The Royal Australasian College of Physicians

Basic Training Mini Clinical Evaluation Exercise

Exploratory Study Report

Updated 18 November 2015

Rebecca Paton
Research Officer
Education Policy, Research and Evaluation
Royal Australasian College of Physicians
Contact: evaluation@racp.edu.au

Reviewers
Libby Newton - Research Officer, Education Policy, Research and Evaluation
Rebecca Udemans - Senior Executive Officer, Education Policy, Research and Evaluation
Julie Gustavs - Manager, Education Development, Research and Evaluation

Table of contents

ACKNOWLEDGEMENTS

The author of this report wishes to acknowledge the support of various RACP staff, in particular, Libby Newton and Nathasha Kugenthiran for coding the patient problem/diagnosis field and providing feedback on the report. The author also wishes to thank members of the College Education Committee and College Research Committee for their support in facilitating this study.

EXECUTIVE SUMMARY

Aim

Since the introduction of the mini-CEX into RACP Basic Training in 2008, there have been approximately 39,500 cases submitted via the Basic Training Portal. An exploratory analysis of this data was performed in order to gain insights into:

- Trends and relationships between trainee and supervisor satisfaction with mini-CEX
- Time reportedly taken for observation and feedback
- The complexity of the case the mini-CEX focussed upon
- Frequency and nature of qualitative comments provided on the assessment form
- Changes over time in the ratings provided to trainees.

Method

Analysis of routine evaluation data collected for mini-CEX assessments in Basic Training, merged with existing data in the College database.

Key results

Sample Characteristics

- The sample included equal proportions of mini-CEX records for each year of training (first year, second year and third year) and calendar year (2010, 2011, 2012 and 2013) and a 75%:25% proportion of assessments taken from the Adult Medicine Division compared to the Paediatrics & Child Health Division.
- 79% of mini-CEX assessments were completed in an in-patient setting
- 20% of mini-CEX assessments were performed with patients presenting with conditions or symptoms relating to the cardiovascular system and 12% with patients presenting with conditions or symptoms relating to the respiratory system.
- Mean time taken to observe the mini-CEX assessment was 21 minutes and the mean time taken to provide feedback on the mini-CEX assessment was 12 minutes.
- The time taken to observe and provide feedback on the mini-CEX assessment increased with case complexity.
- Case complexity and setting of the mini-CEX varied by Division with Adult Medicine trainees performing a higher proportion of their mini-CEX assessments on highly complex cases and within in-patient settings.

- Case complexity and setting of the mini-CEX assessment varied by trainee year of training with more experienced trainees more often completing mini-CEX assessments on more highly complex cases and within in-patient settings.

Clinical Performance Feedback

- The highest clinical performance scores were received for professional qualities/communication and the lowest scores were received for physical examination skills.
- There was a relationship between case complexity and overall clinical performance, professional qualities/communication, clinical judgement and organisation/efficiency. The more complex the case, the higher the overall clinical performance score.
- There was a weak positive correlation between clinical performance scores and Clinical Examination Scores and a weak negative correlation between clinical performance scores and years post Fellowship of the mini-CEX assessor using the full dataset (no correlations were found using the sample dataset).
- The clinical performance scores showed very high internal consistency (Cronbach's $\alpha = 0.964$) and a factor analysis on these scores yielded only one factor.

Written Feedback

- Assessors provided at least one written comment (strengths and/or suggestions for development) in 92% of mini-CEX cases. Comments in both areas were provided in 73% of cases.
- There was a significant change in whether the assessors indicated suggestions for development over time, with the assessors being the most likely to provide suggestions for development in 2010 and less likely to provide comments with each subsequent year.
- Assessors were less likely to provide suggestions for development if they rated the trainee highly for overall clinical performance.
- The majority of comments provided in both open-ended response fields (strengths and suggestions for development) were of moderate quality or above.
- The majority of comments provided in terms of both trainee strengths and suggestions for development focused on medical expertise and communication.

Satisfaction with the Mini-CEX

- Both trainees and assessors were reasonably satisfied with using the mini-CEX (mean satisfaction scores 6.7 and 6.6 respectively, on a 9 point scale).
- There was a moderate positive correlation between assessor satisfaction with using the mini-CEX and trainee satisfaction with using the mini-CEX.
- Approximately 40% of the variance in trainee and assessor satisfaction could be explained by a combination of case complexity, setting, year of training of the trainee, time taken for observation, time taken for feedback and satisfaction (i.e. trainee satisfaction contributed to variation in assessor satisfaction and vice versa).

Recommendations

This evaluation has identified several areas for development. It is recommended that consideration be given to:

- Preparing a journal article documenting the results of this research.
- evaluating mini-CEX data on a regular basis to assess changes over time and the impact of any training or resources provided to supervisors/assessors.
- conducting similar investigations into other tools used in RACP training programs such as the Learning Needs Analysis (LNA) or Professional Qualities Reflection (PQR).
- conducting more in-depth qualitative research focusing on the mini-CEX assessment and the optimum conditions under which the mini-CEX is completed for trainees with differing levels of experience.
- Consider early interventions with trainees performing poorly on the mini-CEX to lower the risk of Clinical Examination failure for trainees.
- examining what training assessors currently receive in terms of completing the mini-CEX assessment and explore opportunities to provide supervisors/assessors with additional support or resources if necessary in order to ensure consistency in the way the mini-CEX assessment is undertaken and the quality of the comments provided on trainees' strengths and suggestions for their development.
- exploring ways to improve the quality of the assessors' written comments on mini-CEX forms.
- encouraging assessors to maintain a broader perspective when commenting on trainees' strengths and providing suggestions for development that go beyond medical expertise and communication skills.

- communicating to assessors the importance of providing written feedback to trainees on the mini-CEX forms and/or the provision of formal training on the use of the mini-CEX assessment.

1. INTRODUCTION

Developed in 1972 by the American Board of Internal Medicine, the Mini-Clinical Evaluation Exercise (mini-CEX) is a formative tool for assessing and providing immediate feedback on trainees' clinical skills. It involves an assessor engaging in a structured observation of how a trainee interacts with a real patient in a clinical setting. Areas for assessment include:

- Medical interviewing skills
- Physical examination skills
- Professional qualities/communication
- Counselling skills
- Clinical judgement
- Organisation/efficiency
- Overall clinical performance.

Using a structured proforma, assessors rate each of the areas of assessment on a nine point Likert scale with three anchor points: 1 (Unsatisfactory), 5 (Satisfactory) and 9 (Superior).

There is also an option to indicate that a particular area has not yet been observed. In addition to these ratings, the assessor provides the trainee with free-text feedback on their strengths and areas for improvement. An example RACP Basic Training mini-CEX rating form can be found in Appendix 2.

It is important that trainees undertake a number of mini-CEX throughout their training, that they are assessed by a range of different assessors in varied clinical settings, and that this assessment focuses on different clinical areas of the curriculum and various stages in the clinical process.

Since the introduction of this assessment tool into RACP Basic Training in 2008, there have been approximately 39,500 mini-CEX submitted via the Basic Training Portal. The College conducted an exploratory analysis of the data from these mini-CEX assessments in order to gain insight into:

- Trends and relationships between trainee and supervisor satisfaction with mini-CEX
- Time reportedly taken for observation and feedback
- The complexity of the case the mini-CEX focussed upon
- Frequency and nature of qualitative comments provided on the assessment form
- Changes over time in the ratings provided to trainees.

This information will help inform decision-making for future educational developments.

Specific research questions include:

Sample Characteristics

- How can we summarise the sample dataset?
- Under what circumstances is the mini-CEX more commonly used?
- How long does the assessment take to conduct?

Clinical Performance Scores

- What do the clinical performance scores look like?
- Have clinical performance scores changed over time? (group level)
- Do clinical performance scores relate to Clinical Examination scores and/or Clinical Examination outcome (pass/fail)?
- Do clinical performance scores relate to the years post fellowship of the assessor? (NB: not all assessors will be fellows)
- How reliable are the clinical performance scores? (group level)

Written Feedback

- What is the frequency, quality and content of written feedback on the forms?
- Has the provision, quality or content of written feedback changed over time?
- Does the provision, quality or content of written feedback relate to clinical performance?

Trainee and Assessor Satisfaction

- How satisfied are assessors with the tool? What factors relate to assessor satisfaction with the tool?
- How satisfied are trainees with the tool? What factors relate to trainee satisfaction with the tool?

2. METHOD

The Basic Training Mini-CEX Exploratory Study involved analysis of routine evaluation data collected for mini-CEX assessments in Basic Training, merged with some existing data in the College database.

Appendix 3 contains a series of research questions explored in the study along with the methodology that was used, potential implications for practice and references (where relevant).

2.1 Data Collection and Analysis

Data from a total of 39,489 mini-CEX assessments were extracted from the College training databases; this represented all mini-CEX assessments completed from January 2008 – June 2014. Records from calendar years 2008, 2009 and 2014 were excluded from this dataset, as were records from trainees beyond year three as there was not enough data to accurately represent these categories within a stratified random sample.[1] Any records which did not provide a Division or in which an inaccurate assessment date was given were also excluded. After this data cleaning process, 31,192 records remained.

A sample size calculation (with 95% confidence level (0.95 power), confidence interval ±5 and population size of 31,192) indicated a sample of 379 was required to perform the analyses (see calculator here: http://www.resolutionresearch.com/results-calculate.html).

A random sample (n=384) was extracted from the cleaned dataset. This random sample was stratified by year of training (1st year, 2nd year, 3rd year), calendar year (2010, 2011, 2012, 2013) and proportionately stratified by Division based on the total proportion of Basic Trainees in the College training database (75% Adult Medicine: 25% Paediatrics & Child Health) in 2013.

The sample was merged with available information from the College database on the year that assessors had become a Fellow of the College and Clinical Examination results for the first attempt taken by a trainee in the 2013 training year (where available). The sample was then analysed using SPSS version 22 and QSR NVivo 9. See Appendix 3 for the specific analyses used for each research question.

Two rubrics (see Appendix 4 and Appendix 5) were developed and used to assess the quality and content of written feedback on the mini-CEX forms; that is, strengths and suggestions for improvement. These rubrics were piloted with a sample of 21 assessments taken from the full dataset (see Appendix 6) and were found to provide a reasonable level of discrimination between quality and content.

The patient problem/diagnosis data was coded by two members of the Education Policy, Research and Evaluation team by grouping together patient problems occurring within the same body system or requiring similar treatment. Any discrepancies were resolved by the author.

[1] According to the 2009 Basic Trainee report, 38.6% of respondents did not yet use the mini-CEX tool in 2009. This decreased to 11.2% in 2010.

2.2 Ethics

Ethics review was required as there is the potential for articles using the study data to be submitted to relevant scholarly publications. De-identified and aggregated data will be used in articles and reports submitted for publication, as well as all internal reports based on the data.

The College Education Committee undertook an initial ethical review of the 2013 Basic Training Evaluation (which originally included the Basic Training Mini-CEX Evaluation) and considered it to be of Negligible Risk. The College Research Committee was asked to undertake an ethical review of the new proposal for the Basic Training Mini-CEX Exploratory Study and determined that there was no foreseeable risk of harm or discomfort associated with the proposal and since trainees and assessors have been notified of the potential for their data to be used for the purposes of evaluating the instrument (this is written on the rating form), that the study be considered to be of Negligible Risk. A guide to Ethical Assessment is included in Appendix 7.

3. RESULTS

Data from a random stratified sample of 384 mini-CEX assessments was analysed. The findings are reported below.

3.2 Sample Characteristics

3.2.1 Description of Sample Characteristics

The characteristics of the stratified random sample are outlined in Table 2 below.

As specified, the sample of data included equal proportions of mini-CEX records for each year of training and calendar year and a 75%:25% proportion of assessments taken from the Adult Medicine Division compared to the Paediatrics & Child Health Division (based on the number of trainees in each of these Divisions in the College training database in 2013).

