



## Longitudinal assessment in an undergraduate longitudinal integrated clerkship: The mini Clinical Evaluation Exercise (mCEX) profile

Denese Playford, Andrew Kirke, Moira Maley & Rhonda Worthington

**To cite this article:** Denese Playford, Andrew Kirke, Moira Maley & Rhonda Worthington (2013) Longitudinal assessment in an undergraduate longitudinal integrated clerkship: The mini Clinical Evaluation Exercise (mCEX) profile, Medical Teacher, 35:8, e1416-e1421, DOI: [10.3109/0142159X.2013.778392](https://doi.org/10.3109/0142159X.2013.778392)

**To link to this article:** <https://doi.org/10.3109/0142159X.2013.778392>



Published online: 02 Apr 2013.



Submit your article to this journal [↗](#)



Article views: 1560



View related articles [↗](#)



Citing articles: 5 View citing articles [↗](#)

## WEB PAPER

# Longitudinal assessment in an undergraduate longitudinal integrated clerkship: The mini Clinical Evaluation Exercise (mCEX) profile

DENESE PLAYFORD, ANDREW KIRKE, MOIRA MALEY & RHONDA WORTHINGTON

The University of Western Australia, Australia

## Abstract

**Aim:** Student and assessor performance were examined over three academic years using the mini Clinical Evaluation Exercise (mCEX) as a continuous feedback tool across all disciplines, in all learning contexts, for an entire integrated undergraduate year.

**Methods:** Students could complete any number of mCEX, but had to submit a minimum number per discipline. Students were free to choose assessors. Assessors were not trained. Data were collected in a customised database, and analysed in SPSS ver 18.0.0.

**Results:** 5686 mCEX were submitted during 2008–2010 (Cronbach's  $\alpha = 0.80$ ). Marks were affected by doctor grade ( $F = 146.6$ ,  $p < 0.000$ ), difficulty of clinical encounter ( $F = 33.3$ ,  $p < 0.000$ ) and clinical discipline ( $F = 13.8$ ,  $p < 0.000$ ). Students most frequently sought harder markers (experienced general practitioner/hospital specialists). Increases in mCEX marks were greatest during the early, formative months ( $F = 42.7$ ,  $p < 0.000$ ). More mCEX were submitted than required, without differentiation between weak or strong students ( $r_{xy} = 0.22$ ,  $p = 0.78$ ).

**Conclusions:** Undergraduate students in longitudinal clerkships acquire most skills during 'formative' learning. They seek 'hard' assessors, consistent with year-long mentoring relationships and educational/feedback value. Assessors mark consistent with a framework of encouraging student performance. Over an entire longitudinal clerkship, students complete mCEX in excess of course requirements. This study confirms the impact of the longitudinal context on assessor and student behaviour.

## Introduction

Current research suggests that longitudinal integrated clerkships in undergraduate education offer a number of educational benefits relative to conventional specialty-rotation learning (Hirsh et al. 2007). That longitudinal clerkships may also confer assessment benefit is an emerging area of research (Epstein 2007).

In contrast to evaluation in one block or rotation, longitudinal undergraduate placements putatively allow sustained monitoring of performance (Cleary 2008). In theory, this kind of feedback is most effective for clinical behaviour change (Van der Vleuten et al. 2000).

But, to date, the psychometric impact of longitudinal integrated clerkships on how undergraduate students are assessed has not been reported, other than to indicate that the overall academic outcome of longitudinal programmes is equivalent to (Zink et al. 2010) or superior to (Worley et al. 2004) rotational clerkships.

The lack of discussion to date may be due to the fact that longitudinal clerkships often simply use the assessment programme already in place of rotational clerkships (Zink et al. 2010). This practice emphasises the reproducibility of students' performance in diverse learning contexts and allows benchmarking of grades. In the Rural Clinical School of Western Australia (RCSWA), the use of urban assessment initially functioned well to show no differences in marks

## Practice points

- Feedback is most strongly related to performance during the development of clinical skills – mCEX feedback should be concentrated at this time.
- Allowing students to complete as many mCEX as they chose (over a required number) to attain top marks does not skew final grade, but does encourage clinical behaviour.
- Students chose hard markers when they are in positive mentoring relationships, and when they are in control of developing their own clinical skill.

between rural-based and urban-based cohorts. However, the urban assessment programme was based on discrete rotations with multiple tools assessing similar skills, and fixed due dates which did not allow for recognition of graduated acquisition of skills over an entire year.

