# OSCE Feedback: A Randomized Trial of Effectiveness, Cost-Effectiveness and Student Satisfaction

Celia A. Taylor, Kathryn E. Green

School of Clinical and Experimental Medicine, University of Birmingham, Birmingham, UK
Email: c.a.taylor@bham.ac.uk

**Purpose:** To develop two new types of clinical feedback for final year medical students using OSCE mark sheets and to evaluate their effectiveness, cost-effectiveness and student satisfaction in a randomized trial. **Methods:** A randomized trial was conducted with two groups (Cohort A and B) of students (n = 350) at the University of Birmingham (UK) participating in a two stage Objective Structured Clinical Examination (OSCE) (November 2011 and April 2012). Students were randomly assigned to receive one of three feedback interventions (skills-based, station-based, or both) after the November OSCE. Multivariate regression analysis was used to test if feedback intervention was a significant predictor of April OSCE score, while controlling for November OSCE score. Secondary outcomes were cost-effectiveness and student satisfaction. **Results:** Feedback group was not a significant predictor of April scores for Cohort B. In Cohort A, the station-based group did better than the group who received both types of feedback (2.8%, 95% CI 0.4% to 5.2%, $p$ = 0.022). There was no difference between the skills-based and station-based groups. The cost of providing the station-based feedback was double of that for the skills-based. Questionnaires were received from 245 students (70%). Students who received both types of feedback were the most satisfied, followed by those in the station-based group. **Conclusion:** There was no consistent difference in effectiveness across the three trial groups. Students tended to prefer station-based feedback over skills-based feedback, but students found elements of the standard feedback more helpful than the feedback evaluated in this trial.

*Keywords*: Assessment; Medical Students; Objective Structured Clinical Examination; Feedback; Randomized Trial

## Introduction

Feedback is an important part of medical education and is necessary for students to improve as clinicians and learners (Ende, 1983; Brown & Cooke, 2009; Chowdhury & Kalu, 2004; Brukner, Altkorn, Cook, Quinn, & Mcnabb, 1999). Contrary to reports from educators who feel feedback is abundant, students often report a lack of and/or dissatisfaction with feedback (Ende, 1983; De, Henke, Ailawadi, Dimick, & Colletti, 2004; Gil, Heins, & Jones, 1984). Feedback should be specific, timely and steer future learning, while reinforcing appropriate behaviors and competency levels (Ende, 1983; Brown & Cooke, 2009; Chowdhury & Kalu, 2004; Brukner et al., 1999; Gigante, Dell, & Sharkey, 2011; Van De Ridder, Stokking, McGaghie, & Ten Cate, 2008). In contrast, feedback not perceived to come from a credible source or that does not conform to the examinees' own self-perception may not be taken seriously (Sargeant, Mann, Van der Vleuten, & Metsemakers, 2009; Bing-You, Paterson, & Levine, 1997; Veloski, Boex, Grasberger, Evans, & Wolfson, 2006; Watling & Lingard, 2012). To aid student receptivity, feedback should also be evaluative, focusing on the task and not the person (Kluger & DeNisi, 1996).

The Objective Structured Clinical Exam (OSCE) is used globally for formative and summative assessments and presents an excellent opportunity to provide students with formal feedback on their clinical skills prior to entering postgraduate training. Some research has explored the use of the OSCE as a way of providing immediate feedback to students by allowing an additional time for examiners to speak with examinees (Black & Harden, 1986; Hodder, Rivington, Calcutt, & Hart, 1989; Hollingsworth, Richards, & Frye, 1994). Though this proved beneficial, time constraints and testing policies do not allow for immediate feedback in many exams. The nearest alternative would be to show students their mark sheets which list the tasks they are expected to complete (e.g. palpate, communicate clearly) and include examiners' comments on each task. The tasks vary by station, but many of the tasks are assessed over multiple stations in different OSCE settings. The logistical difficulties in providing mark sheets are exacerbated by test security restrictions, which would not usually allow students to view mark sheets (University of Washington, 2004). Alternative methods of providing more feedback than marks alone should however be explored, as examiners have a unique opportunity to view students' clinical skills in a controlled setting.

