

**Update in Medical Education
SGIM 36th Annual Meeting
Denver, Colorado**

CUB3

April 25, 2013

3:30 PM to 5:00 PM

Governor Sq 15, Plaza Bldg Concourse

Shobhina G. Chheda, MD, MPH

Carol K. Bates, MD

Kathel Dunn, PhD

Reena Karani, MD

Lisa Willett, MD

Association between a medical school admission process using the multiple mini-interview and national licensing exam scores. Eva KW, Reiter HI, Rosenfeld J, Trinh K, Wood TJ, Norman GR. JAMA. 2012 Dec 5;308(21):2233-40. doi: 10.1001/jama.2012.36914. PMID: 23212501.

This cohort study sought to determine whether students admitted to one Canadian medical school using a 12-station multiple mini-interview (MMI) outperformed those interviewed and rejected by the school on the Canadian national licensing examinations (MCCQE). Examinee performance scores on MCCQE Parts I and II were obtained for all those who interviewed for McMaster University's MD program in the 2004 and 2005 admissions cycle using an MMI focused on ethical issues, communication and collaboration tasks. Part I is a computer based multiple-choice exam measuring general medical knowledge and clinical decision-making skills done immediately following completion of MD training. Part II is a 14-station OSCE testing clinical skills, problem solving, patient interaction and consideration of ethical, legal and organizational aspects of practice generally completed 16 months into residency training.

Of 1071 interviewees, 521 (48.6%) were admitted to McMaster. 751 (70.1%) had available MCCQE Part I scores and of these, 623 had Part II scores. Candidates accepted to the school using the MMI had higher scores than those who were rejected for Part I (mean total score 531 vs. 515, $p=0.003$) and Part II (563 vs. 544, $p=0.007$). In addition, those admitted using the MMI had a higher mean on the "considerations of ethical, legal and organizational aspects of practice" subscore in Part II. Among the accepted group, those who matriculated at McMaster did not outperform those who matriculated at other Canadian medical schools on either Part I or II of the MCCQE. This provides evidence that reported differences did not arise from the school's curriculum.

Limitations: Examination scores on MCCQE Parts I and II are not direct measures of actual practice and thus these results must be interpreted with caution. In addition, demographic variables were not obtained and may have been informative in interpreting results. Finally, the MMI at this institution was designed to emphasize particular things and therefore the impact of other MMI's on future examinee performance cannot be inferred.

Implications: MMI's are now being used by multiple Canadian and US schools in their applicant selection process. This single-institution study showed that the MMI allows for repeated sampling of performance and students who were accepted using an MMI assessment scored higher than those rejected on both steps of the Canadian national licensing examination.

Students' workplace learning in two clerkship models: a multi-site observational study. O'Brien BC, Poncelet AN, Hansen L, Hirsh DA, Ogur B, Alexander EK, Krupat E, Hauer KE. Med Educ. 2012 Jun;46(6):613-24. doi: 10.1111/j.1365-2923.2012.04271.x. PMID: 22626053.

This multi-institutional study used an observational, work-sampling methodology at three medical schools to compare students' daily activities and interactions early and/or late in the year

during core clerkships organized as either longitudinal integrated clerkships (LICs) or block clerkships (BC). At each school, LIC and BC students at one clinical site were invited to participate during their internal medicine (medical) and obstetrics and gynecology (procedural) clinical experiences. Trained research assistants used electronic work-sampling checklists to capture activities, level of participation, individuals present and type of encounter (first or follow-up) at 10-minute intervals during 4-hour clinical sessions in outpatient and inpatient settings.

Twenty-six (16 LIC and 10 BC) and 44 (28 LIC and 16 BC) students were observed 1-3 times early or late in the year for a total of 135 sessions. Only BC students were observed in the inpatient environment as LIC students had completed their inpatient time prior to the study. Outpatient activities (direct patient care, indirect patient care, case discussion, educational activities, in-transit or waiting, personal) and interactions (with preceptors or resident teams) of LIC and BC students were not statistically different early in the year. Later in the year, LIC students spent more time performing direct patient care activities alone (25%) compared with BC students in the outpatient (12%, $p < 0.007$) or inpatient (7%, $p < 0.001$) settings. LIC students saw a significantly higher percentage of returning patients (34%) than BC students (5%) in the outpatient setting ($p < 0.001$) late in the year.