To examine the performance impacts of integrated longitudinal clerkships, holistic assessment programmes need to be designed (van der Vleuten & Schuwirth 2005). To this end, the RCSWA assessment programme was completely revised in 2007. Amongst others, the mini Clinical Evaluation Exercise (mCEX) was introduced as a core tool that could be used formatively and summatively across all disciplines, in all learning contexts, over an entire year (Norcini et al. 2003).

**Correspondence:** Denese Playford, The Rural Clinical School of Western Australia, The University of Western Australia, 35 Stirling Highway, M501, Crawley, WA 6009, Australia. Tel: 61-421-562-872; fax: 618-9346-3120; email: Denese.Playford@rcswa.edu.au

(Davies 2009). Through this tool, characteristics of longitudinal assessment could be monitored.

Since many medical courses are considering the benefits of longitudinal integrated clerkships (Walters et al. 2012), it is timely to examine how longitudinal clinical feedback impacts undergraduate students' performance.

## Methods

Acceptance and development of the mCEX form took place via email discussions that included all RCSWA clinical academics. The final scale was modified from the NHS 6-point scale to an 8-point scale which represented 30 or more clinical academics' marking preferences. The scale points are 0, 3 and 4.5 for below-par performance, 5.5 for a clear pass, 6.5, 7.5, 8.5 for increasing competence and 10 for a faultless performance appropriate to the expectations for this year of education. The criterion was that 'marks should be based on end-of-year expectations for a student completing the second last year of medicine'.

At orientation to the year, all students were given an mCEX triplicate booklet, for use in any context and at any time. Each booklet had a unique, consecutive block of numbers by which the forms were subsequently collated to student name and year.

Students were informed that they could complete as many mCEX as they liked, but had to complete a set minimum per discipline. The minimum was set for both formative and summative assessments: in 2008 and 2009, this comprised 7 formative and 14 summative, in 2010 only summative mCEX were recorded. The required total was 21 mCEX for the year. The period titled 'formative' in the academic calendar started in January and ended the beginning of May, the period titled 'clinical skills' and 'consolidation of clinical skills' continued to year end in November. Students were free to choose any clinician in any setting for their assessment. They were invited to submit all completed forms and were told that the best summative results per discipline would be selected for their final academic record. In 2009 and 2010, submission deadlines mid-year and year-end were instituted to ensure mCEX were not stockpiled.

Assessors were not trained in use of mCEX. In some sites, local doctors were invited to discussions about the assessment, but in most, assessors were simply presented with the form to fill out at point of care. Summative assessments required completion by a medical practitioner.

Clinical settings included general practice, Aboriginal medical services and nursing posts with fly-in medical service as well as the small (secondary) hospitals that comprise the backbone of regional health care in Western Australia.

An electronic database with drop-down selection boxes and automatic calculation was devised so that site administrators could reliably enter the 1000s of forms, and such that the domain scores would be automatically totalled to the final grade. This 'Holistic Information from Workplace Assessment' (HIWAY) database reported multiple assessment data.

Based on the report from the HIWAY database, mCEX data was extracted and entered into an Excel spreadsheet. The subset of best marks used in the final academic record was

extracted from this total dataset. Both sets are reported here. Both were imported into SPSS for analysis.

Validity was examined by examining the performance of the mCEX in the hands of different categories of assessors, and the way case complexity impacted its use. Reliability was examined using Cronbach's  $\alpha$  for repeated measures over the year. The performance of the tool was examined by looking at the progression of marks over the academic year.