Table 2: Characteristics of the Stratified Random Sample

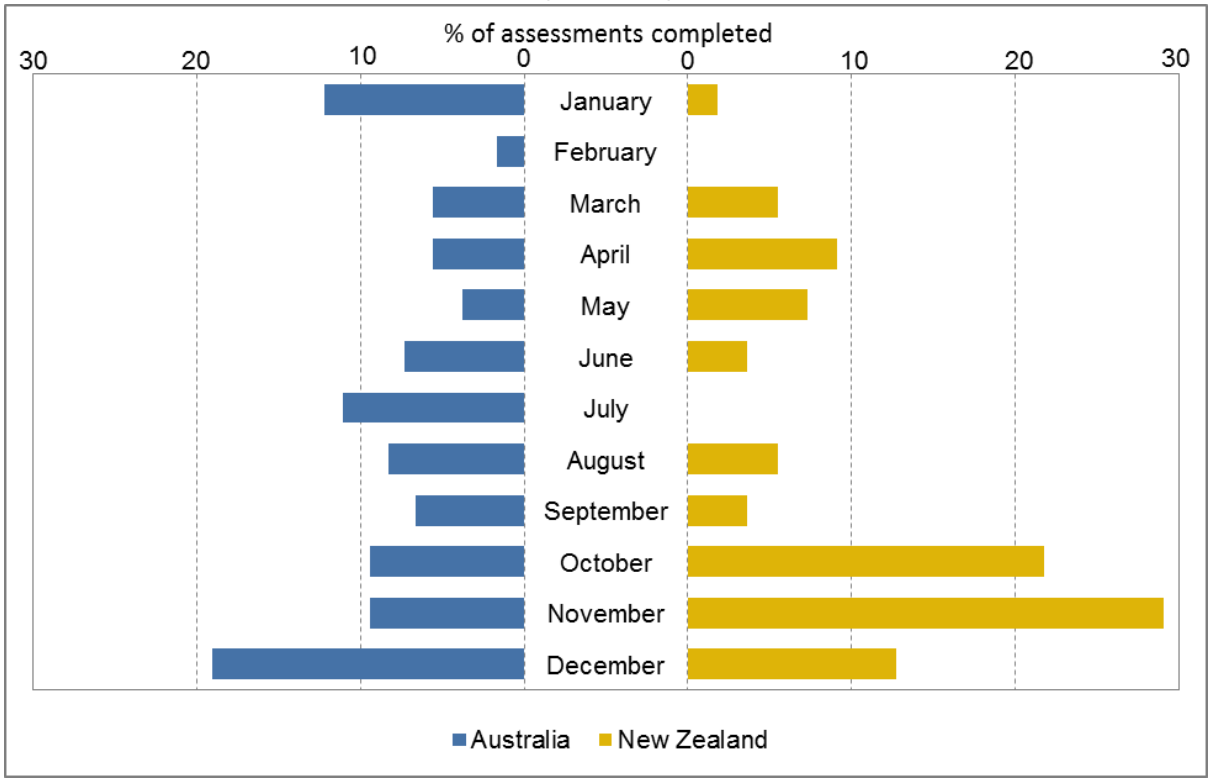
Variable	Level	n (valid %) Random sample (n=384)
Country	Australia	288 (84.0)
	New Zealand	55 (16.0)
Division	Adult Medicine	288 (75.0)
	Paediatrics & Child Health	96 (25.0)
Basic Training year	1st year	128 (33.3)
	2nd year	128 (33.3)
	3rd year	128 (33.3)
Calendar year assessment completed within	2010	96 (25.0)
	2011	96 (25.0)
	2012	96 (25.0)
	2013	96 (25.0)
Month assessment completed within	Jan	40 (10.4)
	Feb	6 (1.6)
	Mar	19 (4.9)
	Apr	23 (6.0)
	May	17 (4.4)
	Jun	27 (7.0)
	Jul	42 (10.9)
	Aug	30 (7.8)
	Sept	23 (6.0)
	Oct	41 (10.7)
	Nov	48 (12.5)
	Dec	68 (17.7)
Setting	In-patient	305 (79.4)
	Out-patient	44 (11.5)
	Emergency	28 (7.3)
	Other[1]	7 (1.8)
Case Complexity	Low	42 (10.9)
	Medium	245 (63.8)
	High	97 (25.3)
Clinical Examination Result	Pass	64 (81.0)
	Fail	15 (19.0)

1. *'Other' response category included: Mock exam, journal club, ALS training, new admission and clinical scenario.*

The majority of mini-CEX assessments took place in Australia (84%) rather than New Zealand (16%).

More than 40% of mini-CEX assessments were completed in the October-December quarter, with approximately 17% being completed in the January-February, March-June and July-September quarters each year. When broken down further, December (18%), November (13%), October (11%) and January (10%) were the most common times for the Mini-CEX to be completed, although this differed by country (see Figure 1).

Figure 1: Month Mini-CEX Completed by Country



Eight out of ten mini-CEX assessments were completed in an in-patient setting (79%) and most often with cases of 'medium' complexity (64%).

Mini-CEX assessments were most often performed with patients presenting with conditions or symptoms relating to the cardiovascular system (20%) followed by clinical conditions that were non-specific (14%), conditions or symptoms relating to the respiratory system (12%), or multiple conditions (11%) (see Figure 2).

Time Taken to Conduct the Mini-CEX Assessment

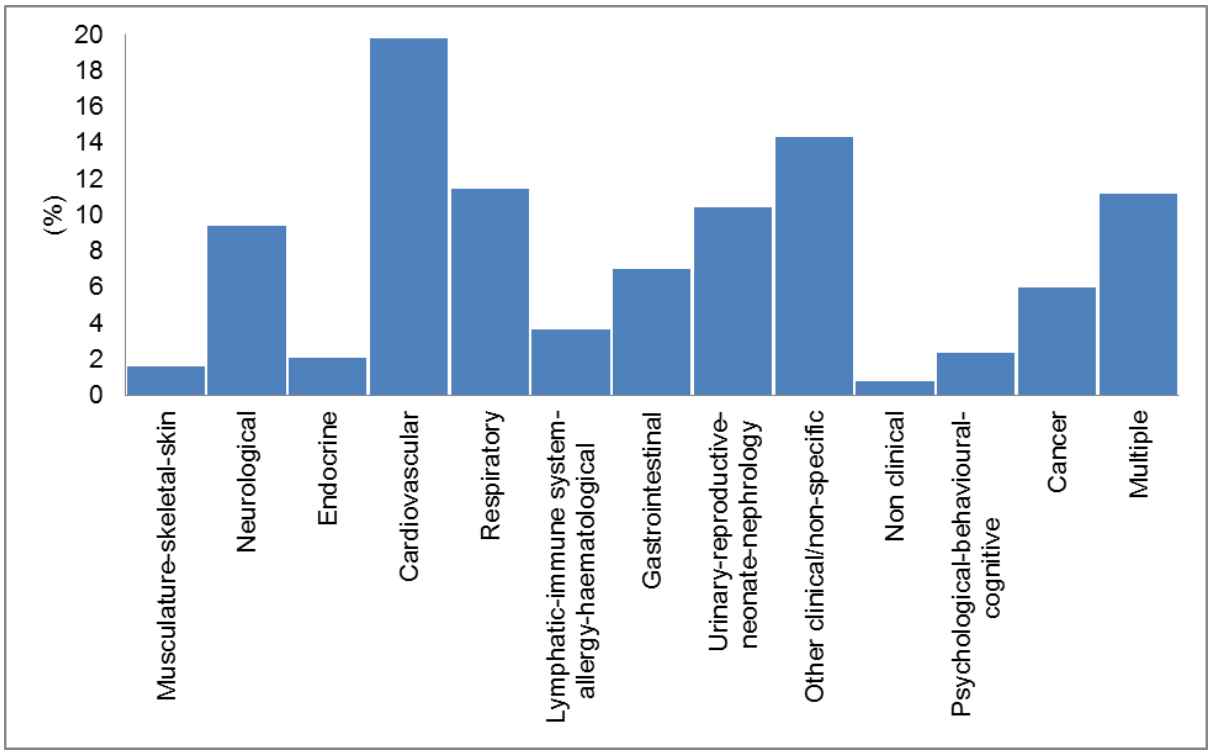
Observation time ranged from 5 minutes to 120 minutes. The mean observation time was 21.4 minutes and the standard deviation was 12.3 minutes.

Feedback time ranged from 5 minutes to 45 minutes. The mean time taken to give feedback was 12.0 minutes and the standard deviation was 6.1 minutes.

Years Post Fellowship of the Assessor

Approximately one third of the stratified random sample (123; 32.0%) had data specified that indicated how many years since the assessor was admitted to Fellowship. The mean number of years since admission to Fellowship for assessors of the mini-CEX was 10.9, with a standard deviation of 10.2.

Figure 2: Mini-CEX Patient Problem/Diagnoses



Note. 'Other clinical/ non-specific' category included things such as fever, falls, dizziness or abdominal pain. Non-clinical category included things such as advanced life support courses, review of a journal article or a guardianship hearing.

Clinical Examination Results

Just over a fifth of the stratified random sample (79; 20.6%) had Clinical Examination result data available. The mean Examination mark was 138.0, with a standard deviation of 16.8. Over 80% of cases with available Clinical Examination data had passed the Clinical Examination (see Table 2).

3.2.2 Relationships between Sample Characteristics

Case complexity and setting of the mini-CEX assessments varied by Division (χ^2=8.366, p=.0015 and χ^2=10.313, p=0.006 respectively). Specifically, Adult Medicine trainees performed a higher proportion of their mini-CEX assessments on highly complex cases compared to Paediatrics & Child Health trainees (see Table 3) and a higher proportion of Adult Medicine trainees performed mini-CEX assessments within in-patient settings compared to Paediatrics & Child Health trainees (see Table 4).

Table 3: Mini-CEX Case Complexity by Division

	Case Complexity			Total
	Low	Medium	High	
Adult Medicine	24 (8.3)	187 (64.9)	77 (26.7)	288 (100.0)
Paediatrics & Child Health	18 (18.8)	58 (60.4)	20 (20.8)	96 (100.0)
Total	42 (10.9)	245 (63.8)	97 (25.3)	384 (100.0)

Table 4: Mini-CEX Setting by Division

	Setting			Total
	In-patient	Out-patient	Emergency	
Adult Medicine	238 (84.7)	26 (9.3)	17 (6.0)	281 (100.0)
Paediatrics & Child Health	67 (69.8)	18 (18.8)	11 (11.5)	96 (100.0)
Total	305 (80.9)	44 (11.7)	28 (7.4)	377 (100.0)

Note. 'Other' category has been excluded from this analysis due to cell counts less than five.

Case complexity of the mini-CEX assessments was also affected by the trainees' experience (χ^2=14.697, p=0.005). Namely, first year trainees were more likely to do their mini-CEX assessments on low complexity cases compared to trainees more advanced in their training (see Table 5).

Table 5: Mini-CEX Case Complexity by Year of Training

	Case complexity			Total
	Low	Medium	High	
1st year	20 (15.6)	83 (64.8)	25 (19.5)	128 (100.0)
2nd year	12 (9.4)	90 (70.3)	26 (20.3)	128 (100.0)
3rd year	10 (7.8)	72 (56.3)	46 (35.9)	128 (100.0)
Total	42 (10.9)	245 (63.8)	97 (25.3)	384 (100.0)

A relationship was present between mini-CEX setting and year of training although this was not interpretable as cell counts were below five in one or more of the categories.

There was a weak to moderate positive correlation between time taken to give feedback and time taken for observation (Pearson's r = 0.307, p<0.0001).Typically, the longer it took to observe the mini-CEX assessments, the longer it took to provide feedback.

There were no significant relationships between observation time and setting or Division. However there were significant relationships between observation time and trainee year of training (F=3.076, p=0.047), calendar year (F=3.481, p=0.016) and case complexity (F=4.877, p=0.008). The time taken to observe the mini-CEX increased with trainee year of training and higher case complexity and decreased over time.

There were no significant relationships between feedback time and trainee year of training, setting or Division. However there were significant relationships between feedback time and calendar year (F=3.409, p=0.018) and between feedback time and case complexity (F=3.930, p=0.020). The time taken to provide feedback on the mini-CEX increased with higher case complexity and varied over time, taking the most amount of time in 2011 and the least amount of time in 2012.

3.3 Clinical Performance Feedback

As part of the mini-CEX assessment, assessors were asked to rate aspects of a trainees' clinical performance against what they would expect of a trainee in that year of training on a 9-point Likert scale ranging from unsatisfactory to superior. Assessors were also given the opportunity to indicate if they had not observed a trainee performing a certain skill. Responses are outlined in Table 6 and Figure 3 below.

3.3.1 Description of Clinical Performance Scores

Assessor Ratings of Trainee Clinical Performance are shown in Table 3. For each of the clinical performance areas, trainees were rated as 'superior' (7-9) at least 60% of the time. Professional qualities/communication skills received the highest percentage of scores in the 'superior' range (77.6%) of any of the clinical performance scores, followed by medical interviewing skills (73.0%). However, the areas with the highest means were found for professional qualities/communication skills and counselling skills (7.2 and 7.1 respectively). Standard deviation was very consistent across all of the clinical performance scores. Very few cases were rated as 'unsatisfactory' (1-3) for any of the clinical performance scores.

