Can a wake-up call improve medical students performance in their first year?

Research Highlights
Ling Wang, PhD-- Michigan State University College of Human Medicine
Heather Laird-Fick, MD, MPH-- Michigan State University College of Human Medicine
Carol Parker, PhD-- Michigan State University College of Human Medicine
Chi Chang, PhD-- Michigan State University College of Human Medicine
Robert Malinowski, PhD-- Michigan State University College of Human Medicine
David Solomon, PhD-- Michigan State University College of Human Medicine

Purpose: In 2016, our medical school launched a new curriculum featuring three segments of integrated science and clinical education called the early, middle, and late clinical experiences. The curriculum includes more traditional seminar-style offerings called intersessions between the early and middle clinical experiences. Science knowledge is assessed through progress tests developed using the NBMEs Customized Assessment Service (CAS) Program as well as the National Board of Medical Examiners (NBME) Comprehensive Basic Science Exam (CBSE) twice per semester. Individualized reports include CAS subject scores corresponding to foundational intersessions (Anatomy/Embryology, Histology/Cell Biology/Pathology, Microbiology/Immunology, and Pharmacology/Toxicology). Students who perform in the lowest quintile for disciplinary topics in the first semester are required to take foundational intersessions designed to help remediate the students weaknesses. There are two time-point interventions: (1) initial notification at the beginning of the second semester that the low performance students are required to take intersessions, and (2) intersession participation 3-5 months later. In this study, we evaluate students performance before and after these two time points of intersession interventions.

Methods: Interrupted time series (ITS) method2 is employed to test for CAS subject scores changes, expressed as slopes, before and after the notification intervention and intersession completion intervention. Six CAS exams are available for analysis: two before and four after notification, and three before and three after intersession completion. The slopes are compared between two groups of students: (A) students who had low performance and completed a required foundational intersession, and (B) students who were not required to and did not complete a foundational intersession. Since the students in Group A are low-performance students in the first semester, we control MCAT scores in the analysis as their latent ability in learning. In addition, the students in Group A are selected based on the first two CAS test results, to avoid potential regression to the mean problem, we control the difference between baseline scores and mean baseline scores.

Results: The total cohort included 191 students: In Anatomy/Embryology, there are 23 in Group A and 89 in Group B; in Histology/Cell Biology/Pathology, 39 in Group A and 40 in Group B; in Microbiology/Immunology, 40 in Group A and 45 in Group B; in Pharmacology/Toxicology, 63 in Group A and 40 in Group B. The difference in slopes for Anatomy/Embryology subject scores is 6.21 points higher for Group A than for Group B for notification (p < 0.01), and 2.54 points higher for Group A than Group B for intersession completion (p = 0.27). Findings are similar for Histology/Cell Biology/Pathology intersession: slope for Group A is 3.64 points higher (p < 0.03) for notification and 2.06 points higher (p = 0.22) for completion than for slope in Group B. In Microbiology/Immunology and Pharmacology/Toxicology, there are no significant differences in slopes for both types of interventions: In Microbiology/Immunology, slope in Group A is 3.96 points higher (p=0.055) for notification and 3.06 points higher (p = 0.14) for completion than slopes in Group B. In Pharmacology/Toxicology slope for Group A is 2.79 points higher (p = 0.16) for notification and 1.29 points higher (p = 0.52) for completion than slopes in Group B.

Discussion: In this observational study, notification of low performance students resulted in statistically significant improvement compared to peers for two CAS subject scores: Anatomy/Embryology and Histology/Cell Biology/Pathology. In contrast, students with low performance did not have statistically significant improvement in all four subjects compared to their peers after a dedicated four-week intersession seminars.

Significance: These findings suggest that early wake-up calls -- counseling students about
implications of their performance -- is a powerful intervention for learning, and may have a greater effect than programmed learning activities.
Creating a Community of Innovation: Reflection on 5 years of the Accelerating Change in Medical Education Consortium

Innovation Highlights
Kimberly Lomis, MD-- American Medical Association
Susan Skochelak, MD, MPH-- American Medical Association
Michael Dekhtyar-- American Medical Association
Sally Santen, MD, PhD-- Virginia Commonwealth University School of Medicine
John Andrews, MD-- American Medical Association
Judee Richardson, PhD-- American Medical Association
Kevin Heckman, MBA-- American Medical Association
Betsy Shea-- American Medical Association