## Results

Over three academic years, a total of 5686 mCEX were recorded in the assessment system. Of the assessments submitted, approximately two-thirds were considered to be of average complexity (Table 1). The distribution across disciplines roughly represented the relative weight of each discipline in the year (Table 2).

The number of assessments completed reflected the discipline matrix of The Rural Clinical School (Figure 1). The majority of assessments were completed by general practitioners (GPs). For the minority completed in the hospital setting, preference was given to completion by medical specialists (local, regular visiting or locum), followed by resident medical officers (RMOs) and finally by junior medical officers (JMOs) who are present in the sites with accredited teaching hospitals. Hundreds of doctors completed assessments, a number far in excess of those who were on staff, or who were paid for precepting these students.

### Validity

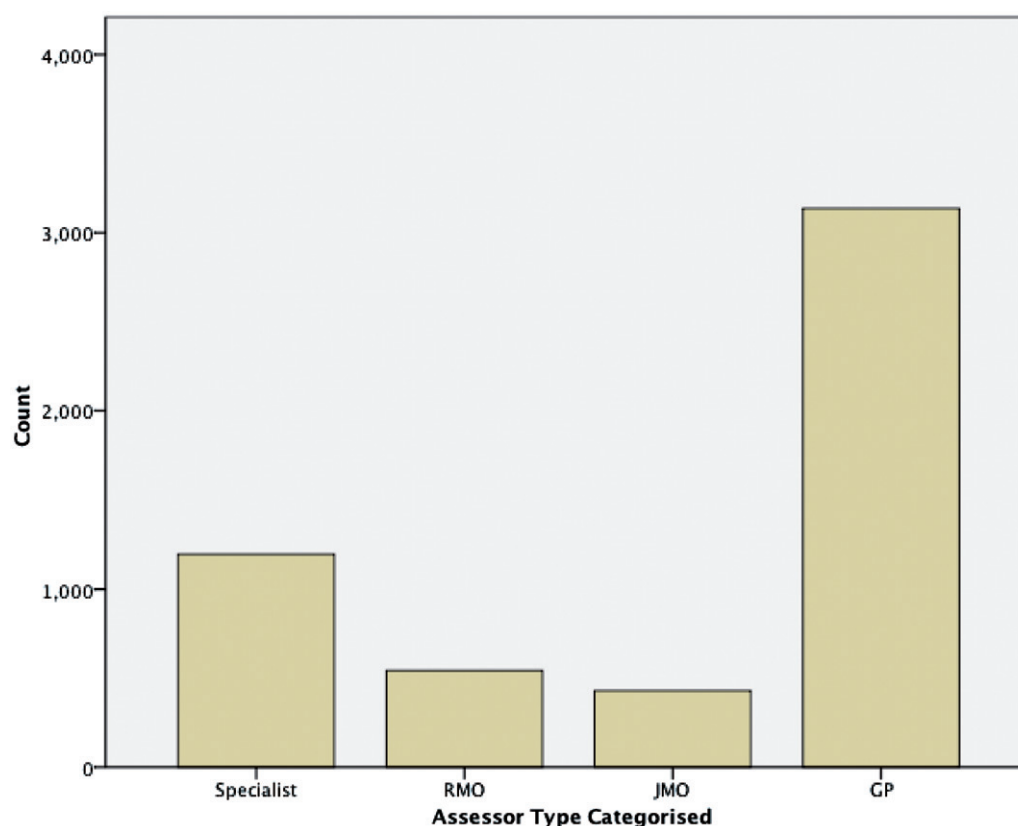
The way that assessments were graded was influenced by the category of doctor who completed the assessment ( $F=146.6$ ,  $p<0.000$ ). Grades were inversely proportional to specialisation, with junior medical officers giving the highest marks, and the specialist consultants giving the lowest marks. GP generalists gave intermediary marks, as shown in Table 3.