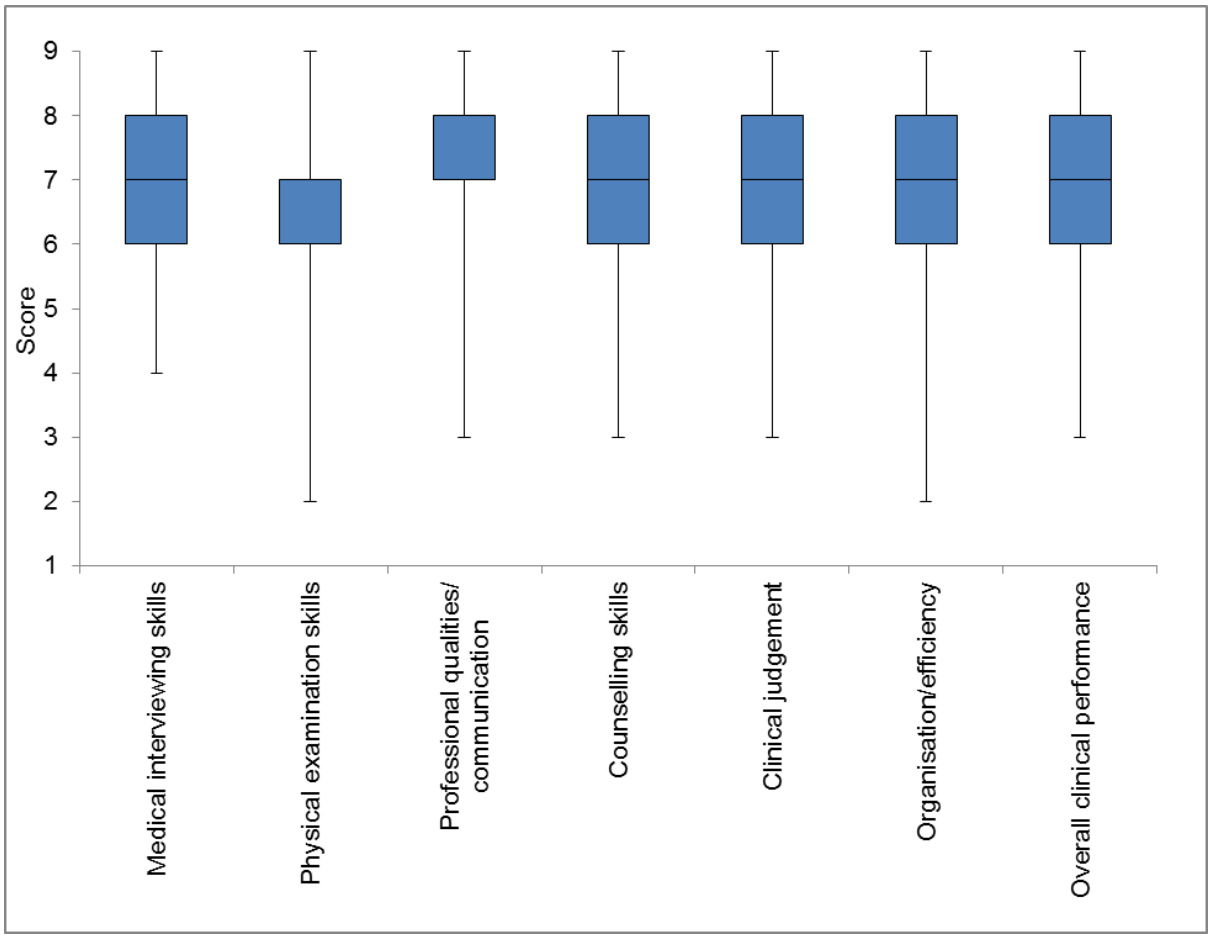
The clinical skills that were most frequently not observed during the mini-CEX interactions were counselling skills (not observed in 46.6% cases) and medical interviewing skills (not observed in 31.8% of cases). An overall clinical performance score was not provided in three cases.

Figure 3 indicates the distribution of scores for each of the clinical performance measures.

Table 6: Assessor Ratings of Trainee Clinical Performance

	Unsatisfactory			**Satisfactory**			**Superior**				
	1	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	Mean (SD)	Total observed
Medical interviewing skills				8 (3.1)	16 (6.1)	47 (17.9)	100 (38.2)	78 (29.8)	13 (5.0)	7.0 (1.2)	262
Physical examination skills		1 (0.3)	1 (0.3)	7 (2.4)	36 (12.1)	72 (24.2)	106 (35.7)	61 (20.5)	13 (4.4)	6.7 (1.2)	297
Professional qualities/ communication			1 (0.3)	12 (3.4)	15 (4.2)	52 (14.6)	136 (38.1)	97 (27.2)	44 (12.3)	7.2 (1.2)	357
Counselling skills			1 (0.5)	4 (2.0)	13 (6.3)	42 (20.5)	67 (32.7)	60 (29.3)	18 (8.8)	7.1 (1.2)	205
Clinical judgement			2 (0.5)	8 (2.1)	27 (7.0)	73 (19.0)	125 (32.6)	89 (23.2)	20 (5.2)	7.0 (1.2)	344
Organisation / efficiency		1 (0.3)	2 (0.6)	10 (2.9)	23 (6.7)	61 (17.7)	121 (35.2)	96 (27.9)	30 (8.7)	7.0 (1.2)	344
Overall clinical performance			2 (0.5)	10 (2.6)	29 (7.6)	74 (19.4)	143 (37.5)	103 (27.0)	20 (5.2)	6.9 (1.1)	381

Figure 3: Trainee Ratings for Measures of Clinical Performance



3.3.2 Relationships between Clinical Performance Scores and Sample Characteristics

No significant relationships were found between measures of clinical performance and:

- calendar year of assessment or calendar quarter of assessment
- year of training
- time lag between observation and completion of the mini-CEX form
- country
- Division
- clinical setting
- number of years since the assessor was admitted to FRACP
- clinical examination score
- clinical examination outcome (pass/fail)

No significant relationships were found between case complexity and medical interviewing skills, physical examination skills or counselling skills. However, significant relationships were found between case complexity and overall clinical performance (χ^2=3.820, p=.023), professional qualities/communication (χ^2=3.807, p=.023), clinical judgement (χ^2=5.624,

p=.004) and organisation/efficiency (χ^2=3.607, p=.028). In each of these cases, the more complex the case the higher the clinical performance score.

3.3.3 Extra analyses using the full dataset

Only a small subset of the data was used to explore the relationship between clinical performance scores and both the Clinical Examination and years post-Fellowship of the assessor data, with only 21% (n=79) of the sample having Clinical Examination data available and 32% (n=123) of the sample having years-post-Fellowship data specified. In order to address this issue, extra analyses were performed using the full cleaned dataset extracted from the Basic Training Portal, supplemented with data from the College database (N=39,484). This dataset included all mini-CEX assessments completed from January 2008 to June 2014.

Clinical Performance Scores and the Clinical Examination

Just over a fifth of the full dataset (8508; 21.5%) had Clinical Examination result data available, as in the original random sample. The mean Clinical Examination score was 134.9, with a standard deviation of 18.9. Almost 75% of cases with available Clinical Examination data had passed the Clinical Examination (6336; 74.47%).

There was a weak positive correlation found between all clinical performance scores on the mini-CEX assessment and Clinical Examination score using the full dataset (see Table 7). As clinical performance scores increased, Clinical Examination scores increased.

Table 7: Correlations between Measures of Clinical Performance in the Mini-CEX and Clinical Examination Score

	Correlation with Clinical Examination Score
Overall clinical performance	.173**
Medical interviewing skills	.187**
Physical examination skills	.145**
Professional qualities/ Communication	.184**
Counselling skills	.174**
Clinical judgement	.156**
Organisation/ Efficiency	.150**

**Correlation is significant at the 0.01 level

Multiple regression determined that a model containing all seven clinical performance scores on the mini-CEX was able to predict 4.1% of the variance in scores on the Clinical Examination *F(7) = 20.897, p<.0001, R^2 = 0.041* (see Appendix 8 for expanded results).

Binary logistic regression analysis revealed that a model including all clinical performance scores was able to predict whether the Clinical Examination was passed or failed. The overall model was significant at the 0.01 level according to the chi-square statistic and was able to predict 75.0% of the responses accurately (see Appendix 8 for expanded results).

Further analyses were performed using only one mini-CEX assessment per Basic Trainee, either the first or the last mini-CEX assessment (n=5525). The results were comparable to the full dataset including all mini-CEX assessments (see Appendix 8 for more information).

Clinical Performance Scores and Years Post Fellowship of the Assessor

Just over a third of the full dataset (11893; 30.1%) had data available that indicated how many years since the assessor was admitted to Fellowship. The mean number of years since admission to Fellowship for assessors of the mini-CEX was 11.5, with a standard deviation of 10.8.

There was a weak negative correlation found between all clinical performance scores on the mini-CEX assessment and years post Fellowship of the mini-CEX assessor (see Table 8). As years post Fellowship of the assessor increased, clinical performance scores decreased.

Table 8: Correlations between Measures of Clinical Performance in the Mini-CEX and Assessor Years post Fellowship

	Correlation with Years Post Fellowship of the Assessor
Overall clinical performance	-.155**
Medical interviewing skills	-.094**
Physical examination skills	-.163**
Professional qualities/ Communication	-.153**
Counselling skills	-.099**
Clinical judgement	-.133**
Organisation/ Efficiency	-.133**

**Correlation is significant at the 0.01 level

A significant relationship was also found between overall clinical performance score on the mini-CEX and years post Fellowship of the mini-CEX assessor when years post Fellowship of the assessor was treated as a categorical variable ($F(4)$=75.883, $p<0.0001$). Assessors who had become RACP Fellows within the last year were the most likely to give trainees higher scores for overall clinical performance (see Table 9).

Table 9: Mean overall clinical performance scores on the mini-CEX assessment by years post Fellowship of the mini-CEX assessor

Years post Fellowship of the assessor	Mean	Std. Deviation	N
0-1	7.13	1.018	1610
2-5	6.93	1.187	3258
6-10	6.75	1.382	2165
11-15	6.72	1.187	1425
16+	6.54	1.340	3435
Total	6.78	1.265	11893

Further analyses were performed using only one mini-CEX assessment per Basic Trainee, the first or the last mini-CEX assessment (n=5525). The results were comparable to the full dataset including all mini-CEX assessments (see Appendix 9).

3.4 Reliability of Clinical Performance Scores

Reliability tests were performed on the clinical performance measures. Reliability tests give an indication of the internal consistency of the survey items; that is, whether several items that propose to measure the same general construct produce similar scores. Reliability was measured with Cronbach's alpha (α), a statistic calculated from the pairwise correlations between items. Cronbach's α was 0.964 which indicates very high internal consistency between the clinical performance measures.

A Cronbach's α greater than 0.7 is generally considered acceptable (1). However, very high reliabilities (0.95 or higher) are not necessarily desirable, as this indicates that the items may not adequately distinguish between different types of clinical performance. The goal in designing a reliable instrument is for scores on similar items to be related (internally consistent), but for each to contribute some unique information as well. On further investigation, measures of clinical performance were found to be very strongly positively correlated (see Table 7).

Table 10: Correlations between Measures of Clinical Performance in the Mini-CEX

Clinical performance measures	Medical interviewing skills	Physical examination skills	Professional qualities/ communication	Counselling skills	Clinical judgement	Organisation/ efficiency	Overall clinical performance
Medical interviewing skills	-	0.810**	0.811**	0.791**	0.799**	0.719**	0.864**
Physical examination skills	0.810**	-	0.758**	0.756**	0.792**	0.756**	0.875**
Professional qualities/ communication	0.811**	0.758**	-	0.814**	0.785**	0.745**	0.839**
Counselling skills	0.791**	0.756**	0.814**	-	0.750**	0.735**	0.855**
Clinical judgement	0.799**	0.792**	0.785**	0.750**	-	0.783**	0.872**
Organisation/ efficiency	0.719**	0.756**	0.745**	0.735**	0.783**	-	0.868**

** Significant at p<0.05*
*** Significant at p<0.01*

Factor analysis is a statistical method used to describe variability between observed variables that are related to one another. A factor analysis was performed on the mini-CEX clinical performance scores as a data education exercise; to group variables into the lowest possible number of unobserved variables or 'factors'. It was revealed that the data yielded from the six clinical performance measures only yielded one factor (see Appendix 10 for full results).

3.5 Written Feedback on Trainees' Strengths and Suggestions for Development

Assessors were asked to provide comments on the strengths and areas for development for the trainee they were completing the mini-CEX assessment for.

3.5.1 Provision of Written Feedback

In 92.2% of cases, assessors provided at least one written comment regarding the trainee being assessed. Table 8 shows the frequency of comments provided.

Table 11: Frequency of Comments Provided on Mini-CEX

Number of comments provided	n (%)
No comments given	30 (7.8)
Strengths only	69 (18.0)
Suggestions for development only	5 (1.3)
Both comments given	280 (72.9)
Total	384 (100.0)

No significant relationship was found between the number of comments provided and calendar year.

There was a very slight negative correlation found between number of comments provided and overall clinical performance score (Pearson's r =-0.154, p=0.003). The more comments that were provided, the lower the overall clinical performance score tended to be.

Strengths

In 349 cases (90.9%) in which at least one comment was provided, assessors provided written feedback indicating the strengths demonstrated by the trainee in question.

The character count of written feedback pertaining to trainee strengths ranged from 0-263 with a mean of 79.7 and a standard deviation of 53.8.

There was no significant relationship found between calendar year of assessment and whether a comment was made on the strengths of a trainee.

Binary logistic regression analysis revealed that a model including all clinical performance scores was not able to predict whether a comment was made on the strengths of a trainee. There was also no significant relationship between overall clinical performance score and whether assessors commented on the strengths of a trainee.

Suggestions for Development

In 285 cases (74.2%) in which at least one comment was provided, assessors provided the trainee with a suggestion for development.

The character count of written feedback pertaining to trainee suggestions for development ranged from 0-270 with a mean of 50.91 and a standard deviation of 50.98.

A significant relationship was found between calendar year of assessment and whether the assessors indicated suggestions for development for the trainees (χ^2=10.493, p=0.015). Assessors were most likely to provide suggestions for development in 2010 and have become less likely to provide comments with each subsequent year (see Table 9).

Table 12: Calendar Year of Assessment by Provision of Written Feedback on Trainee Suggestions for Development

Suggestions for development provided	Calendar year of assessment				Total
	2010	2011	2012	2013	
No	13 (13.1)	27 (27.3)	28 (28.3)	31 (31.3)	99 (100.0)
Yes	83 (29.1)	69 (24.2)	68 (23.9)	65 (22.8)	285 (100.0)
Total	96 (25.0)	96 (25.0)	96 (25.0)	96 (25.0)	384 (100.0)

Binary logistic regression analysis revealed a model including all clinical performance scores was not able to predict whether suggestions for development were made in the assessment. However, there was a significant relationship found between overall clinical performance score and whether the assessor indicated a suggestion for development (χ^2=14.698, p=0.023). Assessors were less likely to provide suggestions for development if they gave the trainee in question a high overall clinical performance score (see Table 10).

