Update in Medical Education

Reena Karani, MD¹, Shobhina G. Chheda, MD, MPH², Kathel Dunn, PhD³, Kenneth Locke, MD⁴, and Carol K. Bates, MD⁵

¹Brookdale Department of Geriatrics and Palliative Medicine, Samuel M. Bronfman Department of Medicine, Department of Medical Education, Mount Sinai School of Medicine, New York, NY, USA; ²Department of Medicine, Section of General Internal Medicine, University of Wisconsin School of Medicine and Public Health, Madison, WI, USA; ³National Library of Medicine, Bethesda, MD, USA; ⁴Department of Medicine, Division of General Medicine, University of Toronto, Toronto, ON, Canada; ⁵Division of General Medicine and Primary Care, Beth Israel Deaconess Medical Center, Boston, MA, USA.

KEY WORDS: advances; innovations; general internal medicine; manuscripts; education.

J Gen Intern Med 26(1):83-7

DOI: 10.1007/s11606-010-1502-y

© Society of General Internal Medicine 2010

INTRODUCTION

This article summarizes 12 manuscripts that represent advances within medical education in 2009 from a general internal medicine perspective.

METHODS

Our search was limited to articles published between January 1, 2009 and December 31, 2009. We reviewed the Table of Contents in 12 publications: Academic Medicine, Annals of Internal Medicine, British Medical Journal, Journal of the American Medical Association, Journal of the American Geriatrics Society, Journal of Continuing Education in the Health Professions, Journal of General Internal Medicine, Lancet, Medical Education, Medical Teacher, New England Journal of Medicine, and Teaching and Learning in Medicine. Our previous literature review consistently found that these journals contained articles most relevant to academic generalists and most likely to impact their teaching practice. We also performed a PubMed search, limited to English language articles published in 2009, using the search strategy of medical education, learners, outcomes and evaluation. More than 7,000 citations were examined. We reached consensus on our final articles based on the questions "Is this relevant to academic general internists?" and "Will this article change the practice of teaching and learning?"

RESULTS

The 12 articles fall into four themes: (1) teaching, (2) simulation, (3) assessment and (4) professionalism.

TEACHING

Wolpaw T, Papp KK, Bordages G. Using SNAPPS to facilitate the expression of clinical reasoning and uncertainties: A randomized comparison group trial.¹

This study determined whether third year medical students (MS), trained to use a learner-centered technique called SNAPPS for case presentations, expressed clinical reasoning and raised learning issues more than students not trained in this technique. The six-step technique required the learner to summarize findings, narrow the differential, analyze the differential, probe the preceptor about uncertainties, plan management and select case-related issues for self-study.

One hundred and eight of 162 eligible students were divided into three groups based on random assignment of their preceptors during the family medicine rotation. SNAPPS students (N=39) received hands-on training about the technique on the first day of the four-week rotation and learning was reinforced twice in the subsequent two weeks. Preceptors also received training about the technique by an author. Comparison students (N=32) received training sessions focused on obtaining preceptor feedback. Preceptors for this group received training on the importance of feedback. Usual-and-customary students (N=37) received no special training and their preceptors received a phone call asking if they had any questions. Outcomes included ten presentation elements in six categories related to clinical reasoning and case-based uncertainties. The last week of the clerkship, students audio-taped all cases presented to faculty. Sixty four of 76 presentations were audible and coded by two authors. Mean number of audiotaped cases was 3.31 (SD 2.1) with no by-group differences. SNAPPS students' case presentations took 1-minute longer than comparison group students ($p = 0.05$). SNAPPS students outperformed all other students across the six outcome categories and were the only ones who identified the need for additional case-related readings.

These results may not be generalizable as the study was conducted in a single outpatient setting. Selection bias and contamination also remain concerns. However SNAPPS, with its

Received May 31, 2010

Revised August 16, 2010

Accepted August 18, 2010

Published online September 18, 2010

emphasis on the learner's responsibility in case presentation, offers preceptors a structured technique to assess trainee competence and diagnostic reasoning abilities during case presentations. Moreover SNAPPS can realistically be incorporated into training environments.

