSCE 2. There is a pool of approximately 20 assessors, all SMOs, either Orthopaedic Surgeons or Rheumatologists, who will perform the assessment. Patients or actors may be used (Personal Communications Dr Nichola Wilson).

At the end of the Year 4 there is an integrated assessment of the clinical skills of the students. These are held in two sites, the Grafton or Tamaki campuses, using a mix of actors, patients and mannequins. Each student will undertake three Medical Clinical Skills Assessments (CSA) (one history taking and two clinical examination) and three Surgical CSAs (one history taking and two clinical examination) and are assessed by six different assessors, all SMOs, from medical and surgical specialties over half-a-day (Personal Communications Dr Kira Bacal).

As outlined in Chapter 4 the mini-CEX assessments are undertaken across many clinical sites, by many different assessors, using patients with a wide-range of clinical scenarios.

Important questions when using this system of programmatic assessment are: for the student is the assessment is reliable, and for staff involved in teaching, where there has been a change from a familiar system, is the combined assessment credible and dependable.

5.2 - Aim.

In line with the concept of programmatic assessment there is a range of assessments, occurring at different sites, throughout the year, with a range of examiners. The overall aim is to determine how this programme of assessment can be optimised. In this chapter I evaluated the contribution of each of these assessments to the final grade of the Communication and Clinical Skills grade in order to explore its reliability.

5.3 - Methods.

Deidentified data from Year 4 medical students enrolled between 2013 and 2017 in the University of Auckland MBChB programme was extracted from the Medical Programme Directorate and Department of Medicine databases.

Previously extracted demographic data and the overall mark for the General Medicine and Specialty Medicine mini-CEX were used. In addition scores for the Musculoskeletal OSCE 1 and 2, Clinical Skills Medical CSA 1, 2 and 3, Clinical Skills Surgical CSA 1, 2 and 3 and the Drug and Alcohol assessment were obtained as well as the overall grade for the Clinical and Communications Skills grade.

Statistical Analysis

Data were analysed using SPSS 25, FACTOR 10.8.04 and SPSS Analysis of Movement Structures (AMOS) 25.

Spearman's rank order correlation coefficient was evaluated to determine the relationship between the assessments and the final Clinical and Communication skills domain grade. Cronbach's alpha and item-to-total correlation was calculated to determine internal consistency.

The mini-CEX and in fact all assessments that contribute to the Clinical and Communications skills grade utilises an ordinal scale. Construct validity is analysed according to ordinal data obtained from the Likert scale. A Cronbach's alpha of > 0.70 was considered to be an acceptable reliability coefficient for determining the internal consistency of the scale (in this case the Clinical and Communication Skills domain grade) [Nunnally, 1978].

To aid in determining the reliability of the current model and how this could be optimized factor analysis was used. Exploratory factor analysis (EFA) was performed using FACTOR to explore the links between the observed variables (the clinical assessments or items) and latent variables (factors) and to identify the factor (programme) structure. Pearson correlation matrices may be utilised to perform factor analysis to evaluate the dimensionality of this data however, this assumes an interval measurement scale which is clearly not the case for our scale. Polychoric correlation estimates the correlation between raters if ratings are made on a continuous scale and was thus deemed more appropriate when evaluating the contribution of these assessments [Holgado-Tello, 2008]. Polychoric correlation was deemed to be appropriate to use if the univariate distribution of the ordinal items were asymmetric or had an excess of kurtosis. If both indices were less than one then Pearson correlation may be deemed more appropriate [Kaplan, 1985]. Eigenvalues of > 1.25 were preferred [Henson, 2006], with a cutoff of 1.0 accepted [Kaiser, 1960]. The impact of removal of the General Medicine and Specialty mini-CEX on internal consistency of the scale was also evaluated. The KMO measure of sampling adequacy and Bartlett's test of sphericity was determined to evaluate the suitability of the data with a KMO of greater than .70 and a Bartlett's test of sphericity less than .05 considered appropriate. Rotation was to

determine the correlation between observed variables and determine the strongest correlations that form factors.

Structural Equations Modeling (SEM) using AMOS 25 was then undertaken to test the relationships between structural paths and latent variables and to determine the universal fit related to the quality of model to either support or reject the structure. Possible models were tested and an ideal model developed. The hypothesised models are illustrated with observed items (the clinical assessments) represented as rectangles, the unobserved factor (the Clinical and Communications Skills grade) represented as ellipses, the arrows between the unobserved variables and the observed variable representing a regression path and the number representing the standardised regression weight. Chi-squared (χ^2) and its subsequent ratio with degrees of freedom (df) was evaluated to determine the fit between the hypothesised model and the observed variables with < 2 considered acceptable [Ullman, 2001]. A number of goodness-of fit indices were calculated including the goodness-of fit index (GFI), comparative fit index (CFI), root mean square error of approximation (RMSEA). A GFI greater than 0.90, a CFI of greater than 0.93 and a RMSEA less than .05 we considered to show acceptable fit to the data [Tabachnick, 2007]. As well as the CFI and RMSEA, Akaike's information criterion (AIC) was used for selecting an optimal model, when considering alternates, with the smallest AIC preferred [Tabachnick, 2007].

5.4 - Results.

5.4.1 - Grades assigned for each assessment.

The grades assigned to each assessment contributing to the Clinical and Communications Skills domain is summarised in Table 6. Only six students failed this domain over the five year period of the study.

Table 6 - Marks achieved by Year 4 medical students at the University of Auckland for assessments contributing to the Clinical and Communication Skills domain grade.

Assessment	Number of Students	Fail	Borderline performance	Pass	Distinction
General Medicine mini-CEX	1196	1	14	900	281
Specialty Medicine mini-CEX	1198	3	18	911	266
Musculoskeletal OSCE 1	1193	26	34	723	410
Musculoskeletal OSCE 2	1193	13	32	960	188
Surgical CSA 1	1188	6	79	837	266
Surgical CSA 2	1188	31	94	851	212
Surgical CSA 3	1188	43	128	824	193
Medical CSA 1	1188	15	84	855	234
Medical CSA 2	1188	30	134	824	200
Medical CSA 3	1188	18	112	841	217
Drug and Alcohol Assessment	1188	3	0	924	272
Overall Clinical & Communication Skills Grade	1201	6		988	135

5.4.2 – Correlation.

A Spearman’s rank order correlation was run to determine the relationship between the 11 clinical assessments and Clinical and Communication Skills domain grade. A statistically significant positive correlation was found for all assessments and the overall Clinical and Communication skills grade (Table 7). A weak correlation was found between the overall grade and the following assessments: General Medicine mini-CEX, Specialty Medicine mini-CEX, Musculoskeletal OSCE 1, Surgical CSA 1, Surgical CSA 2, Surgical CSA 3, Medical CSA 1, Medical CSA 2 and Medical CSA 3. A very weak correlation was found for both the Musculoskeletal OSCE 2 and the Drug and Alcohol assessment and the overall grade. So, no one assessment was strongly predictive of the overall grade that would be awarded to the Year 4 student for the Clinical and Communications Skills domain.

Table 7 - Correlation of clinical assessments contributing to the Clinical and Communication Skills domain grade.

	Clinical and Communications Skill Grade	General Medicine mini-CEX	Specialty Medicine mini-CEX	Musculoskeletal OSCE 1	Musculoskeletal OSCE 2	Surgical CSA 1	Surgical CSA 2	Surgical CSA 3	Medical CSA 1	Medical CSA 2	Medical CSA 3	Drug and Alcohol
Clinical and Communications Skill Grade	1.000											
General Medicine mini-CEX	.286**	1.000										
Specialty Medicine mini-CEX	.246**	.064*	1.000									
Musculoskeletal OSCE 1	.237**	.058*	.099**	1.000								
Musculoskeletal OSCE 2	.181**	0.045	.076**	.215**	1.000							
Surgical CSA 1	.340**	.127**	.110**	.140**	.136**	1.000						
Surgical CSA 2	.312**	.092**	.071*	.074*	.134**	.179**	1.000					
Surgical CSA 3	.350**	.067*	.099**	.122**	.126**	.212**	.160**	1.000				
Medical CSA 1	.251**	0.014	.113**	.111**	.072*	.194**	.116**	.203**	1.000			
Medical CSA 2	.286**	.098**	0.053	.130**	.121**	.224**	.179**	.229**	.124**	1.000		
Medical CSA 3	.317**	.138**	.110**	.130**	.118**	.145**	.144**	.226**	.204**	.134**	1.000	
Drug and Alcohol	.189**	.101**	.114**	.087**	.063*	.086**	.077**	.100**	.092**	.068*	.092**	1.000

N = 1192 to 1201

*Correlation is significant at the .05 level

**Correlation is significant at the .01 level

5.4.3 - Internal Consistency.

The internal consistency for the Clinical and Communications skills grade, estimated by Cronbach's alpha, was .64. Item (assessments) to total (clinical and communication skills) statistics are provided in Table 8. This suggests that the Clinical and Communication skills domain grade is not as reliable an assessment as desired, and that removal of a single clinical assessment will not improve or optimise the reliability of the overall Clinical and Communication skills domain grade.

Table 8 - Item-total statistics for the Year 4 assessments.

Assessment	Mean score	Standard deviation	95% Confidence Interval	Variance	Kurtosis	Corrected Item-total correlation	Cronbach's α if item deleted
General Medicine mini-CEX	3.22	.45	3.19 - 3.25	.21	2.18	.17	.63
Specialty Medicine mini-CEX	3.20	.46	3.16 - 3.23	.21	2.86	.17	.63
Musculoskeletal OSCE 1	3.27	.64	3.21 - 3.31	.42	3.42	.27	.62
Musculoskeletal OSCE 2	3.10	.49	3.06 - 3.14	.25	7.78	.28	.62
Surgical CSA 1	3.14	.56	3.09 - 3.18	.32	3.48	.39	.59
Surgical CSA 2	3.04	.62	2.99 - 3.08	.39	3.43	.30	.61
Surgical CSA 3	2.97	.66	2.92 - 3.02	.44	2.52	.38	.59
Medical CSA 1	3.09	.58	3.05 - 3.13	.34	3.82	.32	.61
Medical CSA 2	3.00	.64	2.95 - 3.04	.41	2.62	.34	.60
Medical CSA 3	3.05	.60	3.00 - 3.09	.37	3.16	.37	.60
Drug and Alcohol assessment	3.22	.43	3.19 - 3.25	.20	3.19	.16	.63

(N = 1189).

5.4.4 - Factor analysis.

The distributions of these (ordinal) items are asymmetric. Polychoric correlation was performed and a standard variance/covariance matrix generated (Table 9). KMO = .81 and significant Bartlett's test of sphericity (χ^2 (55) = 1113.1, $p < .01$) confirmed the adequacy of the correlation matrix.

Table 9 - Standardised variance/covariance matrix of Year 4 Clinical and Communication Skills assessments.

Assessment	General Medicine mini-CEX	Specialty Medicine Mini-CEX	Musculoskeletal OSCE 1	Musculoskeletal OSCE 2	Surgical CSA 1	Surgical CSA 2	Surgical CSA 3	Medical CSA 1	Medical CSA 2	Medical CSA 3
Specialty Medicine Mini-CEX	.13									
Musculoskeletal OSCE 1	.11	.16								
Musculoskeletal OSCE 2	.11	.14	.35							
Surgical CSA 1	.22	.18	.21	.24						
Surgical CSA 2	.15	.11	.13	.23	.27					
Surgical CSA 3	.12	.16	.18	.22	.33	.25				
Medical CSA 1	.04	.18	.17	.14	.31	.19	.31			
Medical CSA 2	.17	.09	.20	.21	.33	.36	.33	.20		
Medical CSA 3	.23	.18	.21	.21	.23	.23	.33	.31	.22	
Drug and Alcohol	.19	.20	.17	.14	.14	.12	.18	.15	.12	.17

N = 1192 to 1201

*Correlation is significant at the .05 level

**Correlation is significant at the .01 level

Exploratory factor analysis

An EFA was undertaken to determine the underlying structure of the 11 assessments contributing to the clinical and communications skills assessment. I hypothesised that a four-factor solution may occur with the two mini-CEX grouping together, the two OSCEs grouping together, the six CSAs grouping together and the Drug and Alcohol assessment grouping separately. Explained variances based on eigenvalues were generated with the advised number of dimensions being one, which explained only 27% of variance (Table 10). Accepting eigenvalues of > 1.0, three factors explained 47% of variance. This suggests that when considering the 11 clinical assessments they group into three clusters (or factors) that are being evaluated, rather than four as was hypothesized.

Table 10 - Explained variance based on eignenvales.

Variable	Eigenvalue	Proportion of Variance	Cumulative Proportion of Variance
General Medicine mini-CEX	3.02	.27	.27
Specialty Medicine mini-CEX	1.08	.10	.37
Musculoskeletal OSCE 1	1.01	.09	.47
Musculoskeletal OSCE 2	.98	.09	
Surgical CSA 1	.82	.07	
Surgical CSA 2	.08	.07	
Surgical CSA 3	.78	.07	
Medical CSA 1	.71	.06	
Medical CSA 2	.65	.06	
Medical CSA 3	.60	.05	
Drug and Alcohol	.54	.05	

In order to determine the strongest correlations of factors the latent factor items were rotated to determine possible clusters. Three clusters were generated with the General Medicine mini-CEX, Specialty Medicine mini-CEX and Drug and Alcohol assessment forming the first cluster, the Musculoskeletal OSCE 1 and Musculoskeletal OSCE 2 forming the second cluster and the Surgical and Medical CSAs forming the third cluster (Table 11). The main difference to the

hypothesised clusters was that the Drug and Alcohol assessment groups with the two mini-CEX assessments.

