

# Effect of Detailed OSCE Score Reporting on Learning and Anxiety in Medical School

Journal of Medical Education and  
Curricular Development  
Volume 8: 1–8  
© The Author(s) 2021  
Article reuse guidelines:  
sagepub.com/journals-permissions  
DOI: 10.1177/2382120521992323



Vijay J. Daniels<sup>1</sup> , Silvia Ortiz<sup>2</sup>, Gurtej Sandhu<sup>1</sup>, Hollis Lai<sup>2</sup>,  
Minn N. Yoon<sup>2</sup>, Okan Bulut<sup>3</sup>  and Tracey Hillier<sup>4</sup>

<sup>1</sup>Department of Medicine, Faculty of Medicine and Dentistry, University of Alberta, Edmonton, Alberta, Canada. <sup>2</sup>School of Dentistry, Faculty of Medicine and Dentistry, University of Alberta, Edmonton, Alberta, Canada. <sup>3</sup>Department of Educational Psychology, Faculty of Education, University of Alberta, Edmonton, Alberta, Canada. <sup>4</sup>Department of Radiology and Diagnostic Imaging, Faculty of Medicine and Dentistry, University of Alberta, Edmonton, Alberta, Canada.

## ABSTRACT

**INTRODUCTION** There is growing literature on increasing feedback from Objective Structured Clinical Examinations (OSCEs) and one approach is a score report. The purpose of this study was to implement and evaluate a score report for a second and fourth-year medical school OSCE.

**METHODS** We developed an electronic OSCE score report that displayed comments and performance by domain within and across stations (checklist items and rating scales were tagged to each domain). Our initial pilot released the score report after pass/fail decisions but subsequent iterations released the score report the same day as the exam. Our evaluation approach included both student surveys and focus groups.

**RESULTS** Students felt the OSCE score report was accurate, identified strengths and weaknesses, and would likely cause them to take future action, with second-year students more likely to act on the report than fourth year students. The thematic analysis revealed barriers and enablers to utilizing feedback as well as the power of the score report to reduce anxiety.

**CONCLUSIONS** Our OSCE score report was simple to develop and implement the same day as an OSCE with an overall positive response from students with respect to accuracy and ability to use the information for future learning.

**KEYWORDS:** OSCE, Computer-based, Feedback, Clinical Skills, Undergraduate

**RECEIVED:** October 8, 2021. **ACCEPTED:** January 12, 2021.

**TYPE:** Original Research

**FUNDING:** The author(s) disclosed receipt of the following financial support for the research, authorship, and/or publication of this article: This work was supported by the University of Alberta's Teaching and Learning Enhancement Fund with Dr. Daniels as Principal Investigator and Drs. Lai, Bulut, Yoon, and Hillier as Co-Investigators.

**DECLARATION OF CONFLICTING INTERESTS:** The author(s) declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

**CORRESPONDING AUTHOR:** Vijay J. Daniels, Division of General Internal Medicine, University of Alberta Hospital, 5-112 Clinical Sciences Building, 11350 83rd Avenue NW, Edmonton, AB T6G 2G3, Canada. Email: vdaniels@ualberta.ca

## Introduction

Health professions education is often focused on the Assessment *of* Learning via summative assessments. Health programs have various assessment formats for this purpose including written exams, simulation-based assessments such as Objective Structured Clinical Examinations (OSCEs), and workplace-based assessments such as the Mini-CEX.<sup>1,2</sup> While these formats facilitate assessment across many domains, they often place the learner solely as the object of the process, instead of recognizing the learner as a user of assessment outcomes. These commonly used assessment formats may fail to provide meaningful and timely feedback that the learner can utilize for subsequent clinical encounters. More recently, we are recognizing the importance of Assessment *for* Learning, where the learner becomes a more active user of assessment data.<sup>3</sup> In order to facilitate this transition, health professions programs are shifting from infrequent high-stakes assessments to more frequent low-stakes assessments. An example of this is the transition to Competency Based Medical Education (CBME) where the same assessment data are used for both summative

and formative purposes.<sup>4</sup> When formative feedback is timely, detailed, and specific, it can guide future learning, motivate learners to investigate other resources, and ultimately improve overall learning.<sup>5,6</sup>

Despite recognizing the importance of assessment for learning, many programs around the world continue to use infrequent higher-stakes assessments such as OSCEs because they are more likely to produce reliable results and can be organized effectively for larger class sizes.<sup>7</sup> However, OSCEs are a source of anxiety for learners and a study by Guraya et al<sup>8</sup> demonstrated that OSCEs were second only to the traditional long case as the most anxiety-provoking assessment amongst various assessment modalities. This anxiety around OSCEs is important as anxiety along with limited feedback, which is common to OSCEs, can reduce learners' ability to connect their performance on an OSCE to future practice.<sup>9</sup> The availability of feedback from OSCE is limited by slower data turnaround time<sup>10</sup> and difficulty in generating and providing feedback to learners due to logistical and transcribing issues.<sup>11,12</sup>



Recently, many studies have investigated using electronic devices to capture OSCE scores (often referred to as eOSCE) to overcome the limitations of the traditional OSCEs, and potentially invoke behavioural change in learners, such as proactively seeking feedback about their performance and making meaningful changes to improve their future performance.