Peets AD, Coderre S, Wright B et al. Involvement in teaching improves learning in medical students: a randomized cross-over study.²

This study examined the effects of teaching on learning outcomes of MS peer educators at one medical school. 135 first year MS were randomly assigned to small groups for the Gastroenterology/Hematology course. For each of 22 mandatory two-hour sessions, two students from each group were randomly selected to be peer educators. Each session included cases focused on a single clinical presentation with questions and discussion. No peer educator training was provided. Faculty facilitators were assigned to groups and instructed to allow peer educators to lead but to correct errors. Following the course, students completed a 94-item MCQ exam and an email questionnaire to estimate time spent preparing as a peer educator versus as a group member. 109 students (72%) completed the email questionnaire. Students spent a mean (SD) of 36 (33) minutes preparing when they were not peer educators versus 99 (60) minutes when they were (Cohen's $d = 1.3$, $p < 0.001$). Reliability of the examination was 0.76 and all students completed it. Mean score (SD) for clinical presentations in which students were peer educators was significantly higher than presentations in which they were not (80.7% (11.8) versus 77.6% (6.9), $p < 0.01$).

The generalizability of these findings is limited since this was a small, single-center study. The use of self-reported preparation time introduces the possibility of recall bias. The absolute difference in exam scores when a student was the peer educator was slight with a small effect size (Cohen's $d = 0.33$). Medical schools across the country have ongoing peer educator programs which have been justified as a way to improve student communication skills, prepare students for their future roles as educators, enhance motivation to learn and reduce faculty burden. This study provides preliminary evidence of benefit of content learning for the peer educators and hence supports implementation of peer assisted learning programs more broadly.

SIMULATION

Fraser K, Peets A, Walker I et al. The effect of simulator training on clinical skills acquisition, retention and transfer.³

This study determined the effect of training using a cardio-respiratory simulator (CRS) on clinical skills and diagnostic performance in 176 1st year MS during a 12 week respiratory/cardiology course. Students were randomly allocated to dyspnea or chest pain as their learning problem for the two-hour small group work with the CRS; two cases within each problem were addressed. Students served as their own controls. Assessment included a similar problem and case on CRS (intervention) and a case from the other problem that was not covered in their training (control). Remote assessment at 6 weeks included a case similar to the training case and a case from the other presentation. To assess skill transfer, students also received two cases that were not exact conditions they were trained in but still had to do with

chest pain and dyspnea with cardiopulmonary physical findings. Students completed an MCQ test to measure skill transfer to a different format. Assessing immediate performance, comparing students on the scenario in which they had trained to the other scenario, students were more able to identify abnormal clinical findings (70.0% versus 52.2% $d = 7.6$, $p < 0.0001$) and had improved diagnostic performance (72.1% versus 55.6% $d = 4.3$, $p = 0.0007$). This effect was irrespective of the clinical problem they received. At the end of the course (retention), students were still better at identifying abnormal findings (57.1% vs. 51.7% $d = 2.5$, $p = 0.004$) and correctly diagnosing (50.0% vs. 38.1% $d = 3.0$, $p = 0.002$) the problem included in their training. However, they showed no difference between training and control presentations in diagnostic performance when transferring skills between cases or performing on the MCQ test.

Limitations impacting generalizability include the narrow focus of training problems, use of novice learners, and the single institution setting. The particular choice of presenting problem to frame physical exam findings may also impact outcomes. However this study suggests that through brief training on a CRS, novice students can acquire and retain clinical skills greater than those acquired with traditional patient-based physical exam training. Regarding lack of transfer of skills to different cases and formats, educators should expect that simulation training in a specific physical finding may not improve identification of another physical finding.

de Giovanni D, Roberts T, Norman G. Relative effectiveness of high- versus low-fidelity simulation in learning heart sounds.⁴

This study compared the effects of training on a high fidelity simulator (Harvey) and a low fidelity simulator (CD-ROM) on recognition of simulated and real patient heart sounds. 37 3rd year MS from a single institution were randomized to receive four hours of identical instruction on a high- or low-fidelity simulator. All students received a one-hour introductory lecture on basic cardiovascular exam technique and were introduced to Harvey to avoid bias in the final assessment. In groups of six, students attended three additional hours of standardized teaching on their assigned simulators. Students were assessed six weeks after training using a six-station OSCE. Five stations consisted of communication with and exam of patients with normal and abnormal heart sounds. The last station included three heart sounds on the CD-ROM and an examination of Harvey on three different settings. All examples had been heard during training. Harvey-trained students had no better skills at identifying heart sounds (3.11 vs. 2.47; $p = 0.06$). There was no difference between Harvey and the CD-ROM-trained groups in diagnosis, communication skills or physical exam skills.

Though this study was small, authors reference a calculated power of 0.65 to detect a 20% difference in accuracy. No control group limits the ability to understand the benefit of any simulation training. This study demonstrated that nearly as much can be gained in learning by using a low fidelity, less expensive simulator when compared to a high fidelity simulator.

