Assessing Practice-Based Learning and Improvement

Deirdre C. Lynch and Susan R. Swing
Accreditation Council for Graduate Medical Education
Chicago, Illinois, USA

Sheldon D. Horowitz
American Board of Medical Specialties
Evanston, Illinois, USA

Kathleen Holt
Accreditation Council for Graduate Medical Education
Chicago, Illinois, USA

Joseph V. Messer
Rush-Presbyterian-St. Luke's Medical Center
Department of Medicine
Chicago, Illinois, USA

Background: *Practice-based learning and improvement (PBLI) is 1 of 6 general competencies expected of physicians who graduate from an accredited residency education program in the United States and is an anticipated requirement for those who wish to maintain certification by the member boards of the American Board of Medical Specialties. This article describes methods used to assess PBLI.*
Summary: *Six electronic databases were searched using several search terms pertaining to PBLI. The review indicated that 4 assessment methods have been used to assess some or all steps of PBLI: portfolios, projects, patient record and chart review, and performance ratings. Each method is described, examples of application are provided, and validity, reliability, and feasibility characteristics are discussed.*
Conclusion: *Portfolios may be the most useful approach to assess residents' PBLI abilities. Active participation in peer-driven performance improvement initiatives may be a valuable approach to confirm practicing physician involvement in PBLI.*

Teaching and Learning in Medicine, *16*(1), 85–92

Practice-based learning and improvement (PBLI) is the experiential bridge between continuous learning and good patient care. PBLI abilities are important considering the growth of medical knowledge and subsequent need for ongoing learning, the observation that current practice is not always the best,[1] and the relative ineffectiveness of widely used didactic approaches to continuing medical education (CME).[2] The broader rationale for PBLI emanates from three assertions. First, physicians should have systematic approaches for monitoring and improving their practice. Second, physicians must be able to recognize the need for positive change[3] and instigate it rather than react to changes made by others.[4] Third, positive changes in small systems, such as an individual physician's practice, may positively affect larger systems.[5,6]

Derived from continuous improvement concepts,[7] PBLI has been described as the ability to complete the following sequential steps: determine improvement needs, identify and apply an intervention, and measure the impact of the intervention.[8] Both the Accreditation Council of Graduate Medical Education[9] and proponents of an experiential approach to CME,[10,11] advance this description by indicating that PBLI is the ability of a physician to identify learning needs relevant to improving practice, engage in and apply new learning to practice, and measure the impact of the intervention (Table 1). Therefore, the latter description, which is

This work was supported in part by Grant No. 034768 from the Robert Wood Johnson Foundation. The views expressed in this article are solely those of the authors.

The text and table in this article are our own work.

Correspondence may be sent to: Deirdre C. Lynch, RhD, Accreditation Council for Graduate Medical Education, 515 North State Street, Suite 2000, Chicago, IL 60610–4322. E-mail: dlynch@acgme.org

Table 1. *Approaches Used to Assess PBLI Steps*

PBLI Steps	Approaches Used to Assess PBLI Steps
Monitor practice	Portfolio,[22–25,27–29] medical record review,[37] performance ratings[38,40,41]
Reflect on or analyze practice to identify learning or improvement needs	Project,[32,33] portfolio
Engage in learning or plan improvement	Project, portfolio
Apply new learning or improvement	Project, portfolio, medical record review
Monitor impact of learning or improvement	Project, portfolio, medical record review, performance ratings

Note: PBLI = practice-based learning and improvement.

used to describe PBLI in this article, emphasizes learning as part of the improvement process and is consistent with models of practitioner learning and change proposed by Slotnick[12] and Grol.[13]

Given the steps involved in PBLI, it is clearly a multicomponent concept that pertains to knowledge and skills in many areas such as problem solving, critical thinking, evidence-based medicine, information technology, population health, error analysis, self-monitoring, reflection, self-directed learning, and systematic approaches to improving practice. As one of six general competencies expected of physicians who graduate from an accredited residency education program in the United States and an anticipated requirement for those who wish to maintain certification by the member boards of the American Board of Medical Specialties (ABMS), the crux for educational programs and certification organizations lies in identifying feasible approaches to assess PBLI that yield valid and reliable data. The purpose of this article, therefore, is to identify and describe approaches to assess PBLI. The concluding section explores the potential of selected methods for assessing the PBLI abilities of residents and practicing physicians.