Denison et al<sup>11</sup> collected examiner-recorded comments from successive first-year medical school formative and summative OSCE examinations and found the eOSCE format provided more timely feedback that was greater in quantity and quality. Munro et al<sup>12</sup> found similar findings in a high-stakes OSCE. However, these studies are limited in that they did not examine if learners used the feedback they were given. In contrast, a study by Harrison et al<sup>13</sup> had examiners use tablets to assess medical students and record audio feedback during an eOSCE. The students were allowed to listen to the audio feedback shortly after the examination. When polled later in the year, 68% of the students who responded stated that they had changed the way they performed a skill based on their audio feedback.

Another form of OSCE-related feedback, other than written or audio feedback, is providing learners with the actual score sheets. A study by Bernard et al<sup>14</sup> found that in a formative OSCE, students variably used feedback in the form of score sheet review, but those who did use them for their first year OSCE, went on to score better in their second year OSCE. A recent study by Daniels et al<sup>15</sup> was conducted using post-graduate trainees in an Internal Medicine residency program who participated in various eOSCE examinations. Following the eOSCE, the trainees were able to view examiner comments and their score sheets (checklists and rating scales) and then were asked to write a learning plan after viewing their score sheets and feedback comments. The study found that this process led to a change of behaviour specific to the learning plan.

However, providing score sheets with case-specific checklist items to learners after exams is not always an option for programs due to exam security. Therefore, an alternative is to provide a score report with information across domains. An effective score report should allow learners to make correct and appropriate inferences from scores to take further actions, and should be easily customizable and automated so that learners can receive unique and immediate feedback regarding their performance.<sup>16</sup> What is not clear from the current body of literature is if providing an eOSCE score report with written comments without the actual score sheet could contribute to a change in behaviour for learners similar to the aforementioned studies.

The purpose of this study is to develop, implement, and evaluate the impact of a score reporting system within an eOSCE environment that can provide learners with timely, detailed, and focused feedback to facilitate their future learning goals, while still maintaining the security and integrity of the exam content.

## Methods

### *Study setting*

The MD program at the University of Alberta has approximately 160 students in each year of its four-year program. Students take a variety of OSCEs throughout the program, but the largest ones are the end-of-Year 2 Physical Exam OSCE (12 stations), and the Year 4 Comprehensive OSCE (11 stations, history or physical exam) that occurs in January of the fourth year to allow enough time to remediate those who fail the OSCE prior to graduation in May. All OSCEs are scored on a locally developed eOSCE system.<sup>17</sup>

### *Score Report*

Score sheets use a combination of checklist items, focusing primarily on clinically discriminating items,<sup>18</sup> as well as rating scales adapted from the Medical Council of Canada's published patient rating scales (MCCQE).<sup>19</sup> Checklist items for history stations that had a diagnostic focus and physical exam stations were either tagged to "essential to the case" (items that discriminated between conditions on the differential diagnosis), or tagged to "routine" (items all physicians would routinely do in that context). History taking stations used rating scales to assess Questioning skills, Listening skills, Organization, Rapport/Empathy, and Information Giving. Physical Exam stations used rating scales for Organization, Attention to Patient's Comfort, and Technical Skills.

Each rating scale was considered a tag and hence history taking stations had seven tags (two for the checklist items and one for each of the five rating scales), and physical exam stations had five tags (two for the checklist items and one for each of the three rating scales).

To allow an immediate, automated report, we used cut points of 1.5 standard deviations (SD) above and below the mean. We chose 1.5 SD because when we ran a pilot with old test data with 2 SD as the cutoff, very few students would receive the feedback message, and so we adjusted it to 1.5 SD. Students were informed the cut points were norm-referenced for generating the feedback message. When a student scored more than 1.5 SD above the mean for a domain, they saw a blue box with "Did Very Well, Good Job!"; when a student scored less than 1.5 SD below the mean for a domain, they saw a yellow box with "Area for Improvement"; when the student scored between 1.5 SD below to 1.5 SD above, they saw a green box with "On Track". Students could view the score report for individual stations and for domain performance across stations. At the station level, students were shown comments provided by the examiner, but the comments were not necessarily related to the domains that were flagged based on SD.