Barsuk JH, Cohen ER, Feinglass J et al. Use of simulation-based education to reduce catheter-related bloodstream infections.⁵

This single institution study assessed the effect of simulation-based education of central venous catheter insertion by resi-

dents on patient care. Over 32 months, 92 non-randomized IM and EM residents received simulation-based education consisting of a pre-test using a simulator and checklist, two two-hour education sessions and three hours of practice on the simulator with feedback.

Baseline data regarding catheter infections (CI) in two ICU settings was collected in the first 16-months of cohort observation prior to the intervention. Follow up observation data was collected in the next 16 months in both control and intervention ICU's. In the pre-intervention period of the study, there were 3.20 infections/1,000 catheter-days in the MICU and 4.86 in the surgical ICU. During the 16-month intervention period there were statistically fewer CI in the MICU than in the surgical ICU (0.50 infections/1,000 catheter days versus 5.26 infections/1,000 catheter days, $p=0.001$). Patient mix in the post-intervention period was comparable across ICU's with no increase in mean Charlson score for the intervention ICU.

Limitations include that this study was conducted at a single institution over a short time period. During the last four months of the study, chlorhexidine-impregnated wipes were introduced to the intervention ICU and this may be a confounder. Simulation-based education can be used to improve patient outcomes. Further studies may clarify the benefits of the didactic component versus the hands-on component of simulation.

ASSESSMENT

Kogan JR, Holmboe ES, Hauer KE. Tools for direct observation and assessment of clinical skills for medical trainees: a systematic review.⁶

This study identified available tools for supervisors' direct observation of trainees' clinical skills. Criteria for inclusion of tools in the review were (1) tools intended for observation of trainees' clinical skills with patients and (2) tools used for direct observation by educational supervisors. An electronic search of English-language studies between 1965 and March 2009 was conducted in five databases in addition to a hand-search of select journals. Articles were independently abstracted and analyzed using validity measures.

Of 10,672 citations, 85 articles met inclusion criteria, yielding 55 unique tools. Tools were used in the US (69%), at a single institution (75%), published in the last decade (50%), and split amongst inpatient (36%), outpatient (24%) or a combination (24%). 62% of learners were residents and 47% were assessed in Internal Medicine. Most tools were developed for formative feedback and had items on history taking, physical exam and communication, and the Mini-Clinical Evaluation Exercise (mini-CEX) was the most frequently studied. Training regarding use of tools was infrequent and the most common outcome measured was trainee satisfaction with the tool. Content validity was most frequently established through consensus and internal tool structure was validated using interrater reliability.

Limitations were the potential for publication bias, search strategy formulation and the limitation to English language articles. Given the cost of standardized patient programs, the need for validated tools that can be used by teaching faculty in settings with real patients is clear. Ensuring that observers are trained may improve tool utility. One commonly used tool—the

mini-CEX—has the strongest validity evidence for direct observation of trainees.

West CP, Dyrbye LN, Sloan JA et al. Single item measures of emotional exhaustion and depersonalization are useful for assessing burnout in medical professionals.⁷

This study determined whether it was possible to assess burnout using two questions from the Maslach Burnout Inventory (MBI) instead of the full 22-item test. Authors assessed the performance of the two questions using data from four previous surveys assessing burnout in medical students, residents, faculty physicians and surgeons. "I feel burned out from my work" assesses emotional exhaustion (EE) and "I have become more callous toward people since I took this job" assesses depersonalization (DP); both have five possible levels of frequency (never to daily).

Original survey response rates ranged from 31.7% among surgeons to 83.6% among faculty physicians, yielding data from 10,525 subjects for analysis. Spearman correlations for both questions ranged from 0.61-0.83 across all four groups relative to the full MBI. Likelihood ratios for both questions showed a clear relationship between the reported frequency and the risk of burnout as predicted by the full MBI. For example, likelihood ratios for EE ranged from 0.01-41.74, correlating with a risk of burnout of 0.4%-95.5%.

The study is limited by response bias. However it does demonstrate that risk of burnout is predicted by these two questions. Researchers, administrators and teachers can probe for EE and DP as triggers to prompt a more in-depth assessment of burnout using the full MBI.

Dyrbye LN, Thomas MR, Harper W et al. The learning environment and medical student burnout: a multicentre study.⁸

This study surveyed 3,080 students in five US medical schools in 2006 to determine the relationship between personal characteristics, aspects of the learning environment and student burnout. The survey included the MBI and items developed by the authors to assess aspects of students' lives associated with distress, including patient and rotational characteristics, workload, learning environment, demographics and life events.