Table 13: Overall Clinical Performance by Provision of Written Feedback on Trainee Suggestions for Development

Suggestion/s for development provided	Overall clinical performance									Total
	Unsatisfactory			Satisfactory			Superior			
	1	2	3	4	5	6	7	8	9	
No					4 (4.0)	13 (13.1)	42 (42.4)	31 (31.3)	9 (9.1)	99 (100.0)
Yes			2 (0.7)	10 (3.5)	25 (8.9)	61 (21.6)	101 (35.8)	72 (25.5)	11 (3.9)	282 (100.0)
Total			2 (0.5)	10 (2.6)	29 (7.6)	74 (19.4)	143 (37.5)	103 (27.0)	20 (5.2)	381 (100.0)

3.5.2 Quality of Written Feedback

Written feedback was coded using a quality rubric adapted from Haffling and Colleagues (2) (see Appendix 4). Poor quality comments were those that were very general or vague (e.g. 'systematic' 'more practice'). Moderate quality comments identified a specific area of strength or an area for development (e.g. 'Comprehensive history presentation' 'needs to improve clinical skills'). Good quality comments made mention of a particular context or targeted a specific behaviour to either continue or develop (e.g. 'Ongoing refining of technique. Need to practice fundoscopy. Work on interpreting findings as performing examination.')

The majority of comments provided in both open-ended response fields (trainee strengths and suggestions for development) were of moderate quality or above. Very few comments were of poor quality (2.3% of strengths comments and 7.5% of suggestions for development comments) (see Table 14).

Table 14: Quality of Written Comments Provided

Quality of written feedback	n (%)	
	Strengths	Suggestions for development
Poor quality	8 (2.3)	19 (7.5)
Moderate quality	195 (55.9)	113 (44.3)
Good quality	146 (41.8)	123 (48.2)
Total	349 (100.0)	255 (100.0)

Note. Written feedback expressing no real content (e.g. 'no concerns' or 'N/A') has been excluded from this data

There were no significant relationships found between the quality of the written feedback provided pertaining to a trainees' strengths or suggestions for development and calendar year of assessment, feedback time, years post Fellowship of the assessor, or any of the clinical scores.

3.5.3 Content of Written Feedback

Written feedback was coded using a content rubric adapted from the draft domain descriptions in the proposed RACP Standards Framework (3) (see Appendix 5). The majority of comments provided in terms of both trainee strengths and suggestions for development focussed on medical expertise and communication (see Table 12). No comments concerned teaching and learning or research.

Table 15: Content of Written Comments Provided

Content	n (%)	
	Strengths	Suggestions for development
Non-specific	13 (3.7)	21 (8.4)
Medical expertise	266 (76.0)	180 (72.3)
Communication	177 (50.6)	49 (19.7)
Quality and safety	3 (0.9)	
Teaching and learning		
Research		
Cultural competence	3 (0.9)	1 (0.4)
Ethics and professionalism	11 (3.1)	
Judgement and decision making	18 (5.1)	6 (2.4)
Leadership, management and teamwork	32 (9.1)	26 (10.4)
Total cases where written feedback was provided	350 (100.0)	249 (100.0)

Note. Content categories are not mutually exclusive. Percentages are calculated based on the total number of cases where written feedback was provided.

3.6 Satisfaction with the Mini-CEX Assessment Tool

Assessors were asked how long the mini-CEX assessment took to complete and both assessors and trainees were asked to indicate their level of satisfaction with using the mini-CEX.

3.6.1 Assessor Satisfaction

Assessor satisfaction with using the mini-CEX was rated on a nine-point Likert scale from 'Low' to 'High'. Scores ranged from 1 to 9. The mean score was 6.6 and the standard deviation was 1.4. In 60% of cases, assessor satisfaction with using the mini-CEX was at the high end of the scale (scores 7-9). In approximately 35% of cases, assessor satisfaction was rated average (scores 4-6). In approximately 2% of cases assessor satisfaction was rated low (scores 1-3).

There was a moderate positive correlation between assessor satisfaction with using the mini-CEX and trainee satisfaction with using the mini-CEX (Pearson's r =0.617, p<0.0001). As assessor satisfaction increased, so did trainee satisfaction.

Very weak positive correlations were found between assessor satisfaction and each of the clinical performance measures (see Table 13). As assessor satisfaction increased so did all of the clinical performance measures.

Table 16: Correlations between Assessor Satisfaction with using the Mini-CEX and Clinical Performance Scores

	Medical interviewing skills	Physical examination skills	Professional qualities/ communication	Counselling skills	Clinical judgement	Organisation/ efficiency	Overall clinical performance
Assessor satisfaction	0.179**	0.181**	0.165**	0.246**	0.201**	0.138*	0.186**

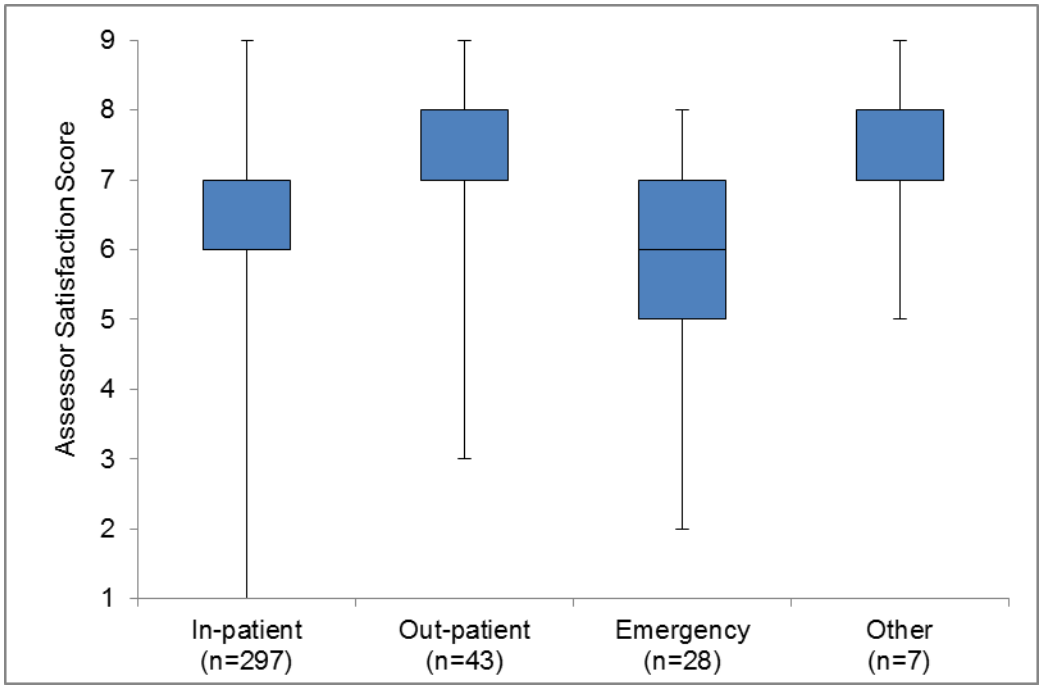
** Significant at p<0.05*

*** Significant at p<0.01*

There were no significant relationships found between assessor satisfaction with using the mini-CEX and the case complexity, year of training of the trainee, time taken for observation or time taken for feedback. However, there was a significant relationship found between assessor satisfaction with using the mini-CEX and the setting the mini-CEX took place within (i.e. in-patient, out-patient, emergency or other - see Figure 4). Assessors were least satisfied with the tool in emergency settings and most satisfied with the tool in 'other' setting such as mock exams, journal club, advanced life support training, new admission and clinical scenarios.

Linear regression analyses showed that approximately 40% of the variance in assessor satisfaction could be explained by a combination of trainee satisfaction, case complexity, setting, year of training of the trainee, time taken for observation and time taken for feedback (F= 37.777, p<0.0001).

Figure 4: Assessor Satisfaction with using the Mini-CEX within Different Clinical Settings



Note. 'Other' response category included: Mock exam, journal club, ALS training, new admission and clinical scenario.

3.6.2 Trainee Satisfaction

Trainee satisfaction with using the mini-CEX was rated on a nine-point Likert scale from 'Low' to 'High'. Scores ranged from 2 to 9. The mean score was 6.7 and the standard deviation was 1.3. In approximately 62% of cases, trainee satisfaction was rated at the high end of the scale. In over 36% of cases assessor satisfaction was rated average. Trainee satisfaction was rated low in less than 2% of cases.

Very weak positive correlations were found between trainee satisfaction and each of the clinical performance measures (see Table 14). As trainee satisfaction increased, so did all clinical performance measures.

Table 17: Correlations between Trainee Satisfaction with using the Mini-CEX and Clinical Performance Scores

	Medical interviewing skills	Physical examination skills	Professional qualities/ communication	Counselling skills	Clinical judgement	Organisation/ efficiency	Overall clinical performance
Trainee satisfaction	0.126**	0.164**	0.180**	0.254**	0.133*	0.111*	0.122*

** Significant at p<0.05*

*** Significant at p<0.01*

There were no significant relationships found between trainee satisfaction with using the mini-CEX and case complexity, setting, year of training of the trainee, time taken for observation or time taken for feedback. However, there was a significant relationship between trainee satisfaction with the mini-CEX and Division, with Paediatrics & Child Health trainees being more satisfied with the tool than Adult Medicine Trainees. There was also a moderate positive correlation found between trainee satisfaction and assessor satisfaction (Pearson's $r = 0.617$, $p<0.0001$). As trainee satisfaction increased, so did assessor satisfaction.

Linear regression analyses showed that approximately 40% of the variance in trainee satisfaction could be explained by a combination of assessor satisfaction, case complexity, setting, year of training of the trainee, time taken for observation and time taken for feedback ($F=35.699$, $p<0.0001$).

4. DISCUSSION AND CONCLUSION

Since the introduction of the assessment tool into RACP Basic Training in 2008, there have been approximately 39,500 mini-CEX cases submitted via the Basic Training Portal (up to 1 July 2014). The current exploratory study is the first evaluation of this data and the results will be able to be used as a baseline for future evaluations.

Overall, this exploratory study revealed some interesting new insights into the context in which the mini-CEX is typically completed, the clinical performance scores given by assessors, the written feedback pertaining to the strengths and suggestions for development for trainees outlined by the assessors and both trainee and assessor satisfaction with the mini-CEX tool. The findings from this study along with comparisons from the literature and recommendations for future developments are discussed below.

4.1 Sample Characteristics

Most mini-CEX assessments were completed in an in-patient setting (79%) and the cases they were performed on were most often of 'medium' complexity (64%) involving patients presenting with conditions or symptoms relating to the cardiovascular system followed by clinical conditions that were non-specific (14%), conditions or symptoms relating to the respiratory system (12%), or multiple conditions (11%).

There was evidence that the mini-CEX assessments were not evenly spaced throughout the training year, especially amongst New Zealand Basic Trainees. When split by country, there were more assessments completed in the months leading up to the end of the training year for each country (December-January for Australian Basic Trainees and October-November for New Zealand Basic Trainees). This finding should be interpreted with caution as the timing of the training year varies slightly between trainees and it is therefore unclear what proportion of assessments completed in the changeover period (approximately two weeks split across two months) were completed at the end of a training year or the start of the next training year. However, the result is consistent with feedback received in the 2013-14 Basic Trainee Survey (4) in which some Basic Trainees described the mini-CEX tool as a 'tick-box

exercise' to complete in order to meet program requirements rather than a useful learning tool.

A number of relationships were uncovered between the sample characteristics that shed more light on how the mini-CEX assessments are typically completed and the circumstances under which they are completed. Namely, Adult Medicine trainees performed a higher proportion of their mini-CEX assessments on highly complex cases compared to Paediatrics & Child Health trainees and a higher proportion of Adult Medicine trainees performed mini-CEX assessments in in-patient settings compared to Paediatrics & Child Health trainees. This may be a reflection of the different training requirements in each Division but it is recommended that these relationships be monitored in the future to establish whether there are changes over time.

The experience of the trainee was also found to influence the complexity of the case chosen for the mini-CEX assessments. This is not surprising given that first year trainees have not has as much exposure to the mini-CEX assessment as trainees beyond their first year and would also not be expected to have as much clinical knowledge or as highly developed skills/competence as trainees in their second or third year.

It took an average of 21 minutes for observation and 12 minutes to provide feedback to the trainee on their performance in the current study. Results for these measures are quite variable in the literature, with some studies reporting higher average times and others reporting lower ones (5-8).

As expected, there was a correlation between time taken to give feedback and time taken for observation. Typically, the longer it took to observe the mini-CEX assessments, the longer it took to provide feedback. There was also a clear link between both observation time and feedback time with case complexity which was also to be expected as there is more to be discussed in more highly complex cases and more that can potentially go wrong.

4.2 Clinical Performance Scores

As part of the mini-CEX assessment, assessors were asked to rate aspects of a trainees' clinical performance, namely their medical interviewing skills, physical examination skills, professional qualities/communication skills, counselling skills, clinical judgement, organisation/efficiency and overall clinical performance.

Average clinical performance scores for Basic Trainees ranged from 6.7 to 7.2 which are on the intersection of 'satisfactory' and 'superior' scores. While these means are reasonably high (given the possible range of 1-9), they are still somewhat lower than what has been reported within other contexts, particularly amongst medical students in the USA (5, 8). This suggests either that Basic Trainees in Australia may not be performing as well as expected by their assessors in comparison to medical students in the USA or that assessors in Australia are less lenient with their scoring.

Trainees received the highest clinical performance scores for professional qualities/communication skills and the lowest scores for physical examination skills. This is

the same result as previously found amongst internal medicine residents (9) and medical students (5) in the USA. A possible reason for this finding is that assessors are less lenient when scoring physical examination skills as these skills are less abstract and easier to measure accurately than other clinical skills such as professionalism.