Purpose: The American Medical Association (AMA) formed the Accelerating Change in Medical Education Consortium (ACE) in 2013 to support rapid innovation among medical schools and disseminate the ideas being tested to additional medical schools. The initiative called upon partners to redesign curricula for flexible, individualized learning pathways, measure achievement of competencies, develop new assessment tools to test readiness for residency, and implement new models for clinical experiences within health care systems. The initial group of 11 schools were each awarded $1 million. The group was expanded via smaller grants in 2016 to a total of 32 institutions, ultimately impacting approximately 19,000 medical students. The AMA supported two meetings per year of the collective, as well as multiple smaller thematic meetings based on shared interests across institutions. As the goal of the consortium is to accelerate change in undergraduate medical education, it is important to assess the consortium’s role in supporting the actual implementation of the proposed projects. The initial grant period ended in fall 2018.

Methods: Throughout the grant period, consortium schools submitted semi-annual progress reports that included routine updates on the specific objectives of each site’s project and captured barriers and facilitators. The final grant reports, submitted in fall 2018, also solicited impacts on faculty and the health system. The AMA ACE staff reviewed interim and final reports to cull key issues related to innovation.

Results: Most institutional objectives were successfully addressed, while some were adjusted or abandoned. Over 150 publications, 250 presentations, and 350 direct consultations were reported by the collective during the grant period. The professional development of educators was advanced through the creation of new faculty roles, many of which were funded. Over 40 reports of significant student impact on the health system were submitted. Common barriers to innovation included issues such as changes in leadership, resources, and organizational structure, as well as competing demands on faculty time. Facilitators included a supportive community, significant grant and in-kind resources, project management provided by AMA staff, external validation for change efforts and shared learning.

Discussion: Interestingly, some of the aspects of the consortium that were cited by members as particularly valuable in facilitating change are uncommon in traditional academic forums. Discussion of works in progress was the norm, which allowed teams to seek feedback and gain recognition for efforts in real time. The consortium deliberately created opportunities to share and discuss failures; this uncharacteristic degree of transparency was cited as particularly valuable in accelerating advances by sparing other institutions from duplicating mistakes. The member institutions reported a significant number of formal consultations with one another beyond the official consortium.
gatherings, indicating the development of a true community. AMA efforts to raise awareness via non-academic media outlets enhanced institutional recognition and spurred additional consultations from outside the consortium. All 32 member institutions have committed to continuing participation in the consortium, despite the lack of ongoing site-specific funding, and all are participating in ongoing evaluation efforts to assess impact.

**Significance:** Creating a community of innovation provides both the resource and moral supports essential to overcoming the significant challenges associated with transformation. Key elements that promoted the success of this consortium differ from traditional academic approaches and can inform future collective efforts for improvement across the medical education system.
Development of an Quality Assessment Scale for Educational Videos for Use in Undergraduate Medical Education

Innovation Highlights
Sean Schooley, BA-- University of Miami Leonard M. Miller School of Medicine
Sean Tackett, MD, MPH-- Johns Hopkins University School of Medicine
Lina Shehadeh, PhD-- University of Miami Leonard M. Miller School of Medicine

Purpose: Medical educators have an array of videos to choose from, and using videos can often make learning more efficient. However previous studies have suggested that a videos popularity is a poor surrogate for its usefulness, and there are no standardized methods for evaluating the quality of medical education videos. Therefore, we developed a scale that can be used by learners, educators, and content developers to evaluate the quality of videos used in medical education.

Methods: Based on a literature search to determine desirable qualities of educational resources and multimedia and curriculum development principles, we drafted 27 items to be included in the quality rating instrument. Items were grouped into five domains of educational design, source reliability, technical quality, learning quality, and accessibility and revised serially by piloting on medical education videos on Youtube.13 Domains were established by linking items known to be associated (e.g. DISCERN criteria, Mayers principles of the Science of Learning/Instruction) or items found through the literature search to be associated, as in the technical quality items. Data collection was performed independently by two raters for 15 videos found through a YouTube search for cardiovascular physiology. The first three categories were assessed to determine baseline reliability. Cronbachs alpha was used to determine internal consistency. Intra-rater reliability was determined using the Intra-class Correlation Coefficient (ICC) two-way random model with single scores. The statistics were calculated using Past3 statistical software package.

