

Nimble Approaches to Curriculum Evaluation in Graduate Medical Education

DARCY A. REED, MD, MPH

Gold-standard approaches to curriculum evaluation in medical education are well established¹⁻⁴ and possess considerable legitimacy among educators. However, despite their apparent validity, these traditional methods are often resource intensive and time consuming and can require specialized training that faculty may lack.^{5,6} Given today's resource-constrained climate, there is a need for speedier and more nimble approaches.

In this issue of the *Journal of Graduate Medical Education*, Willett and colleagues⁷ describe the use of the *ecological momentary assessment* (EMA) to evaluate the internal medicine ambulatory morning report during a period of almost 3 years at the University of Alabama at Birmingham. An evaluation methodology rooted in behavioral medicine, EMA is designed to assess "complex and temporally dynamic psychological, behavioral, and physiological processes in the natural environment."^{8(p35)} The EMA involves repeated sampling of individuals in real time, thus providing immediate evaluation data and minimizing recall bias.⁹ The study by Willett and colleagues⁷ is one of the first reports of the use of EMA in graduate medical education.

Willett et al⁷ conducted a prospective study of 125 internal medicine residents attending ambulatory morning report during a 32-month period.⁷ The authors created an 8-item EMA tool that assessed a resident's views of individual morning report sessions by including their opinion of session content, structure, and learning attained. This tool was administered immediately following each morning report session (3 times per week) and took residents less than 1 minute to complete. Assessments were anonymous and approximately 75% of residents responded, on average, across sessions.

During the first 12 months of the EMA data collection, the investigators discovered that senior residents viewed morning reports as less educationally valuable compared with more-junior residents. In response to these data, teaching faculty implemented a new morning report format with content of a higher cognitive level; senior residents' EMA scores improved during the subsequent 6 months. Thus, this study demonstrates successful use of the EMA to

direct curriculum evaluation and provides "proof of concept" for the use of this approach in graduate medical education.

Important limitations of the study include the inability to account for clustering of assessments within residents (because of the complete anonymity of assessments), and EMA scores were globally high and subject to ceiling effects, as observed for many assessments in medical education.^{10,11}

Rigorous curriculum evaluation is a fundamental responsibility of medical schools and residency programs^{12,13}; therefore, it constitutes a core skill set for medical educators. New understandings about the powerful role the informal and hidden curricula play in the professional development of medical learners¹⁴ raise obvious questions about the relevance of the formal curriculum (to which educators devote most of their efforts) for the development of learner competency. Nonetheless, educators continue to spend countless hours developing, implementing, evaluating, refining, reforming, and continuously improving curricula.

Curriculum evaluation in graduate medical education typically consists of institution-specific evaluation forms completed by residents at the end of clinical rotations. These forms may be administered via a paper or electronic survey and usually contain evaluation items with ordinal scales pertaining to the content of curricula, the quality of the teaching, and other aspects of the learning environment. Sometimes these evaluations are carefully crafted to reflect unique elements of the specific curricula being evaluated (eg, evaluation questions about certain lectures, simulation exercises, procedural training), and this concordance between the curriculum and its evaluation imparts important content-validity evidence for evaluation scores.¹⁵ Furthermore, some institutions have developed sophisticated electronic-evaluation systems^{16,17} that amass enormous amounts of evaluation data that readily allow, because of large sample sizes, assessment of internal consistency and interrater agreement.^{17,18} Linking these evaluation systems to other educational and/or health care databases facilitates the examination of relationships between evaluation scores and other variables, allowing educators to establish the criterion validity of assessments.^{15,19} Such large evaluation data sets also support education research.

Despite these advantages, current approaches to curriculum evaluation also have important limitations. First, expansive evaluation systems require significant time and resources to develop and maintain, and partnerships with psychometricians and information technologists are often required. Second, to increase the number of

Darcy A. Reed, MD, MPH, is Associate Director of the Internal Medicine Residency Program at Mayo Clinic College of Medicine.

Corresponding author: Darcy A. Reed, MD, MPH, Internal Medicine Residency Program, Mayo Clinic College of Medicine, 200 First Street SW, Rochester, MN 55905, 507.284.6391, reed.darcy@mayo.edu

Received April 6, 2011; accepted April 6, 2011.