Most research evaluating feedback to medical students does not explore outcomes beyond student satisfaction (Boehler et al., 2006). A randomized trial examined performance and satisfaction with feedback after a knot-tying task and found that

students rated complementary feedback better than constructive criticism. However, in follow-up the group that received the constructive criticism did significantly better on a subsequent knot-tying task (Boehler et al., 2006). This reiterates the importance of going beyond "happy sheets" and evaluating the real effectiveness of an intervention in terms of knowledge, performance, and behaviors (Hodder et al., 1989; Boet, Sharma, Goldman, & Reeves, 2012). As in the knot-tying study, such evaluations should use random allocation, which helps to reduce bias and enables any causal relationship to be identified. However we were unable to find any studies evaluating feedback provided after OSCEs which fulfilled either of these criteria (random allocation or performance outcomes).

This study considers two methods of using the data on OSCE mark sheets while maintaining test security, by extracting only the data on student performance without giving away the restricted information. We evaluated the effectiveness of feedback derived from the mark sheets in this randomized trial. The research questions we addressed were:

1) Does receiving examiners' comments help students to perform better on a subsequent OSCE? (station-based feedback)

2) Does receiving a grading (good, fair, etc) for important skills, derived from OSCE tasks assessed in multiple stations, help students to perform better on a subsequent OSCE? (skills-based feedback)

3) What is the cost effectiveness of developing and delivering the two types of feedback?

4) Do students prefer one type of feedback over the other?

## Methods

### OSCE Structure and Standard Feedback

Students in the final year of medical school at The University of Birmingham (UB) participate in a two stage OSCE. The OSCE consists of 18, 10 minute stations. Nine stations are taken in November and the remaining nine in April. Students must achieve an overall mean score $\geq 50\%$ to pass the OSCE and are required to pass 12 of the 18 stations with a score $\geq 50\%$ after calibration using the Borderline Group method of standard setting (Livingston & Zieky, 1982).

At the start of the academic year, students are randomly assigned to a cohort (Cohort A or Cohort B). Each cohort embarks on one of two rotation blocks and nine OSCE stations correspond to each block. Cohort A begins with Part 1 subjects (Medicine, Surgery, and the specialities). Cohort B begins with Part 2 subjects (Pediatrics, Obstetrics/Gynaecology, and Community-Based Medicine). After completing their first rotation block, students switch and take the other rotations and OSCE stations. Specific skills are assessed in each station and a number of generic skills are assessed in multiple stations (e.g. communication with patients, history taking).

The standard feedback provided to students after the November OSCE comprises (I) individual station scores, (II) histogram depicting the spread of total scores, and (III) generic station-based feedback for the cohort.

### Trial Design

The authors consulted the CONSORT statement at the planning and reporting phases of this trial (Moher et al., 2010; Schulz et al., 2010). This was a three-arm trial with balanced stratified randomization and colleagues who were not part of

this project determined the allocation to the feedback groups. The allocation was stratified by cohort to ensure a balanced design. One of the authors (KG) collated these allocations and managed the process of delivering the feedback to students. Another author (CT) was blinded from the allocation of the feedback group and conducted the analyses. She was not aware of group allocations until after analyses were complete. Students were aware of their group allocation when they received their feedback, but the examiners for the April OSCE were blinded as to which type of feedback each student had received.

A true control group was not used because of the ethical considerations of depriving one group of students of additional feedback. Instead, a three-arm design was used to evaluate if differences exist between the two new types of feedback (skills-based and station-based) and also if receiving both types of feedback would be beneficial.

### Outcomes

The primary outcome of this trial was April OSCE performance. The secondary outcomes were cost-effectiveness and student satisfaction with feedback.