Response rate to the survey was 55%. Thirty-one percent of students experienced a major negative life event in the previous year; 12% a positive event. Seventy-six percent were satisfied with the overall learning environment. Multivariate analysis for preclinical students showed that their rating of overall learning environment, level of support from the faculty, lack of a positive life event and female gender were associated with burnout. Among clinical students overall learning environment, rotation organization, cynical house-staff, less resident supervision, variety of medical problems, and non-minority status were associated with burnout. Odds ratio for burnout for preclinical students who were very dissatisfied with the overall learning environment was 7.12 when compared with students who were very satisfied. The odds ratio for clinical students was 5.96.

This study's cross-sectional design precludes determination of whether burnout precedes or follows dissatisfaction. However, it shows that important factors in a student's life are related to a

well-validated measure of burnout. Aspects of the learning environment highlighted here are potentially modifiable and administrators and teachers should assess their settings and take action to improve it.

Newman LR, Lown BA, Jones RN et al. Developing a peer assessment of lecturing instrument: Lessons learned.⁹

This article described lessons learned in development and evaluation of a criterion-based instrument for peer evaluation of medical lecturing to supplement student evaluations and provide developmental feedback to faculty. Fourteen expert panelists used a modified Delphi process to develop a list of observable, effective teaching behaviors and achieved consensus on 12 criteria. Behavioral descriptors were developed to anchor 5 rating levels. The same panelists assessed four 1-hour recorded lectures to test instrument reliability.

The tool had high internal consistency (Cronbach's $\alpha=0.87$). The inter-rater agreement was fair (ICC=0.27) across all criteria, with substantial variability for individual criteria. Only one criterion ("Provides a conclusion") showed substantial agreement.

A single institution's experts are a useful initial group to design a new instrument yet further development should involve multi-institutional validation. Strong consensus methods yielded a novel instrument which has applications in faculty development and promotion. This tool may evolve to provide valuable assistance in developing and assessing faculty who give lectures.

Lurie SJ, Mooney CJ, Lyness JM. Measurement of the general competencies of the Accreditation Council for Graduate Medical Education: A systematic review.¹⁰

This review assessed published instruments' validity and reliability in evaluating individual learner competencies. Studies identified using searches of databases and reference lists were included if the stated aim was to develop and test an assessment tool for the ACGME competencies. Authors focused on studies of general competencies, systems-based practice (SBP) and practice-based learning and improvement (PBLI); 56 of 127 articles were included.

No general competency studies separated scores related to individual competencies. In studies of global rating scales, competency scores correlated at the level of individual learners and by year of training. The 14 studies of SBP and PBLI reported clinical outcomes rather than competency measures of trainees, or self reported confidence or knowledge rather than objectively measured competency. In the seven publications cross-referencing assessment tools with competencies, tools mapped to multiple competencies and did not uniquely capture individual competencies.

The authors assumed that concordance in competency measures for trainees reflected tool limitations in separating competencies; alternatively, trainees may be generally concordant in competencies across their individual skill sets. Instruments capturing several competencies may be more efficient and therefore useful. Importantly, the search strategy may have missed instruments specific to patient care, interpersonal and communication skills, or professionalism. Instruments with strong validity and reliability evidence that can discriminate competency specific performance of individual

learners, apart from medical knowledge, are lacking. For individual learners requiring greater performance improvement, evaluating individual competencies would allow precise definition of concerns, but tools to do so are currently lacking.

PROFESSIONALISM

Wilkinson T, Wade WB, Knock LD. A blueprint for professionalism: results of a systematic review.¹¹

This review synthesized definitions and interpretations of professionalism, described a toolbox of assessment methods, created a blueprint matching assessment tools to elements of professionalism, and identified gaps where professionalism elements are not well assessed by existing tools.

A PubMed literature search and manual checks of references yielded >30 definitions of professionalism. Authors completed a thematic analysis of elements of the definitions and reached consensus clustering similar attributes by themes. Five thematic elements of professionalism were identified; 4-10 attributes were identified for each theme. These included adherence to ethical practice principles, effective interactions with patients and others working within the health system, reliability and commitment to autonomous maintenance and continuous improvement of competence in self, others, and systems. Each of seven identified assessment tools were then mapped to these attributes of professionalism. Aspects of professionalism without identified assessment tools included reflectiveness, lifelong learning, dealing with uncertainty, advocacy, balancing availability to others with self care, seeking and responding to audit results and advancing knowledge.