DOI: 10.4300/JGME-D-11-00081.1

TABLE ADVANTAGES AND DISADVANTAGES OF NIMBLE CURRICULUM-EVALUATION MODALITIES

Evaluation Modality	Advantages	Disadvantages
Ecological momentary assessment (EMA)	Provides frequent, immediate, actionable feedback about a specific aspect of a curriculum May reduce recall bias	Evaluation data generally are not standardized across curricula and/or programs
Continuous quality improvement (CQI)	Promotes identification of specific problem areas ("defects") within curricula Allows rapid implementation of curriculum process changes	Not an optimal method for curricula with numerous limitations requiring comprehensive reform
Iterative reflection	Encourages development of reflective practice Engages stakeholders and promotes personal investment in curriculum	May require a longer period to obtain and act on evaluation data

assessments for each evaluation item, evaluation forms are often standardized across curricula within a residency program or institution. Such standardization enhances the ability to draw quantitative conclusions from numeric evaluation data, but may decrease the extent to which evaluation items reflect unique curricular components, and thus decreases the content validity of the evaluations. For example, when evaluations are standardized across clinical rotations, an item like "this rotation met my learning needs" may be selected (because it applies to all rotations) rather than a more specific item like "the central line placement workshop prepared me to place central lines independently" (which may apply only to the specific rotation in which this workshop was offered). The former item is certainly less helpful than the latter in directing curriculum improvement.

The third, and perhaps most important, limitation of standard curriculum evaluation methods is that often the methods are not very nimble. By definition, evaluations that are collected at the end of the curriculum can only be used to inform future curricula; they do not provide real-time feedback to direct improvements when a curriculum is actually being taught. Therefore, an unsuccessful curriculum is typically not discovered until the curriculum is over, the learners have moved on, and the only recourse is to try to improve the program for the next session. This problem is compounded by it taking weeks to months for learners to complete curriculum evaluations, and often, the evaluations then need to be collated and synthesized before they are fed back to teachers. In some instances, data are purposely held for an extended period before they are shared with teachers to collect enough responses to ensure anonymity. Despite allowing an extended period for learners to complete evaluations, response rates to evaluations may still be suboptimal. Waiting until the end of educational experiences also risks recall bias, in which trainees rate only the activities easily recalled. The end result is that teachers often wait a long time to receive evaluation data, and those data may be only partial

representations of the learner opinion, which can cause delayed or misguided curricular improvement. Alternative approaches to curriculum evaluation that may offer greater agility include EMA, continuous quality improvement (CQI), and iterative reflection. As described by Willett et al,⁷ the advantage of EMA is that it provides frequent, immediate feedback about a specific aspect of a curriculum (eg, today's ambulatory morning report session) that teachers can then act on to improve the next session. Because EMA data can be collected during or immediately after the curriculum session, this method is thought to reduce recall bias.⁷⁻⁹ However, the extent to which delayed evaluations are actually influenced by recall bias is uncertain. McOwen and colleagues²⁰ recently examined whether the time elapsed between administration of a curriculum and the students' evaluations of that curriculum affected the students' curriculum ratings. They found that with greater elapsed time, students' mean ratings increased and the variability of ratings decreased, but the magnitude of those differences was so small as to be judged educationally insignificant by the authors.²⁰ However, in that single-institution study, most students returned their evaluations within 4 weeks; therefore, it remains unknown whether the magnitude of effects would be greater in populations of learners with longer periods between curricula and assessments.

The CQI methodology, now commonplace in health care, has also been adapted to curriculum evaluation.²¹⁻²³ Like EMA, CQI involves rapid data collection, small tests of change, and repeat data collection in iterative cycles.²⁴ This method may be advantageous for busy educators who need to quickly identify problem areas within their curricula and make the necessary improvements without allocating substantial resources to expansive data collection processes. Although some may question the legitimacy of CQI relative to traditional curriculum-evaluation approaches, standard guidelines exist for conducting and reporting CQI in health care,²⁵ and those same standards can be easily applied to curriculum evaluation.

Finally, some educators have implemented a process of iterative reflection to evaluate and improve curricula.^{26,27} Reflection in curriculum evaluation is generally a qualitative (or mixed qualitative and quantitative) approach that can provide rich data to direct curriculum improvement. Fetterman and colleagues²⁷ recently described the use of empowerment evaluation to reform the medical school curriculum at Stanford University School of Medicine. Empowerment evaluation includes 5 key tools, but a hallmark of the method appears to be the empowerment of stakeholders to join together in iterative cycles of reflection on curriculum issues. The process includes regular reflection on curriculum assessment data as well as on self-assessment to facilitate the development of individuals as reflective practitioners.²⁷

Educators seeking nimble methods of curriculum evaluation that provide immediate feedback may wish to consider EMA, CQI, and iterative reflection (TABLE). Each individual evaluation method has its strengths and limitations; ideally, educators should employ a combination of methods to attain both meaningful and timely information about curricula to direct curriculum-improvement efforts.

