How do formative objective structured clinical examinations drive learning? Analysis of residents’ perceptions

Debra Pugh, Isabelle Desjardins & Kevin Eva

To cite this article: Debra Pugh, Isabelle Desjardins & Kevin Eva (2018) How do formative objective structured clinical examinations drive learning? Analysis of residents’ perceptions, Medical Teacher, 40:1, 45-52, DOI: 10.1080/0142159X.2017.1388502

To link to this article: https://doi.org/10.1080/0142159X.2017.1388502
How do formative objective structured clinical examinations drive learning? Analysis of residents’ perceptions

Debra Pugh*, Isabelle Desjardins* and Kevin Eva

*Department of Medicine, University of Ottawa, Ottawa, Canada; †Department of Medicine, University of British Columbia and Senior Scientist at the Centre for Health Education Scholarship (CHES), Vancouver, Canada

ABSTRACT

Introduction: Although several studies have explored the relationship between learning and written tests, little is understood about how performance-based examinations influence learning. The purpose of this study was to explore how a formative objective structured clinical examination (OSCE) drives learning.

Methods: We administered surveys to residents (n = 35) at three time points to determine if and how an OSCE influenced their learning: before and immediately following the OSCE, and after the distribution of their results. Differences in quantitative responses between high- and low-performing residents and across time were compared using repeated-measures ANOVA. Thematic analysis was used to analyze narrative comments.

Results: Participants’ goals for the OSCE related to performance, mastery and feedback. Almost all participants reported that they had learned something from the OSCE (94%) and most participants generated learning goals after the OSCE (71%). High performers appeared to recognize the importance of content-related knowledge for scoring well before and after the OSCE, whereas low performers may have underestimated its importance until after the examination.

Discussion: Participants viewed a formative OSCE as both a hurdle to overcome (assessment of learning) and an opportunity to learn (assessment for learning). Understanding how OSCEs influence study behavior can help guide the development of assessments that promote learning.

Background

There is little disagreement among medical educators that assessment drives learning (Newble 2016). However, as the emphasis for the role of assessment shifts purely from assessment of learning (AOL) toward thinking in better ways about how educators can use assessment for learning (AFL), there is an impetus to determine how and when assessments of various forms have educational value.

It has become understood that assessment can have both direct and indirect influences on learners (Larsen et al. 2008). Direct effects arise when the assessment itself prompts learning. For example, the testing effect refers to a well-established phenomenon that testing can have a more powerful influence on learning than when an equal amount of time is devoted to studying (Larsen et al. 2008, 2009; Kromann et al. 2009; Larsen 2013). Indirect effects, in contrast, arise from providing an incentive to study by informing learners that they will be assessed. In addition, feedback provided to learners after a test can help to guide their learning by highlighting areas of strength and weakness that may, in turn, motivate learners to continue exploring the topic after the assessment is over (Agrawal et al. 2012).

Despite these clear benefits of assessment, testing can also be detrimental to learning. For example, the stress caused by a testing event may lead learners to adopt superficial learning strategies (e.g. rote memorization, cramming, studying to the test) that maximize their chances of success in the short term (i.e. on an examination) rather than devoting their efforts to truly understanding the material in a way more likely to yield long-term gain (Pashler et al. 2007). Learners may also use cues on the examination to correctly answer questions they would not otherwise have attempted, which may create an inflated impression of their own abilities (Desjardins et al. 2014). And, finally, examinees may feel pressure to resort to collusion and cheating in an attempt to improve their scores, again detracting from efforts related to actually learning the material (Henning et al. 2013). This creates a considerable problem for health professional educators as better guidance is required if assessments are to be implemented for pedagogical value without falling prey to such unintended consequences.

