Student Perceptions of Formative Assessment Used as Instructional Tools in Medical Education

Research Highlights
Youn Seon Lim, PhD-- Donald and Barbara Zucker School of Medicine at Hofstra/Northwell

Purpose: In self-directed learning environments, students are responsible for monitoring and assessing their learning progress. 1 Formative assessments play a central role as instructional tools to train students in these self-assessment skills. 2 Key to the effectiveness of a formative assessment as a training instrument is the degree of student involvement in the design and implementation of the formative assessment used to train them. 3 Five characteristics are essential to secure a sufficient level of student involvement: (1) congruence, (2) authenticity, (3) consultation, (4) transparency, (5) accommodation. These five characteristics of student involvement were used to evaluate the quality of formative assessments that are employed to train undergraduates in a medical school in self-assessment skills. 4,5

Methods: Different from extant approaches that typically rely on instructors judgments to infer the degree of student involvement, this study used a well-established paradigm in higher educational research: students, not instructors were asked about their perceptions of an instructional tool.5 Students had to evaluate their level of involvement in the design and implementation of the formative assessments used to train them. A survey tailored to the specific requirements of this study was developed and validated. A total of 140 undergraduate medical (UME) students participated in the study. Multivariate statistical methods were used to analyze the data.

Results: The statistical analysis of students responses confirmed, with minor modifications, the five characteristics mentioned above as the major dimensions to describe how students perceive their involvement with the formative assessments employed to train them. The majority of medical students appreciated the use of formative assessments to teach them self-directed learning skills. Female students particularly appreciated the feedback included with formative assessments as it supported their learning. Also, females perceived formative assessments of self-directed learning skills as more realistic and relevant to them than male students. All in all, female students expressed greater preference than male students for formative assessments. Many students also expressed concerns that their input to the consultation process is only marginal. Students wished to be more involved with the decision on how their responses to the items of a formative assessment are evaluated, and which criteria in particular are relevant. This seems especially true for minority students.

Discussion: The results clearly showed that, overall, UME students have positive perceptions of the FAs used to teach them. Still, a few remarks concerning potential limitations of this study seem in order. First, the findings of this study should be interpreted with caution and not be over-generalized, as one should keep in mind that this study was conducted at a rather new medical school that was founded just ten years ago. It is currently unknown to what extent the findings from this study apply to other medical schools in the US with a longer-standing academic heritage. Another limitation concerns the validity of the information provided by the participants through the online-survey portal. The reliability analysis provided satisfactory results, but one should keep in mind that it is certainly possible that participants provided reliably inaccurate information because they might have been concerned about the anonymity of the survey, and that openly expressed critique might harm their academic progress.
Significance: This study provided further evidence of the multidimensional nature of UME students perceptions and preferences concerning the instructional tools used to teach them. The findings obtained in this study can help instructors and researchers in medical education to improve the practice of formative assessment in self-directed learning environments. The results also provide a stepping stone for future studies with larger samples from diverse medical schools for developing a broader evaluation of students perceptions of formative assessments as they are used in medical schools in the US.
Time and Product Goals in Procedural Skills Learning: A Randomized Experiment

Research Highlights
David A. Cook, MD, MHPE-- Mayo Clinic College Of Medicine and Science
Becca Gas -- Center for Clinical and Translational Sciences, Mayo Clinic
V. Shane Pankratz, PhD-- University of New Mexico Health Sciences Center
David Farley, MD-- Mayo Clinic College Of Medicine and Science
Martin Pusic, MD, PhD-- NYU School of Medicine

Purpose: Chronometry the purposeful measurement of time is a potentially important yet understudied feature of medical education.(1) The relative influence of time and product goals on medicine-related procedural task learning and performance remains unclear.(2) To address this gap, we conducted a randomized trial comparing time vs product goals, each customized to the learner's prior performance.

Methods: We conducted a cluster-randomized study in the context of a half-day medical simulation-based activity for secondary school students. Each participant completed eight repetitions of a blood vessel ligation task. Between repetitions four and five, each participant was given a randomly-assigned learning goal, namely either a customized time goal ("Try to complete the remaining repetitions in [20% less your previous time] seconds") or a customized product goal ("[N] of your previous repetitions leaked; please try to have the remaining four not leak"). Primary outcomes were time and ligature tightness on the final repetition; secondary outcomes included ligature placement accuracy and NASA-TLX workload. Analyses were adjusted for school and baseline performance.