Strong positive correlations were found between the seven different clinical performance scores on the mini-CEX forms, ranging from 0.72 to 0.81 excluding overall clinical performance and 0.72 to 0.89 including overall clinical performance. This indicates that the items are very closely interrelated, as found previously (5, 9-12), and may in fact be measuring a one-dimensional construct, rather than six separate clinical abilities and an overall score. This is problematic if the aim of the mini-CEX is to explore different aspects of clinical performance although it may be less of an issue if the primary purpose of the assessment is to prompt trainee-supervisor interaction and targeted feedback.

Predictors of Clinical Performance Scores

In the current study there was no relationship found between clinical performance scores and either trainee year of training or calendar year. This contrasts with results of previous studies that found that clinical performance scores were improved amongst more senior residents (13), or amongst residents over time (5, 9). This may be due to the fact that the RACP mini-CEX forms ask the assessor to rate the trainee against what they would expect of a trainee in that year of training rather than overall. It is unclear whether assessors in different contexts were asked to rate trainees/medical students in the same way.

The only significant predictor of clinical performance scores was case complexity. There were significant relationships between the complexity of mini-CEX cases and overall clinical performance, professional qualities/communication, clinical judgement, and organisation/efficiency. In each of these instances, more complex cases were given higher clinical rating scores. This perhaps points to a tendency for assessors to only give highly performing trainees complex cases for the mini-CEX assessment, while trainees who might not be performing so well are given more straightforward cases. This also fits with the earlier finding that more experienced trainees are more likely to be given more complex cases.

No significant relationships were found between clinical performance scores and years post Fellowship of the assessor using the random sample. However, using the full dataset, a weak negative correlation was found between all clinical performance scores on the mini-CEX assessment and years post Fellowship of the mini-CEX assessor. This suggests that the more experience assessors have, the more critical they get on their scoring. This may perhaps be due to the fact that they have simply assessed more trainees and have more realistic expectations of how trainees should perform at different stages of their training. Alternatively, it could mean that poorer performing trainees are more often matched with more experienced assessors.

Reliability and Validity of Clinical Performance Scores

The mini-CEX assessment has been implemented repeatedly with medical students, residents and Fellows across a range of specialties and its reliability and validity in a range of contexts has been documented. The current study adds to this body of literature.

Previous studies have found the inter-rater reliability of the mini-CEX assessment to be modest (5, 11, 14) but overall reliability to be quite high, with reliability coefficients ranging from 0.79-0.94 (5, 6, 15, 16). The reliability coefficient of the mini-CEX assessment in the current study was slightly higher than previously found, at 0.96. As stated earlier, this may indicate that the mini-CEX measures a single global dimension of clinical competence. If medical educators desire to measure discrete clinical skills, alternative assessment methods may be required.

Multiple sources of evidence for the construct and criterion validity for the mini-CEX have been identified in the literature. High correlations have been found between mini-CEX clinical performance scores and scores on other standardised academic and clinical performance measures or trainee achievements in the past (5, 6, 14, 16, 17).This indicates that the mini-CEX is an important instrument for the direct observation of trainees' clinical performance. It was therefore disappointing to find that no evidence of predictive validity was uncovered using the random sample in this study in terms of clinical examination score or outcome (pass/fail). It is likely that the small sample size is a contributing factor here, with only 21% of the sample having clinical examination data to draw on. The remainder of mini-CEX cases in the sample concerned trainees who had not yet attempted the clinical examination. It may also be worthy of note that there were 16 trainees who have two of their mini-CEX encounters included in the sample which may have slightly diluted the variability.

Validity of Clinical Performance Scores Using the Full Dataset

In order to address the power issues with the sample dataset described above, extra analyses were performed using the full cleaned dataset extracted from the Basic Training Portal, supplemented with data from the College database. These analyses showed a weak positive correlation between all clinical performance scores on the mini-CEX assessment and Clinical Examination score. This finding is consistent with previous literature documenting significant positive correlations between scores on the mini-CEX assessment and other standardised academic and clinical performance measures or trainee achievements (e.g. *(5, 6, 14, 16)*. It also indicates that early intervention with trainees that are performing poorly on mini-CEX may result in them having a better chance of passing the Clinical Examination. Together, all seven clinical performance scores were also able to accurately predict whether the Clinical Examination was passed or failed in 75% of cases. Admittedly, this is at least partially due to the positive skew and low variance in the clinical performance scores in addition to the high pass rate for the Clinical Examination. However, it may still be worth considering early interventions with trainees performing poorly on mini-CEX assessments to lower the risk of Clinical Examination failure for trainees.

4.3 Written Feedback

Provision of Written Feedback

It was encouraging to see that 92% of the mini-CEX assessments examined contained at least one written comment from the assessor and 73% contained two written comments. This rate of provision of written feedback is much higher than found for Australian medical students (18). Interestingly, of the 21% of assessments that contained only one comment, the majority only commented on the trainees' strengths (93%) rather than only providing suggestions for development (7%). This indicates that assessors are aware that a balance between positive and negative feedback is required in order to maintain trainees' motivation and build trainees' confidence.

There was no change over time in the number of comments provided by assessors or whether or not a comment was made on the strengths of a trainee on the mini-CEX assessments. However, there was a significant change in whether the assessors indicated suggestions for development over time, with the assessors being the most likely to provide suggestions for development in 2010 and less likely to provide comments with each subsequent year. This decline in the provision of suggestions for development may indicate a need for the College to communicate to assessors the importance of providing written feedback to trainees on the mini-CEX forms and/or the provision of formal training on the use of the mini-CEX assessment.

In terms of the relationship between written feedback and overall clinical performance score, there was a very slight negative correlation found between the number of comments provided and overall clinical performance score. The more comments were provided, the lower the overall clinical performance score. There was also a relationship found between overall clinical performance score and whether the assessor indicated a suggestion for development. Assessors were less likely to provide suggestions for development if they gave the trainee in question a high overall clinical performance score. However, there was no relationship found between overall clinical performance score and whether assessors commented on the strengths of a trainee. These findings support the earlier assumption that when delivering negative feedback, assessors tend to balance this with positive feedback.

Quality of Written Feedback

The majority of comments provided in both written feedback fields (trainee strengths and suggestions for development) were of moderate quality or above, in that they identified a specific area of strength or an area for development and/or made mention of a particular context or targeted a specific behaviour to either continue or develop. Very few comments were of poor quality, however, there may still be a need to improve the quality of the written comments, so that the majority fall into the 'good quality' category, given that written feedback has a significant part to play in helping students make meaning from their learning experiences over time (18).

There was no change in the quality of the written feedback over time and no relationship between the quality of written feedback and any of the clinical performance scores. There

was also no relationship between the quality of written feedback provided and the time taken to provide feedback on the assessment. The quality of the written feedback may therefore be more influenced by other factors such as the personal inclination of the assessor or the quality of the relationship between the trainee and the assessor.

Content of Written Feedback

The majority of comments provided in terms of both trainee strengths and suggestions for development focussed on medical expertise and communication which is not surprising given the nature of the assessment and the settings in which it is commonly used however, it is recommended that the College encourage the use of the mini-CEX in wider settings and that assessors keep a broad perspective when commenting on trainees' strengths and providing suggestions for development.

4.4 Satisfaction with the Mini-CEX

In almost 62% of cases, trainee satisfaction with the mini-CEX was rated at the high ('superior') end of the scale (scores 7-9). This is comparable to the finding from the 2010 RACP Second Year Basic Trainee survey which found that 67% of trainees were either 'satisfied' or 'very satisfied' with the mini-CEX (19). Trainee satisfaction with the mini-CEX was not explicitly asked about in the 2013-2014 Basic Trainee survey (4) and thus these results cannot be directly compared with the current study.

The mean trainee satisfaction score for the mini-CEX assessment was 6.7 and the standard deviation was 1.3. This mean score is slightly lower than what has been reported in other contexts, particularly amongst medical students in the USA (5, 7, 8).

Trainee satisfaction with the mini-CEX did not change over time, however, it did change with the experience of the trainee. The 2013-2014 RACP Basic Trainee report (4) indicated similar results, with trainee in their first year perceiving the min-CEX as the most useful and third years finding it the least useful.

Assessor satisfaction with the mini-CEX was slightly lower than trainee satisfaction, with 60% of assessors rating their satisfaction with the mini-CEX at the high ('superior') end of the scale (scores 7-9). There was a moderate positive correlation between assessor satisfaction and trainee satisfaction, confirming similar findings in the 2013-14 Basic Trainee survey (4). This may indicate that the assessors' perception of the mini-CEX assessment influences the trainees' perception and could suggest that any improvement in trainee satisfaction will need to start with the assessors/supervisors.

Also worthy of note is the relationship between assessor satisfaction with using the mini-CEX and the setting the mini-CEX took place within (i.e. in-patient, out-patient, emergency or other). Assessors were least satisfied with the tool in emergency settings and most satisfied with the tool in 'other' setting such as mock exams, journal club, ALS training, new admission and clinical scenarios. This interest in using the mini-CEX in non-clinical scenarios such as journal club is interesting given the design and purpose of the mini-CEX

assessment in physician training but this finding should be interpreted with caution given that the 'other' setting included considerably fewer cases than alternative settings (see Figure 4).

4.5 Recommendations for Future Research

This evaluation has identified a number of areas that may merit further exploration. Firstly it may be worth conducting more in-depth qualitative research focusing on the mini-CEX assessment and how it is undertaken and exploring the optimum conditions under which the mini-CEX is completed for trainees with differing levels of experience.

Secondly, it is recommended that the College examine what training assessors currently receive in terms of completing the mini-CEX assessment and explore opportunities to provide supervisors/assessors with additional support or resources if necessary in order to ensure consistency in the way the mini-CEX assessment is undertaken and the quality of the comments provided on trainees' strengths and suggestions for their development.

It is also recommended that the mini-CEX assessment be evaluated on a regular basis to assess changes over time and the impact of any training or resources provided to supervisors/assessors.

Finally, given the insight this exploratory analysis has offered into how the mini-CEX is completed and the conditions under which it is used, it may be useful to conduct similar investigations into other formative assessments used in RACP training programs such as the Learning Needs Analysis (LNA) or Personal Qualities Reflection (PQR).

4.6 Potential Limitations of this Evaluation

It is a limitation of the current study that measures of clinical performance were not standardised and therefore what qualified as 'unsatisfactory', 'satisfactory' or 'superior' performance was open to the interpretation of the assessor/supervisor performing the assessment. The mini-CEX rating form also specified that the assessor/supervisor should rate the trainee's performance based on what they would expect of a trainee in that year of training rather than over the course of training. This prevented any examination of changes over time in individual trainee clinical performance.

Another limitation was the exclusion of a large amount of data from the complete dataset prior to forming the random sample. In order to stratify the sample by training year of the trainee, calendar year and Division, mini-CEX cases from 2008, 2009 and 2014 had to be excluded due to an inadequate amount of cases in comparison to the remaining calendar years.

Finally, analyses involving clinical examination data and analyses involving years post-Fellowship of the assessor data had inadequate power as only 21%of the sample had clinical examination data available and 32% of the sample had years-post-Fellowship data specified. However, this limitation was addressed by performing additional analyses with the full dataset.

4.7 Conclusion

This exploratory study revealed some interesting new insights into the context in which the mini-CEX is typically completed, the clinical performance scores given by assessors, the written feedback pertaining to the strengths and suggestions for development for trainees outlined by the assessors and both trainee and assessor satisfaction with the mini-CEX tool. Of particular interest are the relationships between clinical performance scores and case complexity and the high internal consistency between clinical performance items, suggesting that the separate clinical performance scores may actually measure a unidimensional construct. The significant, if only small, positive correlation between clinical performance scores and the Clinical Examination is also worthy of note and suggests that early intervention with trainees that perform poorly on the mini-CEX may mean less Clinical Examination failures. Analysis of the written feedback also revealed that there was room for improvement in the quality and content of written feedback. While trainee and assessor satisfaction with the mini-CEX assessment were both reasonably high, the correlation between these scores indicates that any intervention or training in the completion of the mini-CEX assessment may be best directed at the assessor as the assessor's satisfaction may closely influence the trainee's satisfaction with the tool.