Table 11- Normalised varimax rotated loading matrix.

Variable	Cluster 1	Cluster 2	Cluster 3
General Medicine mini-CEX	.56	.19	.02
Specialty Medicine mini-CEX	.64	.08	.11
Drug and Alcohol	.70	.05	.13
Surgical CSA 1	.16	.63	.16
Surgical CSA 2	.01	.57	.14
Surgical CSA 3	.14	.67	.08
Medical CSA 1	.19	.60	-.04
Medical CSA 2	-.04	.63	.20
Medical CSA 3	.36	.51	.07
Musculoskeletal OSCE 1	.17	.12	.79
Musculoskeletal OSCE 2	.06	.22	.78

Removing the mini-CEX

To consider the role of the mini-CEX in contributing to the overall Clinical and Communication Skills domain grade for the General Medicine mini-CEX and the Specialty Medicine mini-CEX were removed and only nine assessments were evaluated using EFA. I again hypothesised that a four-factor solution may occur with the two Musculoskeletal OSCEs, the three surgical CSAs, the three Medical CSAs and the Drug and Alcohol assessment grouping separately. KMO = .80 and significant Bartlett's test of sphericity ($\chi^2 (36) = 963.7, p < .01$) confirmed the adequacy of the correlation matrix. Explained variance based on eigenvalues were generated which resulted in only a single significant eigenvalue (= 2.81) with all other values being less than 1.25 (range .59 to 1.03). The advised number of dimensions was one which explained 31.2% of variance. Two factors achieved significance accepting eigenvalues of > 1.0 with all other values being non-significant (range .58 to .95) and these two factors explained 42.6% of variance. The

reliability of the construct using nine assessments was .61 suggesting that removing the mini-CEX assessments made the overall Clinical and Communication Skills domain grade less reliable.

Structural Equations Modeling

The structural model consisting of a single construct was generated (Figure 5). However, analysis of this model using SEM indicated that it did not fit the data well. The χ^2 / df was unsatisfactory ($\chi^2 = 6007$, $df = 55$, $p < .01$). Other markers were all also unsatisfactory (CFI = .004); RMSEA = .912; AIC = 330506.013).

Other theoretical models were tested. Of note in the single construct model was the negative regression weighting for the Drug and Alcohol assessment ($r = -.84$) and the three factors with clusters generated in the rotated EFA.

A proposed model consisting of a single construct (not displayed), but removing the Drug and Alcohol component, indicated that this did not fit the data well ($\chi^2 = .626$, $df = 3$, $p = .89$; CFI = .998; RMSEA = .05; AIC = 753.138).

A valid model using a three factor model was achieved (Figure 6) with the statistically most satisfactory data modeling ($\chi^2 = .626$, $df = 3$, $p = .89$; CFI = 1.00; RMSEA < .01; AIC = 22.63).

This confirms that having the 11 clinical assessments we do not have a reliable assessment for the final grade. However, if the assessments are scaffolded together into three clusters; theoretically, we can generate a more reliable construct.

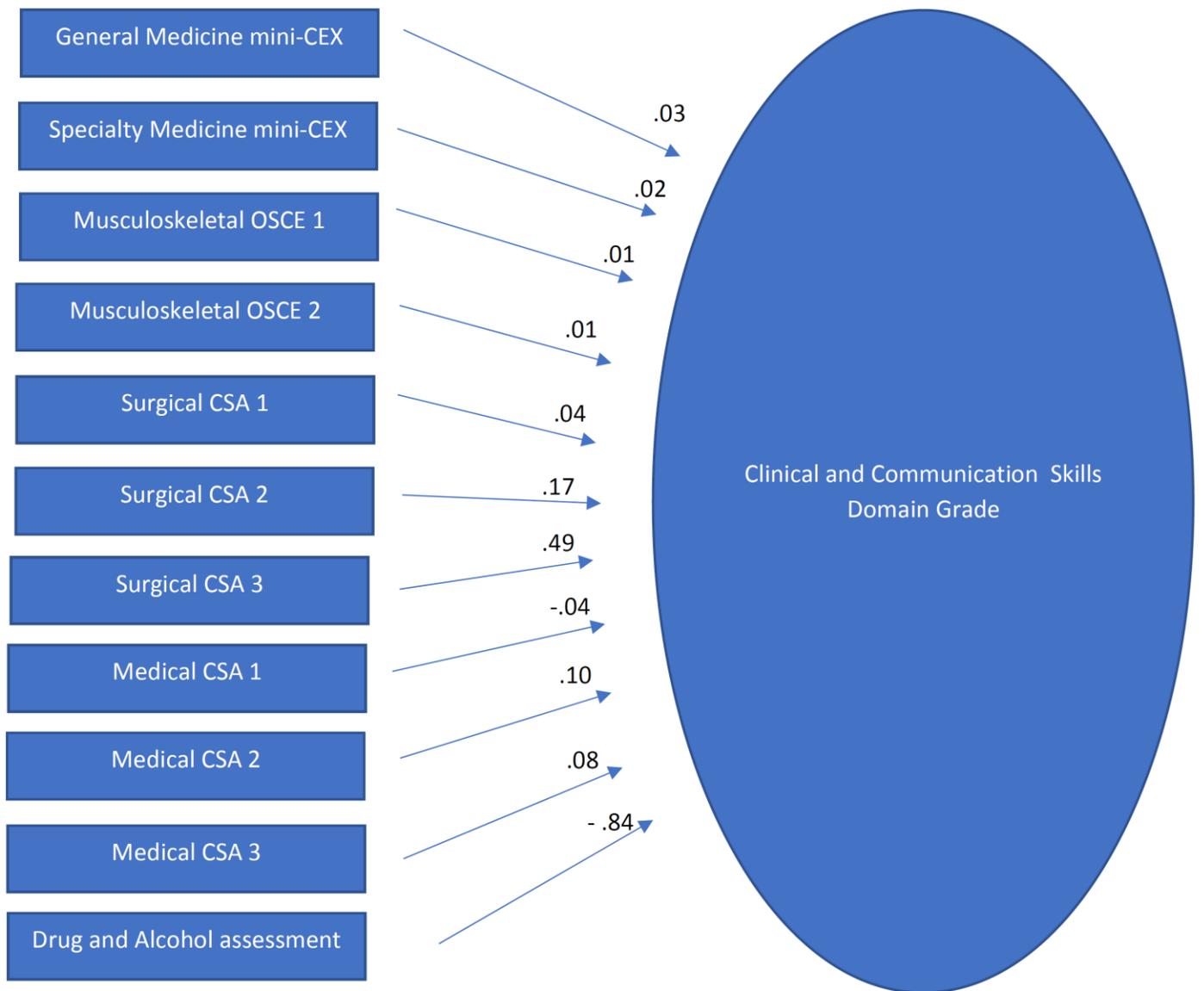


Figure 5 – Factorial Structure of the hypothesized 11 item model of the Clinical and Communication Skills domain grade.

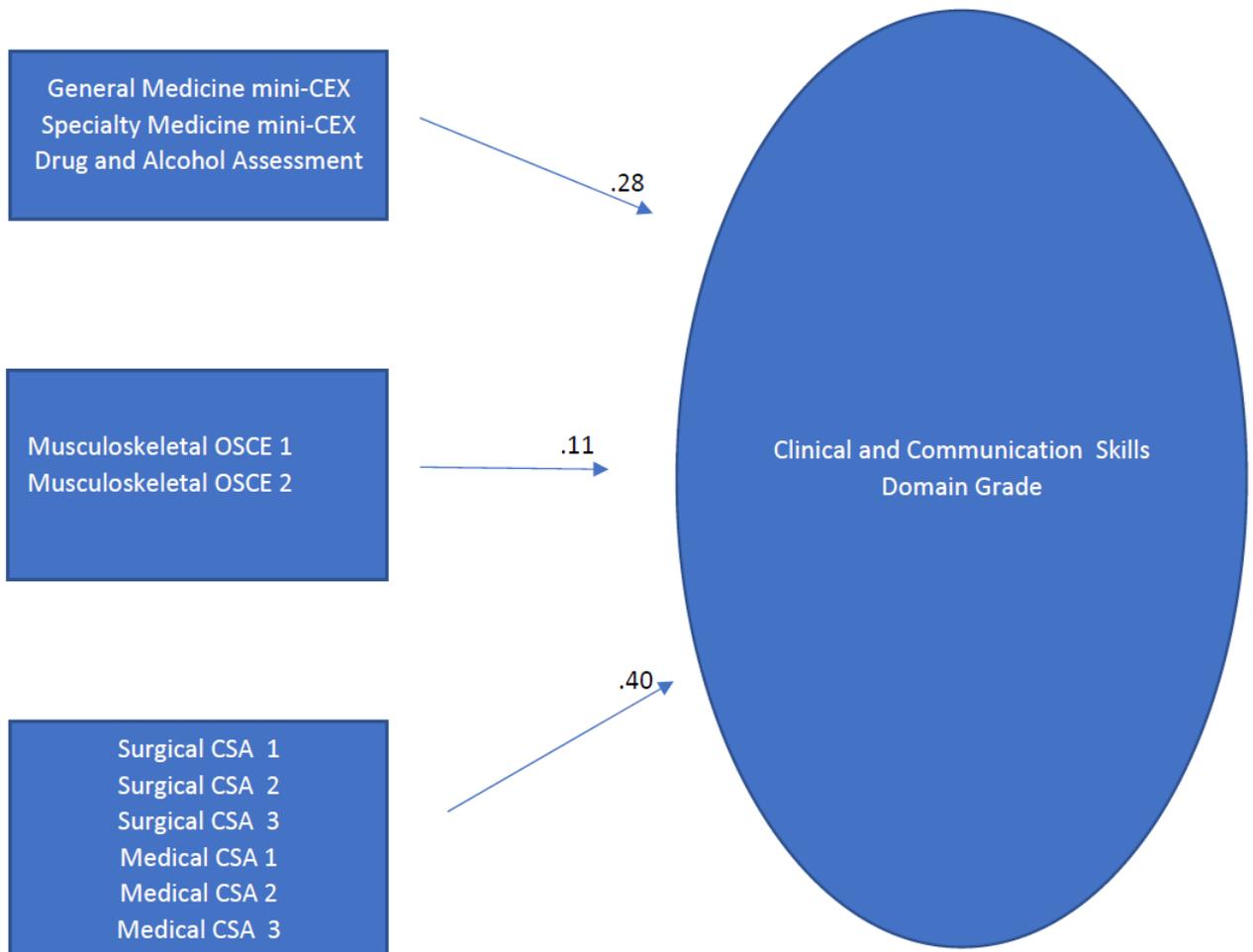


Figure 6 – Factorial Structure of the hypothesised 3 factor model of the Clinical and Communications Skills domain grade.

5.5 - Discussion.

In this study, using the current model of 11 assessments, which all contribute equally to the Clinical and Communication Skills domain final grade, it was determined that this is not a reliable construct (Cronbach's alpha = .64). The item-total correlation is useful in these circumstances and can be used to determine if the reliability can be improved. Of the 11 clinical assessments that contribute to the overall grade the General Medicine mini-CEX, the Specialty Medicine mini-CEX and the Drug and Alcohol assessment correlated the least. However, when removing any of these three assessments, and in fact removing any of the clinical assessments, the reliability of the Clinical and Communication Skills domain grade did not improve. To theoretically determine how the reliability could be improved factor analysis was undertaken. EFA determined that when considering the 11 clinical assessments they grouped into three factors. SEM showed that a theoretically model which was reliable could be generated where the General Medicine mini-CEX, Specialty Medicine mini-CEX and Drug and Alcohol assessment were grouped into one set of assessments, the Musculoskeletal OSCE 1 and the Musculoskeletal OSCE 2 were grouped into a second set of assessments and the three Surgical CSAs and the three Medical CSAs were grouped into a third set of assessments which then contribute to the overall Clinical and Communication Skills final grade.

Spearman's rank correlation may be used to determine the relationship between two variables or two sets of data [Spearman, 1904]. In this study it is used to evaluate the strength and direction of the eleven clinical assessments and the final grade for the Clinical and Communications skills domain. I found that a significant positive monotonic correlation occurred for all eleven assessments (Table 7, $r_s = .189$ to $.350$, $N = 1192$ to 1201 , $p < .05$). This suggests that for each clinical assessment, as the grade for each candidate improved, so did the final overall grade that was achieved. A correlation of less than .20 is considered very weak, .20 to .39 weak, .40 to .59 moderate, .60 to .79 strong and greater or equal to .80 is a very strong [Evan, 1996]. The strength of the relationship for most of the assessment tools was weak or in the case of the Musculoskeletal OSCE 2 and Drug and Alcohol assessment very weak. Thus, although each individual assessment is correlated with final Clinical and Communication Skills

domain grade, it may be concluded that no single assessment alone can predict the overall ability of the candidate. Thus, multiple assessments are required to evaluate the candidates' ability.