These challenges for those striving to employ tests as pedagogical interventions are likely to be compounded when the assessment in question is performance based with observers watching a learner demonstrate their skills. Whereas the influence of testing on learning has been studied primarily in the context of written forms of assessment (Scouller 1998; Cilliers et al. 2012; Davis and Chan 2015), objective structured clinical examinations (OSCEs)
more closely approximate real clinical practice and, therefore, are more likely to connect with the physician’s professional identity. Although they require a foundation of knowledge, well-designed OSCEs are used to assess how learners integrate and apply that knowledge when faced with a clinical problem (Harden et al. 1975). OSCEs are generally used to assess different constructs than those that are targeted by written tests, such as clinical skills and the ability to establish rapport with a patient (Jeffries et al. 2007). Both of these factors suggest that the learning that takes place in OSCEs is less likely to reflect simple declarative knowledge and more likely to reflect complex behaviors that will interact in variably interpreted ways. As such, learning that takes place around an OSCE could provide a better approximation (relative to written tests) of the influence of assessment moments that physicians will experience after their formal training is completed. It is important to keep in mind, however, that the observation required for performance-based assessment leads to “staging” behaviors as learners can feel compelled to “play a role” rather than demonstrate what they would naturally do (Gormley et al. 2016; Shea and Norcini 2017). This creates the potential that learners, especially those operating at an advanced stage of training, will view even a well-designed OSCE as inauthentic, thus engendering cynicism regarding its educational value. All of this sums together to suggest that an OSCE that is ostensibly run for formative purposes in a postgraduate (residency) training context would provide a particularly powerful context within which to better understand the tensions that need to be taken into account when designing assessments to encourage learning.

At present, because our understanding of these influences is incomplete, there is little guidance for educators who wish to optimize their use of performance-based assessments for learning purposes. The aim of this study, therefore, was to explore how learning is promoted or impeded in the context of a formative OSCE during residency. To do this, we administered surveys to residents at three time points that align with a chronologically-based framework (Dochy et al. 2007) regarding the phases during which assessment can have an influence on learning: before an OSCE (pretest effects such as when tests incentivize studying); during or immediately following an OSCE (pure test effects, which represent learning derived from the testing experience itself); and following the release of the OSCE results (post-test effects such as learning engendered by feedback on one’s performance).

In fulfilling this aim, our purpose was not only to explore how a formative OSCE might influence learning in these three phases (and how learners’ perceptions about the value of the OSCE might change over time) but also to determine if and how high and low achievers differ in their approaches to learning in this context. Specifically, we wondered if high and low achievers might differ regarding their: goals for the OSCE; perceptions of the value of the OSCE; approaches to studying; and perceptions regarding how to score well.

Methods

A formative OSCE was administered to Internal Medicine (IM) residents in postgraduate years (PGY) 1–4 at the University of Ottawa as a progress test (i.e. a comprehensive examination administered to learners at different stages of training). Five simultaneous circuits of nine stations were run twice in the same evening to accommodate all residents. This was a mandatory event, but participation in the study was voluntary. Approval for the study was granted by the Ottawa Health Science Network Research Ethics Board.

OSCE progress test format

The OSCE consisted of four physical examination cases that involved observed interactions with standardized patients, four structured oral cases in which a candidate interacted with a physician examiner to gather information and demonstrate an approach to a problem, and one communication/ethics case. Although all stations involved the demonstration of effective communication skills, the latter case was specifically designed to address aspects of communication by having residents demonstrate their capacity to communicate in a challenging situation through interaction with a standardized patient. The blueprint for the OSCE was based on the objectives of training for IM residents established by the Royal College of Physicians and Surgeons of Canada (RCPSC 2011). IM residents from PGY 1-4 all participated in the same OSCE. Prior to the OSCE, examiners were provided with a 30-min orientation during which they were provided with an overview of their role, an introduction to the station types, a review of the rating instruments used, and guidance regarding how to deliver feedback.

Stations were 12 min in length, accompanied by 1 min of verbal feedback. While a single minute may not seem like a long time, our previous research in this context has demonstrated that examiners provide an average of 16 feedback points to residents in 1 min (Humphrey-Murto et al. 2016). Residents were scored by physician examiners using a combination of station-specific checklists and a number of five-point rating scales (e.g. organizational skills, rapport with patient) that were selected based on station type. Additional details about these rating instruments have been previously published in Pugh et al. (2014).

Participants

Participants were recruited from a list of IM residents (PGY1-4) who registered to take the OSCE (n = 86).