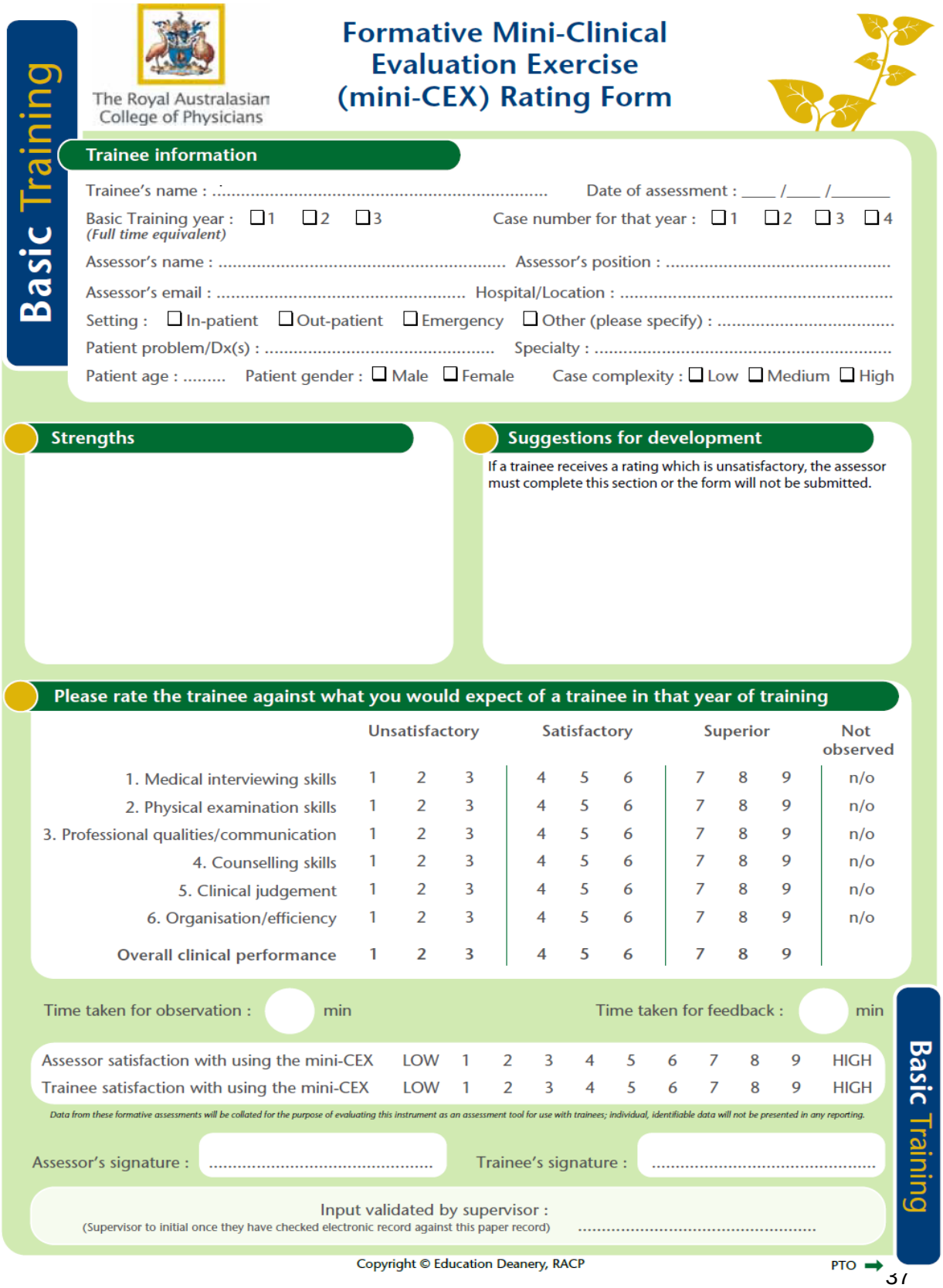
5. APPENDICES

Appendix 1: References

1. Tavacol MaD R. Making sense of Cronbach's Alpha. International Journal of Medical Education. 2011;2:53-5.
2. Haffling AC, Beckman A Fau - Edgren G, Edgren G. Structured feedback to undergraduate medical students: 3 years' experience of an assessment tool. Medical Teacher. 2011;33(1466-187X (Electronic)):e349-e57.
3. The Royal Australasian College of Physicians. RACP Standards Framework (proposed) - Domain Descriptions. 2013.
4. The Royal Australasian College of Physicians. 2013-2014 Basic Trainee Survey Evaluation Report. Sydney, Australia: The Royal Australasian College of Physicians,, 2014.
5. Kogan JR, Bellini LM, Shea JA. Feasibility, Reliability, and Validity of the Mini-Clinical Evaluation Exercise (mCEX) in a Medicine Core Clerkship. Academic Medicine. 2003;78(10):S33-S5.
6. Hatala R, Ainslie M, Kassen BO, Mackie I, Roberts JM. Assessing the mini-Clinical Evaluation Exercise in comparison to a national specialty examination. Medical Education. 2006;40(10):950-6.
7. Kogan JR, Bellini LM, Shea JA. Implementation of the Mini-CEX to Evaluate Medical Students' Clinical Skills. Academic Medicine. 2002;77(11):1156-7.
8. Hauer KE. Enhancing Feedback to Students Using the Mini-CEX (Clinical Evaluation Exercise). Academic Medicine. 2000;75(5):524.
9. Norcini JJ, Blank LL, Duffy FD, Fortna GS. The Mini-CEX: A Method for Assessing Clinical Skills. Annals of Internal Medicine. 2003;138(6):476-81.
10. Norcini JJ, Blank, L. L., Arnold, G. K., & Kimball, H. R. The mini-CEX (clinical evaluation exercise): A preliminary investigation. Annals of Internal Medicine. 1995;123:795-9.
11. Cook DA, & Beckman, T. J. 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,. 2009;14:655-64.
12. Margolis MJ, Clauser, B. E., Cuddy, M. M., Ciccone, A., Mee, J., Harik, P., et al. Use of the miniclinical evaluation exercise to rate examinee performance on a multiple-station clinical skills examination: A validity study. Academic Medicine. 2006;81((10 Suppl)):S56-S60.
13. Liao K-C, Pu S-J, Liu M-S, Yang C-W, Kuo H-P. 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. BMC Medical Education. 2013;13(1):31-.
14. Holmboe ES, Huot S Fau - Chung J, Chung J Fau - Norcini J, Norcini J Fau - Hawkins RE, Hawkins RE. Construct validity of the miniclinical evaluation exercise (miniCEX). Academic Medicine. 2003;78(1040-2446 (Print)):826-30.
15. Cook DA, Beckman Tj Fau - Mandrekar JN, Mandrekar Jn Fau - Pankratz VS, Pankratz VS. Internal structure of mini-CEX scores for internal medicine residents: factor analysis and generalizability. Advances in Health Sciences Education. 2010;15(1573-1677 (Electronic)):633-45.
16. Durning SJ, Cation Lj Fau - Markert RJ, Markert Rj Fau - Pangaro LN, Pangaro LN. Assessing the reliability and validity of the mini-clinical evaluation exercise for internal medicine residency training. Academic Medicine. 2002;77 (9)(1040-2446 (Print)):900-4.
17. Boulet JR, McKinley Dw Fau - Norcini JJ, Norcini Jj Fau - Whelan GP, Whelan GP. Assessing the comparability of standardized patient and physician evaluations of clinical skills. Adv Health Sci Educ Theory Pract. 2002;7(2):85-97.

18. Harvey P, Radomski N Fau - O'Connor D, O'Connor D. Written feedback and continuity of learning in a geographically distributed medical education program. Medical Teacher. 2013;35(1466-187X (Electronic)):1009-13.
19. The Royal Australasian College of Physicians. 2010 Survey of Second Year Basic Trainees Report. 2011.

Appendix 2: Basic Training Mini-CEX Rating Form



Basic Training

The Royal Australasian College of Physicians

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

Trainee information

Trainee's name : ..... Date of assessment : ____ /____ /_____

Basic Training year : ☐1 ☐2 ☐3 *(Full time equivalent)* Case number for that year : ☐1 ☐2 ☐3 ☐4

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

Data from these formative assessments will be collated for the purpose of evaluating this instrument as an assessment tool for use with trainees; individual, identifiable data will not be presented in any reporting.

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

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

Basic Training

PTO ➡

Ratings

Unsatisfactory - gaps in knowledge or skills that you would not expect at this level of training. Some concerns about professionalism or patient safety.

Satisfactory - what you would expect for a trainee at this level at this stage of their training year. Generally clinically competent and with satisfactory communication skills and professionalism.

Superior - performing well above the level they are at. No concerns about their clinical method, professionalism, organisation, communication etc.

The details below outline the skills associated with each domain in this mini-CEX rating form and the mini-CEX framework. Please note that not all skills may be examined during each encounter—this is a guide to show what may be observed and rated.

Medical Interviewing Skills

- Ability to interact with patient
- Ability to direct questions at key problem
- Ability to use second order of questioning to optimise focus
- Ability to incorporate information from questions with other information
- Ability to identify and respond appropriately to non-verbal cues
- Ability to retain a range of diagnostic options

Physical Examination Skills

- Ability to conduct a systematic and structured physical examination
- Shows sensitivity to patient's comfort and modesty
- Ability to detect abnormal signs when present and weigh the significance of these findings
- Informs patient
- Ability to focus the examination on the most important components
- Ability to integrate findings on examination with other information to clarify diagnosis

Professional Qualities / Communication

- Shows respect for patient at all times
- Explains as well as asks
- Listens as well as tells
- Conscious of potentially embarrassing or painful components of interaction
- Shows awareness of issues surrounding confidentiality
- Able to adapt questioning and examination to patient's responses

Counselling Skills

- Explains rationale for test/treatment
- Addresses the transfer of information in a way which is clear and tailored to the patient's needs
- Able to respond to patient and modify or repeat information in a different way
- Recognises patient's own wishes and gives them priority
- Avoids personal opinion and bias

Clinical Judgement

- Ability to weigh importance of potentially conflicting clinical data
- Ability to determine best choice of investigations and management
- Ability to relate management options to the patient's own wishes or situation
- Considers the risks and benefits of the chosen management/treatment options
- Ability to come to a firm decision based on available evidence

Organisation / Efficiency

- Ability to synthesise a collection of data quickly and efficiently
- Demonstrates appropriate judgement and synthesis
- Demonstrates optimal use of time in collection of clinical and investigational data

Appendix 3: Research Questions

Question	Methodology	Implications	References
Clinical Performance Scores			
1. How can we summarise the sample dataset? Under what circumstances is the mini-CEX more commonly used? What do the clinical performance scores look like?	Run basic descriptive statistics on key variables including setting, case complexity, Division and clinical performance scores.	Able to give summaries of the sample and the observations that have been made, especially the most common scenarios in which the tool is used. Action may need to be taken to encourage use of the tool in some settings.	None
2. Have clinical performance scores changed over time? (group level)	Analysis of variance (ANOVA) with calendar year as the independent variable (IV) and overall clinical performance as the dependent variable (DV).	May indicate a change in the way the mini-CEX is used over time (for better or worse).	None
3. How reliable is the mini-CEX as an assessment? (group level)	Internal consistency exploration (Cronbach's Alpha- reliability analysis on clinical performance scores).	Action may need to be taken to improve the assessment if reliability is found to be low.	Reliability of the mini-CEX as an assessment tool has been tested previously (13, 15)
Written feedback on trainees' strengths and suggestions for development			
4(a) What is the frequency of written feedback on the forms?	(a) Count the number of written comments provided for each mini-CEX in the sample (score range 0-2).	Action may need to be taken to improve the frequency, quality and/or content of the comments provided by assessors if any are found to be unsatisfactory. These analyses may indicate the quality of the relationship between assessors and trainees.	Quality rubric adapted from Haffling and colleagues (2) Content rubric adapted from Draft RACP Standards Framework – domain descriptions (3)
4(b) What is the quality of written feedback on the forms?	(b) Rate the quality of the two separate comments (strengths and suggestions for improvement) based on a rubric and run basic descriptive statistics on the findings.		

Question	Methodology	Implications	References
4(c) What is the content of the written feedback on the forms?	(c) Code the content of the two separate comments (strengths and suggestions for improvement) based on a rubric and run basic descriptive statistics on the findings.		
5(a) Has the provision of written feedback changed over time? 5(b) Has the quality of written feedback changed over time?	(a) Binary logistic regression with calendar year as the IV and the provision of written feedback (yes/no for strengths and suggestions for development separately) as DVs. (b) Chi-square tests with calendar year as the IV and quality category for each comment (strengths and suggestions for development separately) as DVs.	May indicate a change in the way qualitative questions are answered over time and potential need for an intervention.	Quality rubric adapted from Haffling and colleagues (2) Content rubric adapted from Proposed RACP Standards Framework –domain descriptions (3)
6(a) Does the provision of written feedback relate to clinical performance? 6(b) Does the quality of the written feedback relate to clinical performance?	(a) Bivariate logistic regression with provision of qualitative comments (yes/no for strengths and suggestions for development separately) as the IV and overall clinical performance score as the DV. (b) ANOVAs with quality category for each comment (strengths and suggestions for improvement separately) as IVs and overall clinical performance score as the DV.	Would indicate whether an intervention targeting the quality, frequency or content of qualitative comments would also result in an improvement in quantitative scores on the forms.	Quality rubric adapted from Haffling and colleagues (2) Content rubric adapted from Proposed RACP Standards Framework –domain descriptions (3)

Question	Methodology	Implications	References
Satisfaction with the Mini-CEX assessment tool			
7. How long does the assessment take to conduct?	Basic descriptive statistics on observation time and feedback time (mean, range and standard deviation)	Able to give summaries of observations that have been made.	None
8(a) How satisfied are assessors with the tool? 8(b) How satisfied are trainees with the tool?	(a) Basic descriptive statistics on assessor satisfaction (mean and standard deviation) (b) Basic descriptive statistics on trainee satisfaction (mean rating and standard deviation)	Able to give summaries of observations that have been made.	None
9(a) What factors relate to assessor satisfaction with the tool? 9(b) What factors relate to trainee satisfaction with the tool?	(a) ANOVAs/ correlations/ multiple regression with assessor satisfaction as the DV and the following IVs: trainee satisfaction, overall clinical performance score, case complexity, setting, trainee stage of training, time taken for observation and time taken for feedback (b) ANOVAs /correlations /multiple regression with trainee satisfaction as the DV and the following IVs: assessor satisfaction, overall clinical performance score, case complexity, setting, trainee stage of training, time taken for observation and time taken for feedback	May help predict how likely assessors are to complete it properly and how accurately the trainees consider themselves to be rated.	None

Question	Methodology	Implications	References
10(a) Do clinical performance scores relate to Clinical Examination scores? 10(b) Do clinical performance scores relate to Clinical Examination outcome?	(a) Correlations between overall clinical performance (IV) and clinical examination score (DV) (b) Binary logistic regression with clinical examination outcome (pass/fail) as the DV and overall clinical performance scores as the IVs.	Predictive validity	None
11. Do clinical performance scores relate to the years post fellowship of the assessor? (NB: not all assessors will be fellows)	Correlation between experience of assessor (in years since 2014) and overall clinical performance score	May suggest a need for standardised training for assessors in administering the tool.	None

Appendix 4: Quality Analysis Rubric

	Strengths		Suggestions for development	
Quality category	**Criteria**	**Examples**	**Criteria**	**Examples**
No written feedback (N)	No written feedback	*"None provided"*	No written feedback	*"None"*
Poor quality (P)	General, non-specific comment	*"Systematic"* *"Comprehensive"* *"Got all the major points"*	General, non-specific comment	*"More practice"* *"Keep enquiring"* *"Maintain your current high standard"*
Moderate quality (M)	Identifies an area/skill that is particularly strong	*"Comprehensive history presentation"* *"Communication with child and family"*	Targets an area/skill in need of development	*"Needs to improve clinical examination skills"* *"Could improve a bit on examination technique"*
Good quality (G)	Identifies how the trainee may have acquired a specific strength. Mention of context.	*"Presented case clearly. Had a good understanding of the issues related to chronic lung disease of prematurity"* *"Saw patient on ward round - reviewed history and came up with a different diagnosis (the correct one) than CED admission diagnosis. Good examination - Explained really well to the patient's family . Mother called her 'awesome'"*	Targets behaviour to either continue or develop to meet a need. Mention of specific actions to improve.	*"Practise RACP style clinical exam"* *"Practice concise presentations with a more definitive style"*

This rubric has been adapted from Haffling et al 2011[3]
Categories are mutually exclusive.