Limitations: The small number of students observed early and late in the year at each school and the fact that no early observations were conducted at one school limits the generalizability of this study. Inpatient observations were only conducted in the morning and hence activities and interactions may have been impacted by the sampling methodology. Finally, the Hawthorne effect due to the use of observers must be taken into consideration when interpreting the results of this study.

Implications: LICs afford students more independent patient care opportunities and more longitudinal follow up with patients through return visits later in the year. The LIC model should be strongly considered by medical schools.

Vicarious learning during simulations: is it more effective than hands-on training?
Stegmann K, Pilz F, Siebeck M, Fischer F. Med Educ. 2012 Oct;46(10):1001-8.
doi: 10.1111/j.1365-2923.2012.04344.x. PMID: 22989134

This single institution study aimed to determine whether simulation based learning by doing vs. learning by watching (with an observation script or a feedback formulation script) and the order of these activities would affect knowledge acquisition regarding patient-doctor communication. The study utilized a 2x2x2 pre-post design that varied the order of the role of the student (as doer or observer) and the provision of an observer script (with or without) and of a feedback script (with or without). 200 undergraduate medical students were randomly assigned to one of the eight conditions. Each student had two learning sessions (approximately 45 minutes), divided into 3 segments including a 20 minute simulated patient encounter in which they were directly interacting with the standardized patient or observing, followed by a 10 minute feedback segment where they were either receiving feedback from the SP or preparing feedback for their peer, and concluded with a 15 minute

segment where the examining student received peer feedback. The observation script contained a list of important steps for the encounter and included six separate sections each containing between 4 and 13 yes/no checklist items and a comment area. Observers without the script were asked only to take notes to allow for feedback. The feedback script required the observer to pre-structure the feedback according to a 4 -step procedure; observers without the feedback script were asked only to prepare feedback based on their observations. Outcomes were assessed through use of open-ended questions assessing communication knowledge at three time points- before any sessions, immediately following the second session and immediately following the last session.

The results demonstrated a significant effect of type of learning-based activity ($F=12.56$, $p<0.001$), with vicarious learners demonstrating higher means on knowledge assessment than those that were learning by doing (Estimated mean (EM)=21.84 vs. EM 18.58, respectively.) Only the observation script had significant impact on knowledge of patient-doctor communication ($F=35.13$, $p< 0.001$), with students with the script learning significantly more than students without the script (EM 22.85 vs. EM 17.84, respectively). The feedback formulation script did not show an effect.

Limitations: The outcome of this study was knowledge, and not skill, related to patient-doctor communication.

Implications. This study demonstrated that learners could benefit, as measured on a knowledge outcome, from observing others in simulation-based learning. Greater benefit to the observer was demonstrated through provision of an observation script. Further studies are needed to determine if this knowledge acquisition can translate to skill and if results are reproducible in other domains besides doctor-patient communication.

A comparison of two methods of teaching reflective ability in Year 3 medical students. Med Educ. 2012 Aug;46(8):807-14. Aronson L, Niehaus B, Hill-Sakurai L, Lai C, O'Sullivan PS. doi: 10.1111/j. 1365-2923.2012.04299.x. PMID: 22803758.

This single institution study sought to determine the impact of reflective writing guidelines and process feedback on assignment scores as assessed by a previous validated rubric. 149 third year medical students completing a reflective writing assignment during intersessions at months 2 and 12 of their third year were included. During the first intersession all students received definitions of reflection and critical reflection and were then sorted into small groups that were randomly assigned to receive guidelines for reflection, feedback on reflective skill, both or neither. All students additionally received content feedback. Blinded, trained faculty scored reflective writing assignments with inter-rater reliability (intraclass correlation coefficient) of 0.91.