Cronbach's alpha is classically used to determine the internal consistency of an assessment [Cronbach, 1951]. The theoretical value of alpha can range from 0 to 1, with the higher the score the more reliable the generated scale. A perfectly reliable test is typically not achieved; a Cronbach's alpha of greater than .70 is considered an acceptable reliability coefficient [Nunnally, 1978]. However, even higher acceptable values up to .95 have been proposed [Bland, 1997; DeVellis, 2003] and for a high stakes assessment, a higher Cronbach's alpha to support reliability may be preferred.

An item-total correlation test may also be used to determine whether any item (in this case a clinical assessment), is inconsistent with the measured behaviours of the other items and could potentially be discarded to improve reliability [Field, 2013]. A corrected item-total correlation of less than .20 suggests that the item does not correlate well with the overall score and may need to be discarded.

In this study, the Clinical and Communication Skills domain grade internal consistency was .64, suggesting that the current model of 11 assessments is not a reliable construct. In other words, this suggests that if the same clinical assessments were repeated by a student that the final grade assigned to the student may not be the same, or that if two students of the same clinical ability undertook the 11 assessments that they may not be awarded the same final grade. A minimum Cronbach's alpha of 0.7 is desirable. The ability to assess and communicate with patient and whanāu is a core skill and it could be argued that an even greater cutoff, for reliability, would be preferred.

The item-total correlation is useful in considering the overall Clinical and Communications Skills domain grade and can be used to determine if the reliability may be improved. The corrected

item-total correlation was positive for all the assessments suggesting that each item correlated with the final grade. However, a small corrected item-total correlation suggests that the item may not be measuring the same construct as measured by other items, with a value of less than .20 considered significant [Kline, 1999]. The corrected item-total for the General Medicine mini-CEX, Specialty Medicine mini-CEX and the Drug and Alcohol assessment, were each less than .20, suggesting that these assessments do not correlate well with the overall grade. This provides a signal, as to which assessments it may be useful to remove, to improve the overall reliability. If the General Medicine mini-CEX, the Specialty Medicine mini-CEX or the Drug and Alcohol assessment were removed I found that there was little impact on the overall reliability of the Clinical and Communications Skills domain grade; Cronbach's alpha reduced to .63. In fact, when removing any single clinical assessment, the internal consistency did not improve (the Cronbach's alpha range was .59 to .63 when any single item was deleted). Thus, removing one of the assessments does not significantly improve the reliability of the overall grade. The follow on from this is to consider how the reliability of the final Clinical and Communications Skills domain grade may be achieved.

To better understand the relationship between the 11 clinical assessments (the observed variables) and the Clinical and Communications Skills domain grade (the latent variable) factor analysis was undertaken. Factor analysis is a statistical procedure that allows the relationship between observed and latent variable to be evaluated. First, EFA is undertaken to determine the connections between the observed variables and the latent construct [Yong, 2013]. From this a number of factors can be generated and this is termed the factor structure. Factors are rotated to identify which of the observed variables cluster together and allow an optimal simple structure to be developed. SEM can determine the relationship between the observed and latent variables whilst accounting for measurement error [Beran, 2010]. Using the information from EFA the number of factors can be confirmed and different models may be tested using SEM.

In this study EFA was undertaken and only a single factor was retained using an eigenvalue of greater than 1.25 which is the preferred cutoff [Henson, 2006]. However, it was not possible to generate a construct with acceptable fit indices using a single dimension. Of the 11 assessments, the Drug and Alcohol component had the greatest negative correlation ($r = -.84$), with the Medical CSA 1 the only other assessment with a small negative correlation ($r = -.04$). Of note, the Drug and Alcohol assessment also had the smallest corrected-item correlation ($r_s = .16$). A theoretical model which removed the Drug and Alcohol assessment was generated and again, was unable to achieve satisfactory goodness-of-fit indices. This supports the finding that the reliability could not be achieved when considering a Cronbach's alpha of greater than .70 for all eleven assessments or with ten clinical assessments.

Using a more permissible eigenvalue, of greater than 1 as the cutoff [Kaiser, 1960], three factors were generated. However, although this tells us there are three factors, how the clinical assessments group into these three factors is ambiguous. To determine the loading of each variable to each factor, rotation of the axes of the factor was undertaken. Using rotation, it was found, that the General Medicine mini-CEX, Specialty Medicine mini-CEX and Drug and Alcohol assessment form the first cluster; the Musculoskeletal OSCE 1 and the Musculoskeletal OSCE 2 form the second cluster and the three Surgical CSAs and the three Medical CSAs form the third cluster (Table 10). There are several possible explanations for these clusters. It is possible that different aspects of the students' (clinical and communication skills) ability are being evaluated in each cluster. Each cluster is occurring at a similar time so perhaps these are assessing the students' ability at that time point. Finally, it is possible, that these reflect the assessment method being used.

This information was used to generate theoretically reliable models. Utilising three factors, and grouping the assessments, a valid model was generated that retains all 11 clinical assessments currently used. With this model (Figure 6) the General Medicine mini-CEX, Specialty Medicine mini-CEX and Drug and Alcohol assessment group together, the Musculoskeletal OSCE 1 and 2 group together, and the six CSAs group together. With this model a new rubric would need to

be developed for the Clinical and Communication Skills domain grade rather than the current rubric (Table 1). In this the student would have three parts to the Clinical and Communications Skills domain: Component 1 (comprising of the General Medicine mini-CEX, Specialty Medicine mini-CEX, and Drug and Alcohol assessment), Component 2 (comprising of the Musculoskeletal OSCE 1 and Musculoskeletal OSCE 2) and Component 3 (comprising of Medical CSA 1, Medical CSA 2, Medical CSA 3, Surgical CSA 1, Surgical CSA 2 and Surgical CSA 3) and would need to pass each *component*, but not every assessment. A possible rubric of how this could be used is described (Figure 7).

Clinical and Communications Skills Domain.

This consists of 11 assessments through the year which are grouped into three components.

Clinical and Communication Skills Component 1.

This consists of the General Medicine mini-CEX 1, Specialty Medicine mini-CEX 2, and Drug and Alcohol assessment. The grade assigned will be:

Grade	Requirement
Distinction	Distinction in three assessments of clinical skills.
Pass	Not meeting the criteria for fail or distinction.
Fail	Fail is assigned if any one of the following occurs: <ul style="list-style-type: none"> One fail in assessments of clinical skills or Two or more borderline performances assessments of clinical skills.

Clinical and Communication Skills Component 2.

This consists of the Musculoskeletal OSCE 1 and Musculoskeletal OSCE 2. The grade assigned will be:

Grade	Requirement
Distinction	Distinction in two assessments of clinical skills.
Pass	Not meeting the criteria for fail or distinction.
Fail	Fail is assigned if any one of the following occurs: <ul style="list-style-type: none"> One fail in assessments of clinical skills or Two or more borderline performances assessments of clinical skills.

Clinical and Communication Skills Component 3.

This consists of the Medical CSA 1, Medical CSA 2 Medical CSA 3, Surgical CSA 1, Surgical CSA 2 and Surgical CSA 3. The grade assigned will be:

Grade	Requirement
Distinction	Distinction in four assessments of clinical skills.
Pass	Not meeting the criteria for fail or distinction.
Fail	Fail is assigned if any one of the following occurs: <ul style="list-style-type: none"> Two or more fails in assessments of clinical skills One fail in assessments and two borderline performance Three or more borderline performances in assessments of clinical skills.

Overall Clinical and Communication Skills grade.

The final Clinical and Communication Skills domain grade will be based on marks from components 1, 2 and 3. The final grade assigned will be:

Grade	Requirement
Distinction	Distinction in all three components.
Pass	Not meeting the criteria for fail or distinction.
Fail	Fail in one or more component.

Figure 7 - Possible rubric for marking of the hypothesised model for the Clinical and Communication Skills domain.

However, a problem with the hypothetical system, is that there are only three assessments in Component 1 and two assessments in Component 2. With this proposed system, if the student fails one assessment in either Component 1 or Component 2, then the student fails that section and in turn will fail the overall Clinical and Communication Skills domain. With the proposed systems the mini-CEX, OSCE and Drug and Alcohol assessments have now moved from “low stakes”: where even if the student fails the assessment can guide development and the student can pass the year, to “high stakes”: if the student fails they fail the year.

A possible solution would be to add a mini-CEX assessment; so that there are three: one done in the General Medicine rotation and one done in each Specialty Medicine rotation and unbundle the Musculoskeletal evaluation so we have four OSCE and each station is marked separately. Thus, there would now be 14 clinical assessments which contributed to the overall Clinical and Communication Skills domain grade and the rubric for marking would change (Figure 8).

It should be noted that if the student has a fail or has a borderline pass awarded, in any of the mini-CEX assessments, then a repeat evaluation during the rotation could still be offered, with the better assessment submitted. This structure would have a minimal impact on workload. Currently, there are two Specialty Medicine rotations and a mini-CEX assessment can be undertaken in both rotations. The Year 4 programme would still have a number of low stakes assessments which could provide feedback and guide the student’s development. An early signal as to the student performance will be available and would allow the opportunity to provide additional support for a candidate who is struggling. In turn, the student would have a reasonable idea of their ability and progress through the year. This of course, would need to be considered and implemented by the Assessment subcommittee and the Board of Studies at the University of Auckland.

Clinical and Communications Skills Domain.

This consists of 14 assessments: General Medicine mini-CEX 1, Specialty Medicine 1 mini-CEX 2, Specialty Medicine 2 mini-CEX, Drug and Alcohol assessment, Musculoskeletal OSCE 1, Musculoskeletal OSCE 2, Musculoskeletal OSCE 3 and Musculoskeletal OSCE 3, Medical CSA 1, Medical CSA 2 Medical CSA 3, Surgical CSA 1, Surgical CSA 2 and Surgical CSA 3. These are grouped into three components.

Clinical and Communication Skills Component 1.

This consists of the General Medicine mini-CEX 1, Specialty Medicine 1 mini-CEX 2, Specialty Medicine 2 mini-CEX and Drug and Alcohol assessment. The grade assigned will be:

Grade	Requirement
Distinction	Distinction in four assessments of clinical skills.
Pass	Not meeting the criteria for fail or distinction.
Fail	Fail is assigned if any one of the following occurs: <ul style="list-style-type: none"> • Two fails in assessments of clinical skills or • One fail and one borderline performance in assessments of clinical skills or • Three or more borderline performances assessments of clinical skills.

Clinical and Communication Skills Component 2.

This consists of the Musculoskeletal OSCE 1, Musculoskeletal OSCE 2, Musculoskeletal OSCE 3 and Musculoskeletal OSCE 3. The grade assigned will be:

Grade	Requirement
Distinction	Distinction in four assessments of clinical skills.
Pass	Not meeting the criteria for fail or distinction.
Fail	Fail is assigned if any one of the following occurs: <ul style="list-style-type: none"> • Two fails in assessments of clinical skills or • One fail and one borderline performance in assessments of clinical skills or • Three or more borderline performances assessments of clinical skills.

Clinical and Communication Skills Component 3.

This consists of the Medical CSA 1, Medical CSA 2 Medical CSA 3, Surgical CSA 1, Surgical CSA 2 and Surgical CSA 3. The grade assigned will be:

Grade	Requirement
Distinction	Distinction in four assessments of clinical skills.
Pass	Not meeting the criteria for fail or distinction.
Fail	Fail is assigned if any one of the following occurs: <ul style="list-style-type: none"> • Two or more fails in assessments of clinical skills or • One fail in assessments and two borderline performance or • Three or more borderline performances in assessments of clinical skills.

Overall Clinical and Communication Skills grade.

The final Clinical and Communication Skills domain grade will be based on marks from components 1, 2 and 3. The final grade assigned will be:

Grade	Requirement
Distinction	Distinction in all three components.
Pass	Not meeting the criteria for fail or distinction.
Fail	Fail is assigned if any one of the following occurs: <ul style="list-style-type: none"> • Fail in one or more Component or • Fail in three clinical assessments of clinical skills or • Fail in two clinical assessments and two or more borderline performances or • Fail in one clinical assessments and four or more borderline performances or • Six or more borderline performances in clinical assessments.

Figure 8 - Rubric for marking of a possible revised model of the Clinical and Communication Skills domain grade.

A number of limitations for this study exist. Cronbach’s alpha is the measure used to determine the reliability of the overall Clinical and Communications Skills domain grade. Cronbach’s alpha is the most commonly used method of determining reliability, but like any test is subject to limitations. Other methods of assessing reliability exist, most notably G-theory, and were not evaluated. Factor analysis, of the Clinical and Communications Skills domain, did not support

the current structure. The study uses a Cronbach's alpha of 0.7 as the cutoff for a reliable test. It has been argued that any assessment method, whether structured (like the OSCE) or unstructured (like the mini-CEX) has utility, and that increasing sampling is the key for reliability, with between 7 and 11 assessments and several hours of testing time required [van der Vleuten, 2005]. The current format fulfills these recommended parameters and in these circumstances a Cronbach's alpha 0.64 may be viewed as quite acceptable. It has been argued that when considering an assessment, this should not just be viewed from classical criteria such as reliability and validity, but also should consider concepts such as educational impact, acceptability of the method and resource cost [van der Vleuten, 2005]. This study does not evaluate validity, acceptability, the cost of assessment or the impact on learning. Evaluating these concepts, in the University of Auckland MBChB programme are potential areas for future research. When developing a valid theoretical model I was working with the concept of maintaining the "status quo" i.e. to only use assessments that are currently available. For instance, the item-total correlation for the Drug and Alcohol assessment is low and when using EFA this has a negative regression weighting. So the options considered were to remove or scaffold this assessment with others. What is not considered is whether an alternate assessment could be used to evaluate the Clinical and Communications Skills of the student, and whether that could improve the reliability of the overall grade.