The score report was first implemented after the January 2018 Year 4 OSCE. The score report system was originally set for release after students had their official pass/fail marks

**Table 1.** Survey statements and Kirkpatrick outcomes.

STATEMENT [SHORT DESCRIPTOR]	KIRKPATRICK STAGE
"My OSCE score report provides an accurate reflection of my performance on the Year 4 OSCE." [Accurate Information]	Satisfaction
"My OSCE score report provides me with sufficient information to identify the specific clinical skills domain(s) (e.g., organization, questioning skills, etc. . .) in which I did well." [Identified Strengths]	Satisfaction
"My OSCE score report provides me with sufficient information to identify the specific clinical skills domain(s) (e.g., organization, questioning skills, etc. . .) in which I need to improve." [Identified Areas for Improvement]	Satisfaction
"As a result of viewing my OSCE score report, I am more likely to review one or more content areas." [Review Content]	Action-Knowledge
"As a result of viewing my OSCE score report, I am more likely to change my approach to certain clinical skills domain(s) (e.g., organization, questioning skills, etc. . .)." [Change Approach]	Action-Behaviour
"As a result of viewing my OSCE score report, I am more likely to get help from residents and/or attending physicians with certain clinical skills domain(s) (e.g., organization, questioning skills, etc. . .)." [Seek Help]	Action-Behaviour

(which the score report itself does not provide). This was done to ensure students interacted with it solely for formative purposes but resulted in a four-week delay before the score report was released as a committee had to finalize marks given the high-stakes nature of this exam. Based on feedback from students after the first iteration, we decided to release the score report immediately after the subsequent three OSCEs before students received their official pass/fail result.

### Survey Feedback

In an email that provided instructions on how to access the score report, students were provided an information letter and invited to fill out a six-item survey built around Kirkpatrick's<sup>20</sup> outcomes framework (see Table 1 for the survey questions) administered in Google Forms. For each statement, students were asked to rate their agreement on a six-point Likert scale: 1=Strongly Disagree, 2=Disagree, 3=Slightly Disagree, 4=Slightly Agree, 5=Agree, 6=Strongly Agree. There was also an open text box with the prompt "Any comments (good or bad) about the OSCE score report?". There was no incentive provided to complete the voluntary survey and the information letter and survey stated clearly consent was implicit by submitting the survey.

### Focus Group Feedback

Subsequent to the above survey questions, students were invited to provide their email address if they were willing to be contacted by our research team for a follow-up focus group. A research assistant not involved in the education or assessment of the students recruited medical students after each OSCE with a small lunch as the incentive. An information letter was provided at the focus group with time for questions and then written consent was collected from each participant including permission to record the participants. Focus group participants were asked to elaborate on each of the six survey items (see Table 1). Focus groups were audio recorded and then transcribed.

### Data analysis

Two-tailed independent samples t-tests were conducted using SPSS 26.0 (IBM)<sup>21</sup> to compare if there was any difference between the 2018 and 2019 Year 2 OSCE Score Report survey responses, and between the 2018 and 2019 Year 4 OSCE Score Report survey responses prior to combining all survey results specific to an OSCE. We then used a one-sample t-test to compare survey item scores to the neutral Likert scale score of 3.5 (half way between Slightly Disagree and Slightly Agree). A one-tailed independent t-test was conducted to see if Year 2 students were more likely to act on feedback than Year 4 students (our presumption being an OSCE earlier in the program would be more likely to be used for further learning). Two authors (VJD and GS) conducted a thematic content analysis on narrative data from the open-ended survey question and the focus group data first by coding independently and then differences were resolved by consensus.

### Ethical review

The University of Alberta Research Ethics Board approved all of the above data collection and analysis procedures (reference number: Pro00076415; November 17, 2017).

## Results

### Assessing Score Report

On average, 97% of students viewed their score report with the vast majority of views in the first week. In addition to initial views, 22% of students came back to their score report more than 30 days after being released with some coming back up to 16 months after the OSCE.