**Table 1.** Number of mCEX per category of complexity.

Complexity	Number	Percentage
Average	3836	67.4
High	987	17.3
Low	646	11.3
Unstated	218	3.8

**Table 2.** Number of mCEX per clinical discipline.

Discipline	Number	Percent
General practice	920	16.2
Ophthalmology	452	7.9
Internal medicine	1058	18.6
Cancer	457	8.0
Obs & Gynae	985	17.3
Paediatrics	1049	18.4
Surgery	765	13.4
Total	5686	100.0



**Figure 1.** The number of assessments completed by differing types of assessor.

**Table 3.** Grades given by differing types of assessor.

Assessor type	N	Mean	SE	95% CI – lower	95% CI – upper
JMO	430	8.2202	0.05231	8.1174	8.3230
RMO	543	7.71	0.04202	7.6281	7.7932
GP-Generalist	3136	7.4699	0.01839	7.4338	7.5060
Specialist	1197	7.079	0.02837	7.0234	7.1348
Total	5306	7.4672	0.01455	7.4387	7.4957

Grades were also impacted by the level of rated difficulty of the clinical encounter. Students with the most difficult encounters were graded higher than students with the least difficult encounters ( $F=33.3$ ,  $p < 0.000$ ).

The grade was further impacted by the clinical discipline. Disciplines considered 'special', with relatively less teaching time in the RCSWA year (Surgery, Cancer, Ophthalmology), were given higher grades than the other core disciplines ( $F=13.8$ ,  $p < 0.000$ ).

#### Reliability

The relationship between each mCEX mark in the final academic record for the total of 215 students showed a reasonable level of consistency between mCEX over the year (Cronbach's  $\alpha = 0.8$ ).

#### Progression of marks over the year

The average marks given to mCEX increased significantly over the course of the academic year. Marks during the formative

**Table 4.** Marks given during the initial formative and subsequent clinical development phases in the academic years 2008–2010.

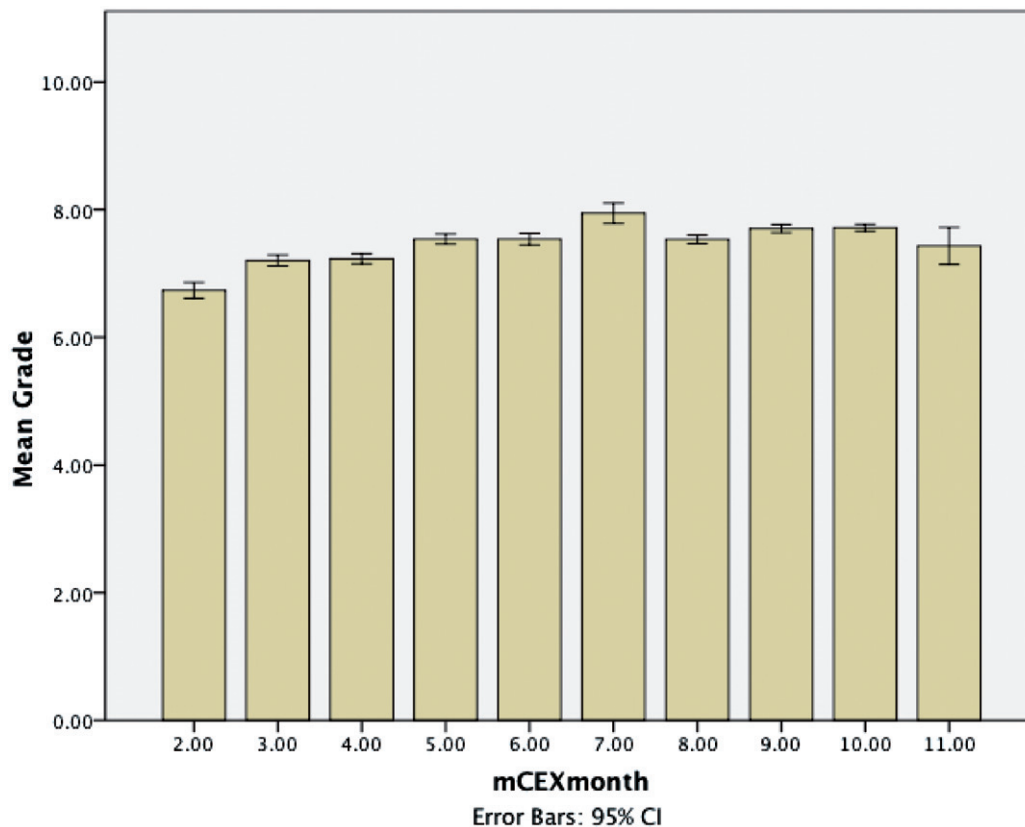
Phase	Number	Average	SD	t-test
Formative	1454	7.12	1.03	16.34
Summative	4231	7.64	1.00	$p < 0.000$