Surveys

Surveys were chosen as an expeditious method through which to gather exploratory data from a relatively large number of participants. Participants were asked to complete three surveys: before the examination (pre-examination), immediately following the examination (postexamination), and after the examination, results were made available (postresults); see Supplementary Appendix. Survey questions were developed iteratively using the literature on assessment-driven behavior as well as the researchers’ own experiences. The goal of each question was to explore participants’ perceptions regarding the educational value of the OSCE (AFL vs. AOL) as well as the OSCE’s influence on
learning (in both positive and negative ways). Questions were drafted, debated and revised by all members of the research team until consensus was achieved regarding the interpretability and relevance of each item. Four selected response questions were deliberately written in a manner that allowed them to be repeated on all three surveys to enable understanding of how perspectives on learning might change over the course of the examination cycle (e.g. if poor performance leads to more negative impressions of the learning experience). The final surveys consisted of 15, 16 and 23 items on the pre-examination, postexamination and postresults surveys, respectively.

**Administration**

Surveys were distributed using an online survey service (Fluid Surveys®) and included both selected-response (i.e., Likert-like rating scales) and constructed response (i.e. narrative comments) questions. The pre-examination survey was distributed approximately one week prior to the examination out of anticipation that residents began thinking about the examination deliberately only when it became imminent. The postexamination survey was distributed the day after the examination to try to gather their perceptions about their experience in the OSCE itself with minimal delay (and accompanying risk of forgetting). Administering the survey the evening of the OSCE was infeasible due to the logistics of running the OSCE for a large number of residents. The post-results survey was distributed the day after the release of examination results, one month after the OSCE, to again capture residents’ perceptions as soon as possible after they received formal notification of their results.

**Analyses**

Responses to selected response items were treated as continuous variables with scores ranging from 1 to 5 (e.g. strongly disagree to strongly agree). Negatively worded items were recoded to align with positively worded items. Descriptive statistics were calculated and responses for each question were grouped by themes. Themes were generated independently by the three coinvestigators with each item being coded and differences discussed after collaboration during a face-to-face meeting through an iterative process until consensus was reached. Using these themes, composite scores were derived by averaging the ratings given to questions within the theme.

Differences between high and low performers were explored using median splits on total OSCE score with residents classified as either high or low performers within their particular postgraduate year. We hypothesized that low achievers would adopt different learning strategies when compared to high achievers. A mixed design ANOVA was used to examine differences in low and high performers’ ratings as well as changes in those ratings over the examination cycle (i.e. to see if participants’ perceptions changed after the OSCE and/or after they received their results). Where assumptions of sphericity were violated, Greenhouse–Geisser corrections were used. Effect sizes were calculated using partial eta squared ($\eta^2_p$) with 0.01, 0.06 and 0.14 being treated as small, medium, and large effect cutoffs, respectively (Cohen 1988). *Post hoc* independent t-tests using Fisher’s LSD were used to further explore any differences found.

**Thematic analysis,** namely the framework method, was applied to responses to open-ended comments as a supplement to the quantitative findings that served as the primary focus of this study (Gale et al. 2013). The research team reviewed all data to become familiar with the dataset, then the data were coded and an analytical framework was developed and applied to all data to allow for interpretation.

**Results**

Of the 86 residents who participated in the OSCE, 39 consented to participate and 35 (41%) completed all three surveys. One participant dropped out because he missed the OSCE due to illness. Three consented, but did not complete all three surveys and so their data were excluded from analysis.

Participants were relatively evenly distributed across all levels of training, as expected based on the distribution of residents enrolled in the OSCE; (Table 1). The mean overall OSCE score for participants ($m = 65.0, SD = 7.7$) was not significantly different than the mean for non-participants ($m = 63.0, SD = 7.9$) ($t = 1.168, df = 84, p = 0.246$).

<table>
<thead>
<tr>
<th>Resident level</th>
<th>N</th>
<th>Mean overall OSCE score</th>
</tr>
</thead>
<tbody>
<tr>
<td>PGY-1</td>
<td>10</td>
<td>58.1</td>
</tr>
<tr>
<td>PGY-2</td>
<td>9</td>
<td>65.2</td>
</tr>
<tr>
<td>PGY-3</td>
<td>9</td>
<td>67.5</td>
</tr>
<tr>
<td>PGY-4</td>
<td>7</td>
<td>71.4</td>
</tr>
<tr>
<td>Combined</td>
<td>35</td>
<td>65.0</td>
</tr>
</tbody>
</table>

**Theme generation for selected response questions**

Four selected response questions (each with multiple statements that participants rated using a five-point Likert-like scale) were repeated across all three surveys (Supplementary Appendix). For each question, statements were divided into themes after coding by the coinvestigators. The validity of these themes was then explored by considering their relationship to the narrative comments participants offered in response to open-ended questions.