In conclusion, this chapter determined that there is a positive correlation between all 11 clinical assessments and the final grade assigned to the Clinical and Communications Skill domain, however, the current rubric of each of these 11 assessments contributing equally to the final grade is not as reliable as may be desired using Cronbach's alpha. Factor analysis determined that most, but not all assessments, contributed to the Clinical and Communications Skill domain. The Drug and Alcohol assessment has the weakest correlation using Spearman's rank test and also factor analysis demonstrated that this assessment showed negative regression. However, removing this or any other assessment, did not improve reliability. Using factor analysis, a theoretical construct, which groups the 11 assessments into three components, was been developed which could provide a reliable overall assessment. However, this model would mean that if the student failed one assessment then they would fail the year. A further revised

model is proposed that would involve more resource cost but provides the benefits of programmatic assessment and should be considered by the Assessment committee of the University of Auckland MBChB programme.

Chapter 6 - Discussion.

6.1 - Introduction.

The mini-CEX is a workplace based assessment tool, developed over 20 years ago to assess Internal Medicine residents [Norcini, 1995]. The utility of this instrument has been recognized and the tool is used widely to evaluate postgraduate trainees. Currently, in NZ, it is used in the assessment of Anaesthesia, Emergency Medicine, General Practice, Internal Medicine, Orthopaedic, Paediatric, Psychiatry, and Surgical trainees. It is an assessment at the peak of Miller's Pyramid of Competence, evaluating how trainees interact with patients, and promotes direct and immediate feedback [Miller, 1990]. The mini-CEX was first described as being used in undergraduate student Internal Medicine rotation core clerkship [Hauer, 2000] and has also been adopted for undergraduate assessment. In the USA it is reported to be used in 22% of Medicine rotations [Hawthorne, 2014]. The uptake in Australasia is greater; in 2018 both NZ medical schools and 17 of 20 Australian medical schools were using the mini-CEX.

In 2013, the University of Auckland MBChB programme was refreshed, and the use of the mini-CEX was expanded. It is now used in years 4, 5 and 6 of the programme in multiple specialties. For Year 4 medical students two mini-CEX assessments are undertaken, in clinical rotations in General Medicine (1 assessment) and Specialty Medicine (1 assessment). These same two assessments, along with nine other assessments (2 OSCEs, 1 Drug and Alcohol assessment and 6 CSAs), contribute to the year-long, longitudinal assessment of the students' Clinical and Communication Skills. The use of the mini-CEX as a component of programmatic assessment, in an undergraduate programme is novel and not described in the literature.

This chapter of the thesis will summarise the key findings of this research into the use of mini-CEX in Year 4 medical students in the University of Auckland MBChB programme both during Medicine rotations and as component of Clinical and Communications Skills domain placing this work in the context of the wider literature. It will describe the limitations of the study, touch on options for future research and provide recommendations for the ongoing use of the assessment tool within the MBChB programme.

6.2 - Use in the General Medicine and Specialty Medicine Rotations.

As summarised in the narrative review of the literature, eight to eleven assessments with the mini-CEX are needed for reliability to be achieved. The mark awarded to a student reflects not only candidate factors such as their ability, and their aptitude for the attachment, but also external factors such as the case difficulty, and the examiners' characteristic. By increasing the number of cases the student evaluates, it is likely that the variance that occurs from each individual case and a specific examiner is nullified and the combined score of the assessments are more likely to reflect the ability of the candidate. In the University of Auckland programme the mini-CEX is used twice in Year 4, with one assessment occurring during the General Medicine rotation and one during a Specialty Medicine rotation.

In this thesis it is demonstrated that the majority of students complete and pass two mini-CEX assessments. In our cohort, patients with wide variety of clinical problems are evaluated, more male than female patients are seen and those patients evaluated in General Medicine patients are a significantly older than those seen in Specialty Medicine. Although it was found that that a significantly higher mini-CEX score is awarded to candidates in the General Medicine attachment than Specialty Medicine (3.22 versus 3.20, $p < .01$), reassuringly, the effect of this is trivial. Overall, most students received a pass for the rotation and effectively there is no difference in the mark based on the attachment.

There is no difference in case complexity between General Medicine and Specialty Medicine patients, with about one-fifth of all patients (92/446), being highly complex. Studies in undergraduate students [Hill, 2009, Playford, 2013] have demonstrated that more complex cases are associated with higher scores. However, in this study a difference in grades awarded was not seen based on the difficulty of the case. A number of possibilities could explain this observation. However, the most likely reason is that the University of Auckland mini-CEX use a 4-point scale, and that the compressed scale is minimizing the potential for assessors to compensate for a difficult case when a grade for the mini-CEX.

The characteristics of the assessor are noted to also play a role in the grade awarded. Similar to other undergraduate studies [Kogan, 2003; Hill, 2009; Playford, 2013; Rogausch, 2015] this thesis demonstrated that more senior assessors awarded significantly lower scores than junior assessors (3.20 versus 3.45, $p < .01$). However, the effect size for this, in our cohort, was found to be negligible. Thus, whether a student was assessed by an SMO or RMO neither advantaged nor disadvantaged the student. At the Waitematā clinical campus, there has been a move to have all Year 4 mini-CEX done by SMOs. This work demonstrates that, from the perspective of grade awarded, this is not necessary. Of note, many assessors use the mini-CEX infrequently; over 80% (204/254) of assessors performed two or less assessments annually. Of some concern is how the assessors may be using the assessment tool. We see a huge variation in the time spent assessing the student suggesting there may be confusion in how the assessment should be undertaken. Anecdotally, students have reported to me quite significant variations in the execution of the mini-CEX. For instance they have been asked to evaluate the patient without being observed and then present and discuss the case and others have been asked to take a full history and complete a systematic examination in several organ systems i.e. to undertaking case-based discussion or a long case respectively. Although a clear explanation of how to undertake the mini-CEX is provided with the evaluation form (Appendix 1) this may not be adequate. Scripted, videotaped performances have been used as a training tool for assessors of students undertaking the mini-CEX in Year 6 primary practice [Eggleton, 2016] in the University of Auckland MBChB programme and a similar format could also be developed as an optional training resource for Year 4 Medicine assessors. Limiting the assessors to those with a faculty commitment is another potential option, however, this would increase the strain on stretched staff. Mandating that all assessor underwent a dedicated workshop or training session is also a possible option, however, this may discourage people from becoming assessors.

One study has highlighted that the size of the institution which trains the students may also contribute to variation in scores, with larger institutions noted to award higher grades than small institutions [Rogausch, 2015]. There has been a steady and deliberate expansion in the number of medical students trained at the University of Auckland, to manage the increased

need for doctors in NZ [Poole, 2019]. This is not unique to NZ, and this has led to new clinical sites being used to teach medical students [Perkins, 2013]. In Year 4 of the University of Auckland MBChB programme, students spend their clinical year at one of six different sites, with Rotorua and Tauranga, considerably smaller than Waitematā, Auckland, South Auckland and Waikato. It was found that for the Specialty Medicine mini-CEX there was a significantly lower score awarded to students who undertook their training in Tauranga (3.07) compared to Waitemata (3.25, $p < .01$) and a significantly lower score than South Auckland (3.24, $p < .01$). Of note, the effect of this was negligible. In addition this finding was not seen between Tauranga and other sites or Rotorua and other sites. This finding was not replicated when evaluating the General Medicine mini-CEX alone, nor was it seen when the General Medicine and Specialty Medicine mini-CEX were combined. Thus, we can be reassured that in our institution, the clinical training site is not influencing grades that students are awarded.

The student can enter medical school via several pathways. The NZ population has 15% Māori, indigenous to NZ, and 7.4% Pacific Peoples. However, these groups are significantly underrepresented in the healthcare workforce with only 2.6% Māori and 1.6% Pacific Peoples being doctors [Curtis, 2013]. MAPAS is designed to address the lack of Māori and Pacific healthcare professionals. This thesis demonstrates that MAPAS students have significantly lower mini-CEX scores than General and RRAS students. Although this effect is trivial, and does not impact on the overall grade awarded to MAPAS students, it does reinforce the need for space for Māori and Pacific students to learn and connect, and the need for the dedicated academic support provided by the MAPAS scheme [Curtis, 2013].

Reassuringly, when considering significant variables such as the seniority of the assessor, gender, graduate status, cohort, entry scheme and clinical site in combination, we do not see a difference in the final grade awarded for the mini-CEX.

6.3 – Use as a component of Programmatic Assessment.

The decision to reinvigorate of the MBChB programme in 2013 was driven by a paradigm shift towards assessment *for* learning, rather than assessment *of* learning [Schuwirth, 2011]. When considering the Clinical and Communications Skills domain, from this perspective it is seen that four assessment tools are used to evaluate this aspect of the students' ability: the mini-CEXs, OSCEs (Musculoskeletal), a simulated patient evaluation (Drug and Alcohol assessment), and CSAs. There are five assessments that occur throughout the year, with six that occur as part of an end-of-year assessment. These tools should guide learning and provide the student with direct and specific feedback [van der Vleuten, 2012]. As highlighted in the narrative review, a key benefit of the mini-CEX, valued by both assessors and students, is immediate and direct feedback [Dewi, 2010; Pernar, 2011; Suhoyo, 2014, Sudarso, 2016, Yanting, 2016]. The Drug and Alcohol assessment also provides the student with feedback; the student critiques their own performance, and receives feedback from an assessor. However, for the other assessments there is less feedback to guide the students' development. Only students who fail or receive a borderline mark in the Musculoskeletal OSCE are provided with feedback by the course coordinator, and this typically occurs several days after the assessment is completed (Personal Communications, Dr Nichola Wilson). Personalized feedback following a summative OSCE has been described [Harrison, 2015] and could be considered, although this many prove challenging [Harrison, 2017]. If a student fails a single CSA station but passes all other assessment for the year they are not provided with feedback for several weeks (Personal Communications Dr Kira Bacal). Thus, we see that the mini-CEX is a particularly valuable assessment tool, in supporting assessment for learning and the design of a programme of assessment.

The key finding of this thesis was that the Clinical and Communications Skills domain, with the 11 clinical assessments, not as reliable a construct as desired (Cronbach's alpha = .64), if the commonly accepted standard for moderate reliability, a cutoff of .7 for Cronbach's alpha, is used [Nunnally, 1978]. On one hand, an even higher Cronbach's alpha has been proposed for high-stakes exams such as selection for admission to college or professional licensing [Sackett,

2001]. On the other hand it is important to note that reliability, in this context, should be interpreted cautiously. For instance it has been highlighted that the communication skills in IMGs are context specific and so achieving a “satisfactory reliability” may be less important [Baig, 2009]. In fact, it has been argued that if an adequate sample has been obtained, typically by performing 7-11 assessments, using both naturalistic and structured assessments, then an appropriate evaluation has occurred and a consideration of reliability is not mandatory [van der Leuten, 2005].

It was demonstrated that all 11 clinical assessments had a positive correlation with the overall Clinical and Communications Skills domain, albeit these correlations were weak or very weak. A simple solution to improve reliability would be to remove a single assessment. However, using an item-total correlation test, it was demonstrated removing any single assessment made the reliability worse.

Using factor analysis, the current construct of 11 clinical assessments equally contributing to the Clinical and Communications Skills domain, could not be supported using standard goodness-of-fit indices. EFA was used to determine that the 11 clinical assessments grouped into three clusters and SEM was used to generate a theoretical construct which grouped the eleven assessments into three components: Component 1 (the General Medicine mini-CEX, Specialty Medicine mini-CEX, and Drug and Alcohol assessment), Component 2 (the Musculoskeletal OSCE 1 and 2) and Component 3 (the six CSAs). However, this construct means that the student could fail the year if they fail one clinical assessment which defeats a key feature of the design of the programme of assessment. Thus, an alternate system was proposed, which involved an addition mini-CEX being performed and revision to the rubric for marking of the for the Clinical and Communications Skills domain.

6.4 - Limitations.

A number of significant limitations exist for this thesis. Individual mini-CEX forms for 2013 and 2014 were reviewed to determine characteristics of assessors, which patients were being seen

and how much time was spent with the student. There is a reasonably stable senior medical workforce in the clinical teaching sites. However, it is possible that how the mini-CEX is being used has changed. I was unable to determine how many students had an initial assessment, which was a fail or borderline mark, and were offered a repeat assessment. Only the final grade was available for use.

When consider the reliability of the overall Clinical and Communications Skills domain, only Cronbach's alpha, the most commonly used measure of reliability, was used. Initially, G-theory was considered, as another method to assess reliability, and would also have evaluated the variance that was contributing to the reliability of the mini-CEX. However, as there were a large number of assessors, many of whom would only have evaluated the student in only one case, this was not possible.

Other important measures at the value of an assessment tool [van der Vleuten, 1996] such as its utility, validity, education impact, cost, and acceptability were not evaluated.

6.5 - Future Research.