**Table 5.** Marks given over the course of the academic year.

mCEX	Feb	Mar	Apr	May	Jun	Jul	Aug	Sep	Oct	Nov
Average	6.74	7.20	7.23	7.54	7.54	7.94	7.53	7.70	7.72	7.43

phase (January to beginning May) were significantly lower than those for the subsequent phases of the year, for all years combined ( $t=16.34$ ,  $p < 0.000$ , Table 4) and for each year individually ( $t=13.1$ ,  $p < 0.000$  in 2008;  $t=12.19$ ,  $p < 0.000$  in 2009;  $t=6.63$ ,  $p < 0.000$  in 2010).

Month-by-month analysis of marks showed a significant progression over the year for all three years collectively ( $F=42.7$ ,  $p < 0.000$ , Table 5) and for each year separately ( $F=24.3$ ,  $p < 0.000$  in 2008;  $F=23.5$ ,  $p < 0.000$  in 2009;  $F=5.2$ ,  $p < 0.000$  in 2010). However, the profile was not even. As shown in Figure 2 for all years combined, the increase was sequential from month to month during the formative phase of the year, from February to beginning May ( $F=24.3$ ,  $p < 0.000$ ). Thereafter marks developed relatively less ( $F=8.7$ ,  $p < 0.000$ ).



**Figure 2.** The month-by-month progression of mCEX marks over the academic year.

### Impact on clinical behaviour

Many more mCEX were submitted than were required for the final assessment portfolio (Figure 3). Over the three years, only 17 students submitted no excess. For the remainder, the excess ranged from 5% to 180% of the required number. As an average for all students, the annual excess per student was 4.25 in 2008, 5.5 in 2009 and 6.4 in 2010. There was no correlation between the number of mCEX a student submitted and their final year mark ( $r \times y = 0.22$ ,  $p = 0.78$ ).