References

- Kern DE, Thomas PA, Howard D, Bass E. *Curriculum Development for Medical Education: A Six-Step Approach*. Baltimore, MD: Johns Hopkins University Press; 1998.
- Green ML. Identifying, appraising, and implementing medical education curricula: a guide for medical educators. *Ann Intern Med*. 2001;135(10):889–896.
- Morrison J. ABC of learning and teaching in medicine: evaluation. *BMJ*. 2003;326(7385):385–387.
- Musick DW. A conceptual model for program evaluation in graduate medical education. *Acad Med*. 2006;81(8):759–765.
- Huwendiek S, Mennin S, Dern P et al. Expertise, needs and challenges of medical educators: results of an international web survey. *Med Teach*. 2010;32(11):912–918.
- Windish DM, Gozu A, Bass EB et al. A ten-month program in curriculum development for medical educators: 16 years of experience. *J Gen Intern Med*. 2007;22(5):655–661.
- Willett LL, Wellons MF, Hartig JR et al. Do women residents delay childbearing due to perceived career threats? *Acad Med*. 2010;85(4):640–646.
- Smyth JM, Stone AA. Ecological momentary assessment research in behavioral medicine. *J Happiness Stud*. 2003;4:35–52.
- Shiffman S, Stone AA, Hufford MR. Ecological Momentary Assessment. *Annu Rev Clin Psychol*. 2008;4:1–32.
- Cacamese SM, Elnicki M, Speer AJ. Grade inflation and the internal medicine subinternship: a national survey of clerkship directors. *Teach Learn Med*. 2007;19(4):343–346.
- Roman BJ, Trevino J. An approach to address grade inflation in a psychiatry clerkship. *Acad Psychiatry*. 2006;30(2):110–115.
- Accreditation Council for Graduate Medical Education. Common program requirements. Available at: http://www.acgme.org/acwebsite/home/Common_Program_Requirements_07012011.pdf. Accessed March 1, 2011.
- Liaison Committee for Graduate Medical Education. Functions and structure of a medical school: standards for accreditation of medical education program leading to the M.D. degree. Available at: <http://www.lcme.org/functions2010jun.pdf>. Accessed March 1, 2011.
- Hafferty FW. Beyond curriculum reform: confronting medicine's hidden curriculum. *Acad Med*. 1998;73(4):403–407.
- Downing SM. Validity: on meaningful interpretation of assessment data. *Med Educ*. 2003;37(9):830–837.
- McOwen KS, Bellini LM, Morrison G, Shea JA. The development and implementation of a health-system-wide evaluation system for education activities: build it and they will come. *Acad Med*. 2009;84(10):1352–1359.
- Beckman TJ, Mandrekar JN, Engstler GJ, Ficalora RD. Determining reliability of clinical assessment scores in real time. *Teach Learn Med*. 2009;21(3):188–194.
- Beckman TJ, Cook DA, Mandrekar JN. What is the validity evidence for assessments of clinical teaching? *J Gen Intern Med*. 2005;20(12):1159–1164.
- Cook DA, Beckman TJ. Current concepts in validity and reliability for psychometric instruments: theory and application. *Am J Med*. 2006;119(2):166.e7–166.e16.
- McOwen KS, Kogan JR, Shea JA. Elapsed time between teaching and evaluation: does it matter? *Acad Med*. 2008;83(10)(suppl):S29–S32.
- Friedman CP, Krams DS, Mattern WD. Improving the curriculum through continuous evaluation. *Acad Med* 1991;66(5):257–258.
- Rose SH, Long TR. Accreditation Council for Graduate Medical Education (ACGME) annual anesthesiology residency and fellowship program review: a “report card” model for continuous improvement. *BMC Med Educ*. 2010;10:13.
- Heard JK, O'Sullivan P, Smith CE, Harper RA, Schexnayder SM. An institutional system to monitor and improve the quality of residency education. *Acad Med*. 2004;79(9):858–864.
- Cleghorn GD, Headrick LA. The PDSA cycle at the core of learning in health professions education. *Jt Comm J Qual Improv*. 1996;22(3):206–212.
- Davidoff F, Batalden P, Stevens D, Ogrinc G, Mooney S. Publication guidelines for improvement studies in health care: evolution of the SQUIRE Project. *Ann Intern Med*. 2008;149(9):670–676.
- Spratt C, Walls J. Reflective critique and collaborative practice in evaluation: promoting change in medical education. *Med Teach*. 2003;25(1):82–88.
- Fetterman DM, Deitz J, Gesundheit N. Empowerment evaluation: a collaborative approach to evaluating and transforming a medical school curriculum. *Acad Med*. 2010;85(5):813–820.