### Quantitative Survey Results

Of the 616 students who viewed their score report, 329 (53%) responded to the survey with response rates amongst cohorts

**Table 2.** Accessing score reports.

OSCE (YEAR)	TOTAL # STUDENTS	# STUDENTS WHO VIEWED REPORT (%)	# STUDENTS WHO VIEWED SCORE REPORT AGAIN AFTER 30 D (%)	# STUDENTS WHO RESPONDED TO SURVEY
Year 2 OSCE (2018)	164	161 (98%)	30 (19%)	77 (48%)
Year 2 OSCE (2019)	155	153 (99%)	23 (15%)	62 (41%)
Year 4 OSCE (2018)	154	144 (94%)	38 (26%)	109 (76%)
Year 4 OSCE (2019)	161	158 (98%)	45 (28%)	81 (51%)
All OSCEs	634	616 (97%)	136 (22%)	329 (53%)

**Table 3.** Comparison between years.

SURVEY ITEM	YEAR 2 2018 MEAN SURVEY RESPONSE (SD)	YEAR 2 2019 MEAN SURVEY RESPONSE (SD)	TWO-TAILED T-TEST P	YEAR 4 2018 MEAN SURVEY RESPONSE (SD)	YEAR 4 2019 MEAN SURVEY RESPONSE (SD)	TWO-TAILED T-TEST P
Accurate Information	4.82 (0.88)	4.89 (0.70)	.62	4.81 (0.57)	4.72 (0.86)	.40
Identified Strengths	4.58 (1.27)	4.61 (1.23)	.89	4.50 (1.00)	4.60 (1.09)	.51
Identified Areas for Improvement	4.61 (1.19)	4.44 (1.23)	.40	4.36 (1.03)	4.59 (1.17)	.15
Review Content	4.61 (1.18)	4.76 (1.17)	.46	4.40 (1.20)	4.37 (1.31)	.86
Change Approach	4.39 (1.36)	4.50 (1.18)	.62	4.10 (1.09)	4.32 (1.29)	.21
Seek Help	4.44 (1.32)	4.53 (1.21)	.68	3.86 (1.13)	3.90 (1.20)	.82

**Table 4.** Comparison of Year 2 and Year 4 score report perceptions.

SURVEY ITEM	YEAR 2 MEAN SURVEY RESPONSE (SD)*	YEAR 4 MEAN SURVEY RESPONSE (SD)*	ONE-TAILED T-TEST FOR YEAR 2 > YEAR 4
Accurate Information	4.85 (0.81)	4.77 (0.71)	0.17
Identified Strengths	4.60 (1.25)	4.55 (1.04)	0.35
Identified Areas for Improvement	4.53 (1.21)	4.46 (1.10)	0.28
Review Content	4.68 (1.18)	4.39 (1.25)	0.02
Change Approach	4.44 (1.28)	4.19 (1.18)	0.04
Seek Help	4.48 (1.27)	3.88 (1.16)	<0.01

\*Note all mean scores were statistically significantly different than the neutral score of 3.5 on the Likert scale ( $P < .001$ ).

varying between 41-76% (see Table 2). There was no difference in average survey responses between the 2018 Year 2 students compared with the 2019 Year 2 students (p-values ranged 0.4-0.89), nor between the 2018 Year 4 students and the 2019 Year 4 students (p-values ranged 0.15-0.86) (see Table 3). Therefore, 2018 and 2019 data were combined for further analyses comparing Year 2 and Year 4 students. The mean agreement with the six statements ranged from 3.9 to 4.9 and all were statistically significantly different than the neutral score on the Likert scale of 3.5 ( $p < 0.001$ ) (see Table 4). For the three questions related to satisfaction with the score report's accuracy and ability to identify strengths and

areas for improvement, there was no difference between Year 2 students and Year 4 students. However, for statements regarding taking action related to seeking out knowledge or changing behaviour, Year 4 students were statistically less likely to use the score report for this purpose.

### *Thematic Analysis*

Out of 329 survey respondents, 193 (59%) left comments in the open comment box. We conducted two focus groups after each OSCE and across the eight focus groups we had a total of 30 participants. Initial coding from the focus groups matched

those from the survey comments so all narrative data were combined for the thematic analysis. Our thematic analysis identified five themes (note comments are annotated based on year of administration and level of OSCE such that 2019-Yr-2 indicates a comment came from the 2019 administration of the Year 2 Physical Exam OSCE):

*Theme 1: Timeliness of feedback mostly appreciated.* Many students commented on the positive aspect of immediate score report feedback.

*I love that . . . it's in the same day, so I can vividly remember what went on in each station. And you know, sometimes it's just good to get that immediate feedback. Then you can actually reflect and say 'yeah, ok, that makes sense so going forward, I'm going to pay attention to these two or these three [comments]. . . (2018-Yr-2)*

*. . . I think that kind of reinforced those stations as, yes, that's something I should work on, going forward, more strongly than if it was just wait until the end of the week, then you get your score report. (2019-Yr-2)*

Conversely, another student indicated that feedback was more useful to look at later and not immediately after the exam.