For the question asking participants about their goals for the OSCE, three themes were identified: performance-oriented goals (i.e. goals indicating the participant wanted to score well), mastery-oriented goals (i.e. goals indicating the participant wanted to improve their clinical skills) and receiving feedback (i.e. goals indicating the participant was seeking guidance about their performance from the OSCE).

Items related to participants’ perceptions about the value of the OSCE were coded into the themes of: valid means of assessment and educational value. Items associated with both the question about the types of study activities used to prepare for the OSCE and the question about participant perceptions of how to score well on the OSCE were divided into two themes: content-focused and test-taking strategies.

**Goals for the OSCE**

In terms of participants’ goals in anticipation of the OSCE, participants’ comments aligned well with the three themes...
Performance-oriented goals: Since it will be my first time doing the practice OSCE, my main goal is to familiarize myself with the format of a Royal College OSCE and hopefully improve the next time I take the exam. (C3Q; PGY-2)

Mastery-oriented goals: Since the OSCE is designed for the PGY-4 level in preparation for the Royal College, my goal is to learn from this experience as an R1 and not expecting (sic) to pass the exam. (A1P; PGY-1)

Desire for feedback: I would like feedback on my ‘exammanship’ (sic) and Royal College preparation from that point of view, rather than just content alone. (A1E; PGY-3)

Most participants claimed to have generated learning goals as a result of participating in the OSCE (n = 25, 71%) and more than half (n = 19, 54%) generated learning goals after receiving their results. The majority of learning goals listed related to mastering specific topics that had been covered on the OSCE:

- Practice pre op scenarios - especially around insulin, pain management and risk stratification as per guidelines. (C3J; PGY-4)
- Practice MSK exams multiple times for each joint before the OSCE. (A1D; PGY-4)

As shown in Table 2, both high- and low-performing residents wanted to perform well. Mean ratings related to performance-oriented statements were generally high, although they changed over time (F = 4.1, p = 0.03). Post hoc tests indicated that mean ratings were higher in the pre-examination survey (mean = 4.6, SE = 0.11) when compared to the postexamination (mean = 4.2, SE = 0.15) and postresults (mean = 4.3, SE = 0.10) surveys. No main effect of high versus low performers was observed (F = 0.52, p = 0.48) and occasion (i.e. pre-examination, postexamination, postresults) did not interact with performance group (F = 3.2, p = 0.07).

Ratings for mastery-oriented statements, indicating participants’ desire to use the OSCE to improve their clinical skills, went down immediately after the OSCE but then increased after participants received their results (F = 123, p < 0.001). There was no difference by performance level (F = 0.01, p = 0.91) and, again, occasion did not interact with performance level (F = 0.64, p = 0.53).

Ratings for statements related to the importance of receiving feedback were greatest in anticipation of the OSCE as these ratings decreased between the pre-examination survey and subsequent surveys (F = 8.7, p = 0.001), with no difference between performance level (F = 1.3, p = 0.26) and no interaction (F = 3.0, p = 0.07).

**Perceptions of the OSCE**

Despite the dynamic changes in participants’ responses regarding the goals they set over time, their ratings related to the validity of the OSCE were stable (F = 3.4, p = 0.05) as were their ratings of the educational value of the OSCE (F = 0.56, p = 0.57), as illustrated in Table 3. Again, these ratings did not differ by performance level (F < 0.50, and p > 0.45 for both themes) and occasion and performance level did not interact (F < 2.0, p > 0.15 for both themes).