### Participants and Setting

All final year medical students at UB in October 2011 were eligible for participation (n = 351). An opt-out method was used so that students who may have forgotten to "opt-in" would not be deprived of the additional feedback. One student who was only required to take nine stations was excluded. We undertook a power analysis to determine the effect size that could be detected in each regression model. Assuming 10% of students opted-out of the study or did not complete both parts of the OSCE, we estimated the likely sample size of each cohort to be 157. At alpha = 0.05 and 80% power, this would enable us to detect an effect size of $f^2 = 0.05$ (between Cohen's standards for a "small" and "medium" effect), where $f^2$ is the proportion of variance explained by a predictor (Faul et al., 2009; Cohen, 1988).

The trial took place at UB between November 2011 and May 2012. The OSCEs took place in six different National Health Service (NHS) sites in the Birmingham area.

### Feedback Interventions

*Skills-Based Feedback*: The skills-based feedback was designed to give students an overall rating of how they did on generic skills in the November OSCE. Each mark sheet has between three and nine tasks that students are expected to complete and each task was rated by examiners on a 4-point rating scale from "0" not done/very poor to "3" very good. The authors independently mapped each task on the OSCE mark sheets to a Tomorrow's Doctors competency (General Medical Council, 2009). For example, auscultation or palpation tasks mapped to the "perform an examination" competency. All discrepancies were discussed until consensus was reached. The competencies were then collapsed into seven skills (**Table 1**) and Excel 2007 was used to calculate the feedback ratings. Skills were averaged within stations to prevent any one station from skewing the score. Skills that were assessed in three or more stations were then averaged across stations and an overall percentage score calculated. Students were told whether their

**Table 1.**
Skills-based feedback categories.

| Skill Category | Range of number of stations/9 included in each skill category[a] | |
| --- | --- | --- |
| | Cohort A | Cohort B |
| Taking a patient history | 3 - 4 | 5 - 6 |
| Communicating with patient/role player | 6 - 7 | 6 |
| Conducting a physical or mental examination | 5 - 7 | 3 - 4 |
| Planning investigations | 4 - 7 | N/A (<3 common stations) |
| Interpreting history, examination and investigations and/or making a (differential) diagnosis | 8 - 9 | 5 - 7 |
| Treatment and management, including prescribing skills and providing immediate care in an emergency | 4 - 5 | 6 |
| Communicating scientific and/or critical knowledge to a professional | 5 - 6 | 4 - 6 |

Note: [a]Different scenarios are used within some stations in each half-day OSCE session; there is some variation in the skills assessed across the different scenarios.

performance in each skill was excellent (85%+), very good (75% - 84.9%), good (65% - 74.9%), fair (55% - 64.9%), or needs improvement (<55%). Students were also given a rank ordering of their relative performance in each skill.

*Station-Based Feedback*: In addition to the tasks on the mark sheets, there are boxes where examiners can comment on how a student performed on each task. For the station-based feedback students were given verbatim transcriptions of the examiners' comments. Examiners were trained prior to the OSCE and were made aware that their comments might be transcribed for students. If there were no comments on a mark sheet, "No comments" was listed as the feedback for that station.

On December 16, 2011 all students received an email with their personalized feedback. This was the earliest possible date, as all OSCE marks had to be approved by an exam board before the feedback could be generated.

### Analysis of the Primary Outcome

Multiple linear regression was performed separately for Cohort A and Cohort B using STATA v. 12 to examine the effect of feedback group on April performance. In addition to a dummy coded variable for feedback group, the mean percentage score for the November OSCE was included in the model in order to control for regression to the mean. The outcome variable was the mean percentage score for the April OSCE stations. To account for the use of two regression models, $p$-values < 0.025 would be considered statistically significant.

### Cost-Effectiveness

The authors kept track of the time they spent developing and executing each type of feedback so that cost-effectiveness could be assessed. The total hours were extrapolated to the time required to provide each type of feedback to the entire cohort. We assumed that the station-based feedback could be executed by

an administrator, while the skills-based feedback would require academic input. The time commitment was then costed using 2011/12 UK Higher Education pay scales, at the bottom end of administrative band 500 and academic grade 7, uplifted for employer National Insurance and pension contributions, assuming a working week of 37.5 hours and 46 working weeks per year.