*I wish I had been prepared for the fact that it was mostly constructive criticism, because it was tough to look at it right before bed and then try to fall asleep, but then when I looked at it the second time, a week or so later, when I was ready for the feedback, then that was a lot more helpful. (2019-Yr-4)*

*Theme 2: Timing of the OSCE may impact utilization of feedback.* While students liked the immediate feedback, another theme identified was the timing of the OSCE as a barrier to utilizing the feedback. For some, it was not close enough to when they would apply it.

*. . . even if I did have some areas that . . . were not [on track], I would probably not think about it for the rest of the summer, and by the time that the new year rolls around (laughs). . . (2019-Yr-2)*

For others, receiving the feedback during a stressful period was a barrier to using it.

*I think it could be used as a tool, later on. Right away, no, 'cause, contextually, our week – that week was so kooky busy that, we had other acute fires to kind of put out and exams to write, but I think if there had been a latent area that had been flagged or something, then it could have been useful to, go back and have that to look back on. (2019-Yr-2)*

*I felt like there were definitely areas of improvement for me, but I don't know what the right timing is or if there is a right timing, but the timing of the OSCE and the feedback and [residency interviews] and finishing clerkship . . . there is no way I was going to review anything related to the OSCE now. (2019-Yr-4)*

*Theme 3: Comments are useful when they are specific and correlate to either the domain flags and/or and perceived OSCE experience.* Students provided examples of specific comments they were provided and why they were useful.

*. . . I had the one comment 'don't lift your fingers in a lymph node exam', – but because that was a really specific comment – then I'm like, okay, maybe I could ask someone to show me how to do a lymph node exam again (2019-Yr-4)*

For some students, the lack of specificity was an issue.

*I scored low on the [ . . . ] station for specific questions pertinent to that case, but there were no comments on my OSCE letter describing the specifics of that. I was actually most keen to see what those specifics were that I missed, but couldn't find them anywhere. I'm not sure if releasing the check lists would help with that, but at the very least, examiners should be writing comments for each station, especially if there are short comings. (2018-Yr-4)*

*. . . not so much like 'don't flag me' but if I'm flagged, it would be helpful to have a little bit of more clarity of what I can work on. Like, bruise my ego, that's fine, but just help me with what to fix so that I can go work on it. (2018-Yr-2)*

Students acknowledged when the feedback comments were congruent with the score report and their perceived OSCE experience, they were much more likely to utilize the feedback to improve clinically. Statements like these were more common from Year 2 students.

*. . . [if] the notes or the comments that were given were kind of incongruent with whatever the report had, I guess I'd be less likely to use the report to kind of direct where I should go 'cause I put, definitely, much more stock in the comments, but I think that if the comments and the report did line up, I do think the report would add kind of an extra useful piece. . . (2019-Yr-2)*

*I felt that the overall feedback aligned with how I felt I did on the stations. I got stations where I struggled; I saw that it said, you know, needs improvement or whatever else. On stations where I felt I did really well, like I felt like it correlated well. It gave me positive reinforcement and also directed me where I might need to work on some things. . . (2019-Yr-2)*

Some students found there was a discrepancy between the score report and their perception of the OSCE with their perception having a greater impact on what they would do. Such statements were more common from Year 4 students.

*I didn't feel like my thoughts on how I did in the OSCE correlated very well with what my score report said, and then I'll probably base my improvement more based off of what I thought I experienced in my OSCE than the score report, 'cause some of the score report things were very either very, very specific, and yeah, I missed that and I'll change that, or kind of, too vague to really go up to a specialist and go 'hi, how can I work on communication?' or something like that. Like, for example, I thought my coma one was done poorly in my OSCE and I might work on that – in patients who are comatose – with someone there, or someone more senior to give me feedback, but I don't think my score report reflected that. (2019-Yr-4)*

*Theme 4: The score report was first used to flush out pass/fail which had variable effects on anxiety.* For the last three groups who received their score report prior to pass or fail decisions, there

was a clear signal that students were trying to figure out if they passed or failed in order to lower anxiety.

*. . . and I know the goal is to, like, okay, where are we struggling, or where can, say, I improve? But even when I'm looking at [the score report], I'm like, is this a pass or a fail? Like, that's the first thing on my mind. (2019-Yr-4)*

*And I think that we got all that feedback, and it was all divided into colors or whatever, and then it was immediately trying to decipher the colors to see if you passed. (2019-Yr-4)*

*. . . it also quelled a lot of worries I had because I was particularly worried about a few stations, and when I got it back and saw that everything was on track for those stations, I was like, okay, you can stop panicking now . . . and so having the report, like the immediate feedback . . . was integral to being able to get through the next few days of other assessments. (2019-Yr-2)*

*I like knowing if I'm on track or not. I think this is one of the most significant sources of anxiety as a medical student. We compare ourselves to residents and staff without ever really knowing what we're actually expected to know. From a mental health and resiliency point of view I think this breakdown with the added "on track vs need to work on" will help alleviate some of these insecurities. (2018-Yr-4)*

But for students who could not decipher if they passed or failed, this increased anxiety.