Appendix 5: Content Analysis Rubric

Category	Criteria	Examples - Strengths	Examples - Suggestions for development
No suggestions (N)	No written feedback at all or no strengths/suggestions for improvement listed	*"None provided"*	*"None"* *"Maintain your current high standard"* *"Continue"*
Non Specific (NS)	Non-specific strengths **OR** Non-specific suggestions for improvement including: - At appropriate level - Further training/experience (non-specific) - More practice needed (non-specific)	*"Good approach"* *"Comprehensive"* *"Very competent and capable"*	*"Further experience"* *"Continue to develop clinical knowledge"* *"Ongoing development of skills"*
Medical expertise (ME)	- Diagnostic skills - Selecting appropriate investigations - Interpreting results of investigations - Making decisions about patient care and management - Prioritising patient care - Managing complex or unexpected clinical situations - Managing long term conditions	*"Good knowledge base"* *"Good preparation of short and long term issues"* *"Comprehensive history and examination "*	*"Work on diagnostic synthesis and approach to a 'short case'"* *"Check list thinking rather than diagnostic thinking"* *"Refining a methodical approach to the complex patient"*

Communication (C)	- Apply communication skills to engage and reassure the patient in first encounters, when taking a history, when providing counselling and when breaking bad news - Empower patients and be respectful of their rights in all aspects of communication - Apply communication skills in encounters with a patient's family (including extended family) and/or carers - Communicate effectively with referring doctors and when referring a patient to another specialist - Facilitate effective clinical handover and transfer of care - Communicate effectively with health administration - Communicate effectively with support organisations, administrative bodies, governments and others in the wider community - Apply specific medico-legal communication practices	*"Communication"* *"Good communication and explanation to patient re test"* *"Caring sensitive approach to patient and family"*	*"Consider 'aim' of written communication"* *"Clearer communication"* *"Improve handovers "*
Quality and safety (QS)	- Use evidence to inform quality improvement - Safe work practices - Safe prescribing and administration of medication - Safe continuity of care for patients - Recognise, report on, and manage adverse events and error - Comply with relevant risk-management/minimisation procedures - Follow processes for managing patient complaint - Use patient complaints to enhance medical care	*"Good discussion regarding safety"* *"Good risk assessment"* *"Take trouble over patient safety"*	*"Consider safety first when dealing with complex patient history and presentation"* *"Outline of critical safety issues"* *"Become more involved in audit and quality assurance"*

Teaching and learning (TL)	- Participate in educational and CPD - Recognise the importance of health education and the role of the physician as a teacher to patients, other physicians and in the wider community	*"Good teaching skills"* *"Actively involved in teaching"* *"Some education around the disease process"*	*"Teaching role"* *"To be more involved in teaching"* *"He does not allow himself to attend enough teaching sessions"*
Research (R)	- Contribute to the development of new knowledge by active involvement in research - Adhere to the principles of evidence-based medicine in daily clinical practice - Present research findings	*"Well researched talk (Presentation on tumour lysis syndrome)"* *"Good research into the underlying aetiology and where the evidence lies"* *"Involvement in medical research"*	*"More research around the subject matter"* *Continue to understand research methodology and limitations of published research, clinical guidelines"* *"Research"*
Cultural competence (CC)	- Communicate effectively with people from diverse backgrounds - Apply knowledge of the patient's cultural and religious background, attitudes and beliefs in managing and treating that patient - Recognise how the special history of Māori & Pacific peoples (NZ) and Aboriginal & Torres Strait Islander peoples (Australia) impacts on their current health status	*"Judicious about management of the elderly patient in a broader cultural context"* *"Culturally appropriate approach"* *"Observes the cultural boundaries"*	*"Practice cross-cultural communication"* *"Appreciation of cultural differences to the perception of death"* *"Think about holistic assessment and cultural impact on disease management"*
Ethics and professionalism (EP)	- Practice within ethical frameworks - Ethical research - Sound professional and personal conduct - Reflect on personal beliefs, biases and behaviours - Apply legal and ethical frameworks to physician–patient relationships	*"Good work ethics"* *"Ethical decision making"* *"Ethical"*	*"Probably needs a bit more emphasis on ethics"* *"Ethical issues regarding euthanasia and psychiatric presentation"* *"Ongoing development of ethical decision making"*

Judgement and decision making (JDM)	- Diagnostic reasoning - Risk prediction - Maximum patient benefit and acceptance - Use evidence to inform decision making	*"Good clinical decision making"* *"Sensible decision making"* *"Good clinical judgement"*	*"Needs to develop further skills in clinical judgement"* *"Independent decision making"* *"Clinical Judgement"*
Leadership, management and teamwork (LMT)	- Self-management - Multidisciplinary teamwork - Leadership and management	*"Good organisation skills and planning"* *"Provides leadership"* *"Intelligent at management"*	*"Confidence regarding management"* *"Working on time management"* *"Leadership experience"*

Adapted from Proposed RACP Standards Framework – draft domain descriptions[4].
Categories are NOT mutually exclusive. A single comment can be coded multiple times but only once for each category but the 'non-specific' category should only be used when a comment has no other codes.

Appendix 6: Example Coding Using the Quality and Content Rubrics (n=21)

Case no.	Strengths	Strengths Quality	Strengths Content	Areas for Development	Develop-ment Quality	Developm-ent Content
1	history taken Procedure skills decision making management skills	M	C ME JDM LMT	communication skills	M	C
2	quite through reasonably systematic good baseline medical knowledge	M	ME	allow for more open question synthesis could be more refined communicate more clearly	M	C JDM
3	Good interpersonal style and communication technique. Examination skills are well developed for stage. Good intelligent discussion.	M	C ME	Ongoing evaluation and review of clinical scenarios. Practice and revision as expected for level of training.	G	ME
4	In depth understanding of case ,excellent presentation with succinct summary of case and differential diagnosis	G	ME JDM	NIL	N	N
5	- Focused interview in context of complex history. - Overall comprehensive examination	G	C ME	- Continue to practise short case examinations, including "failure to thrive". - Further reading about GORD investigations and treatment.	G	ME
6	methodical history confirmed info good rapport with parents good clinical judgement and management plan	M	C JDM	none identified	N	N
7	Empathetic, calm and great communication. Looks caring	M	C	speak slightly slower, think a bit harder when using medical terms to mother	G	C
8	Thorough Organised Good theoretical knowledge	M	ME		N	N
9	Good communication skills in a difficult situation Good bedside manner	G	C	Communication style could be more concise	M	C
10	full and detailed history aware of relevant co-morbidities Appropriate examinations and investigations	G	C ME		N	N

11	Easy rapport with parent. Gentle examination of young child, not causing distress. Covered all important areas.	G	C	Asked to do Cardiovascular Examination. Remember to remove clothing. Some more knowledge of physical organs.	G	ME
12	Good approach to patient Enthusiasm Eagerness to learn	M	C TL	Structuring presentation in easily understandable blocks	G	C
13	Very calming reassuring manner Good clinical skills - able to anticipate clinical problems Good clinical summary at end of examination	G	C ME	Just a function of more experience in conjunction with her abilities	P	NS
14	Handled a difficult patient very well, clear management plan	G	C JDM		N	N
15	Good rapport with mother Detailed general history	G	C	Structured approach for examination Not to repeat same question asked before	G	ME C
16	Very good history taking. Excellent documentation. Excellent medical knowledge.	M	C QS ME	Needs to be more confident	M	LMT
17	Comprehensive history taking Good time management Keen learner & good team member Fluent in presentation and examination	M	C LMT	-	N	N
18	Good approach to patient Thorough history Exam included all elements of upper limb neuro exam	G	C ME	Further practice in obtaining history - more important questions to be asked earlier in history	G	C
19	Good history taking from patient and family Physical examination good Identify issues satisfactory	M	C ME	Precise geriatric issues will improve with time and experience	P	NS
20	Jason has good communication skills, is hardworking, punctual and reliable. His is a good trainee.	M	C LMT	Continued reading and study for exams will broaden his knowledge base.	M	ME TL
21	Information gathering through verbal and non-verbal developed skills Pain assessment skills developing appropriately High degree of empathy, enabling rapport	M	C ME	Framework for information gathering strengthening Decision making around how much detail to cover at initial assessment Increased awareness of non-medical cues	G	C JDM

TOTAL		N= 0 P= 0 M=11 G=10	N=0 NS=0 ME=12 C=18 QS=1 TL=1 R=0 CC=0 EP=0 JDM=4 LMT=3		N=6 P=2 M=5 G=8	N=6 NS=2 ME=5 C=8 QS=0 TL=1 R=0 CC=0 EP=0 JDM=2 LMT=1

Appendix 7: Guide to Ethical Assessment of the BT Mini-CEX Exploratory Study

Steps for assessment of ethical acceptability of risks	Comment	Assessment
(a) identifying the risks, if any	Nil identified as trainees and assessors of the mini-CEX have been notified of the potential for the data collected on the rating forms to be collated for the purpose of evaluating the instrument. Individual, identifiable data will not be presented in any reporting.	Negligible risk
b) assessing the likelihood and severity of the risks;	N/A	N/A
(c) identifying whom (participants and/or others) the risks may affect	N/A	N/A
(d) establishing the means for minimising the risks;	N/A	N/A
(e) identifying the potential benefits;	This study will inform: • development of an evidence base for future educational developments • ways to improve Basic Training mini-CEX • content for supervisor training	N/A
(f) identifying to whom benefits are likely to accrue.	The outcomes of this project will benefit Basic Trainees, assessors, College staff, College committees making educational decisions and the broader community.	N/A

Appendix 8: Expanded results - Validity of clinical performance Scores in predicting Clinical Examination Scores Using the Full Dataset

1. All mini-CEX cases in the full dataset

Analyses were performed using the full cleaned dataset extracted from the Basic Training Portal, supplemented with data from the College database (N=39484). This dataset includes all mini-CEX assessments completed from January 2008 to June 2014. Just over a fifth of the full cleaned dataset (8508; 21.5%) had Clinical Examination result data available. Clinical Examination results included were those for the first attempt taken by a trainee in the 2013 training year. The mean Clinical Examination score was 134.9, with a standard deviation of 18.9. Almost 75% of cases with available Clinical Examination data had passed the Clinical Examination (6336; 74.47%).

There was a weak positive correlation found between all clinical performance scores on the mini-CEX assessment and Clinical Examination score (see Table 1). As clinical performance scores increased, Clinical Examination scores increased.

Table 1: Correlations between Measures of Clinical Performance in the Mini-CEX and Clinical Examination Score

	Correlation with Clinical Examination Score
Overall clinical performance	.173**
Medical interviewing skills	.187**
Physical examination skills	.145**
Professional qualities/ Communication	.184**
Counselling skills	.174**
Clinical judgement	.156**
Organisation/ Efficiency	.150**

**Correlation is significant at the 0.01 level

Multiple regression determined that a model containing all seven clinical performance scores on the mini-CEX was able to predict 4.1% of the variance in scores on the Clinical Examination $F(7) = 20.897$, $p<.0001$, $R^2 = 0.041$. The Professional qualities/Communication score within this model had a *t* statistic equal to 3.337 and was significant at the 0.01 level. This means that for every one unit increase in the Professional qualities/ Communication skills score on the mini-CEX assessment, there was a 1.968 increase in the Clinical Examination score (all else remaining constant). The Physical Examination Skills score had a *t* statistic equal to -2.610 and was also significant at the 0.01 level. This means that that for every one unit increase in the Professional qualities/ Communication skills score on the mini-CEX assessment, there was a 1.968 decrease in the Clinical Examination score. None of the other variables in the model were significant at the 0.01 level (see Table 2). This indicates that while clinical performance scores on the mini-CEX assessment can explain some of the variance in Clinical Examination results, there may be more influential factors at play that have not been accounted for by this model.