Students overall did not improve with respect to reflective ability as measured on their two assignments. There was no interaction between guidelines and process feedback ($F=1.36$, d.f. 1,144, $p=0.24$). The provision of guidelines improved reflective ability compared with the provision of a definition of critical reflection only ($F = 147.1$, d.f. = 1, 145, $p < 0.001$). Feedback

also improved reflective ability, but only when provided for both content and reflective skills as compared to feedback on content alone; this effect was smaller ($F = 6.5$, d.f. = 1, 145, $p = 0.012$).

Limitations: This study was limited by unintentional randomization outcomes with a disproportionate number of students that had taken time out of school assigned to the group that did not receive guidelines or process feedback. Also two- month gap between assignment and receiving process feedback may have contributed to the small effect size of this feedback. Lastly, though the tools were developed independently from the reflection literature, there is a possibility that there was interaction between the guidelines and the scoring tool for reflective ability.

Implications. Medical student performance on reflective exercises can be improved through the use of critical reflection guidelines as well as by provision of feedback on reflective ability in addition to feedback on content. Building a curriculum and assignments using these components may improve reflective skills of learners.

Seen through their eyes: residents' reflections on the cognitive and contextual components of diagnostic errors in medicine. Ogdie AR, Reilly JB, Pang WG, Keddem S, Barg FK. Von Feldt JM, Myers JS. Acad Med. 2012 Oct;87(10):1361-7. PMID: 22914511.

This single institution study used reflective writing and narrative discussion to assess internal medicine residents' experiences with diagnostic error caused by cognitive bias and to identify related contextual factors. The 41 participating second year residents received an one- hour overview of cognitive bias and diagnostic error 3 months prior to the reflective writing and narrative session. In faculty -facilitated groups of 4-6 residents, participants were given 10 minutes to write about a specific case of diagnostic error with cognitive bias that they had witnessed or participated in. They were asked to include all circumstances that may have contributed to the error, what they learned, and how this might change future practice. Narratives were read aloud individually, followed by a group discussion facilitated by the faculty leader, who had attended a one- hour training session and received a training guide. The entire session was audiotaped and subsequently coded by two coders using a predetermined coding scheme that included cognitive biases, debiasing strategies, and contextual factors (interrater agreement ranged from 80-95%).

All residents reported a case in which they had cared for a patient and experienced a diagnostic error or a delay in diagnosis due to cognitive bias. 85% (35 of 41) provided strategies to prevent similar cognitive errors in the future. The most commonly described bias was anchoring (87.8%) followed by availability bias (76%) and framing effect (56%). Group discussions focused on contextual factors that were subsequently classified into environmental factors, patient factors and team or provider factors.

Limitations: The descriptive nature of the study limits the ability to draw strong conclusions regarding the comparative importance of various biases. The prior didactic sessions may have directed residents to recognize some cognitive biases over others.

Implications. The use of reflective writing and narrative discussion can be used to teach the recognition and analysis of diagnostic error in medical education. The impact of such teaching is yet to be determined, but seems to be an important component of life-long, self-directed learning.

Fostering and evaluating reflective capacity in medical education: developing the REFLECT rubric for assessing reflective writing. Wald HS, Borkan JM, Taylor JS, Anthony D, Reis SP. Acad Med. 2012 Jan;87(1):41- 50. doi:10.1097/ACM.0b013e31823b55fa. PMID: 22104060.

Many institutions are using reflective writing to enhance learners understanding of complexity, clinical reasoning, and ethical dilemmas and to enhance self- awareness. This single institution study describes the iterative development and interrater reliability of an assessment tool to evaluate reflective writing. The authors reviewed existing instruments and then developed their own tool through an iterative process.

Their final tool evaluates writing samples in 5 domains using a 4 point scale. Iterative versions of the developing evaluative tool were assessed by measuring inter rater reliability on pared scores of writing samples. The penultimate version of the rubric was evaluated with 2 raters evaluating each narrative with cronback alpha of 0.774. The final rubric which is provided in detail was not evaluated; the authors ultimately decided to use the rubric for formative rather than summative evaluation.