Results: 80 students were randomly assigned. As hypothesized, those receiving a time goal performed faster on their final repetition (unadjusted mean [standard deviation], 49.0 [16.9] seconds) than did those receiving a product goal (66.7 [28.5]), with an adjusted fold-change 0.81 (95% confidence interval [CI], 0.67, 0.99; p=.04), reflecting 19% faster time. The time goal group performed slightly fewer tight ligatures (20/40 [50%]) than the product goal group (26/40 [65%]), but this did not reach statistical significance (adjusted odds ratio 0.63 [95% CI, 0.37, 1.08; p=.09]). The time goal group also performed fewer accurately placed ligatures than did the product goal group (13/40 [33%] vs 18/40 [45%]; adjusted odds ratio 0.47 [95% CI, 0.22, 0.98; p=.04]). For the NASA-TLX item, "How hurried or rushed was the pace of the task?", the time goal group showed significantly higher scores (16.3 [3.3]; range 1=very low to 20=very high) than did the product goal group (12.8 [4.7]; adjusted difference 3.5 [95% CI, 0.10, 6.9; p=.04]). Between-group differences in NASA-TLX domains of mental demands, level of success, level of work, and frustration were relatively small (<1.5 points) and not statistically significant (p>.24).

Discussion: Customized time-oriented learning goals led to improved time in a simulated surgical task, and customized product-oriented goals led to improved product quality. Those receiving time goals also felt more hurried or rushed in their task, but no different in self-reported mental demands, level of work, or frustration.

Significance: Time and product goals promote differential learning outcomes. Use of both time and learning goals may facilitate more effective and efficient procedural skills learning. (3)
Short Answer Progressive Disclosure Clinical Reasoning Exams May Improve Medical Student Knowledge Acquisition

Purpose: Short answer progressive disclosure clinical reasoning examinations are rarely used in conjunction with multiple choice question (MCQ) examinations in the assessment of preclinical medical students. By mimicking a patient encounter and requiring recall rather than recognition, this method of examination invites students to approach exam preparation in a clinically oriented fashion. However, limited evidence is available to support the efficacy of this type of examination in facilitating medical knowledge acquisition (1). In this study, we investigated the effect of a short answer progressive disclosure clinical reasoning examination on medical student knowledge acquisition as assessed by performance on associated MCQ exams.

Methods: In conjunction with MCQ exams, Period 2 students in the Class of 2021 (CO 2021, n=118) at the Herbert Wertheim College of Medicine (HWCOM) were exposed to two short answer progressive disclosure clinical reasoning examinations during the Cardiopulmonary Systems Course and one during the Gastrointestinal System course. Students in the Class of 2020 (CO 2020, n=130) were exposed to MCQ exams alone. Course content and pedagogy were unchanged over the two years. Aggregate mean performance on internally developed and National Board of Medical Examiners (NBME) MCQ exams was compared for unchanged exam questions that were asked of both CO 2020 and CO 2021. Two-tailed T-tests were used to compare means.

Results: Aggregate mean performance on internally developed MCQs in the Cardiopulmonary and Gastrointestinal Systems courses (n=136 questions) was 83.13% for CO 2020 and 86.67% for CO 2021 (absolute difference=3.54%, p<0.0001). Aggregate mean performance on NBME MCQs (n=109 questions) was 84.42% for CO 2020 and 85.64% for CO 2021 (absolute difference=1.22%, p=0.031). Aggregate mean performance on NBME MCQs including a clinical vignette (n=90) was 84.14% for CO 2020 and 85.49% for CO 2021 (absolute difference=1.34%, p=0.035). The mean cumulative GPA for CO 2020 students was 87.13% ± 3.64% at the end of Period 1, as compared with 87.52% ± 4.00% for CO 2021 students (p=0.42).

Discussion: Students exposed to short answer progressive disclosure clinical reasoning exams performed significantly better on MCQ exams than students exposed to MCQ exams alone. This improvement extended both to internally developed and NBME exams. A comparison of the mean GPA at the of Period 1 suggests no significant difference in baseline performance between classes to explain the performance gap. In addition, there were no major changes made in curriculum or pedagogy between the classes. Because these exams supplemented but did not replace MCQ exams, a testing effect can not be ruled out.