*I think maybe some ambiguity if you were to have multiple 'could work on's', or like kind of could almost instill more anxiety if it's not – if you don't understand how to interpret. (2019-Yr-2)*

*. . . like what I said earlier was it honestly stressed me out more than it made me feel at ease because the colours are good, but these comments I really don't know. (2019-Yr-4)*

## Discussion

The purpose of this study was to examine the impact on student learning of receiving an eOSCE score report, especially those features that helped learners use the feedback. Our first iteration was intentionally delayed until after pass/fail decisions were communicated to students given research warning that releasing a report before the summative decision will cause students to focus on whether they passed or failed rather than on how they can improve.<sup>22</sup> However, based on feedback from students collected independent of this study by the program, we were advised to release the score report immediately following the OSCE. Receiving the score report immediately was one of the most applauded strengths of this project (see Theme 1). But unsurprisingly, we did see the expected comments about interacting with the score report to figure out if they passed or failed.

That said, the survey results did not show any difference between the 2019 Year 4 students who received their score report before the pass/fail decision and the 2018 Year 4 students (see Table 3). This reassures us that even if students start engaging with their score report to find out if they passed or failed, they still benefited from the report for the reasons we

had hoped (e.g., behavioural change for enhancing future performance), and hence the added value of immediate feedback is worth the possible distraction of trying to figure out if they passed or failed.

Another somewhat expected finding was the impact of immediate score reporting on anxiety, especially when the score report was released before the pass/fail determination. Many students used the score report to calm themselves before they had to tackle other high-stakes events around that time, such as other high-stakes assessments, or for our Year 4 OSCE, their residency interviews. This is consistent with previous literature in both health professions education<sup>23</sup> and broader education literature<sup>24,25</sup> in which the score report itself lessened post-exam anxiety beyond the relief of completing the exam. While there is literature on the impact of test anxiety on performance in an OSCE with a recent systematic review by Martin and Naziruddin,<sup>26</sup> the literature focuses on how pre-OSCE anxiety does or does not impact performance. Our study demonstrated that a post-OSCE score report can lessen anxiety for many students prior to other high stakes events such as the next high stakes assessment or residency interviews. We would like to think that less anxiety for these events would be beneficial, but this would be conjecture.

But while the report was able to lower anxiety for many students, a key finding is how such a report can potentially worsen anxiety when there is discrepant information between the score report and the comments. This corroboration between data sources and the importance of multiple sources is a staple of good feedback. In a recent scoping review on best practices for effective feedback, Ossenbert et al<sup>27</sup> identified "multiple forms and sources of evidence" as one of 11 feedback attributes that support uptake of feedback. Hence when it is present, it is highly valued.

Another key finding is who will use and when they will use the score report. As illustrated in Table 4, the second-year students were statistically more likely to change their behaviour in response to the score report, compared to the fourth-year students, especially with respect to seeking help from more senior physicians. Although it did not come out as a distinct theme, one comment from a year 4 student alluded to a greater trust of their memory of the OSCE than their score report. While this seems to devalue the score report, we would argue that this is still a benefit of the entire OSCE process as it stimulates future growth, but more senior students may prioritize this self-assessment over data from an OSCE score report. This would be supported by other research. Sargeant et al<sup>28</sup> examined what learners focus on in assessment data to make an informed self-assessment. In their work, they found that for OSCEs, there was often a sense that the standards in the exam were not reflective of real life, and hence why it is quite possible more experienced students who have experienced real life medicine in their clerkship, might value their own construction of the OSCE experience over a score report to help determine a plan going forward. However, again, we did not have a lot of comments attesting to this so we are careful not to infer too much.

Finally, the timing of when to interact with a score report and motivations behind this would prove an interesting area of future study. As one student stated, “*when I looked at it the second time, a week or so later, when I was ready for the feedback, then that was a lot more helpful*” (2019–Yr-4). We found students returning to their score reports up to 16 months after their OSCE, suggesting the optimal time to use it may be both shortly after an OSCE while the context specific deficiencies are fresh in one’s memory, but also again weeks to months later when one can reflect on their experiences since the exam. This is a strength of a score report that can be accessed later.