Limitations: There is not an evaluation on the final rubric, and the authors own reservations about numerically evaluating reflective learning.

Implications: This study adds another assessment tool to be considered in evaluating reflective writing, which would be easily exported.

Validity evidence for a new checklist evaluating consultations: the 5Cs model. Kessler CS, Kalapurayil PS, Yudkowsky R, Schwartz A. Acad Med. 2012 Oct;87(10):1408-12. PMID: 22914527.

Effective consultation depends upon clear communication. While trainees frequently request consultations, there are no previously validated tools to evaluate resident competence in communication regarding consultations.

The authors developed and validated the 5C checklist with domains of Contact, Communicate, Core Question, Collaboration, and Closing the Loop containing a total of 12 measurable items. 47 internal medicine or emergency medicine residents at a single institution were randomized into an intervention or control group. The intervention group had a 90- minute teaching session focused on the 5Cs model and were given note cards describing the model. Control residents

had a 90 minute session describing consultation medicine but without the 5C model. Learners were evaluated by 5C checklists scored by 3- trained faculty who assessed their performance on audiotaped conversations requesting a psychiatric consultation on a patient with psychosis and surgical consultation on a patient with worsening abdominal pain. Intervention residents scored higher than control residents in 9 of 12 domains. The 5C instrument was favorably evaluated for validity on 5 separate measures.

Limitations: This study lacks clarity as to whether or not faculty evaluators were blinded to resident intervention or control group assignment. Simulated consultations appear to have been performed immediately after the teaching, so the durability of the learning is not clear. The study was done in a single institution, and generalizability may be limited to other types of trainees and to consultative communication beyond the initial conversation requesting consultation.

Implications: This validated tool could easily be exported to other institutions and may assist in both teaching and evaluating resident ability in communicating an initial consult request.

Pocket card and dedicated feedback session to improve feedback to ward residents: A randomized trial. Peccoraro L, Karani R, Coplit L, Korenstein D. J Hosp Med. 2011;7(1):35-40. doi: 10.1002/jhm.934. PMID: 22038842.

Despite a variety of previous literature on feedback, residents often note insufficient feedback, little constructive feedback on weaknesses; faculty often note insufficient time to provide feedback, as well as discomfort in providing criticism and lack of direct observation of residents in clinical settings.

Investigators designed a pocket card that included 5 of 6 ACGME competencies (excluding practice-based learning) with 3 to 6 items per competency all rated on a 5 point Likert scale. Intervention hospitalist attendings on 4 -week rotations were given an added 5 minutes orientation at the start of ward rotations in which the card was introduced and they were instructed to dedicate one mid-rotation attending rounds to feedback using the card for 10-15 minutes with each of 3-4 residents. For any resident who scored less than 3 in any domain, the attending was to provide examples and offer suggestions for improvement.

34 attendings (94%) and 93 residents (85%) completed end of rotation surveys. Intervention residents received more positive feedback (89.7 vs 64.3%) and more feedback suggesting improvement (51.3 vs 21.5%) as compared to control residents. Intervention residents reported improving their clinical skills (61.5 vs 27.8%) and their professionalism skills (51.3 vs 29.1%) more often in response to feedback. The majority of intervention attendings positively evaluated the feedback card and the concept of time set aside for feedback. Attendings planned to use the card again, and would recommend the card to others.

Limitations: This study was completed at a single site. As interns and students were not included though were on the ward teams, the generalizability to these learners, and the ability to use this method to evaluate an entire team is uncertain. Several study design features may have made

control residents aware of the study, in one case biasing them toward a positive result (where they had previously experienced feedback and then not); and in another case by contamination, having been on service simultaneously with the intervention group, possibly underestimating the impact of the study. It is unclear whether the feedback card or the allotted time for feedback done alone would have been effective.

Implications: This study uses a simple card and time strategy that is easily exported and could help other programs struggling with provision of feedback.