Table 2: Regression results for a model including all clinical performance scores on the mini-CEX

	B	S.E.	Beta	t	Sig.
Medical interviewing skills	1.079	.653	.063	1.654	.098
Physical examination skills	-1.427	.547	-.087	-2.610	.009
Counselling skills	1.020	.550	.063	1.854	.064
Clinical judgement	-.264	.601	-.016	-.439	.661
Professional qualities/ communication	1.968	.590	.120	3.337	.001
Organisation/ efficiency	.423	.533	.028	.793	.428
Overall clinical performance	.481	.904	.029	.532	.594
Constant	111.436	2.141		52.050	.000

Binary logistic regression analysis revealed that a model including all clinical performance scores was able to predict whether the Clinical Examination was passed or failed. The overall model was significant at the 0.01 level according to the chi-square statistic and was able to predict 75.0% of the responses accurately. The Professional Qualities/ Communication score within this model had a Wald statistic equal to 12.588 which was significant at the 0.01 level. This means that for every one unit increase in the Professional qualities/ Communication skills score on the mini-CEX assessment, there was a 0.767 increase in the likelihood of passing the Clinical Examination (all else remaining constant). None of the other partial effects in the model were significant at the 0.05 level (see Table 3).

Table 3: Logistic regression results for a model including all clinical performance scores on the mini-CEX

	B	S.E.	Wald	df	Sig.	Exp(B)
Medical interviewing skills	-.072	.084	.743	1	.389	.930
Physical examination skills	.183	.072	6.538	1	.011	1.201
Counselling skills	-.111	.071	2.456	1	.117	.895
Clinical judgement	.002	.077	.001	1	.976	1.002
Professional qualities/ communication	-.265	.075	12.588	1	.000	.767
Organisation/ efficiency	-.115	.068	2.802	1	.094	.892
Overall clinical performance	.000	.116	.000	1	.999	1.000
Constant	1.604	.267	36.073	1	.000	4.975

Overall clinical performance score on its own was also able to predict whether the Clinical Examination was passed. This model was able to predict 74.4% of the responses accurately. The overall clinical performance score within this model had a Wald statistic equal to 185.173 which was significant at the 0.01 level. This means that for every unit increase in the overall clinical performance score, there was a 0.740 increase in the likelihood of passing the Clinical Examination (see Table 4).

Table 4: Logistic regression results for a model including only overall clinical performance score on the mini-CEX

	B	S.E.	Wald	df	Sig.	Exp(B)
Overall clinical performance	-.301	.022	185.173	1	.000	.740
Constant	.961	.150	41.328	1	.000	2.615

This indicates that the mini-CEX is an important instrument for the direct observation of trainees' clinical performance.

Table 6: Mean overall clinical performance scores on the mini-CEX assessment by years post Fellowship of the mini-CEX assessor

Years post Fellowship of the assessor	Mean	Std. Deviation	N
0-1	7.13	1.018	1610
2-5	6.93	1.187	3258
6-10	6.75	1.382	2165
11-15	6.72	1.187	1425
16+	6.54	1.340	3435
Total	6.78	1.265	11893

2. Last mini-CEX assessment completed per trainee

To check the robustness of the results above, the analyses above were performed again using only one mini-CEX assessment per trainee (the last, most recently completed mini-CEX assessment). This dataset contained 5,525 records.

Fewer than 15% of cases in the last mini-CEX assessment per trainee dataset (763; 13.81%) had Clinical Examination result data available. The mean Clinical Examination score was 135.22, with a standard deviation of 18.91. Over three quarters of cases with available Clinical Examination data had passed the Clinical Examination (580; 76.02%).

There was a weak positive correlation found between all clinical performance scores on the mini-CEX assessment and Clinical Examination score (see Table 7). As clinical performance scores increased, Clinical Examination scores increased. Each of these correlations was slightly stronger than that found using the full dataset.

Table 7: Correlations between Measures of Clinical Performance in the Mini-CEX and Clinical Examination Score

	Correlation with Clinical Examination Score
Overall clinical performance	.195**
Medical interviewing skills	.191**
Physical examination skills	.172**
Professional qualities/ Communication	.175**
Counselling skills	.176**
Clinical judgement	.225**
Organisation/ Efficiency	.190**

**Correlation is significant at the 0.01 level

Multiple regression determined that a model containing all seven clinical performance scores on the mini-CEX was able to predict 6.3% of the variance in scores on the Clinical Examination *F(7) = 2.986, p=0.005,* R^2 *=.063.* The Clinical Judgement score within this

model had a *t* statistic equal to 2.497 and was significant at the 0.05 level. This means that for every one unit increase in the Clinical Judgement score on the mini-CEX assessment, there was a 5.642 increase in the Clinical Examination score (all else remaining constant). None of the partial effects were significant at the 0.05 level in this model (see Table 8).

Table 8: Regression results for a model including all clinical performance scores on the mini-CEX

	B	S.E.	Beta	t	Sig.
Medical interviewing skills	.391	2.307	.022	.170	.865
Physical examination skills	-2.905	2.024	-.169	-1.435	.152
Counselling skills	3.059	2.032	.179	1.505	.133
Clinical judgement	5.642	2.259	.335	2.497	.013
Professional qualities/ communication	-.493	2.060	-.029	-.239	.811
Organisation/ efficiency	.681	1.853	.042	.367	.714
Overall clinical performance	-2.925	3.240	-.164	-.903	.367
Constant	111.845	7.406		15.102	.000

Binary logistic regression analysis revealed that a model including all clinical performance scores was able to predict whether the Clinical Examination was passed or failed. The overall model was significant at the 0.01 level ($p<0.0001$) according to the chi-square statistic and was able to predict 80.1% of the responses accurately. Clinical Judgement score within this model had a Wald statistic equal to 4.978 which was significant at the 0.05 level. This means that for every one unit increase in the Clinical Judgement score on the mini-CEX assessment, there was a 0.496 increase in the likelihood of passing the Clinical Examination (all else remaining constant). None of the other partial effects in the model were significant at the 0.05 level (see Table 9).

Table 9: Logistic regression results for a model including all clinical performance scores on the mini-CEX

	B	S.E.	Wald	df	Sig.	Exp(B)
Medical interviewing skills	-.079	.318	.062	1	.803	.924
Physical examination skills	.278	.295	.891	1	.345	1.321
Counselling skills	-.325	.276	1.390	1	.238	.723
Clinical judgement	-.700	.314	4.978	1	.026	.496
Professional qualities/ communication	.028	.284	.010	1	.922	1.028
Organisation/ efficiency	-.510	.262	3.776	1	.052	.601
Overall clinical performance	.761	.439	3.008	1	.083	2.141
Constant	2.358	.994	5.625	1	.018	10.569

However, overall clinical performance score on its own was able to predict whether the Clinical Examination was passed. This model was able to predict 75.9% of the responses accurately. The overall clinical performance score within this model had a Wald statistic equal to 28.205 which was significant at the 0.01 level. This means that for every unit increase in the overall clinical performance score, there was a 0.654 increase in the likelihood of passing the Clinical Examination (see Table 10).

Table 10: Logistic regression results for a model including only overall clinical performance score on the mini-CEX

	B	S.E.	Wald	df	Sig.	Exp(B)
Overall clinical performance	-.425	.080	28.205	1	.000	.654
Constant	1.730	.540	10.252	1	.001	5.643

3. First mini-CEX assessment completed per trainee

Additional analyses performed using the first assessment completed per trainee yielded similar if not slightly stronger results than the last assessment completed by each trainee in relation to the Clinical Examination. This confirms the robustness of the results and indicates that early intervention with trainees that are performing poorly on mini-CEX may result in them having a better chance of passing the Clinical Examination.

Appendix 9: Expanded results – Years Post Fellowship of the Assessor using the full dataset

1. All mini-CEX cases in the full dataset

These additional analyses have been performed using the full cleaned dataset extracted from the Basic Training Portal, supplemented with data from the College database (N=39484).

Just over a third of the full cleaned dataset (11893; 30.1%) had data that indicated how many years since the assessor was admitted to Fellowship available. The mean number of years since admission to Fellowship for assessors of the mini-CEX was 11.5, with a standard deviation of 10.8.

There was a weak negative correlation found between all clinical performance scores on the mini-CEX assessment and years post Fellowship of the mini-CEX assessor (see Table 5). As years post Fellowship of the assessor increased, clinical performance scores decreased.

Table 5: Correlations between Measures of Clinical Performance in the Mini-CEX and Assessor Years post Fellowship

	Correlation with Years Post Fellowship of the Assessor
Overall clinical performance	-.155**
Medical interviewing skills	-.094**
Physical examination skills	-.163**
Professional qualities/ Communication	-.153**
Counselling skills	-.099**
Clinical judgement	-.133**
Organisation/ Efficiency	-.133**

**Correlation is significant at the 0.01 level

A significant relationship was also found between overall clinical performance score on the mini-CEX and years post Fellowship of the mini-CEX assessor when years post Fellowship of the assessor was treated as a categorical variable ($F(4)$=75.883, $p<0.0001$). Assessors who had become RACP Fellows within the last year were the most likely to give trainees higher scores for overall clinical performance (see Table 6).

Table 6: Mean overall clinical performance scores on the mini-CEX assessment by years post Fellowship of the mini-CEX assessor

Years post Fellowship of the assessor	Mean	Std. Deviation	N
0-1	7.13	1.018	1610
2-5	6.93	1.187	3258
6-10	6.75	1.382	2165
11-15	6.72	1.187	1425
16+	6.54	1.340	3435
Total	6.78	1.265	11893

2. Last mini-CEX assessment completed per trainee

To check the robustness of the results above, the analyses above were performed again using only one mini-CEX assessment per trainee (the last, most recently completed mini-CEX assessment). This dataset contained 5,525 records.

Just under a third of the last mini-CEX assessment per trainee dataset (1516; 27.43%) had data that indicated how many years since the assessor was admitted to Fellowship available. The mean number of years since admission to Fellowship for assessors of the mini-CEX was 12.6, with a standard deviation of 10.9.

There were very weak negative correlations found between Overall Clinical Performance, Physical Examination Skills, Professional Qualities/ Communication, Counselling skills, Clinical Judgement and Organisation/Efficiency scores and years post Fellowship of the mini-CEX assessor (see Table 11). As years post Fellowship of the assessor increased, these clinical performance scores decreased. There was no significant correlation between Medical Interviewing Skills and years post Fellowship of the assessor at the 0.05 level.

Table 11: Correlations between Measures of Clinical Performance in the Mini-CEX and Assessor Years post Fellowship

	Correlation with Years Post Fellowship of the Assessor
Overall clinical performance	-.120**
Medical interviewing skills	-.050
Physical examination skills	-.122**
Professional qualities/ Communication	-.126**
Counselling skills	-.080*
Clinical judgement	-.117**
Organisation/ Efficiency	-.102**

**Correlation is significant at the 0.01 level

A significant relationship was found between overall clinical performance score on the mini-CEX and years post Fellowship of the mini-CEX assessor when years post Fellowship of the assessor was treated as a categorical variable ($F(4)=4.959$, $p=0.001$). This relationship was linear, with assessors who had become RACP Fellows within the last year the most likely to give trainees higher scores for overall clinical performance (see Table 12).

Table 12: Mean overall clinical performance scores on the mini-CEX assessment by years post Fellowship of the mini-CEX assessor

Years post Fellowship of the assessor	Mean	Std. Deviation	N
0-1	7.07	1.099	175
2-5	7.03	1.185	352
6-10	6.99	1.205	288
11-15	6.87	1.152	204
16+	6.69	1.297	497
Total	6.89	1.221	1516

3. First mini-CEX assessment completed per trainee

Additional analyses performed using the first assessment completed per trainee yielded similar if not slightly stronger results than the last assessment completed by each trainee in relation to the years post Fellowship of the assessor.

Appendix 10: Factor Analysis Results for Clinical Performance Measures

Component Matrix[a]

	Component
	1
Medical_interviewing_skills_new	.916
Physical_examination_skills_new	.899
Professional_qualities_communication_new	.905
Counselling_Skills_new	.892
Clinical_judgement_new	.925
Organisation_efficiency_new	.866

Extraction Method: Principal Component Analysis.

a. 1 components extracted.

Glossary of terms

Clinical performance scores - As part of the mini-CEX assessment, assessors are asked to rate aspects of a trainees' clinical performance, namely their medical interviewing skills, physical examination skills, professional qualities/communication skills, counselling skills, clinical judgement, organisation/efficiency and overall clinical performance. These scores are known as their 'clinical performance scores'.

Cronbach's alpha (α) - Cronbach's α is a statistic calculated from the pairwise correlations between question items. This statistic is used to measure internal consistency.

Factor analysis - Factor analysis is a statistical method used to describe variability between observed variables that are related to one another.

Linear regression - In statistics, **l**inear regression attempts to model the relationship between two variables by fitting a linear equation to observed data. One variable is considered to be an independent variable, and the other is considered to be a dependent variable.

Logistic regression – In statistics, binary logistic regression is used to predict a categorical variable (usually dichotomous) from a set of predictor variables.

Random sampling - Random sampling is a method of sampling in which a group of subjects or cases (a sample) for study are chosen from a larger group (a population). Each subject/case is chosen entirely by chance and each subject/case has an equal chance of being selected from the population.

Reliability - Reliability refers to the degree to which an assessment tool produces stable and consistent results.

Stratified random sampling - Stratified random sampling is a method of sampling that involves the division of a population into smaller groups, or 'strata'. These strata are formed based on members' shared attributes or characteristics. A random sample from each stratum is taken in a number proportional to the stratum's size when compared to the population. These subsets of strata are then combined to form a random sample.

Validity - Validity refers to how well an assessment tool measures what it is purported to measure.