Results: The average score in the technical quality section was 7.9/10, while in the educational design and reliability sections together was 3.4/12. Only 13% of the videos made it clear what information sources were used, 7% made it clear when the video content was last updated, and 0% acknowledged areas of uncertainty related to the topic. Also, no videos provided opportunities for self-assessment during or at the end of the video. Cronbachs Alpha of (0.939, CI (0.690, 1.000)) for the scale total indicated the internal consistency was high. ICC(2,1) of (0.7084, CI (0.6247, 0.7700)) for the scale total indicated the intra-rater reliability was good.

Discussion: Results from this pilot seem to indicate that producers of videos put more emphasis on technical quality of videos than on educational design or reliability. This is concerning because studies have shown that video consumers cannot distinguish videos with useful content from other videos, and suggests that content and sources of educational videos deserve greater scrutiny. Likewise, assessment drives learning, but videos commonly lack opportunities for assessing knowledge to be gained from videos. Finally, pilot results from our instrument development are promising and warrant further efforts to generate validity evidence for its use.
Significance: The development of our video quality rating instrument has created a summary of best practices in video creation that can guide content development and be used by content consumers when selecting videos to be used in medical education. Ongoing data collection using the instrument on a larger number of videos, including videos from multiple topics, and assessment of how quality ratings relate to learning, will provide further validity evidence for the scale and recommend its use more broadly.
Application of Propensity Score Matching in Evaluation of Longitudinal Clerkships

Yuanyuan Zhou, PhD-- Texas A&M University
William H. Pieratt, DO, FACP-- Texas A&M Health Science Center College of Medicine
Cayla R. Teal, PhD-- Texas A&M Health Science Center College of Medicine

Purpose: The A&M Integrated Medicine (AIM) program is a longitudinal integrated clerkship. In contrast to the traditional program in which each of the core clerkships is taken individually and sequentially, the AIM program has students work with physicians in core specialties continuously throughout a year while simultaneously following a panel of patients representing a wide spectrum of medical conditions. After a small pilot in the academic year (AY) 13-14, about 20 students in each years cohort have enrolled in the AIM program with the remainder enrolled in the traditional program. Prior research has suggested that longitudinal integrated program are more effective,1-2 which prompts preceptors to know students over the clerkship duration to provide more constructive feedback,1 and results in greater maintenance of patient-centered attitudes.2 However, the data regarding student performance on summative evaluations of knowledge, clinical performance, and clinical skills have been mixed.3 The methods used to evaluate programs have varied, but the choice of methods needs to meet with the increasingly rigorous evaluation standards so that the evidence used to generate conclusions can stand up to close examination.4 Any comparative evaluation of the AIM and traditional programs must account for the absence of student random assignment, as students self-select into the program of their choice. Therefore, direct comparison of learning outcomes between the two programs may not be the best practice.

Methods: Propensity score matching (PSM) is a statistical technique developed to estimate the treatment effect in quasieperimental and even observational study.5 In this case, a student who receives the treatment, that is, participates in the AIM program, can be matched to a student in the traditional program using the calculated propensity score, which simulates a counterfactual of the same person as an AIM student and a traditional student, thus enabling the calculation of the AIM program (or treatment) effect. This study included 726 students from four cohorts (2016, 2017, 2018, and 2019), in which 77 belonged to the AIM program, and 649 were from the traditional program. Six clerkships NBME shelf exams results were used as indicators of students learning outcomes. Average core preclerkship course grades (introduction to disease, neuroscience, and multiple organ systems courses) and Step 1 were selected as the covariates. Four types of mean differences were calculated: direct comparison of means between the two programs, PSM using preclerkship course grades as the covariate, PSM using Step 1 as the covariate, and PSM using preclerkship course grades and Step 1 as covariates.

Results: In direct comparison, except Psychiatry, students in AIM program on average performed higher (2-3 points in equated percent corrected score) in all other five clerkships NBME shelf exams. However, after using the PSM analytical method, the positive differences were enhanced (0-2 points) no matter which covariate or covariate combination (step 1 or preclerkship grade) were used to create a match.

Discussion: Results from direct comparison and PSM both indicated the superiority of the AIM program in producing stronger NBME shelf exam performance. Since PSM is able to address the non-random assignment in the observational study, PSM results are more scientifically
defensible. Additional analyses have indicated that as a group, the AIM students actually were poorer preclerkship performers overall compared to traditional students overall. The PSM analysis suggests that this difference was overcome by the AIM experience.

**Significance:** In medical education program evaluation, there are few experimental studies that include randomization into treatment conditions. For the sake of students best development, the evaluation almost always deals with data that have many other confounding effects. Applying advanced statistical methods such as PSM in program evaluation may be a good practice in improving the quality of program evaluation.