Effect of exposure to good vs poor medical trainee performance on attending physician ratings of subsequent performances. Yeates P, O'Neill P, Mann K, Eva KW. JAMA. 2012 Dec 5;308(21):2226-32. doi: 10.1001/jama.2012.36515. PMID: 23212500.

This national study from the United Kingdom sought to determine whether attending physician rater scores were biased by recently observed Mini-CEX video performances of interns. Two groups of physicians (n=41) with prior experience with Mini-CEX viewed and scored performance of six scripted videos on a six-point Likert scale (1=well below expectations; 3=borderline; 6 well above expectations). Group one was primed with scoring 3 good performances; group two was primed with scoring 3 poor performances; both groups subsequently scored 3 identical borderline performances. Mean scores of the borderline performances were compared to assess bias.

Mean scores of borderline performances were lower following exposure to good performances as compared to poor (2.7 vs 3.4). Borderline performances were given failing scores over twice as often (55% vs. 24%) from assessors exposed to good performances compared to scores from assessors exposed to poor performances. With hierarchical regression, stringency index (stringent versus lenient raters) accounted for 18%; the addition of priming condition (good vs poor performances) accounted for a further 24% of the observed variation.

Limitations: Raters viewed 3 consistently good or poor videos prior to evaluating borderline videos. In real clinical practice it is more likely that evaluators view "mixed performance". It is unclear if an effect would occur if a single good or poor performance would bias scoring of a subsequent candidate.

Implications: With a goal for criterion referenced competency-based evaluation, raters may be subject to contrast bias. This contrast bias may have greater impact than stringency of an evaluator, affecting summative evaluation.

Effects of 2- vs 4-week attending physician inpatient rotations on unplanned patient revisits, evaluations by trainees, and attending physician burnout: a randomized trial. Lucas BP, Trick WE, Evans AT, Mba B, Smith J, Das K, Clarke P, Varkey A, Mathew S, Weinstein RA. JAMA. 2012 Dec 5;308(21):2199-207. doi: 10.1001/jama.2012.36522. PMID: 23212497.

This single institution study sought to compare the effects of 2 vs 4 weeks of attending rotation schedules on patient care, trainee evaluations, and attending burnout. This was a cluster randomized crossover noninferiority trial of 62 attending physicians over one academic year. Primary outcome was unplanned revisit to the hospital within 30 days. Secondary outcomes included patient length of stay; trainee (student and housestaff) evaluations of the attendings (ability to evaluate, clinical care, professionalism, teaching, and combined score); and four facets of burnout and emotional exhaustion of attendings. Housestaff rotation periods coincided with crossover periods though medical student rotations did not.

There was no difference in patient care outcomes. The percentage of unplanned revisits for 2-week rotations was 21.2% compared with 21.5% for 4-week rotations (mean difference, -0.3%; 95% CI, -1.8% to +1.2%). Average length of stay was not significantly different. Trainees perceived attendings on the 2-week rotation as less able to perfectly evaluate them (students 82% vs. 69% rated less than perfect adjusted relative odds 1.41, $p=0.045$; housestaff 41% vs. 28% ; adjusted relative odds 2.10, $p<0.001$). On 2-week rotations medical students spent a median of 7 days with attending physicians and on 4-week rotations a median of 17 days were spent aligned. Attendings favored the 2 week rotation for burnout with OR 0.39 ($p<0.001$) and emotional exhaustion 0.45 ($p<0.001$).

Limitations: The study was powered for the single patient outcome of 30 day unplanned visits, and the statistical significance of the other outcomes should be interpreted with increased potential for chance findings. Other important patient outcomes (quality of care, adherence to treatment, patient satisfaction) were not measured. Trainee evaluation of attendings was not focused on attending performance but on a validated measure of trainee perception of performance. Results may be related to coincided days spent together, especially demonstrated by less time with the medical students. This might be mitigated by scheduling to maximize overlap. Attending burnout assessments were not validated measures.

Implications: Two-week inpatient attending rotations compared with 4-week did not result in worse patient outcomes. There was less attending burnout and emotional exhaustion, but at the expense of worse trainee perception of ability to evaluate them.