## Discussion

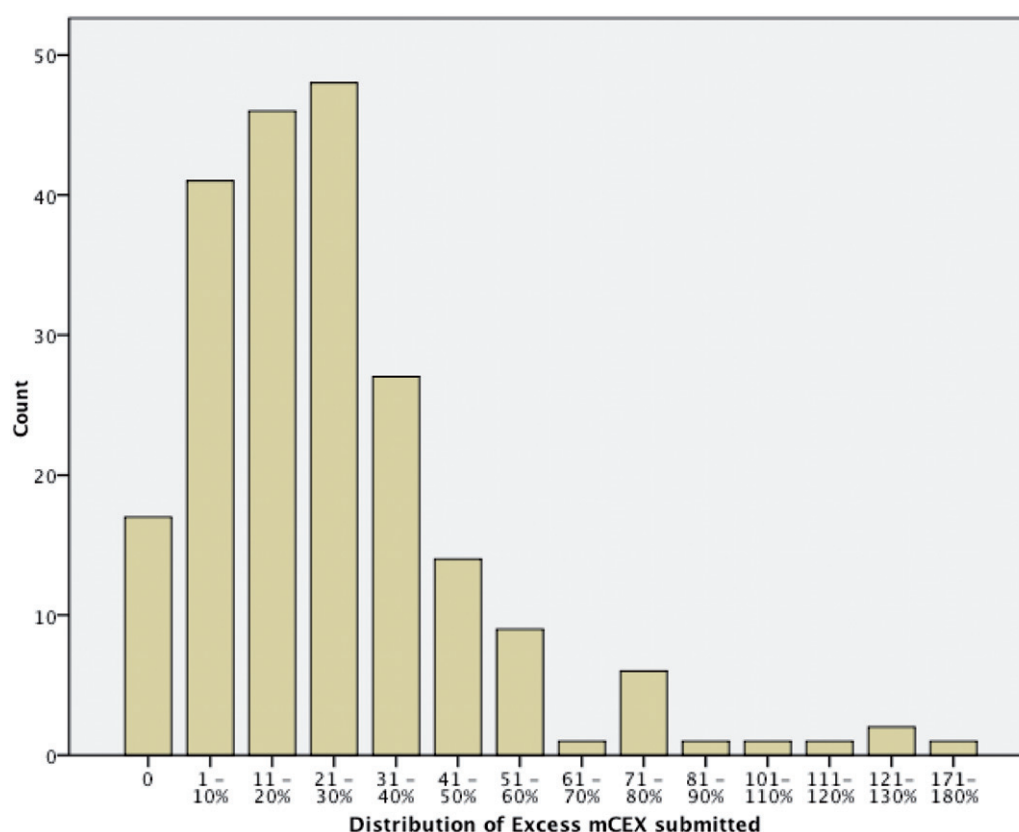
We provide first detailed description of acquisition of performance skills for undergraduate students in an integrated longitudinal clerkship. Continuous monitoring using the mCEX tool showed significant gains in performance over the year. These data confirm mCEX data from post-graduates in similar longitudinal training contexts. Davies et al. (2009) showed that interns' mCEX marks increased between the first and second halves of the UK F1 year. In more detail, Norcini et al. (2003) showed increase in marks per quarter year for American residents, whose level of postgraduate training was not specified. However, both these studies suggest an even profile of skills acquisition. Our data indicated that, for RCSWA undergraduates, the profile was less even. Development of skill was concentrated in the early, formative months.

A similar early pattern of skills acquisition was shown for general practice vocational trainees' development of communication skills. Using an internally developed tool,

trainees' gains were significant when given feedback at various points during training. But their skill showed maximal increase during the earliest phase of training (Greco et al. 2001). An older study looking at undergraduate students' history-taking skills over three years in medical school indicated that history skills improved most during the initial phase of training, with more gradual improvements in clinical years thereafter (Barbee & Feldman 1970).

These data suggest that, for novices, learning may not be a smooth process. Feedback appears to be particularly critical during the acquisition stage of learning (Wood et al. 1976). During this time, the characteristics of mCEX in a longitudinal setting – repeated performance stress, reinforcement and reward – have the characteristics described as effective for neurobiology of learning (Friedlander et al. 2011). The fact that it took a matter of months for performance scores to peak suggests that locating undergraduate learners in the same context with the same set of assessors could be particularly beneficial to the acquisition of clinical skills. This interpretation would be consistent with reports that clinical performance may be better at the end of longitudinal integrated clerkships (Worley 2004).

A limitation of this analysis is the likelihood that students will submit only those mCEX with the best marks, and that these best marks may accrue over the year leading to an apparent but spurious increase in mCEX marks. This was addressed in 2009 and 2010 by having a mid-year mandatory submission date to reduce the likelihood of stockpiling, and to enable mCEX to be submitted co-incident with their phase of



**Figure 3.** Distribution of excess submitted mCEX.

learning. The issue of stockpiling was also addressed by inviting submission of excess mCEX. The fact that students did so, with the clear strategic intention of aiming for top marks, suggests that they trusted the system sufficiently to submit more than their optimal assessments.

These data also give an account of the assessors who are approached to make assessments in a longitudinal integrated clerkship setting. Although students were at liberty to choose any assessors, they most frequently approached local GPs, many of whom were also their designated longitudinal clerkship mentors. When appraisal is linked to mentorship, a trusting relationship is said to further develop (Levy & Williams 2004). This kind of engagement between student and preceptor has been asserted as key to the perceived benefit of feedback in longitudinal clerkships (Bell et al. 2008; Mazotti et al. 2011). Our numerical data demonstrates this qualitative relationship.