Method

Six electronic databases, from 1980 to September 2002, were searched: MEDLINE, TIMELIT, ERIC, PsycINFO, CINAHL, and Health and Psychosocial Instruments. Several search terms were used, including practice-based learning, practice improvement, self-directed learning, lifelong learning, reflection, and variations of these terms and their combination with assessment, evaluation, and medical education. Also, reference lists of retained articles were manually searched. Four criteria were used to select assessment techniques. Assessment had to (a) address learning by residents or practicing physicians, (b) pertain to learning applied to medical practice, (c) have a history of use as evidenced by data reported in the peer-reviewed literature, and (d) assess one or more PBLI steps. Although cognitive tests, such as knowledge of statistics, critical appraisal, and population-based medicine have been used to assess PBLI, they were not included because these assessments typically target isolated components of PBLI rather than the linked steps involved in PBLI.

The following information was gleaned from retained articles: description of assessment method and its application, validity, reliability, and feasibility. Validity is the extent to which evidence supports the inferences that are made from assessment results.[14] Different types of evidence may be used to infer validity. Content validity may be inferred when assessment items adequately represent the domain or topic being assessed and is key to determining the breadth of items that should be included in a performance rating scale and portfolio, and the specific diagnoses or procedures that should be addressed by a medical record review. Construct validity may be inferred when an assessment captures the concepts targeted for assessment. This is an important consideration in determining the criteria for PBLI projects and how portfolio content reflects selected abilities. Criterion validity may be inferred when assessment results correlate with results obtained from another assessment designed to measure a similar attribute (concurrent validity) or with a key future behavior (predictive validity).[15]

Reliability refers to the consistency of assessment results.[16] There are different types of reliability and the relative importance of each depends on the assessment method used. Interrater or interobserver reliability is the extent of agreement across two or more observers rating the same performance and is relevant to judging portfolios, projects, and medical record reviews. Another approach to examining reliability is to determine the number of data samples required to obtain a stable estimate of performance. This approach is relevant to performance rating forms and medical record reviews. Feasibility pertains to the practical aspects of implementing assessment, such as cost and ease of use.

Results

Of approximately 4,500 abstracts, 460 were selected for further review. Only 13 articles met selection criteria. They described assessments that fit into the following four categories: portfolios, projects, medical record review, and performance ratings. The paucity of techniques used to assess PBLI may be due to its relative novelty. Indeed, some examples were not designed to address all PBLI steps, but are included, nonethe-

less, to stimulate consideration of possible approaches to assess PBLI.

Portfolios

A portfolio is "a purposeful collection of student work done over time."[17,18] It functions as both a stimulus for self-directed learning and a repository for evidence of reflection, judgment,[17] learning and achievement.[19] Portfolios may be paper-based or electronic. The exact content of a portfolio depends on the kind of learning and assessment questions it is designed to address. Many portfolios include a self-monitoring component that requires an individual to observe occurrences of specific behaviors or events and keep a record of them.[20] Diaries, journals, logs, and recordings of critical incidents (significant events[21]) are examples of self-monitoring techniques.

In Britain, for instance, physicians have used portfolios to facilitate and document CME activities,[22,23] and residents have used them to aid reflective learning.[24,25] Emergency medicine, obstetric and gynecology[24] and family medicine[25] trainees used portfolios to log critical incidents designed to trigger learning and initiate dialog with their mentors. The family medicine residents also kept a diary. The emergency medicine and obstetric and gynecology residents were asked to document learning achieved, but very few did. As part of CME, physicians were required to place the following items in their paper-based portfolios: a learning plan (which is a list of learning needs or objectives and approaches to addressing them),[26] a set of critical incidents (to prompt reflection on learning or improvement needs), a second learning plan based on the critical incidents, a description of how this new learning had been used in practice, a self-appraisal of their learning, and a learning plan for the next year.

In Canada, residents have used an Internet-based portfolio to encourage reflection and self-directed learning.[27] Named the KOALA™, this structured relational database program includes a conduit for searching online resources and was initially available for use by obstetric and gynecology residents at four medical centers. The portfolio was designed to collect information about the following: a log of patient encounters, a list of critical incidents, a list of clinical questions derived from the critical incidents, the information used to answer these questions, and a statement indicating whether the resident planned to change practice behavior as a result of new learning. The KOALA™ has since expanded to about 16 medical centers and has been tailored for use by radiology, maternal–fetal medicine, and obstetric and gynecology specialty and subspecialty programs.

Physicians who participate in the Maintenance of Competence Program of the Royal College of Physicians and Surgeons in Canada may document their CME activities by completing either paper- or computer-based portfolios (initially called PCDiary® but now known as Web Diary®).[28,29] Designed to assist self-directed learning, physicians are asked to respond to the following questions (paraphrased): What do you want to learn? What stimulated you to consider this? What resources did you use to help you learn? What do you think the outcome of your learning will be? Recording date? Time spent on learning activity? References used?