### Student Satisfaction

A short questionnaire was emailed to students via Survey Monkey on January 11, 2012. The survey was designed to ascertain student opinion on the helpfulness of the feedback provided as part of this trial and the standard feedback they received (in regards to future OSCE performance and clinical practice) using a 4-point rating scale. The percentage of students rating each type of feedback as somewhat or very helpful was calculated by trial group using SPSS v. 18.

## Results

### Participation

No students opted-out of the study. A flow diagram of the trial is shown in **Figure 1**. One student did not take the April stations due to ill health and was excluded from the analysis. Summary statistics for each feedback group are shown in **Table 2**.

**Table 2.**
Summary Statistics.

| | Feedback Group | | |
| --- | --- | --- | --- |
| | Skills | Station | Both |
| Number of Students | 116 | 116 | 117 |
| Nov. OSCE, N (%) | | | |
| Part 1 subjects | 58 (50.0) | 58 (50.0) | 58 (49.6) |
| Part 2 subjects | 58 (50.0) | 58 (50.0) | 59 (50.4) |
| Nov. OSCE Score[a], Mean (SD) | | | |
| Part 1 subjects | 67.4 (8.23) | 66.4 (8.03) | 67.5 (7.96) |
| Part 2 subjects | 66.9 (8.68) | 68.7 (8.01) | 69.3 (8.61) |
| Nov. # passed OSCE stations, Median (IQR) | | | |
| Part 1 subjects | 8 (7 - 9) | 8 (7 - 9) | 8 (7 - 9) |
| Part 2 subjects | 8 (7 - 9) | 8 (7 - 9) | 9 (7 - 9) |
| Apr. OSCE Score[a], Mean (SD) | 65.6 (7.86) | | |
| Part 1 subjects | 70.0 (7.45) | 65.9 (7.48) | 65.6 (6.82) |
| Part 2 subjects | | 70.5 (7.69) | 68.2 (7.66) |
| Apr. # passed OSCE stations, Median (IQR) | | | |
| Part 1 subjects | 8 (7 - 9) | 8 (7 - 9) | 8 (7 - 9) |
| Part 2 subjects | 8 (7 - 9) | 8 (7 - 9) | 8 (7 - 9) |
| Improvement in OSCE (Apr.-Nov.), Mean (SD) | | | |
| Part 1→ Part 2 | 2.30 (7.99) | 4.11 (8.09) | 0.75 (6.90) |
| Part 2→ Part 1 | −1.29 (8.52) | −2.80 (8.67) | −3.71 (7.95) |