When surveyed immediately following the OSCE, almost all participants (n = 34, 97%) reported that they had learned something. Most of the comments on the postexamination survey focused on what was learned in relation to improving test-taking strategies:

- I need to practice how to ask the proper questions. I found it difficult to have to jump on the HPI questions without any background information on the patient as this is contrary to what we do clinically. (A1M; PGY-3)
- Ways to prepare for Royal College; the styles of questioning; the styles of OSCE stations (e.g. two stations in one, with no buzzer to indicate halfway through). (B2H; PGY-1)
- Approach to Royal College OSCE - style of asking for specific history/exam findings and stating what I would expect instead of having the examiner tell me everything. (B2N; PGY-2)

In contrast, when asked about what they had learned from the OSCE following the release of the results, participants’ narrative comments highlighted the fact that the OSCE had allowed them to identify significant gaps in their knowledge:

- Parkinson’s exam, B12 deficiency, approach to pre-op assessment. (B2N, PGY-2)

**Table 2. Goals for the OSCE.**

<table>
<thead>
<tr>
<th>Score</th>
<th>Mean pre-examination rating (SD)</th>
<th>Mean postexamination rating (SD)</th>
<th>Mean postresults rating (SD)</th>
<th>For main effect of survey occasion</th>
</tr>
</thead>
<tbody>
<tr>
<td>Performance oriented</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Low (n = 16)</td>
<td>4.6 (0.51)</td>
<td>4.3 (0.70)</td>
<td>4.0 (0.36)</td>
<td>F (df = 2.66) p = 0.111</td>
</tr>
<tr>
<td>High (n = 19)</td>
<td>4.6 (0.76)</td>
<td>4.1 (1.6)</td>
<td>4.5 (0.70)</td>
<td></td>
</tr>
<tr>
<td>Mastery oriented</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Low (n = 16)</td>
<td>4.4 (0.53)</td>
<td>2.6 (0.44)</td>
<td>4.0 (0.62)</td>
<td>123 &lt;0.001 0.788</td>
</tr>
<tr>
<td>High (n = 19)</td>
<td>4.3 (0.71)</td>
<td>2.5 (0.40)</td>
<td>4.1 (0.77)</td>
<td></td>
</tr>
<tr>
<td>Feedback</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Low (n = 16)</td>
<td>4.4 (0.57)</td>
<td>4.0 (0.69)</td>
<td>3.8 (0.60)</td>
<td>8.7 &lt;0.001 0.208</td>
</tr>
<tr>
<td>High (n = 19)</td>
<td>4.4 (0.39)</td>
<td>4.0 (0.52)</td>
<td>4.2 (0.45)</td>
<td></td>
</tr>
</tbody>
</table>

**Table 3. Perceptions of the OSCE.**

<table>
<thead>
<tr>
<th>Score</th>
<th>Mean pre-examination rating (SD)</th>
<th>Mean postexamination rating (SD)</th>
<th>Mean postresults rating (SD)</th>
<th>For main effect of survey occasion</th>
</tr>
</thead>
<tbody>
<tr>
<td>Valid means of assessment</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Low (n = 16)</td>
<td>3.4 (0.63)</td>
<td>3.6 (0.52)</td>
<td>3.6 (0.55)</td>
<td>3.4 &lt;0.05 0.093</td>
</tr>
<tr>
<td>High (n = 19)</td>
<td>3.5 (0.50)</td>
<td>3.7 (0.58)</td>
<td>3.6 (0.55)</td>
<td></td>
</tr>
<tr>
<td>Educational value</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Low (n = 16)</td>
<td>4.0 (0.51)</td>
<td>4.2 (0.45)</td>
<td>4.0 (0.35)</td>
<td>0.56 &lt;0.001 0.017</td>
</tr>
<tr>
<td>High (n = 19)</td>
<td>4.1 (0.61)</td>
<td>4.1 (0.60)</td>
<td>4.3 (0.65)</td>
<td></td>
</tr>
</tbody>
</table>
...Pre-operative cardiac risk using MET criteria… (B2P, PGY-1)

...the physical exam for vitamin B12 deficiency… (C3H; PGY-1)

When asked what factors were useful in guiding their learning, participants’ ratings were relatively neutral regarding examiner feedback (mean = 3.4, SD = 0.79). Narrative comments reflected a dichotomy in which some participants discounted the validity of examiner feedback, while others called for more feedback:

The main issue with the OSCE is that feedback is limited. It would be helpful to either have longer feedback sessions or have the staff demonstrate how to do the station properly after we do it so we can learn. (B2N, PGY-2)

For the most part, the OSCE evaluators are our staff, but some are random faculty members and I don’t know if they are the best people to evaluate us on those stations. (C3B, PGY-3)

It was not helpful to be told by the examiner in the ‘medical error’ station that the responses did not feel (sic)genuine - this was a subjective assessment and it was with an actor… so OSCE limitations need to be respected. (A1N, PGY-4)

Time spent on study activities

Although this was a formative examination, about half the participants (n = 18, 51%) reported spending time specifically preparing for the OSCE.