As is characteristic of rural work anywhere, Western Australia's rural primary care clinicians deliver care in many contexts and draw the student from their private general practices into public hospitals, where students also interact with their precepting GP's network. This practice of inclusion into clinical community, which is considered core to the success of the Harvard Integrated Clerkship (Bell et al. 2008), was demonstrated in this study by the students' extensive assessment community of hospital doctors.

Within this wider assessment community, students made clear choices based on doctors' specialisation rather than the marks they could get. The most undifferentiated junior doctors gave the highest marks. But early concern that students would

seek out the most lenient markers was clearly not demonstrated by the data: prevocational JMOs and RMOs were least approached. In contrast, the most specialised clinicians were the hardest markers and yet were the most frequent hospital doctors approached. The differences in doctors' assessment decisions confirm work by Kogan et al. (2011) who show that assessing doctors use multiple frames of reference, including their own expertise, their experience with other learners and the provision of adequate patient care, in making assessment judgements (Kogan et al. 2011). Govaerts et al. (2011) show that experts use contextual cues and make more inferences than novices (Govaerts et al. 2011). In the present case, whereas more junior doctors tended to affirm current performance, the most experienced doctors encouraged better performance by harder marks. Students' selection of the latter showed their learning priorities in a longitudinal clerkship.

According to social-learning theory, learning is strongly situated in a social context (Lave & Wenger 1991). In the Rural Clinical School longitudinal context, both appraiser and appraisee are mutually accountable for their performance, and are influenced by their ongoing subjective relationship to each other. Elsewhere, these factors have been shown to influence appraisal (Levy & Williams 2004). The same holds true for mCEX, since further characteristics of the data suggest that the mCEX was being used as a tool in a community of practice sense (Wenger 1998). Across all assessor types, students undertaking cases rated as 'high difficulty' were given higher 'encouragement' marks for their performance. Assessments in the less frequently taught parts of the curriculum were also given higher marks. These data extend Norcini



et al. (2003) original work, and results for an undergraduate programme in Southampton, where examiner status, case complexity and attachment specialty all influenced ratings (Hill et al. 2009). The data collectively show that, although assessor judgements may not be standardised, nevertheless they were systematic. In a community of practice, tools are developed as a way of increasing and managing knowledge. In this sense, in a longitudinal clerkship setting, mCEX were being systematically used to motivate more expert performance in a junior colleague by a senior assessor.

The longitudinal assessment programme that allowed students to complete as many mCEX as they wanted also provided clear motivation for increased clinical behaviour. The majority received feedback on considerably more patients than they were required to. This was the case whether they were strong or weak students, as indicated by the lack of correlation between number of mCEX and final mark. Although this excess of clinical encounters was pragmatic and aimed at receiving higher marks, nevertheless it created a feedback culture which, elsewhere, has been associated with positive appraisal experience (Levy & Williams 2004). Students' taking this kind of responsibility for performance demonstrates intrinsic motivation associated with adult learning (ten Cate et al. 2004), and endorses students' ownership of assessment as a learning experience.

In conclusion, these data show that longitudinal integrated assessment in a naturalistic setting particularly identifies and contributes to skill development in the formative phase of learning. Naïve assessors use the mCEX in a psychometrically unreliable but systematic way to develop clinical behaviours, and are willing to conduct thousands of assessments to this end. Further, students actively participate in their own development by using the tool in excess of course requirements. These data demonstrate Van der Vleuten's observation (Van der Vleuten et al. 2000) that good assessment is inclusive of, but broader than, the statistical characteristics of assessment tools. This study confirms the impact of the longitudinal context on assessor and student behaviour.

## Notes on contributors

A/Prof DENESE PLAYFORD, PhD, is a medical educator in The RCSWA.

A/Prof ANDREW KIRKE, MBBS, is a medical co-ordinator in The RCSWA.

A/Prof MOIRA MALEY, PhD, is a medical educator (technology) in The RCSWA.

Mrs RHONDA WORTHINGTON is the administrative manager for The RCSWA.