Note: [a]All scores are first sit scores. Some students had extenuating circumstances that would entitle them to a further attempt if they failed the OSCE overall. However all students taking both parts of the OSCE are included in the analysis (one student declared herself unfit to sit in April so is excluded, although did receive feedback on her November performance).
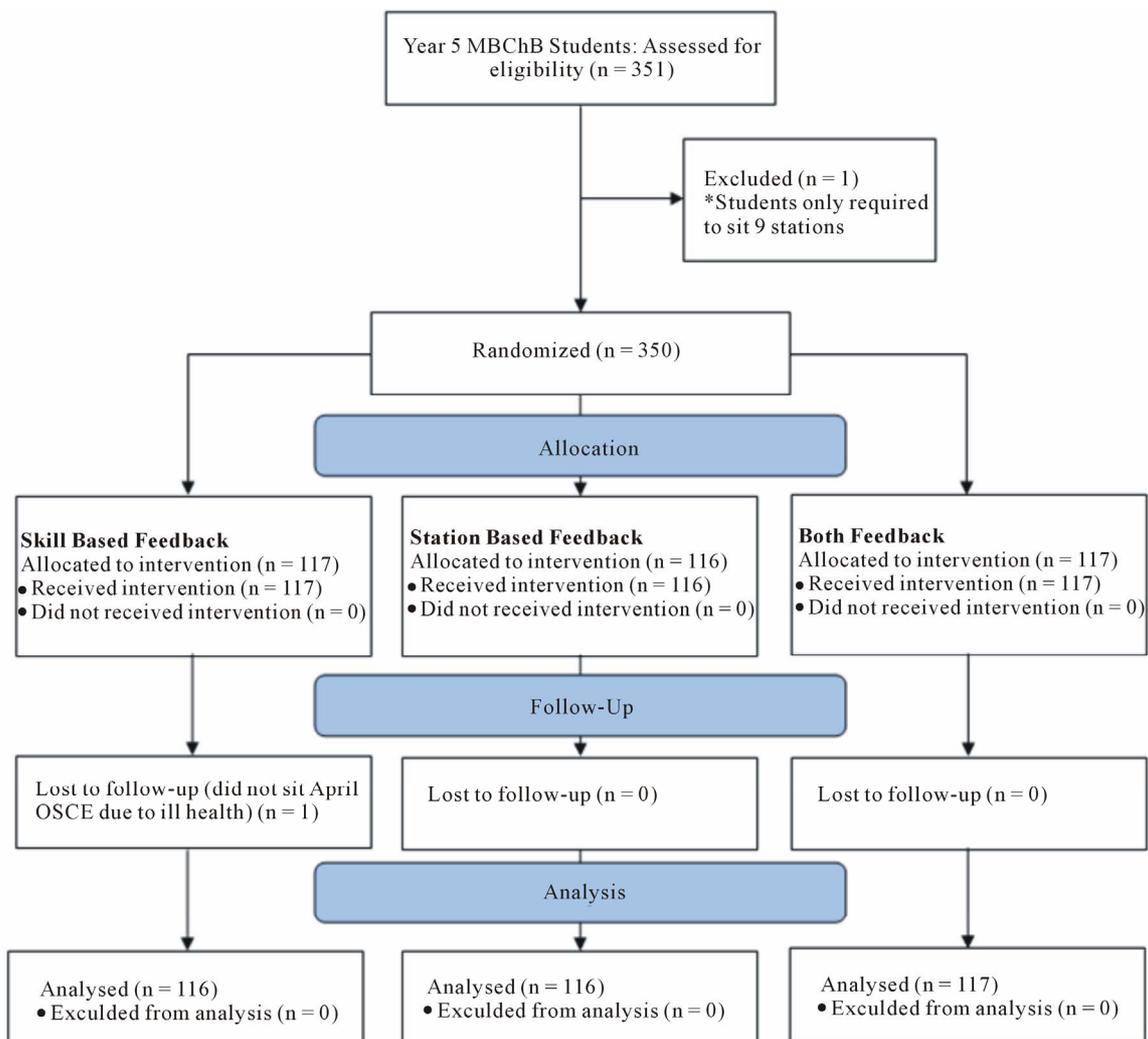
**Figure 1.**
Trial flow diagram.

## OSCE Performance

The internal consistency of the 18 station-level scores in the OSCE (calculated using Cronbach's alpha) was 0.72 for Cohort A and 0.71 for Cohort B. The coefficients for the regression models predicting April OSCE performance are shown in **Table 3**. Feedback group was not a statistically significant predictor of April OSCE score for Cohort B. For Cohort A, students who received station-based feedback had higher April scores than the group that received both types (mean difference 2.8%, 95% CI 0.4 to 5.2, $p = 0.022$). There was no statistically significant difference between the April scores of the station- and skills-based groups. The effect sizes ($f^2$) of the skills-only and station-only feedback groups compared to the group receiving both types were 0.014 and 0.032 (Cohort A); and 0.004 and 0.001 (Cohort B), respectively.

## Cost-Effectiveness

The skills-based feedback took 25 hours to complete and would not take extra time for more students. The station-based feedback would have taken 75 hours had it been done for the entire cohort. There are no economies of scale in providing both feedback interventions, so the total time spent providing both for the whole cohort would have been approximately 100 hours. The total cost of providing the feedback would be approximately £500 ($790) for the skills-based and £1050 ($1660) for the station-based. This amounts to around £1.40 ($2.20) per student for the skills-based feedback and £3.00 ($4.75) per student for the station-based feedback. No further analysis of cost-effectiveness was undertaken because there was no difference in effectiveness between the station- and skills- based groups in either cohort.