Narrative comments suggested that the OSCE had altered their study habits:

I have centered my studies on subjects I struggle more with. (A1M; PGY-3)

I have been reading a lot more on physical examination and doing more group studying than usual. (A1P; PGY-1)

The impending OSCE has influenced my study habits because of me to specifically prepare for the OSCE itself, instead of doing more general studying, reading around cases during the day or learning guidelines. I believe that with inexperience with the upcoming OSCEs and not knowing what to expect, we spend much more time trying to find out WHAT to study, instead of actually studying. (C3O; PGY-1)

The scores illustrated in Table 4 suggest that, with regards to time spent on study activities, differences existed between participants’ intentions and their experience. Pre-examination, when trainees were asked how much time they would plan to devote to studying for future OSCEs, their estimates came up again to the pre-examination level (mean = 2.3, SE = 0.25). These ratings did not differ by performance level (F < 0.6, p > 0.4 for both themes) nor did performance level and occasion interact (F < 0.6, p > 0.5 for both themes).

How to score well on the OSCE

Ratings related to participants’ perceptions that more content knowledge is required to score well on the OSCE changed over time (F = 5.0, p = 0.01). Post hoc tests indicated that this difference arose between the pre-examination (mean = 3.6, SE = 0.09) and postresults (mean = 3.9, SE = 0.10) surveys, as illustrated in Table 5. There was no main effect of performance level (F = 0.099, p = 0.75). However, there was an interaction between performance and occasion (F = 6.5, p = 0.005). Post hoc tests indicated that low performers’ impressions of the importance of content knowledge increased after the pre-examination survey, whereas high performers’ ratings were higher than low performers’ ratings pre-OSCE and remained high over time. There was no difference in participants’ ratings of the importance of test-taking strategies over time (F = 0.17, p = 0.81).

When asked how they had prepared for the OSCE, the most common study strategies involved improving content knowledge: reading around areas of content (n = 18), examining patients during regular clinical rotations (n = 13) and watching videos related to clinical examinations (n = 10). Strategies related to testmanship were less commonly expressed: review of old OSCE checklists (n = 5) and OSCE prep courses (n = 1).

When asked what, if anything, they would do differently to prepare for future OSCEs, participants’ comments reflected a need to improve both content knowledge and test-taking strategies:

I would definitely prepare and will start from now making a notebook of approaches to common topics in internal medicine. (B2T; PGY-1)

Practice sessions with friends and go over the royal college requirements. Do more reading on my own. (C3T; PGY-1)

Table 4. Time spent on study activities.

<table>
<thead>
<tr>
<th>Score</th>
<th>Mean pre-examination rating (SD)</th>
<th>Mean postexamination rating (SD)</th>
<th>Mean postresults rating (SD)</th>
<th>F (main effect) df = 2.66</th>
<th>p</th>
<th>np²</th>
</tr>
</thead>
<tbody>
<tr>
<td>Content related</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Low (n = 16)</td>
<td>2.9 (1.1)</td>
<td>9.6 (1.2)</td>
<td>2.4 (1.5)</td>
<td>9.6</td>
<td>&lt;0.001</td>
<td>0.225</td>
</tr>
<tr>
<td>High (n = 19)</td>
<td>2.0 (1.3)</td>
<td>1.3 (1.3)</td>
<td>2.2 (1.5)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Test-taking strategies</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Low (n = 16)</td>
<td>0.88 (0.96)</td>
<td>0.19 (0.54)</td>
<td>1.2 (1.5)</td>
<td>10.7</td>
<td>&lt;0.001</td>
<td>0.245</td>
</tr>
<tr>
<td>High (n = 19)</td>
<td>1.4 (1.3)</td>
<td>0.26 (0.65)</td>
<td>1.2 (1.4)</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Table 5. How to score well on the OSCE.