Overall, despite differences between Year 2 and 4 students, students in all cohorts reported that they would likely use the eOSCE score report to change behaviour (see Table 4). Most studies that look at the effects of educational interventions tend to focus on Kirkpatrick level 1 (satisfaction) and 2 (knowledge),<sup>29</sup> but some argue that we should aim for level 3 (behaviour),<sup>30</sup> and our study suggests an eOSCE score report may be used to impact future behaviour. We are only aware of two previous studies that have shown a change in behaviour from feedback related to an OSCE.<sup>13,15</sup> What is common in these studies and ours is that the learner was involved in determining the feedback that was important to them. This is described as the essential characteristic of “meaning-making” by Garino<sup>31</sup> in her study exploring successful use of feedback. While our goal was to provide accurate OSCE feedback that helped guide future learning, we hand over the baton to the learner at this stage for them to decide how they will use the information.

#### *Limitations and Future Directions*

This study was conducted at one university in one medical school program and as such, generalizations to other programs may be limited. Second, our study did not investigate how the process of taking the OSCE alone may motivate change compared with an OSCE that added the score report. Lastly, we did not follow-up with our students to see if they did change their behaviour as they had intended. Future research could examine a similarly structured score report in different contexts and/or timing in the curriculum and in different institutions, could ask motivation to change before and after seeing the score report, and could follow students to ask how they used the information months later, and/or how they would perform on subsequent OSCEs as compared to historical controls.

#### *Practical Implications*

Despite the above limitations, our study has revealed some key issues that should be kept in mind when developing an OSCE score report. First, rater training is paramount so that examiners give specific, behavioural-based comments. Second, the timing of not only when feedback is available relative to the OSCE, but the timing of the OSCE and feedback relative to the educational context impacts uptake. For example, OSCE

feedback will be used more if the exam and feedback is more proximate to when this feedback can be applied. Lastly, early access to feedback at the risk of learners engaging with the score report to figure out if they passed or failed seems worth the trade-off as there was no negative impact on utilizing the feedback, and our study and other literature suggests this can be an effective strategy to lower immediate anxiety.

#### **Conclusion**

In conclusion, we were able to develop a score report that could be implemented immediately after an electronic OSCE, that provided information that was perceived as accurate in identifying strengths and weaknesses, and could be used by students to enhance their future performance down the road in the clinical environment. We have also enhanced our understanding of what aspects of the process of an OSCE score report can increase student uptake of feedback and lower anxiety at the same time.

#### **Acknowledgements**

Dr. Daniels would like to acknowledge the Alberta Academic Medicine and Health Services Program for its financial support.

#### **Notes on Contributors and Affiliations**

All contributors are at the University of Alberta.

Vijay Daniels, MD MHPE, is a Professor in the Department of Medicine and the Assistant Dean of Assessment for the MD Program in the Faculty of Medicine and Dentistry

Silvia Ortiz B.Sc., is a graduate student in the School of Dentistry, Faculty of Medicine and Dentistry

Gurtej Sandhu, MD, is a fifth year resident in the General Internal Medicine subspecialty residency program, Faculty of Medicine and Dentistry

Hollis Lai, PhD, is an Associate Professor in the School of Dentistry, Faculty of Medicine and Dentistry.

Minn Yoon, PhD, is an Associate Professor in the School of Dentistry, Faculty of Medicine and Dentistry

Okan Bulut, PhD, is an Associate Professor in the Department of Educational Psychology, Faculty of Education

Tracey Hillier, MD MEd, is an Associate Professor in the Department of Radiology and Diagnostic Imaging, Faculty of Medicine and Dentistry

#### **ORCID iDs**

Vijay J. Daniels  <https://orcid.org/0000-0002-6350-3129>

Okan Bulut  <https://orcid.org/0000-0001-5853-1267>

#### **REFERENCES**

1. Bandiera G, Sherbino J, Frank JR. *The CanMEDS Assessment Tools Handbook. An Introductory Guide to Assessment Methods for the CanMEDS Competencies*. The Royal College of Physicians and Surgeons of Canada; 2006.
2. POGOe - Portal of Geriatrics Online Education. ACGME competencies: suggested best methods for evaluation. 2009. Accessed June 5, 2020. <https://pogoe.org/productid/20385>
3. Stiggins RJ. Assessment crisis: the absence of assessment for learning. *Phi Delta Kappan*. 2002;83:758-765.