## Student Satisfaction

Questionnaires were received from 245 students (70%). **Table 4** shows the percentage of students rating each type of feedback as somewhat or very helpful, by feedback group. Only 35% of students in the skills-based only feedback rated this as somewhat or very helpful, compared to 73% of students who received both interventions. The "both" group were also more positive about the station-based feedback, with 92% rating this

**Table 3.**
Predictors of April OSCE performance.

| Variables in Regression Model | Cohort A | Cohort B |
|---|---|---|
| | Coef. (95%CI), *p*-value | Coef. (95%CI), *p*-value |
| Constant | 35.1 (26.7 - 43.5), <0.001 | 38.5 (30.1 - 46.9), <0.001 |
| Nov. OSCE Score | 0.49 (0.37 - 0.61), <0.001 | 0.39 (0.27 - 0.51), <0.001 |
| Feedback group (cf Both) | | |
| Skills-based | 1.83 (−0.55 - 4.22), 0.131 | 0.90 (−1.54 - 3.34), 0.467 |
| Station-based | 2.80 (0.41 - 5.18), 0.022 | 0.52 (-1.90 - 2.95), 0.671 |
| n | 174 | 175 |
| $R^2$ | 0.28 | 0.20 |

**Table 4.**
Questionnaire results[a] by feedback group.

| Type of Feedback | Feedback Group | | |
|---|---|---|---|
| | Skills N (%)[a] | Station N (%)[a] | Both N (%)[a] |
| Individual Station Scores | 63 (92.6) | 78 (95.1) | 69 (90.8) |
| Histogram | 40 (75.5) | 49 (80.3) | 52 (78.8) |
| Generic Station-Based | 40 (54.1) | 55 (66.3) | 50 (67.6) |
| Skills-Based | 25 (34.7) | N/A | 55 (73.3) |
| Station-Based | N/A | 65 (77.4) | 69 (92.0) |

Note: [a]Percent of students who rated each mode of feedback somewhat helpful or very helpful by Feedback Group (4-point rating scale).

as somewhat or very helpful compared to 77% of students who only received this type of feedback. At least 90% of students in all three groups rated their individual station scores as somewhat or very helpful.

## Discussion

This randomized trial evaluated the effectiveness of using OSCE mark sheets to deliver two new types of feedback to students on their clinical performance. The group that received both skills- and station-based feedback were the most satisfied with the additional feedback they received. However, they did not perform better on their subsequent OSCE and actually those in Cohort A did not do as well as the station-only group. This divergence between satisfaction and effectiveness has been noted previously (Boehler et al., 2006). While the average cost per student of providing either type of additional feedback is low, the resources required still have an opportunity cost and may be more productively employed elsewhere at the medical school.

The effect sizes were very low and differences amongst trial groups in the regression models lacked educational significance. Furthermore, only one cohort saw a statistically significant result for trial group. Receiving station feedback alone was better than receiving both for Cohort A, but not for Cohort B. This may be due to moving from Part 1 subjects to Part 2 sub-

jects, or vice versa. The examiners in Cohort A and B are different, but if the comments of the November Cohort A examiners negatively affected the students, we would also have expected to see the station-based group do worse than the skills-based group. Cohort A also received feedback on one more skill (planning investigations) compared to Cohort B (**Table 1**), but it seems unlikely this would negatively impact scores for only those students receiving both types of feedback.