To better understand the mini-CEX grade awarded it would be interesting to: 1.) have the students' entry GPA, UMAT and Year 3 GPA and undertake a regression analysis to provide a clearer understanding of the impact of prior academic achievement on the mini-CEX score; and 2.) identify students that had a fail or a borderline pass in the mini-CEX and determine if this predicted their final Year 4 grade, not just the grade for the Clinical and Communications Skills domain and 3.) track changes in the mini-CEX that occur in Years 5 and 6 and determine their impact on the Clinical and Communication Skills in subsequent year.

A qualitative study of assessors' understanding of the mini-CEX, and how it is executed, would be of interest to understand the large variation in time spent with the student. If a video teaching resources of the mini-CEX was developed then inter-rater reliability could be explored.

Qualitative research into how feedback to the student could be enhanced, in OSCEs, could be undertaken.

6.6 - Conclusion.

This thesis demonstrates that the mini-CEX is being effectively used in Year 4 of the University of Auckland MBChB programme. In both the General Medicine and Specialty Medicine rotations, the majority of students achieve a pass or distinction, and a wide range of clinical problems are evaluated. There are significant differences in the grade awarded between the General Medicine and Specialty Medicine mini-CEX and significant differences based on the seniority of the assessor, and entry characteristics of the student. However, the effect size of these differences is minimal. Reassuringly, when combining potential sources of variance, no difference in mini-CEX grade is seen attributable to assessor, gender, graduate status, cohort, entry scheme or clinical site. The mini-CEX is one of several assessment tools used to evaluate the student's Clinical and Communication Skills. Importantly, it involves direct assessment of patients in a "real world" setting and provides immediate feedback to guide the student in addressing areas for development or consolidating skills. Consideration should be given to reviewing the rubric used to assess the Clinical and Communication Skills domain and recognising that this would impact on resourcing, consider increasing the number of mandatory mini-CEX undertaken in Year 4.

References.

- Alves de Lima A, Henquin R, Thierer J, Paulin J, Lamari S, Belcastro F, Van der Vleuten C. 2005. A qualitative study of the impact on learning of the mini clinical evaluation exercise in postgraduate training. *Medical Teacher*, 27(1): 46–52.
- Alves de Lima A, Barrero C, Baratta S, Costa YC, Bortman G, Carabajales J, Conde D, Galli A, Degrange G, Van der Vleuten C. 2007. Validity, reliability, feasibility and satisfaction of the Mini-Clinical Evaluation Exercise (Mini-CEX) for cardiology residency training. *Medical Teacher*, 29(8): 785–790.
- Alves de Lima A, Conde D, Costabel J, Corso J, Van der Vleuten C. 2013. A laboratory study on the reliability estimations of the mini-CEX. *Advances in Health Sciences Education: Theory and Practice*, 18(1): 5–13.
- Arora VM, Berhie S, Horwitz LI, Saathoff M, Staisiunas P, Farnan JM. 2014. Using standardized videos to validate a measure of handoff quality: the handoff mini-clinical examination exercise. *Journal of Hospital Medicine*, 9(7): 441-446.
- Baig LA, Violato C, Crutcher RA. 2009. Assessing clinical communication skills in physicians: are the skills context specific or generalizable. *BioMedCentral Medical Education*, 9(5): 22.
- Baig L, Violato C, Crutcher R. 2010. Construct Validity Study of Clinical Competence: A Multitrait Multimethod Matrix Approach. *Journal of Continuing Education in the Health Professions*, 30(1): 19–25.
- Bajwa NM, Yudkowsky R, Belli D, Vu NV, Park YS. 2017. Improving the residency admissions process by integrating a professionalism assessment: A validity and feasibility study. *Advances in Health Sciences Education: Theory and Practice*, 22(1): 69-89.

Behere R. 2014. Introduction of Mini-CEX in Undergraduate Dental Education in India. *Education for Health*, 27(3): 262-268.

Beran TN, Violato C. 2010. Structural equation modeling in medical research: a primer. *BioMed Central research notes* 3(10): 267.

Bland J, Altman D. 1997. Statistics notes: Cronbach's alpha. *British Medical Journal*, 314(7098): 275.

Bok HGJ, Jaarsma DADC, Spruijt A, Van Beukelen P, Van Der Vleuten CPM, Teunissen PW. 2016. Feedback-giving behaviour in performance evaluations during clinical clerkships. *Medical Teacher*, 38(1): 88-95.

Bond WF, Gonzalez HC, Funk AM, Fehr LS, McGarvey JS, Svendsen JD, Sawicki R. 2017. Deliberate Practice with Standardized Patient Actors and the Development of Formative Feedback for Advance Care Planning Facilitators. *Journal of Palliative Medicine*, 20(6): 631-637.

Boulet JR, McKinley D, Norcini JJ, Whelan GP. 2002. Assessing the Comparability of Standardized Patient and Physician Evaluations of Clinical Skills. *Advances in Health Sciences Education: Theory and Practice*, 7(2): 85-97.

Butterworth K. 2010. Developing self-assessment skills amongst doctors in Nepal. *Medical Teacher*, 32(2): e85-e95.

Castanelli DJ, Jowset T, Chan Y, Weller JM. 2016. Perceptions of purpose, value, and process of the mini-Clinical Evaluation Exercise in anesthesia training. *Canadian Journal of Anaesthesia* 63(12): 1345-1356.

Chandra M. 2013. Initiating formative assessment of postgraduate students in obstetrics and gynecology. *National Journal of Integrated Research in Medicine*, 4(5): 132-137.

Chang YC, Lee CH, Chen CK, Liao CH, Ng CJ, Chen JC, Chaou CH. 2017. Exploring the influence of gender, seniority and specialty on paper and computer-based feedback provision during mini-CEX assessments in a busy emergency department. *Advances in Health Sciences Education: Theory and Practice*, 22(1): 57-67.

Cilliers FJ, Schuwirth LW, Herman N, Adendorff HJ, van der Vleuten CP. 2012. A model of the preassessment learning effects of summative assessment in medical education. *Advances in Health Sciences Education: Theory and Practice*, 17(1): 39-53.

Cohen LH. 1988. Measurement of life events. In: LH Cohen (Ed.), *Life Events and Psychological Functioning: Theoretical and Methodological Issues* (pp. 11-30). Newbury Park, California: Sage.

Cohen SN, Farrant PB, Taibjee SM. 2009. Assessing the assessments: U.K. dermatology trainees' views of the workplace assessment tools. *The British Journal of Dermatology*, 161(1):34-39.

Cook DA, Dupras DM, Beckman TJ, Thomas KG, Pankratz VS. 2009a. Effect of Rater Training on Reliability and Accuracy of Mini-CEX Scores: A Randomized, Controlled Trial. *Journal of General Internal Medicine*, 24(1): 74–79.

Cook DA, Beckman TJ. 2009b. Does scale length matter? A comparison of nine- versus five-point rating scales for the mini-CEX. *Advances in Health Sciences Education: Theory and Practice*, 14(5): 655–664.

Cook DA, Beckman TJ, Mandrekar JN, Pankratz VS. 2010. Internal structure of mini-CEX scores for internal medicine residents: factor analysis and generalizability. *Advances in Health Science Education: Theory and Practice*, 15(5): 633–645.

Cook DA, Brydges R, Ginsburg S, Hatala R. 2015. A contemporary approach to validity arguments: a practical guide to Kane's framework. *Medical Education*, 49(6): 560–75.

Cronbach L. 1951. Coefficient alpha and the internal structure of tests. *Psychometrika*, 16(3): 297–334.

Cronbach LJ, Meehl PE. 1955. Construct validity in psychological tests. *Psychological Bulletin*, 52(4): 281-302.

Cruess R, Mcllroy JH, Cruess S, Ginsburg S, Steinert Y. 2006. The Professionalism Mini-evaluation Exercise: a preliminary investigation. *Academic Medicine*, 81(10): S74-S78.

Crossley J, Johnson G, Booth J, Wade W. 2011. Good questions, good answers: construct alignment improves the performance of workplace-based assessment scales. *Medical Education*, 45(6): 560-569.

Curtis E, Reid P. 2013. Indigenous Health Workforce Development: challenges and successes of the Vision 20:20 programme. *ANZ Journal of Surgery*, 83 (1-2): 49–54.

Curtis E, Wikaire E, Jiang Y, McMillan L, Loto R, Poole P, Barrow M, Bagg W, Reid P. 2017. Examining the predictors of academic outcomes for indigenous Māori, Pacific and rural students admitted into medicine via two equity pathways: a retrospective observational study at the University of Auckland, Aotearoa New Zealand. *BMJ Open*, 7 (8): e017276.

Davies H, Archer J, Southgate L, Norcini J. 2009. Initial evaluation of the first year of the Foundation Assessment Programme. *Medical Education*, 43(1): 74–81.

Daelmans HE, Mak-van der Vossen MC, Croiset G, Kusurkar RA. 2016. What difficulties do faculty members face when conducting workplace-based assessments in undergraduate clerkships? *International Journal of Medical Education*, 7(1): 19-24.

DeVellis R. 2003. *Scale development: theory and applications* (2nd Ed). Thousand Oaks, California: Sage.

Dewi SP, Achmad TH. 2010. Optimising feedback using the mini-CEX during the final semester programme. *Medical Education*, 44(5): 509.

Downing S. 2004. Reliability: on the reproducibility of assessment data. *Medical Education*, 38(9): 1006–1012.

Dreyfus SE, Dreyfus HL. February 1980. *A five-stage model of the Mental Activities Involved in Directed Skill Acquisition*. Washington, District of Columbia: Storming Media.

Durning SJ, Cation LJ, Markert RJ, Pangaro LN. 2002. Assessing the Reliability and Validity of the Mini–Clinical Evaluation Exercise for Internal Medicine Residency Training. *Academic Medicine*, 77(9): 900–904.

Eggleton K, Goodyear-Smith F, Paton L, Falloon K, Wong C, Lack L, Kennelly J, Fishman T; Moyes SA. 2016. Reliability of Mini-CEX Assessment of Medical Students in General Practice Clinical Attachments. *Family Medicine*, 48(8): 624-630.

Evans JD. 1996. *Straightforward Statistics for the Behavioral Sciences*. Pacific Grove, California: Brooks/Cole Publishing.

Farnan JM, Paro JAM, Rodriguez RM, Reddy ST, Horwitz LI, Johnson JK, Arora VM. 2010. Hand-off education and evaluation: Piloting the Observed Simulated Hand-off Experience (OSHE). *Journal of General Internal Medicine*, 25(2): 129-134.

Ferguson E, James D, Madley L. 2002. Factors associated with success in medical school: systematic review of the literature. *British Medical Journal*, 324(7343):952–957.

Fernando N, Cleland J, McKenzie H, Cassar K. 2008. Identifying the factors that determine feedback given to undergraduate medical students following formative mini-CEX assessments *Medical Education*, 42(1): 89–95.

Ferrari R. 2015. Writing narrative style literature reviews. *Medical Writing*, 24(4): 230-235.

Field A. 2013. *Discovering Statistics Using IBM SPSS* (4th Ed). Singapore: Sage Publications.

Fuller R, Coulby C, Homer-Vanniasinkam S, Hennessy S, Davis N. 2009. e-Learning in medical education: Guide supplement 32.4 – Practical application. *Medical Teacher*, 31(4): 368-369.

Gauthier S, Cavalcanti R, Goguen J, Sibbald M. 2015 Deliberate practice as a framework for evaluating feedback in residency training. *Medical Teacher* 37(6):551-557.

Gingerich A, van der Vleuten CPM, Eva KW, Regehr G. 2014. More Consensus Than Idiosyncrasy: Categorizing Social Judgments to Examine Variability in Mini-CEX Ratings. *Academic Medicine*, 89(11): 1510-1519.

Gingerich A, Ramlo SE, van der Vleuten CPM, Eva KW, Regehr G. 2017 Inter-rater variability as mutual disagreement: identifying raters' divergent points of view. *Advances in Health Sciences Education: Theory and Practice*, 22(4): 819-838.

Goel A, Singh T. 2015. The usefulness of Mini Clinical Evaluation Exercise as a learning tool in different pediatric clinical settings. *International Journal of Applied and Basic Medical Research*, 5(1): S32–S34.

Golnik KC, Goldenhar LM, Gittinger JW Jr, Lustbader JM. 2004. The Ophthalmic Clinical Evaluation Exercise (OCEX). *Ophthalmology*, 111(7): 1271-1274.

Golnik KC, Goldenhar L. 2005. The Ophthalmic Clinical Evaluation Exercise: reliability determination. *Ophthalmology*, 112(10): 1649-1654.

Greenburg DL, Durning SJ, Cohen DL, Cruess D, Jackson JL. 2007. Identifying medical students likely to exhibit poor professionalism and knowledge during internship. *Journal of General Internal Medicine*, 22(12): 1711–1717.

Han PK, Keranen LB, Lescisin DA, Arnold RM. 2005. The palliative care Clinical Evaluation Exercise (CEX): an experience-based intervention for teaching end-of-life communication skills. *Academic Medicine*, 80(7): 669-676.

Harrison CJ, Molyneaux AJ, Blackwell S, Wass VJ. 2015. How we give personalised audio feedback after summative OSCEs. *Medical Teacher*, 37(4): 323- 326.

Harrison CJ, Könings KD, Dannefer EF, Schuwirth LW, Wass V, van der Vleuten CP. 2016. Factors influencing students' receptivity to formative feedback emerging from different assessment cultures. *Perspectives of Medical Educatio*, 5(5): 276-284.