<table>
<thead>
<tr>
<th>Score</th>
<th>Mean pre-examination rating (SD)</th>
<th>Mean postexamination rating (SD)</th>
<th>Mean postresults rating (SD)</th>
<th>F (main effect) df = 2.66</th>
<th>p</th>
<th>np²</th>
</tr>
</thead>
<tbody>
<tr>
<td>Content knowledge</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Low (n = 16)</td>
<td>3.5 (0.55)</td>
<td>4.0 (0.45)</td>
<td>3.9 (0.62)</td>
<td>5.0</td>
<td>0.01</td>
<td>0.132</td>
</tr>
<tr>
<td>High (n = 19)</td>
<td>3.8 (0.35)</td>
<td>3.7 (0.68)</td>
<td>4.0 (0.61)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Test-taking strategies</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Low (n = 16)</td>
<td>4.0 (0.32)</td>
<td>4.0 (0.28)</td>
<td>4.0 (0.32)</td>
<td>0.17</td>
<td>0.81</td>
<td>0.005</td>
</tr>
<tr>
<td>High (n = 19)</td>
<td>3.9 (0.31)</td>
<td>3.8 (0.26)</td>
<td>3.9 (0.32)</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
I would spend more time practicing with my colleagues with regards to approaches to OSCE. I found that I got hung up in the details and did not place enough emphasis on having a good approach. (A1P; PGY-1)

Discussion

In an assessment for learning model, tests are viewed as valuable learning opportunities to be used to promote continuous improvement. While it is known that assessment can influence learning in both positive and negative ways, we do not yet fully understand how performance-based assessments promote or impede learning. The current study provides some preliminary insights into how a formative OSCE influences learning. For the purposes of this discussion, we have divided these into the pre, pure and post-test phases of assessment.

Pretest effects

Although it was a formative examination, the OSCE influenced study habits in the pre-assessment phase. Participants reported that the OSCE led them to engage in study behaviors aimed at both improving content knowledge and maximizing their test scores, which parallels what is seen in written tests (Cilliers et al. 2012). Interestingly, although learners are often categorized as being either performance oriented or mastery oriented (Chen et al. 2016), participants’ responses suggested that they viewed the OSCE as both a hurdle to overcome and an opportunity to learn and receive feedback, as reflected by their goals and study strategies.

This may not be surprising, given that medical residents have spent much of their academic career pursuing both knowledge and high marks in order to be successful. Therefore, performing well and mastering the material may more accurately be conceptualized as complementary goals that are more or less foregrounded in the minds of learners, depending on context. That learners appeared to view such a low stakes OSCE as both an assessment of their learning and an assessment for their learning suggests that there is considerable functional overlap between these constructs and that it may be artificial to think of assessment moments as necessarily falling into one category or the other. Certainly, there is literature to suggest that learners do not differentiate between formative and summative assessment the way educators do, as all assessment may be viewed as a hurdle to be overcome to some extent (Bok et al. 2013; Heeneman et al. 2015).

Pure test effects

The act of testing itself has been shown to lead to learning with written tests (test-enhanced learning). Several studies on the use of written tests have found that testing leads to more learning and better transfer of learning when compared to studying alone, but it is not clear if this transfers to performance-based tests (Butler 2010; Larsen et al. 2013). While there have been indications that similar effects occur as a result of sitting performance-based tests, that phenomenon is less well established (Kromann et al. 2009). In the current study, almost all participants reported that they had learned something by participating in this formative OSCE.

Based on their ratings, it appears that participants’ relative perceptions of the importance of performing well decreased over time, perhaps because the stress associated with the impending examination had been removed. In contrast, ratings of the importance of mastering the content decreased immediately following the examination and then bounced back up following the release of the results. This can be understood by examining participants’ responses to qualitative questions. Immediately following the examination, many of the narrative comments reflected the perception that the OSCE helped them to gain insight into testmanship strategies to allow them to be successful on future OSCEs. In contrast, when asked about what they had learned in the follow-up survey completed after the release of the results, participants’ comments tended to focus on content knowledge. This may reflect the fact that removal of the immediate threat of being assessed allowed participants to better focus on their learning rather than simply aiming to perform well, while also suggesting that any tendencies assessments have toward inducing performance-focused shortsightedness may be transient. It is not clear how this would differ with a summative OSCE.