Limitations include missed studies due to the search strategy used. This study provides a valuable source of tools to assess specific behaviors and professionalism in general.

Green M, Zick A, Makoul G. Defining professionalism from the perspective of patients, physicians and nurses.¹²

This study identified and prioritized behaviorally-based signs of professionalism relevant to patients, physicians, and nurses. To generate specific professionalism attributes, 22 focus groups of these constituents from IM, general surgery, psychiatry and pediatrics were videotaped and analyzed. Patients, physicians, and nurses then ranked the relative importance of the 68 behaviorally-based attributes from the focus groups to judge their own ability to observe and assess the professional behavior in a survey. Threshold for deeming items "very important and observable" was set at 75% of respondents rating items as such.

18% of surveyed patients participated, yielding 415 surveys. 52% of 415 physician leaders in the American Board of Medical Specialties responded to the physician survey and there were 237 Illinois nurse respondents. Patients deemed 30 of 35 behaviors as "very important", but only 16 of 30 were observable. Nine of the 30 very important behaviors identified by patients were not labeled as "very important" by physicians. Two behaviors "very important" to >75% of patients but not meeting threshold for nurses or physicians were: "pays attention to the cleanliness and comfort of patient areas" (48.5% nurses; 45.0% physicians) and "is open to patient getting a second opinion" (59.6% nurses; 51.6%

physicians). Professional dress was very important to only 35.4% of patients, 53.0% of nurses and 37.1% of physicians.

A limitation was the high threshold for labeling behaviors as “very important”; most items categorized as “unimportant” with this approach were still labeled as “very important” by >50% of respondents. Educators should consider behaviors important to patients and nurses as they re-design and modify tools to assess professionalism. Physicians and nurses should understand that criteria such as cleanliness and comfort of patient areas may have a stronger influence on their patients than expected.

Conflict of Interest: None disclosed.

Corresponding Author: Reena Karani, MD; Brookdale Department of Geriatrics and Palliative Medicine, Samuel M. Bronfman Department of Medicine, Department of Medical Education, Mount Sinai School of Medicine, 1 Gustave L. Levy Place, New York, NY 10029, USA (e-mail: reena.karani@mssm.edu).

REFERENCES

1. **Wolpaw T, Papp KK, Bordages G.** Using SNAPPS to facilitate the expression of clinical reasoning and uncertainties: A randomized comparison group trial. *Acad Med.* 2009;84:517–524.
2. **Peets AD, Coderre S, Wright B, et al.** Involvement in teaching improves learning in medical students: a randomized cross-over study. *Med Educ.* 2009;9:55.
3. **Fraser K, Peets A, Walker I, et al.** The effect of simulator training on clinical skills acquisition, retention and transfer. *Med Educ.* 2009;43(8):784–9.
4. **de Giovanni D, Roberts T, Norman G.** Relative effectiveness of high-versus low-fidelity simulation in learning heart sounds. *Med Educ.* 2009;43(7):661–8.
5. **Barsuk JH, Cohen ER, Feinglass J, et al.** Use of simulation-based education to reduce catheter-related bloodstream infections. *Arch Intern Med.* 2009;169(15):1420–3.
6. **Kogan JR, Holmboe ES, Hauer KE.** Tools for direct observation and assessment of clinical skills for medical trainees: a systematic review. *JAMA.* 2009;302(12):1316–1326.
7. **West CP, Dyrbye LN, Sloan JA, et al.** Single item measures of emotional exhaustion and depersonalization are useful for assessing burnout in medical professionals. *J Gen Intern Med.* 2009;24:1318–21.
8. **Dyrbye LN, Thomas MR, Harper W, et al.** The learning environment and medical student burnout: a multicentre study. *Med Educ.* 2009;43:274–282.
9. **Newman LR, Lown BA, Jones RN, et al.** Developing a peer assessment of lecturing instrument: Lessons learned. *Acad Med.* 2009;84:1104–1110.
10. **Lurie SJ, Mooney CJ, Lyness JM.** Measurement of the general competencies of the Accreditation Council for Graduate Medical Education: A systematic review. *Acad Med.* 2009;84:301–9.
11. **Wilkinson T, Wade WB, Knock LD.** A blueprint for professionalism: results of a systematic review. *Acad Med.* 2009;84:551–58.
12. **Green M, Zick A, Makoul G.** Defining professionalism from the perspective of patients, physicians and nurses. *Acad Med.* 2009;84:566–73.