Harrison CJ, Könings KD, Schuwirth LWT, Wass V, van der Vleuten CPM. 2017. Changing the culture of assessment: the dominance of the summative assessment paradigm. *BioMedCentral Medical Education*, 17(1): 73.

Harvey P, Radomski N, O'Connor D. 2013. Written feedback and continuity of learning in a geographically distributed medical education program. *Medical Teacher*, 35(12): 1009–1013.

Hatala R, Ainslie M, Kassen BO, Mackie I, Roberts JM. 2006. Assessing the mini-Clinical Evaluation Exercise in comparison to a national specialty examination. *Medical Education*, 40(10): 950–956.

John Hattie. University of Auckland. 1999 Inaugural Professorial Lecture.

Hawthorne MR, Chretien KC, Torre D, Chheda SG. 2014. Re-demonstration without remediation - a missed opportunity? A national survey of internal medicine clerkship directors. *Medical Education Online*, 19(12): 25991.

Hauer KE. 2000. Enhancing Feedback to Students Using the Mini-CEX (Clinical-Evaluation Exercise). *Academic Medicine*, 75(5): 524.

Hauer KE, Soni K, Cornett P, Kohlwes J, Hollander H, Ranji SR, ten Cate O, Widera E, Calton B, O'Sullivan PS. 2013. Developing Entrustable Professional Activities as the Basis for Assessment of Competence in an Internal Medicine Residency: A Feasibility Study. *Journal of General Internal Medicine*, 28(8): 1110–1114.

Hawkins RE, Margolis MJ, Durning SJ, Norcini JJ. 2010. Constructing a Validity Argument for the Mini-Clinical Evaluation Exercise: A Review of the Research. *Academic Medicine*, 85(9): 1453-1461.

Henson R, Roberts J. 2006. Use of exploratory factor analysis in published research: Common errors and some comment on improved practice. *Educational and Psychological Measurement*, 2006, 66(3): 393-416.

Hill F, Kendall K. 2007. Adopting and adapting the mini-CEX as an undergraduate assessment and learning tool. *Clinical Teacher*, 4(4): 244–8.

Hill F, Kendall K, Galbraith J, Crossley J. 2009. Implementing the undergraduate mini-CEX: a tailored approach at Southampton University. *Medical Education*, 43(4): 326–334.

Holgado–Tello FP, Chacón–Moscoso S, Barbero–García I, Vila-Abad E. 2010. Polychoric versus Pearson correlations in exploratory and confirmatory factor analysis of ordinal variables. *Quality and Quantity* 44(1): 153-166.

Hojat M, Paskin DL, Callahan CA, Nasca TJ, Louis DZ, Veloski J, Erdmann JB, Gonella JS. 2007. Components of postgraduate competence: analyses of thirty years of longitudinal data. *Medical Education*, 41(10): 982– 989.

Holmboe ES, Huot S, Chung J, Norcini JJ, Hawkins RE. 2003. Construct Validity of the Mini Clinical Evaluation Exercise (MiniCEX). *Academic Medicine*, 78(8): 826–830.

Holmboe ES, Hawkins RE, Huot SJ. 2004a. Effects of training in direct observation of medical residents' clinical competence: a randomized trial. *Annals of Internal Medicine*, 140(11): 874-881.

Holmboe ES, Yepes M, Williams S, Huot SJ. 2004b. Feedback and the mini-clinical evaluation exercise. *Journal of General Internal Medicine*, 19(5): 558–561.

Homer M, Setna Z, Jha V, Higham J, Roberts T, Boursicot K. 2013. Estimating and comparing the reliability of a suite of workplace-based assessments: An obstetrics and gynaecology setting. *Medical Teacher*, 35(8): 684-691.

Horwitz LI, Dombroski J, Murphy TE, Farnan JM, Johnson JK, Arora VM. 2013. Validation of a handoff assessment tool: The Handoff CEX. *Journal of Clinical Nursing*, 22(9-10): 1477-1486.

Jackson D, Wall D. 2010. An evaluation of the use of the mini-CEX in the foundation programme. *British Journal of Hospital Medicine*, 71(10): 584 – 588.

Jenkins GD, Taber TD. 1977. A Monte Carlo study of factors affecting three indices of composite scale reliability. *The Journal of Applied Psychology*, 62(4): 392–398.

Joshi MK, Singh T, Badyal DK. 2017. Acceptability and feasibility of mini-clinical evaluation exercise as a formative assessment tool for workplace-based assessment for surgical postgraduate students. *Journal of Postgraduate Medicine*, 63(2): 100-105.

Kaiser HF. 1960. The application of Electronic Computers to Factor Analysis. *Educational and Psychological Measurement*, 20 (1): 141-151.

Kane MT. 2001. Current concerns in validity theory. *Journal of Educational Measurement*, 38(4): 319–342.

Kapoor H, Tekian A, Mennin S. 2010. Structuring an internship programme for enhanced learning. *Medical Education*, 44(5): 501-502.

Khalil S, Aggarwal A, Mishra D. 2017. Implementation of a Mini-Clinical Evaluation Exercise (Mini-CEX) Program to Assess the Clinical Competence of Postgraduate Trainees in Pediatrics. *Indian Paediatrics*, 54(4): 284-287.

Kline P. 1999. *A Handbook of Psychological Testing* (2nd Ed.) London: Routledge.

Kim S, Willett LR, Noveck H, Patel MS, Walker JA, Terregino CA. 2016. Implementation of a Mini-CEX Requirement Across All Third-Year Clerkships. *Teaching and Learning in Medicine*, 28(4): 424-431.

Kogan JR, Bellini LM, Shea JA. 2002. Implementation of the Mini-CEX to Evaluate Medical Students' Clinical Skills. *Academic Medicine*, 77(11): 1156 – 1157.

Kogan JR, Bellini LM, Shea JA. 2003. Feasibility, Reliability, and Validity of the Mini-Clinical Evaluation Exercise (mCEX) in a Medicine Core Clerkship. *Academic Medicine*, 78(10): S33 – S35.

Kogan JR, Hess BJ, Conforti LN, Holmboe ES. 2010. What Drives Faculty Ratings of Residents' Clinical Skills? The Impact of Faculty's Own Clinical Skills. *Academic Medicine*, 85(10): S25-S28.

Kogan JR, Conforti LN, Bernabeo EC, Durning SJ, Hauer KE, Holmboe ES. 2012. Faculty staff perceptions of feedback to residents after direct observation of clinical skills. *Medical Education*, 46(2): 201–215.

Kogan JR, Conforti LN, Bernabeo EC, Iobst W, Holmboe ES. 2015. How faculty members experience workplace-based assessment rater training: a qualitative study. *Medical Education*, 49(7): 692–708.

Knorr M, Klusmann D. 2015. The trouble with validity: what is part of it and what is not? *Medical Education*, 49(5): 550-552.

Liao KC, Pu SJ, Liu MS, Yang CW, Kuo HP. 2013. Development and implementation of a mini-Clinical Evaluation Exercise (mini-CEX) program to assess the clinical competencies of internal medicine residents: from faculty development to curriculum evaluation. *BioMedCentral Medical Education*, 13(2): 31

London Z, Schuh L, Gelb DJ, Schultz L. 2013. Education Research: unsatisfactory NEX rating correlations: searching for the reasons. *Neurology*, 80(13): e142-e145.

Malhotra S, Hatala R, Courneya CA. 2008. Internal medicine residents' perceptions of the mini-clinical evaluation exercise. *Medical Teacher*, 30(4): 414-419.

Margolis MJ, Clauser BE, Cuddy MM, Ciccone A, Mee J, Harik P, Hawkins RE. 2006. Use of the Mini-Clinical Evaluation Exercise to Rate Examinee Performance on a Multiple-Station Clinical Skills Examination: A Validity Study. *Academic Medicine*, 81(10): S56-S60.

MBCHB Phase 2 (Year 4) Guidebook 2017. The University of Auckland.

Medvedev ON, Krägeloh CU, Hill EM, Billington R, Siegert RJ, Webster C.S, Booth RJ, Henning MA. 2019. Rasch analysis of the Perceived Stress Scale: Transformation from an ordinal to a linear measure. *Journal of Health Psychology*, 24(8): 1070-1081.

Miller GE. 1990. The assessment of clinical skills/competence/performance. *Academic Medicine*, 65(9): S63-S67.

Milner KA, Watson SM, Stewart JG, DeNisco SJ. 2014. Use of Mini-CEX tool to assess clinical competence in family nurse practitioner students using undergraduate students as patients and doctoral students as evaluators. *Journal of Nursing Education*, 53(12): 719-720.

Ministry of Health Manatū Hauora. <https://www.health.govt.nz>. Accessed 19th November 2018.

Mitchell C, Bhat S, Herbert A, Baker P. 2011. Workplace-based assessments of junior doctors: do scores predict training difficulties? *Medical Education*, 45(12): 1190-1198.

Moore K, Vaughan B. 2017. Students today ... Educators tomorrow. *The Clinical Teacher*, 14(5): 325-329.

Montagne S, Rogausch A, Gemperli A, Berendonk C, Jucker-Kupper P, Beyeler C. 2014. The mini-clinical evaluation exercise during medical clerkships: are learning needs and learning goals aligned? *Medical Education*, 48(10): 1008–1019.

Moonen-van Loon JMW, Overeem K, Donkers HHLM, van der Vleuten CPM, Driessen EW. 2013. Composite reliability of a workplace-based assessment toolbox for postgraduate medical education. *Advances in Health Science Education*, 18(5): 1087–1102.

Mortaz Hejri S, Jalili M, Shirazi M, Masoomi R, Nedjat S, Norcini J. 2017. The utility of mini-Clinical Evaluation Exercise (mini-CEX) in undergraduate and postgraduate medical education: protocol for a systematic review. *Systematic Reviews* 6(1): 146.

Muthén B, Kaplan D. 1985. A comparison of some methodologies for the factor analysis of non-normal Likert variables. *British Journal of Mathematical and Statistical Psychology*, 38(2):171-189.

Nair BR, Alexander HG, McGrath BP, Parvathy MS, Kilsby EC, Wenzel J, Frank IB, Pachev GS, Page GG. 2008. The mini clinical evaluation exercise (mini-CEX) for assessing clinical performance of international medical graduates. *Medical Journal of Australia*, 189(3): 159-161.

Nair BR, Moonen-van Loon JMW, Parvathy M, Jolly BC, van der Vleuten CPM. 2017. Composite reliability of workplace-based assessment of international medical graduates. *Medical Journal of Australia*, 207(10): 453-457.

Nasca TJ, Gonnella JS, Hojat M, Veloski J, Erdmann JB, Robeson M, Brigham TP, Callahan C. 2002. Conceptualization and measurement of clinical competence of residents: A brief rating form and its psychometric properties. *Medical Teacher*, 24(3), 299–303.

Newble DI, Jaeger K. 1983. The effect of assessments and examinations on the learning of medical students. *Medical Education*, 17(3): 165–171.

Ney EM, Shea JA, Kogan JR. 2009. Predictive Validity of the mini-Clinical Evaluation Exercise (mCEX): Do Medical Students' mCEX Ratings Correlate With Future Clinical Exam Performance? *Academic Medicine*, 84(10): S17–S20.

Nishisato S, Torii Y. 1970. Effects of categorizing continuous normal variables on product-moment correlation. *The Japanese Psychological Research*, 13(1): 45–49.

Norman G, Keane D, Oppenheimer L. 2008. Compliance of Medical Students with Voluntary use of personal data assistants for Clerkship Assessments. *Teaching and Learning in Medicine*, 20(4): 295-301.

Norcini JJ, Blank LL, Arnold GK, Kimball HR. 1995. The mini-CEX (clinical evaluation exercise): a preliminary investigation. *Annals of Internal Medicine*, 123(10):795–799.

Norcini JJ, Blank LL, Arnold GK, Kimball HR. 1997. Examiner Differences in the Mini-Cex. *Advances in Health Sciences Education: Theory and Practice*, 2(1): 27–33.

Norcini JJ. 1999. Standards and reliability in evaluation: when rules of thumb don't apply. *Academic Medicine*, 74(10): 1088-1090.

Norcini JJ. 2002. The death of the long case? *British Medical Journal*, 324(7334): 408 – 409.

Norcini JJ, Blank LL; Duffy FD, Fortna GS. 2003. The Mini-CEX: A Method for Assessing Clinical Skills. *Annals of Internal Medicine*, 138(6): 476-481.

Nunnally JC. 1978. *Psychometric theory*. New York: McGraw-Hill.

Olupeliyawa AM, O'Sullivan AJ, Hughes C, Balasooriya CD. 2014. The Teamwork Mini-Clinical Evaluation Exercise (T-MEX): a workplace-based assessment focusing on collaborative competencies in health care. *Academic Medicine*, 89(2): 359-365.

Olupeliyawa A, Balasooriya C, Hughes C, O'Sullivan A. 2014. Educational impact of an assessment of medical students' collaboration in health care teams. *Medical Education*, 48(2): 146-156.

Paravicini I, Peterson CK. 2015. Introduction, development, and evaluation of the miniclinical evaluation exercise in postgraduate education of chiropractors. *Journal of Chiropractic Education*, 29(1): 22-28.