4. Lockyer J, Carraccio C, Chan MK, et al. Core principles of assessment in competency-based medical education. *Med Teach*. 2017;39:609-616.
5. Merry S, Price M, Carless D, Taras M, eds. *Reconceptualising Feedback in Higher Education: Developing Dialogue with Students*. Routledge; 2013.
6. Tett L, Hounsell J, Christie H, Cree VE, McCune V. Learning from feedback? Mature students' experiences of assessment in higher education. *Res Post-Compulsory Edu*. 2012;17:247-260.
7. Patrício MF, Julião M, Fareleira F, Carneiro AV. Is the OSCE a feasible tool to assess competencies in undergraduate medical education? *Med Teach*. 2013;35:503-514.
8. Guraya SY, Guraya SS, Habib F, AlQuiliti KW, Khoshhal KI. Medical students' perception of test anxiety triggered by different assessment modalities. *Med Teach*. 2018;40(supp 1):S49-S55.
9. Cazzell M, Rodriguez A. Qualitative analysis of student beliefs and attitudes after an objective structured clinical evaluation: implications for affective domain learning in undergraduate nursing education. *J Nurs Educ*. 2011;50:711-714.
10. Brown CW. Tablet- or iPad-based marking of OSCEs and MMIs: an imaginative cost-saving approach. *Med Teach*. 2016;38:211-212.
11. Denison A, Bate E, Thompson J. Tablet versus paper marking in assessment: feedback matters. *Perspect Med Educ*. 2016;5:108-113.
12. Munro AJ, Cumming K, Cleland J, Denison AR, Currie GP. Paper versus electronic feedback in high stakes assessment. *J R Coll Physicians Edinb*. 2018;48:148-152.
13. Harrison CJ, Molyneux AJ, Blackwell S, Wass VJ. How we give personalised audio feedback after summative OSCEs. *Med Teach*. 2015b;37:323-326.
14. Bernard AW, Ceccolini G, Feinn R, et al. Medical students review of formative OSCE scores, checklists, and videos improves with student-faculty debriefing meetings. *Med Educ Online*. 2017;22:1324718.
15. Daniels VJ, Strand AC, Lai H, Hillier T. Impact of tablet-scoring and immediate score sheet review on validity and educational impact in an internal medicine residency Objective Structured Clinical Exam (OSCE). *Med Teach*. 2019;41:1039-1044.
16. Zenisky AL, Hambleton RK. A model and good practices for score reporting. In: Lane S, Raymond MR, Haladyna TM, eds. *Handbook of Test Development*. 2nd ed. Routledge; 2016: 585-602.
17. Daniels V, Surgin C, Lai H. Enhancing formative feedback of an OSCE through tablet scoring: OA2-5. *Med Educ*. 2016;50(Suppl. 1):28.
18. Daniels VJ, Bordage G, Gierl MJ, Yudkowsky R. Effect of clinically discriminating, evidence-based checklist items on the reliability of scores from an Internal Medicine residency OSCE. *Adv Health Sci Educ*. 2014;19:497-506.
19. Medical Council of Canada. Patient Interaction Rating Scale Items [Internet]. 2013. Accessed May 2014. Available from <https://mcc.ca/media/MCCQP-PART-II-RATING-SCALES.pdf>
20. Kirkpatrick D. Evaluation of training. In: Craig R, Bittel L, eds. *Training and Development Handbook*. McGraw-Hill; 1967: 131-167.
21. IBM Corp. *IBM SPSS Statistics for Windows*, Version 26.0; Released 2019.
22. Harrison CJ, Könings KD, Molyneux A, Schuwirth LW, Wass V, van der Vleuten CP. Web-based feedback after summative assessment: how do students engage? *Med Educ*. 2013;47:734-744.
23. Harrison CJ, Könings KD, Schuwirth L, Wass V, van der Vleuten C. Barriers to the uptake and use of feedback in the context of summative assessment. *Adv Health Sci Educ*. 2015a;20:229-245.
24. Bulut O, Cutumisu M, Aquilina AM, Singh D. Effects of digital score reporting and feedback on students' learning in higher education. *Front Educ*. 2019; 4:1-16.
25. Daniels LM, Gierl MJ. The impact of immediate test score reporting on university students' achievement emotions in the context of computer-based multiple-choice exams. *Learn Instr*. 2017;52:27-35.
26. Martin RD, Naziruddin Z. Systematic review of student anxiety and performance during objective structured clinical examinations. *Curr Pharm Teach Learn*. 2020;12:1491-1497.
27. Ossenberg C, Henderson A, Mitchell M. What attributes guide best practice for effective feedback? A scoping review. *Adv Health Sci Educ Theory Pract*. 2019;24: 383-401.
28. Sargeant J, Eva KW, Armson H, et al. Features of assessment learners use to make informed self-assessments of clinical performance. *Med Educ*. 2011;45: 636-647.
29. Yardley S, Dornan T. Kirkpatrick's levels and education 'evidence'. *Med Educ*. 2012;46:97-106.
30. Moreau KA. Has the new Kirkpatrick generation built a better hammer for our evaluation toolbox? *Med Teach*. 2017;39:999-1001.
31. Garino A. Ready, willing and able: a model to explain successful use of feedback. *Adv Health Sci Educ*. 2019;25:337-361.