Post-test effects

In the current study, participants were divided in terms of how they viewed the utility of the feedback provided, despite reporting that they learned something from the experience.

Some discounted the value of the feedback provided, while others sought more feedback. This raises questions about the extent to which examiner feedback is an important modulator of learning in a formative OSCE (Eva et al. 2010). Even those who did not highly value the feedback received, however, reported self-regulating their learning through the development of learning goals. This reinforces the notion that the incorporation of feedback by learners is a complex process influenced by many social and environmental factors (Telio et al. 2015).

High versus low performers

Surprisingly, there were no differences between low and high performers in terms of their goals for the OSCE, their perception of the value of the OSCE, or time spent on study activities. We had hypothesized that low performers might employ a more performance-oriented approach to learning or that they might have more negative views with regards to this form of assessment, but we did not find any evidence of this, suggesting that the leading influences of assessment practices may not be dependent on the overall proficiency of the examinee, at least in a formative context.

However, it is notable that high performers’ recognition of the importance of content-related knowledge in order to score well was stable throughout the examination period, whereas low performers’ ratings increased after the OSCE. This may suggest that, although low performers did not fully appreciate the value of content-related knowledge
initially, the OSCE experience may have provided them with some insights into their weaknesses.

Limitations
There are several limitations to this study. Because it included a relatively small sample from a single institution, with residents from a single program, the results may not be generalizable to other settings. Also, whenever research participants are volunteers there is risk of a selection bias, but our sample appears to have been representative of the overall cohort with comparable OSCE scores and PGY distribution. In addition, because web-based surveys comprised of pre-determined questions were used, data interpretation was limited due to it not being possible to ask follow-up or clarifying questions. This is certainly a threat to the validity of the interpretation of the results, but we are reassured by the consistency provided by the mixed methods analyses. In other words, the qualitative analysis of the narrative comments facilitated our interpretation of the quantitative findings even though follow-up inquiry, in the form of focus groups, interviews or other forms of member checking would be beneficial. Finally, this study was conducted in the context of a formative OSCE, so it is unclear if the results would apply to a summative setting.

Conclusions
Understanding how a formative OSCE drives learning can help educators to better harness its educational potential. This study represents a preliminary step in explaining when and how OSCE-related learning occurs. It would seem that the residents in this study valued the opportunity to complete an OSCE as a learning experience (AFL), but this was tempered somewhat by a desire to perform well (AOL) despite the fact that it was a formative, low stakes, examination. As has been shown with written tests (Cilliers et al. 2010), OSCEs seem to influence learning in both positive and negative ways in the pre- (i.e. by changing study habits) pure- (i.e. by directly leading to learning) and post-test (i.e. by leading to the development of learning goals) phases. However, our understanding remains incomplete. Questions remain, such as: How do learners’ balance conflicting goals related to mastery and performance in an OSCE? What features of an OSCE optimize direct learning?; and What variables influence the type of feedback that is most effective in stimulating learning in this setting? To those ends, future studies, using focus groups or interviews, aimed at gaining a deeper understanding of the role of OSCEs in promoting or impeding learning are warranted.

Disclosure statement
The authors report no conflicts of interest. The authors alone are responsible for the content and writing of this article.

Funding
This work was supported by Department of Innovation in Medical Education (DIME), University of Ottawa; Department of Medicine, The Ottawa Hospital.

Glossary
Formative assessment: A discussive form of assessment that guides future learning, provides reassurance, promotes reflection, and shapes future values.

Notes on contributors
Debra Pugh, MD, MHPE, FRCP, is an Associate Professor, Department of Medicine, University of Ottawa.
Isabelle Desjardins, MD, MMEd (candidate), FRCP, is an Assistant Professor, Department of Medicine, University of Ottawa.
Kevin Eva, PhD, is a Professor, Department of Medicine, University of British Columbia and Senior Scientist at the Centre for Health Education Scholarship (CHES).

References