Pelgrim EA, Kramer AW, Mokkink HG, Van der Vleuten CP. 2012. Quality of written narrative feedback and reflection in a modified mini-clinical evaluation exercise: an observational study. *BioMedCentral Medical Education*, 12: 97.

Perkins D, Daly M. 2013. What is the evidence for clinical placements in underserved areas? *Medical Education*, 47(10): 958-960.

Pernar LIM, Peyre SE, Warren LEG, Gu X, Lipsitz S, Alexander EK, Ashley SW, Breen EM. 2011. Mini-clinical evaluation exercise as a student assessment tool in a surgery clerkship: Lessons learned from a 5-year experience. *Surgery*, 150(2): 272-277.

Phillips AW, Madhavan A. 2013. A Critical Evaluation of the Intercollegiate Surgical Curriculum and Comparison With its Predecessor the “Calman” Curriculum. *Journal of Surgical Education*, 70(5): 557-562.

Phillips AW, Madhavan A, Bookless LR, Macafee DAL. 2015. Surgical Trainers’ Experience and Perspectives on Workplace-Based Assessments. *Journal of Surgical Education*, 72(5): 979-984.

Playford D, Kirke A, Maley M, Worthington R. 2013. Longitudinal assessment in an undergraduate longitudinal integrated clerkship: The mini Clinical Evaluation Exercise (mCEX) profile. *Medical Teacher*, 35(8): e1416–e1421.

Poole P, Shulruf B, Rudland J, Wilkinson T. 2012. Comparison of UMAT scores and GPA in prediction of performance in medical school; a national study. *Medical Education*, 46(2): 163-171.

Poole P, Wilkinson TJ, Bagg W, Freegard J, Hyland F, Jo E, Kool B, Roberts E, Rudland J, Smith B, Verstappen AC. 2019. Developing New Zealand’s medical workforce: realising the potential of longitudinal career tracking. *New Zealand Medical Journal*, 132(1495): 65-73.

Preston CC, Colman AM. 2000. Optimal number of response categories in rating scales: Reliability, validity, discriminating power, and respondent preferences. *Acta Psychologica*, 104(1): 1–15.

Ramaprasad A. 1983. On the definition of feedback. *Behavioral Science*, 28(1): 4-13.

Rasch G. 1960. *Probabilistic Models for Some Intelligence and Attainment Test*. Copenhagen: Danish Institute for Educational Research.

Rethans JJ, Norcini JJ, Barón-Maldonado M, Blackmore D, Jolly BC, LaDuca T, Lew S, Page GG, Southgate LH. 2002. The relationship between competence and performance: implications for assessing practice performance. *Medical Education*, 36(10): 901-909.

Rogausch A, Beyeler C, Montagne S, Jucker-Kupper P, Berendonk C, Huwendiek S, Gemperli A, Himmel W. 2015. The influence of students' prior clinical skills and context characteristics on mini-CEX scores in clerkships--a multilevel analysis. *BioMed Central Medical Education*, 15: 208-217.

Sackett PR, Schmitt N, Ellingson JE, Kabin MB. 2001. High-stakes testing in employment, credentialing, and higher education: Prospects in a post-affirmative-action world. *American Psychologist*, 56(4): 302-318.

Sadanand DM, Mohan J. 2013. Use Of Mini-Cex As An Effective Tool Of Assessment In Second B.A.M.S. Students In Rognidan. *National Journal of Integrated Research in Medicine*, 4(5): 114-117.

Sadler DR. 1989. Formative assessment and the design of instructional systems. *Instructional Science*, 18(2): 119-144.

Schuh LA, London Z, Neel R, Brock C, Kissela BM, Schultz L, Gelb DJ. 2009. Education research: Bias and poor interrater reliability in evaluating the neurology clinical skills examination. *Neurology*, 73(11): 904-908.

Schuwirth LW, van der Vleuten CP. 2011a. General overview of the theories used in assessment: AMEE Guide No. 57. *Medical Teacher*, 33(10):783-97.

Schuwirth LWT, van der Vleuten CPM. 2011b. Programmatic assessment: From assessment of learning to assessment for learning. *Medical Teacher*, 33(6): 478-485.

Shaw B, Fox J, Brown J, Hart A, Mamelok J. 2014. An investigation of factors affecting the outcome of the clinical skills assessment (CSA) in general practice specialty training. *Education for Primary Care*, 25(2): 91-95.

Shayne P, Heilpern K, Ander D, Palmer-Smith V. 2002. Protected clinical teaching time and a bedside clinical evaluation instrument in an emergency medicine training program. *Academic Emergency Medicine*, 9(11): 1342-1349.

Sidhu RS, Hatala R, Barron S, Pachev G, Page G. 2009. Reliability and Acceptance of the Mini-Clinical Evaluation Exercise as a Performance Assessment of Practicing Physicians. *Academic Medicine*, 84(10): S113–S115.

Singh T, Sharma M. 2010. Mini-clinical examination (CEX) as a tool for formative assessment. *National Medical Journal of India*, 23(2): 100–102.

Siddaway AP, Wood AM, Hedges LV. 2019. How to do a systematic review: A best practice guide to conducting and reporting narrative reviews, meta-analyses, and meta-syntheses. *Annual Review of Psychology*, 70(1): 747-770.

Silber CG, Nasca TJ, Paskin DL., Eiger G, Robeson M, Veloski JJ. 2004. Do global rating forms enable program directors to assess the ACGME competencies? *Academic Medicine*, 79(6): 549–556.

Spearman C. 1904. The Proof and Measurement of Association between Two Things. *American Journal of Psychology*, 15(1): 72–101.

Streiner DL. 2003. Starting at the beginning: an introduction to coefficient alpha and internal consistency. *Journal of Personality Assessment* 80(1): 99-103.

Sударso S, Rahayu GR, Suhoyo Y. 2016. How does feedback in mini-CEX affect students' learning response? *International Journal of Medical Education*, 7: 407-413.

Suhoyo Y, Schönrock-Adema J, Rahayu GR, Kuks JBM, Cohen-Schotanus J. 2014. Meeting international standards: A cultural approach in implementing the mini-CEX effectively in Indonesian clerkships. *Medical Teacher*, 36(10): 894–902.

Sweet LP, Glover P, McPhee T. 2013. The midwifery miniCEX--a valuable clinical assessment tool for midwifery education. *Nurse Education in Practice*, 13(2): 147-153.

Tabachnick B, & Fidell L. 2007. *Using multivariable statistics* (5th Ed). New York: Pearson/Allyn & Bacon.

Ullman JB. 2001. Structural equation modeling. In BG Tabachnick & LS Fidell (Ed), *Using Multivariate Statistics* (4th Ed) (pp. 653- 771). Boston: Allyn & Bacon.

The Royal Australasian College of Physicians. Basic Training Curriculum Adult Internal Medicine. <https://www.racp.edu.au/trainees/basic-training/adult-internal-medicine>. Accessed August 2018

The Royal Australasian College of Physicians. Nephrology Advanced Training Curriculum. <https://www.racp.edu.au/docs/default-source/default-document-library/nephrology-advanced-training-curriculum.pdf>. Accessed May 2019.

Thomas PA, Gebo KA, Hellmann DB. 1999. A pilot study of peer review in residency training. *Journal of General Internal Medicine*, 14(9): 551–554.

Tighe J, McManus IC, Dewhurst NG, Mucklow J. 2010. The standard error of measurement is a more appropriate measure of quality for postgraduate medical assessments than is reliability: an analysis of MRCP(UK) examinations. *BioMedCentral Medical Education*, 10(6): 40

Torre DM, Simpson DE, Elnicki DM, Sebastian JL, Holmboe ES. 2007. Feasibility, Reliability and User Satisfaction With a PDA-Based Mini-CEX to Evaluate the Clinical Skills of Third-Year Medical Students. *Teaching and Learning in Medicine*, 19(3): 271-277.

Torre DM, Treat R, Durning S, Elnicki M. 2011. Comparing PDA- and Paper-based Evaluation of the Clinical Skills of Third-Year Students. *Wisconsin Medical Journal*, 110(1): 9-13.

Tsugawa Y, Tokuda Y, Ohbu S, Okubo T, Cruess R, Cruess S, Ohde S, Okada S, Hayashida N, Fukui T. 2009. Professionalism Mini-Evaluation Exercise for medical residents in Japan: a pilot study. *Medical Education*, 43(10): 968-978.

Tsugawa Y, Ohbu S, Cruess R, Cruess S, Okubo T, Takahashi O, Tokuda Y, Heist BS, Bito S, Itoh T, Aoki A, Chiba T, Fukui T. 2011. Introducing the Professionalism Mini-Evaluation Exercise (P-MEX) in Japan: results from a multicenter, cross-sectional study. *Academic Medicine*, 86(8): 1026-1031.

van der Vleuten C. 1996. The assessment of professional competence: developments, research and practical implications. *Advances in Health Sciences Education: Theory and Practice*, 1(1):41–67.

van der Vleuten C. 2000. Validity of final examinations in undergraduate medical training. *British Medical Journal*, 321(7270): 1217–1219.

van der Vleuten CP, Schuwirth LW. 2005. Assessing professional competence: from methods to programmes. *Medical Education*, 39(3): 309-317.

van der Vleuten CP, Schuwirth LW, Driessen EW, Dijkstra J, Tigelaar D, Baartman LK, van Tartwijk J. 2012. A model for programmatic assessment fit for purpose. *Medical Teacher* 34(12): 205– 214.

Vaughan B, Moore K. 2016. The mini Clinical Evaluation Exercise (mini-CEX) in a pre-registration osteopathy program: Exploring aspects of its validity. *International Journal of Osteopathic Medicine*, 19(3): 61 – 72.

Volkan K, Simon SR, Baker H, Todres ID. 2004. Psychometric structure of a comprehensive objective structured clinical examination: A factor analytic approach. *Advances in Health Sciences Education: Theory and Practice*, 9(2): 83–92.

Walsh E, Foley T, Sinnott C, Boyle S, Smithson WH. 2017. Developing and piloting a resource for training assessors in use of the Mini-CEX (mini clinical evaluation exercise). *Education for Primary Care*, 28(4): 243-245.

Wang PW, Cheng CC, Chou FH, Tsang HY, Chang YS, Huang MF, Yen CF. 2011. Using multiple assessments to evaluate medical students' clinical ability in psychiatric clerkships. *Academic Psychiatry*, 35(5): 307-311.

Wass V, Van Der Vleuten C. 2004. The long case. *Medical Education*, 38(11): 1176-1180.

Weijs CA, Coe JB, Hecker KG. 2015. Final-Year Students' and Clinical Instructors' Experience of Workplace-Based Assessments Used in a Small-Animal Primary-Veterinary-Care Clinical Rotation. *Journal of Veterinary Medical Education*, 42(4): 382-392.

Weller JM, Jolly B, Misur MP, Merry AF, Jones A, Crossley JG, Pedersen K, Smith K. 2009a. Mini-clinical evaluation exercise in anaesthesia training. *British Journal of Anaesthesia*, 102(5):633– 641.

Weller JM, Jones A, Merry AF, Jolly B, Saunders D. 2009b. Investigation of trainee and specialist reactions to the mini-Clinical Evaluation Exercise in anaesthesia: implications for implementation. *British Journal of Anaesthesia*, 103 (4):524–30.

Weller JM, Misur M, Nicolson S, Morris J, Ure S, Crossley J, Jolly B. 2014. Can I leave the theatre? A key to more reliable workplace-based assessment. *British Journal of Anaesthesia*, 112(6): 1083–091.

Weller JM, Castanelli DM, Chen Y, Jolly B. 2017. Making robust assessments of specialist trainees' workplace performance. *British Journal of Anaesthesia*, 118(2): 207–214.

Weston PSJ, Smith CA. 2014. The use of mini-CEX in UK foundation training six years following its introduction: Lessons still to be learned and the benefit of formal teaching regarding its utility. *Medical Teacher*, 36(2): 155-163.

Wilkinson TJ, Frampton CM. 2003. Assessing performance in final year medical students. Can a postgraduate measure be used in an undergraduate setting? *Medical Education*, 37(3): 233–240.

Wilkinson JR, Crossley JGM, Wragg A, Mills P, Cowan G, Wade W. 2008. Implementing workplace-based assessment across the medical specialties in the United Kingdom. *Medical Education*, 42(4): 364–373.

Willett LL, Heudebert GR, Palonen KP, Massie FS, Kiefe CI, Allison JJ, Richman J, Houston TK. 2009. The Importance of Measuring Competency-Based Outcomes: Standard Evaluation Measures are not Surrogates for Clinical Performance of Internal Medicine Residents. *Teaching and Learning in Medicine*, 21(2): 87-93.

Yanting SL, Sinnathamby A, Wang DB, Heng MTM, Hao JLW, Lee SS, Yeo SP, Samarasekera DD. 2016. Conceptualizing workplace based assessment in Singapore: Undergraduate Mini-Clinical Evaluation Exercise experiences of students and teachers. *Tzu-Chi Medical Journal*, 28(3): 113-120.

Yeates P, O'Neill P, Mann K, Eva KW. 2012. Effect of Exposure to Good vs Poor Medical Trainee Performance on Attending Physician Ratings of Subsequent Performances. *Journal of the American Medical Association*, 308(21): 2226-2232.