Significance: Our results suggest that utilizing short answer progressive disclosure clinical reasoning exams may contribute to improved medical knowledge acquisition as measured by performance on MCQ examinations. While the difference detected is significantly significant, the absolute difference in performance was small and may not be educationally meaningful. Further investigation will be needed to explore the role of these exams as an adjunct to MCQ exams, to
elucidate why and how they improve MCQ performance, and to assess for any observed translation into improvements in clinical care.
Do medical students behave or perform differently on open vs closed book exams?

Research Highlights
Suzy McTaggart, BS—University of Michigan
Josh Kurtz, BS—University of Michigan Medical School
Jill Cherry-Bukowiec, MD—University of Michigan Medical School
Patrick D. Bridge, PhD—University of Michigan Medical School
Seetha U. Monrad, MD—University of Michigan Medical School

Purpose: The availability of on-line resources at the point of care requires physicians to access and interpret information quickly. These requisite skills identify a need to re-evaluate methods by which medical students are assessed. Open-book assessments (OBAs) have been proposed to help develop the skill of interpreting information rather than testing rote memorization. [2, 3, 5] This study explores differences in students behaviors and performance between OBAs and closed-book assessments (CBAs).

Methods: In 2017-2018, the University of Michigan Medical School (UMMS) administered 39 course assessments during the preclinical phase to 177 students. These included 36 Multiple-Choice (MCQ) CBAs, 2 MCQ OBAs, and 1 Very Short Answer Question (VSAQ) OBA, all administered via an ExamSoft assessment platform. No limits were placed on resources students were permitted to access for the OBAs. Students take assessments during a flexible testing window (generally Friday - Sunday).

Student performance (percentage of items answered correct) and average time used per question were recorded by ExamSoft built-in functionality. To study the impact of assessment modality on these areas, performance and average time per item were reviewed by Kruskal Wallis and Conover-Iman testing.

The time students took their assessment was also recorded by ExamSoft, then recoded to generate a categorical variable [1-5] relating to one-fifth of the testing window. Association between assessment modality and testing time was analyzed graphically and by Chi-Squared testing.

Results: Students scored higher on MCQ OBAs (M=94.58, SD=3.67) than the VSAQ OBA (M=89.66, SD=5.28) and performed better on both OBAs than CBAs (M=86.08, SD=7.61 (Kruskal Wallis X2 (2) = 517.99, pairwise comparisons p<0.0001).

Students showed significant differences in the duration and timing of OBAs compared with CBAs. Review of the time-on-task distributions shows that students used only slightly more time per question on the MCQ OBA (M=1.71, SD=0.55) as compared to CBAs (M=1.57, SD=0.74). Students did, however, utilize more time per question for the VSAQ OBA (M=3.25, SD=1.07) than both the MCQ OBA and CBAs (Kruskal Wallis X2 (2) = 391.1, p<0.0001).

The distribution of when students test during flexible weekend testing is remarkably divergent for OBAs and CBAs. The majority of students test earlier in the flexible testing window for either OBA (X2 (10) = 360.35, p<0.0001) compared to CBAs.

Discussion: Our study adds to the literature in several significant ways [1]. To the best of our knowledge, this is the first medical education study of OBAs that utilized testing platform analytics, rather than student questionnaires, to assess how student behaviors differ between assessment modalities. Our study supports the idea that students prepare less for OBAs, based on when they take the assessment in a flexible testing window, and take more time to answer questions, based
on time spent per question. However, we found that students did not perform worse on OBAs compared with CBAs. Areas for future study include demographic differences, resource utilization, data search strategies, long-term retention of knowledge, and wellness measures with OBAs compared to CBAs.

**Significance:** Open-book assessments can help students develop skills needed to engage with clinical environments enriched by the availability of technology. They also may have the ability to play a role in designing assessment strategies that promote student wellness. This study suggests that previous information regarding OBAs from the literature needs additional review (impact of OBAs on student performance and time on task specifically). Additionally, this is a first step in understanding how OBAs can impact the learning environment and experience for medical students at an institution with flexible testing.