Our findings suggest that the additional feedback derived from OSCE mark sheets was not effective in improving performance. One of the reasons for this could be weaknesses in both types of additional feedback: feedback should be specific, non-evaluative, timely, and provide guidance on current and future behaviour (Ende, 1983; Brown & Cooke, 2009; Chowdhury & Kalu, 2004; Brukner et al., 1999; Kluger & DeNisi, 1996). Despite our best efforts to map OSCE tasks to important clinical skills, deriving those skills from tasks on OSCE mark sheets may not have produced specific enough areas for students to improve upon. Likewise, examiner comments may not have been specific enough to be useful, or perhaps were too evaluative (about the student instead of the task). Examiners could instead be asked to provide specific information about each student's strengths and areas in which they could improve.

The lag time between the November OSCE and provision of the feedback (5 weeks) could have been one reason for its ineffectiveness (Gigante et al., 2011). This delay was the result of having to wait until all marks had been processed and ratified by the exam board to begin transcribing the station-based feedback and calculating skills-based scores. Other methods of delivering feedback as immediately as possible should be explored. For example, digitised voice recordings or comments typed electronically by examiners could be sent to students much sooner than transcriptions of paper mark sheets.

While this was a randomized trial, it was not possible to blind students to which feedback intervention they had received and this may have affected their study behavior for the April OSCE. However both April examiners and the statistician were blinded as to students' group allocations and it is unlikely that students receiving both types of feedback would have relied on this to ensure they passed this high-stakes exam. Ethical considerations prevented us from having a "no additional feedback" group, although it does not appear that either type of additional feedback was effective in improving OSCE performance. Students are inevitably concerned with passing their examinations, but from a patient perspective, the most important outcome—for which April OSCE performance is a surrogate of unknown quality—is whether the feedback helps students perform better in clinical practice.

In the planning phases of this project, it was hoped that we could use OSCE mark sheets to provide valuable feedback to graduating medical students and that the resources required would be justified. What we learned is that OSCE mark sheets in their current form did not include the information necessary for students to improve performance, but we feel that with a few amendments, OSCE mark sheets could be used to provide useful feedback in the future. Despite a lack of improved OSCE performance, students seemed to appreciate the additional feedback. It would be useful to undertake some qualitative work to explore how students implement different types of feedback and perhaps to identify ways in which students could be supported in their use of feedback. We would also recommend faculty or institutions focus resources into ensuring feedback is

used effectively by students.

## Acknowledgements

## REFERENCES

Bing-You, R. G., Paterson, J., & Levine, M. A. (1997). Feedback falling on deaf ears: Residents' receptivity to feedback tempered by sender credibility. *Medical Teacher, 19,* 40-44. doi:10.3109/01421599709019346

Black, N. M. I., & Harden, R. M. (1986). Providing feedback to students on clinical skills by using the objective structured clinical examination. *Medical Education, 20,* 48-52. doi:10.1111/j.1365-2923.1986.tb01041.x

Boehler, M. L., Rogers, D. A., Schwind, C. J., Mayforth, R., Quin, J., Williams, R. G., et al. (2006). An investigation of medical student reactions to feedback: A randomised controlled trial. *Medical Education, 40,* 746-749. doi:10.1111/j.1365-2929.2006.02503.x

Boet, S., Sharma, S., Goldman, J., & Reeves, S. (2012). Review article: Medical education research: An overview of methods. *Canadian Journal of Anesthesia, 59,* 159-170. doi:10.1007/s12630-011-9635-y

Brown, N., & Cooke, L. (2009). Giving effective feedback to psychiatric trainees. *Advances in Psychiatric Treatment, 15,* 123-128. doi:10.1192/apt.bp.106.003293

Brukner, H., Altkorn, D. L., Cook, S., Quinn, M. T., & Mcnabb, W. L. (1999). Giving effective feedback to medical students: A workshop for faculty and house staff. *Medical Teacher, 21,* 161-165. doi:10.1080/01421599979798

Chowdhury, R. R., & Kalu, G. (2004). Learning to give feedback in medical education. *The Obstetrician & Gynaecologist, 6,* 243-247. doi:10.1576/toag.6.4.243.27023

Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (2nd ed.). Hillsdale, NJ: Earlbaum.