Yeates P, O'Neill P, Mann K, Eva K. 2013. Seeing the same thing differently. Mechanisms that contribute to assessor differences in directly-observed performance assessments. *Advances in Health Sciences Education: Theory and Practice*, 18(3): 325- 341.

Yeates P, Cardell J, Byrne G, EVA KW. 2015. Relatively speaking: contrast effects influence assessors' scores and narrative feedback. *Medical Education*, 49(9): 909-919.

Young AG, Pearce S. 2013. A Beginner's Guide to Factor Analysis: Focusing on Exploratory Factor Analysis. *Tutorials in Quantitative Methods for Psychology*, 9(2): 79-94.

Appendix 1 - Versions of the mini-CEX used in NZ and selected Australian Medical Schools.

University of Auckland					
Year 4 Medicine / Mini-CEX student evaluation					
Assessor's name: _____		Date: _____			
Student Name: _____		ID: _____			
Location: _____					
Patient problem / diagnosis: _____					
Patient: _____		Age: _____		Sex: _____	
Setting: Outpatient <input type="radio"/> Inpatient <input type="radio"/> Acute <input type="radio"/> Other <input type="radio"/>					
Complexity (for Year 4): Low <input type="radio"/> Medium <input type="radio"/> High <input type="radio"/>					
Please tick the appropriate column:					
	Major Deficiencies	Some Reservations	Good	Excellent	Not assessed
Interviewing Skills / History Taking					
Physical Examination Skills					
Professional Behaviour					
Clinical Reasoning					
Overall clinical performance. Please circle:	Fail	Borderline Performance	Pass	Distinction	
Time spent observing student (mins): _____					
Feedback to student:					
Assessor's signature: _____					

Instructions

This assessment should take 15-20 minutes. After direction to main system the student should be asked to take an appropriate history and focused examination. They should at the end summarise their findings and identify problem(s) and differential diagnosis (3-5). The assessment may take place on the ward or in clinic. Student may be invited to explain findings to patient to test counselling skills.

The following guides should be used to fill in each of the sections of the form:

These standards against which the student is being judged are that of a Year 4 student. Please tick the appropriate column:

Major Deficiencies: Several significant gaps or omissions. Unsafe or inappropriate actions

Some Reservations: A few gaps but no major deficiencies

Good: Very good performance, 1 or 2 minor gaps only

Excellent: Outstanding performance, no gaps or deficiencies

- **Interviewing skills / History Taking:** Was there effective history taking showing logical and empathetic approach to the patient's problems. Student should use a hypothetico-deductive, and reasoning approach.
- **Physical Examination Skills:** Was the appropriate examination carried out in a logical and methodical manner, indicating efficiency and prior experience. Student should adapt approach to context and history.
- **Professional Behaviour:** Student was sensitive, courteous and respectful and used appropriate communication techniques.
- **Clinical Reasoning:** Was the student able to summarise the history/examination findings in context, define problem(s), and propose a simple differential diagnosis.
- **Overall clinical performance:** A global professional judgement on the overall integrated evaluation of the student's performance.

Figure 9 – Initial version of the University of Auckland mini-CEX form.

This was initially used in 2013 and part of 2014.

MBChB Standard Mini-CEX Student Assessment

Assessor: _____ Date: _____
 Student: _____ Year: _____
 Teaching Site: _____
 Setting: Inpatient Outpatient ED GP Other _____
 Patient Presenting Complaint/ Diagnosis: _____
 Patient Age: _____ Patient Gender: _____

Please tick the appropriate boxes:

	Major Deficiencies	Some Reservations	Good	Excellent
History Taking / Communication Skills				
Physical Examination Skills				
Clinical Judgement / Reasoning				
Humanistic Qualities / Professionalism				

Overall Clinical Competence: (Circle the appropriate)

* FAIL	BORDERLINE PERFORMANCE	PASS	DISTINCTION
--------	------------------------	------	-------------

Feedback to student:

What aspects were done well	Areas requiring improvement

* If the student fails the Mini-CEX, a second assessment is to be completed before the end of the attachment.

Time spent directly observing student: _____ (minutes) Assessor's Signature: _____

Instructions for the Assessor

General

This mini-CEX evaluation should take about 15 to 20 minutes in total. The patient-student encounter should take 10 to 15 minutes followed by a brief discussion and provision of feedback to the student.

At the patient's bedside, the student should be directed to take an appropriate history and perform a focused and relevant physical examination. Away from the bedside (preferably), the student then presents a summary of the key examination findings, proposes a logical working diagnosis, and lists reasonable differential diagnoses. A discussion should follow, exploring the student's clinical knowledge and reasoning. Ask the student to clarify the purpose of their actions rather than make assumptions about their intentions.

Feedback should be provided verbally to the student and also as brief written comments using the boxes provided. Focus on one or two major points and explicitly describe how your feedback connects to the student's observed actions.

Criteria Descriptions

Excellent	Comprehensive performance, no gaps or deficiencies
Good	Sound performance, 1-2 minor gaps only
Some Reservations	A few gaps but no major deficiencies
Major Deficiencies	Several significant gaps or omissions. Unsafe or inappropriate actions

Your assessment of performance should take into consideration the complexity level of the chosen case/patient, and the level of the student.

Criteria Descriptions

- *History Taking / Communication Skills:* Was the patient's storytelling facilitated in an efficient, thorough, and accurate manner? Were relevant questions asked in a smooth, effective, and logical progression? Were appropriate listening skills demonstrated?
- *Physical Examination Skills:* Was the appropriate examination carried out in a logical and systematic manner, indicating efficiency and prior experience? Did the student attempt to elicit additional clinical signs pertinent to the case/condition?
- *Clinical Judgement / Reasoning:* Was the student able to synthesize history and physical examination findings into a clear case presentation, suggest reasonable working and differential diagnoses, and formulate a plan for further investigations (and initial management if appropriate)?
- *Humanistic Qualities / Professionalism:* Was the patient treated appropriately and with respect? Did the student communicate professionally and demonstrate an empathic approach to the patient?
- *Overall Clinical Competence:* A global assessment of the student's performance based on an integrated evaluation of the above competency domains.

Figure 10 – Current University of Auckland mini-CEX form.

This was introduced partway through 2014 and remains the current version.

Formative Mini-Clinical Evaluation Exercise (mini-CEX) Rating Form



The Royal Australasian
College of Physicians

Trainee information

Trainee's name : Date of assessment : ____ / ____ / ____
 Basic Training year : 1 2 3 Case number for that year : 1 2 3 4
(Full time equivalent)
 Assessor's name : Assessor's position :
 Assessor's email : Hospital/Location :
 Setting : In-patient Out-patient Emergency Other (please specify) :
 Patient problem/Dx(s) : Specialty :
 Patient age : Patient gender : Male Female Case complexity : Low Medium High

Strengths

Suggestions for development

If a trainee receives a rating which is unsatisfactory, the assessor must complete this section or the form will not be submitted.

Please rate the trainee against what you would expect of a trainee in that year of training

	Unsatisfactory			Satisfactory			Superior			Not observed
1. Medical interviewing skills	1	2	3	4	5	6	7	8	9	n/o
2. Physical examination skills	1	2	3	4	5	6	7	8	9	n/o
3. Professional qualities/communication	1	2	3	4	5	6	7	8	9	n/o
4. Counselling skills	1	2	3	4	5	6	7	8	9	n/o
5. Clinical judgement	1	2	3	4	5	6	7	8	9	n/o
6. Organisation/efficiency	1	2	3	4	5	6	7	8	9	n/o
Overall clinical performance	1	2	3	4	5	6	7	8	9	

Time taken for observation : min

Time taken for feedback : min

Assessor satisfaction using mini-CEX LOW 1 2 3 4 5 6 7 8 9 HIGH

Trainee satisfaction using mini-CEX LOW 1 2 3 4 5 6 7 8 9 HIGH

Data from formative assessments is collated for the purpose of evaluation. Individual, identifiable data will not be presented in any published reporting.

Assessor's signature : Trainee's signature :

Input validated by supervisor :
(Supervisor to initial once they have checked electronic record against this paper record)

Figure 11 - Page 1 of the Royal Australasian College of Physicians mini-CEX.



Copyright © Education Services, RACP

Figure 12 - page 2 of the Royal Australasian College of Physicians mini-CEX.

Figure



**6th Year Paediatrics, UOC
Mini CEX form**

STUDENT NAME:..... DATE OF ASSESSMENT

ASSESSOR NAME..... ASSESSOR'S POSITION

SETTING WARDS POPD CAA OTHER..... PATIENT PROBLEM

CLINICAL SKILLS OBSERVED

History taking Clinical examination of a child Clinical case presentation

Things you did well:	Something you can improve on

Please rate the student against what you would expect of a 6th year student (circle one value)

	Unsatisfactory			Satisfactory			Superior		
Quality of clinical skills	1	2	3	4	5	6	7	8	9
Interaction with Child / use of child-friendly skills	1	2	3	4	5	6	7 N/A	8	9

Examples:

1-3 would indicate inability to follow the basic steps of history taking or examination i.e. below what a competent 5th year student should do
4-6 indicates varying levels of ability to follow the basic steps of history taking or clinical examination. They need to demonstrate good interaction with children

7-9 indicates excellent approach to all aspects of the exam including interpretation of subtleties, and exceptional engagement with the child, (9 approaching a registrar level exam)

Student may add comments about the process or assessment here, if desired:

Assessor's signature..... Trainee's signature.....

Please upload this at the end of your Paediatric run

Department of Paediatrics
University of Otago, Christchurch, PO Box 4345, Christchurch 8140, New Zealand
Department Manager – Nicola_drake@otago.ac.nz Tē: +64 3 372 6713

Figure 13 - The University of Otago mini-CEX form.

**FORMATIVE CLINICAL EXAMINATION – STUDENT (CEX-S)
MEDI603 - GRADUATE MEDICINE**



1. STUDENT
NAME:

STUDENT
NUMBER:

2. LOCATION (GP PRACTICE/AGED
CARE/HOME VISIT/HOSPITAL):

3. PATIENT
PROBLEM/DIAGNOSIS:

4. MAJOR SKILL(S)
ASSESSED:

1)

2)

CHOOSE FROM THE FOLLOWING SKILLS LIST: HISTORY TAKING, EXAMINATION, COMMUNICATION,
PROFESSIONAL BEHAVIOURS, CLINICAL JUDGEMENT, AND PROCEDURAL SKILLS.
NB: THE ACTIVITY BEING ASSESSED MUST HAVE BEEN DIRECTLY OBSERVED BY THE ASSESSOR

5. STRENGTHS: WHAT DID THE STUDENT DO WELL?

6. IMPROVEMENTS: WHAT COULD THE STUDENT DO BETTER OR DIFFERENTLY TO IMPROVE?

7. ACTION PLAN: WHAT HAVE YOU AND THE STUDENT AGREED THEY SHOULD WORK ON TO
EFFECT IMPROVEMENTS?

8. TIME SPENT IN OBSERVATION AND
FEEDBACK: [] MINS

Figure 15 – The University of Wollongong’s formative mini-CEX.

Appendix 2 – University of Auckland Ethics Approval.

Research Office
Post-Award Support Services



The University of Auckland
Private Bag 92019
Auckland, New Zealand
Level 10, 49 Symonds Street
Telephone: 64 9 373 7599
Extension: 83711
Facsimile: 64 9 373 7432
ro-ethics@auckland.ac.nz

UNIVERSITY OF AUCKLAND HUMAN PARTICIPANTS ETHICS COMMITTEE (UAHPEC)

03-May-2018

MEMORANDUM TO:

Prof Jennifer Weller
Medicine

Re: Application for Ethics Approval (Our Ref. 020974): Approved

The Committee considered your application for ethics approval for your study entitled **Evaluating the Feasibility, Validity, Reliability and Relevance of the Mini-Clinical Evaluation Exercise for Fourth Year Medical Students at the University of Auckland.**

We are pleased to inform you that ethics approval has been granted for a period of three years.

The expiry date for this approval is 03-May-2021.

If the project changes significantly, you are required to submit a new application to UAHPEC for further consideration.

If you have obtained funding other than from UniServices, send a copy of this approval letter to the Activations team in the Research Office at ro-awards@auckland.ac.nz. For UniServices contracts, send a copy of the approval letter to the Contract Manager, UniServices.

The Chair and the members of UAHPEC would be happy to discuss general matters relating to ethics approvals. If you wish to do so, please contact the UAHPEC Ethics Administrators at ro-ethics@auckland.ac.nz in the first instance.

Please quote Protocol number **020974** on all communication with the UAHPEC regarding this application.

(This is a computer generated letter. No signature required.)

UAHPEC Administrators
University of Auckland Human Participants Ethics Committee

c.c. Head of Department / School, Medicine
Dr Janak de Zoysa
Prof Warwick Bagg
Dr Oleg Medvedev
Dr Elana Curtis

Additional information:

1. Do not forget to fill in the 'approval wording' on the Participant Information Sheets, Consent Forms and/or advertisements, giving the dates of approval and the reference number. This needs to be completed, before you use them or send them out to your participants.
2. At the end of three years, or if the study is completed before the expiry date, please advise the Ethics Administrators of its completion.
3. Should you require an extension or need to make any changes to the project, please complete the online Amendment Request form associated with this approval number giving full details along with revised documentation. If requested before the current approval expires, an extension may be granted for a further three years, after which a new application must be submitted.