**Declaration of interest:** The authors report no conflicts of interest. The authors alone are responsible for the content and writing of the article.

## References

- Barbee RA, Feldman SE. 1970. A three year longitudinal study of the medical interview and its relationship to student performance in clinical medicine. *Acad Med* 45:770–776.
- Bell S, Krupat E, Fazio S, Roberts D, Schwartzstein R. 2008. Longitudinal pedagogy: A successful response to the fragmentation of the third-year medical student clerkship experience. *Acad Med* 83:467–475.
- Cleary L. 2008. 'Forward feeding' about students' progress: The case for longitudinal, progressive, and shared assessment of medical students. *Acad Med* 83:800.
- Davies H, Archer J, Southgate L, Norcini J. 2009. Initial evaluation of the first year of the foundation assessment programme. *Med Educ* 43:74–81.
- Epstein RM. 2007. Assessment in medical education. *N Engl J Med* 356:387–396.
- Friedlander M, Andrews L, Armstrong E, Aschenbrenner C, Kass J, Ogden P, Schwartzstein R, Viggiano T. 2011. What can medical education learn from the neurobiology of learning? *Acad Med* 86:415–420.
- Govaerts M, Schuwirth L, Van Der Vleuten C, Muijtjens A. 2011. Workplace-based assessment: Effects of rater expertise. *Adv Health Sci Educ Theory Pract* 16:151–165.
- Greco MB, Brownlea A, McGovern J. 2001. Impact of patient feedback on the interpersonal skills of general practice registrars: Results of a longitudinal study. *Med Educ* 35:748–756.
- Hill F, Kendall K, Galbraith K, Crossely J. 2009. Implementing the undergraduate mini-CEX: A tailored approach at Southampton University. *Med Educ* 43:326–334.
- Hirsh D, Ogur B, Thibault G, Cox M. 2007. 'Continuity' as an organizing principle for clinical education reform. *N Engl J Med* 356:858–866.
- Kogan J, Conforti L, Bernabeo E, Iobst W, Holomboe E. 2011. Opening the black box of clinical skills assessment via observation: A conceptual model. *Med Educ* 45:1048–1060.
- Lave JA, Wenger E. 1991. *Situated learning: Legitimate peripheral participation*. Cambridge: Cambridge University Press.
- Levy PA, Williams J. 2004. The social context of performance appraisal: A review and framework for the future. *J Manag* 30:881–905.
- Mazotti L, O'Brien B, Tong L, Hauer K. 2011. Perceptions of evaluation in longitudinal versus traditional clerkships. *Med Educ* 45:464–470.
- Norcini J, Blank L, Duffy F, Fortna G. 2003. The mini-CEX: A method for assessing clinical skills. *Ann Intern Med* 138:476–481.
- Ten Cate O, Snell L, Mann K, Vermunt J. 2004. Orienting teaching toward the learning process. *Acad Med* 79:219–228.
- Van Der Vleuten C, Scherpbier A, Dolmans M, Schuwirth L, Verwijnen G, Wolfhagen H. 2000. Clerkship assessment assessed. *Med Teach* 22:592–603.
- Van Der Vleuten CPM, Schuwirth LWT. 2005. Assessing professional competence: From methods to programmes. *Med Educ* 39:309–317.
- Walters L, Greenhill J, Richards J, Ward H, Campbell N, Ash J, Schuwirth L. 2012. Outcomes of longitudinal integrated clinical placements for students, clinicians and society. *Med Educ* 46:1028–1041.
- Wenger E. 1998. *Communities of practice*. New York: Cambridge University Press.
- Wood D, Bruner JS, Ross G. 1976. The role of tutoring in problem solving. *J Child Psychol Psychiatr* 17:89–100.
- Worley P, Esterman A, Prideaux D. 2004. Cohort study of examination performance of undergraduate medical students learning in community settings. *Br Med J* 328:207–209.
- Zink T, Power D, Finstead D, Brookes K. 2010. Is there equivalency between students in a longitudinal, rural clerkship and a traditional urban-based programme? *Fam Med* 42:702–706.