De, S. K., Henke, P. K., Ailawadi, G., Dimick, J. B., & Colletti, L. M. (2004). Attending, house officer, and medical student perceptions about teaching in the third-year medical school general surgery clerkship. *Journal of the American College of Surgeons, 199,* 932-942. doi:10.1016/j.jamcollsurg.2004.08.025

Ende, J. (1983). Feedback in clinical medical education. *Journal of the American Medical Association, 250,* 777-781. doi:10.1001/jama.1983.03340060055026

Faul, F., Erdfelder, E., Buchner, A., & Lang, A. G. (2009). Statistical power analyses using G*Power 3.1: Tests for correlation and regression analyses. *Behavior Research Methods, 41,* 1149-1160. doi:10.3758/BRM.41.4.1149

General Medical Council (2009). *Tomorrow's doctors.* London: General Medical Council.

Gigante, J., Dell, M., & Sharkey, A. (2011). Getting beyond "good job": How to give effective feedback. *Pediatrics, 127,* 205-207. doi:10.1542/peds.2010-3351

Gil, D. H., Heins, M., & Jones, P. B. (1984). Perceptions of medical school faculty members and students on clinical clerkship feedback. *Academic Medicine, 59,* 856-864. doi:10.1097/00001888-198411000-00003

Hodder, R. V., Rivington, R. N., Calcutt, L. E., & Hart, I. R. (1989). The effectiveness of immediate feedback during the Objective Structured Clinical Examination. *Medical Education, 23,* 184-188. doi:10.1111/j.1365-2923.1989.tb00884.x

Hollingsworth, M., Richards, B. F., & Frye, A. W. (1994). Description of observer feedback in an objective structured clinical examination and effects on examinees. *Teaching and Learning in Medicine, 6,* 49-53. doi:10.1080/10401339409539643

Kluger, A., & DeNisi, A. (1996). The effects of feedback interventions on performance: A historical review, a meta-analysis, and a preliminary feedback intervention theory. *Psychological Bulletin, 119,* 254-284. doi:10.1037/0033-2909.119.2.254

Livingston, S., & Zieky, M. (1982). *Passing scores: A manual for setting standards of performance on educational and occupational tests.* Princeton, NJ: Educational Testing Service.

Moher, D., Hopewell, S., Schulz, K. F., Montori, V., Gøtzsche, P. C., Devereaux, P. J., et al. (2010). CONSORT 2010 Explanation and Elaboration: Updated guidelines for reporting parallel group randomised trials. *British Medical Journal, 340,* c.869. doi:10.1136/bmj.c869

Sargeant, J. M., Mann, K. V., Van der Vleuten, C. P., & Metsemakers, J. F. (2009). Reflection: A link between receiving and using assessment feedback. *Advances in Health Science Education, 14,* 399-410. doi:10.1007/s10459-008-9124-4

Schulz, K. F., Altman, D. G., & Moher, D. for the CONSORT Group (2010). CONSORT 2010 statement: Updated guidelines for reporting parallel group randomised trials. *PLoS Medicine, 7,* e1000251. doi:10.1371/journal.pmed.1000251

University of Washington (2004). Second year OSCEs: Preparing for the second year OSCEs. URL (last checked 14 September 2012). http://www.uwmedicine.org/Education/MD-Program/Current-Students/Curriculum/OSCE/Pages/Second-Year-OSCEs.aspx

Van De Ridder, J. M. M., Stokking, K. M., McGaghie, W. C., & Ten Cate, O. T. (2008). What is feedback in clinical education? *Medical Education, 42,* 189-197. doi:10.1111/j.1365-2923.2007.02973.x

Veloski, J., Boex, J. R., Grasberger, M. J., Evans, A., & Wolfson, D. B. (2006). Systematic review of the literature on assessment, feedback and physicians' clinical performance: BEME Guide No. 7. *Medical Teacher, 28,* 117-128. doi:10.1080/01421590600622665

Watling, C., & Lingard, L. (2012). Toward meaningful evaluation of medical trainees: The influence of participants' perceptions of the process. *Advances in Health Science Education, 17,* 183-194. doi:10.1007/s10459-010-9223-x