These examples indicate that portfolios have been used to obtain evidence of the following PBLI steps: monitor, reflect and analyze practice experience; and identify, engage in, and apply improvements or new learning.

The validity of portfolio data was addressed in three studies. Residents who used the KOALA™ had significantly higher scores on the Self-Directed Learning Readiness Scale than residents who did not use it (i.e., concurrent validity).[27] The construct validity of Web Diary® was examined in a follow-up survey of and group meetings with users.[28] Physicians who attended the meetings reported that Web Diary® helped them to focus their learning activities and be more systematic in their approach to learning. Over 70% of survey respondents reported enhanced reflection about learning activities and how they relate to practice. Physicians who intended to modify their practice as a result of new learning were significantly more likely ($p < .0001$) to agree with the following statements: "Entering items of learning into my diary encourages me to think about my practice" and "Entering items of learning into my diary encourages me to think in terms of questions that guide my learning." The concurrent validity of residents' personal learning logs was examined by comparing specialty-specific knowledge and confidence of those who completed logs and those who did not.[24] There were no significant differences between groups. The authors were not surprised by these findings, however, given residents' relatively poor participation in completing logs.

None of the aforementioned studies addressed reliability. Research indicates that the variability of portfolio content results in unreliable ratings. In one study, even the use of relatively global rating criteria resulted in a poor to moderate interrater reliability range of 0.1 to 0.41 and a moderate intrarater reliability range of 0.38 to 0.54.[30] Subsequent research in this area, however, indicated that pairs of raters who discussed rating criteria prior to judging portfolios increased interrater reliability to 0.5.[31] In terms of feasibility, portfolios have been reported as being relatively time consuming for both learners and assessors.[23,30] This disadvantage, however, seems to be counterbalanced by the educational value of portfolios[17] and the flexibility they afford learners.[23] In their evaluation of portfolio use,

Snadden and Thomas[25] concluded that the effective use of portfolios by residents depended on having a supportive mentor and no impending examinations. Following a disappointing pilot of learning logs with residents, Kelly and Murray[24] also emphasized the importance of active engagement between residents and their mentors and a systematic approach to using portfolios for formative assessment.

Project

According to Lough and Murray,[32] projects have four characteristics: they aid learner understanding of an issue through active involvement in a real-life problem, prompt learner responsibility for planning and completing learning activities, comprise work carried out over a period of time, and result in a written product.[33] The topic for a PBLI project is stimulated by physician awareness of a practice behavior that needs improvement. The purpose of the project is to provide evidence of the physician's ability to analyze and improve his or her practice of medicine. Completion of a PBLI project (known as an audit project in Britain) constitutes one of four parts of the summative assessment of general practitioner registrars (residents) in Britain. The projects, which begin in the final year of training, have to be completed within 1 year and are graded according to a checklist that has been modified and improved since its initial use in 1997.[32,33] Checklist criteria require the following characteristics of each project be examined: project rationale, choice of intervention (and supporting evidence), standard setting, preparation, implementation, data collection, and conclusions.

The project thus provides evidence of all PBLI steps, from monitoring practice, to identifying and applying improvements, to ascertaining the effects of planned changes in practice behavior. Because the project itself is a permanent product that reflects all steps of a completed PBLI cycle, it could also be a portfolio item.

Two rounds of expert review, with an intervening implementation period, were conducted to address the validity of the PBLI project checklist used to rate projects.[32,33] Because these projects are part of a summative assessment, much attention has been directed toward development of a rating system that yields perfect interrater reliability between pairs of raters (i.e., 1.0), does not miss projects "below minimum competence" (sensitivity), and passes projects "above minimum competence" (specificity).[32] To be rated "pass," two raters must independently agree that a project meets eight criteria. If the raters disagree (i.e., interrater reliability < 1.0), then the project is referred to a second pair of raters, who must also demonstrate an interrater reliability of 1.0 for the project to pass. If the second pair of raters disagrees, projects are sent back to the residents with recommendations for improvement. When 24 new raters examined previously judged projects, the results indicated that the rating system was sufficiently sensitive and specific.[32] Because raters would have to read entire projects to grade them, this could be considered a resource-intensive assessment technique especially for programs with relatively large numbers of residents, and thus preclude projects as a feasible approach to PBLI assessment.

Medical Record Review

Review of medical records is not a new approach to obtaining information about practice behaviors or patient outcomes. Many health care centers now do this as part of a quality assurance process,[34–36] and review of regional practices and outcomes for subgroups of practitioners has also begun.[37] The use of medical record review to ascertain PBLI, however, may be distinguished from these activities in two ways. In PBLI, individual physicians who wish to improve or change their clinical behaviors or their patient outcomes initiate the review. In addition, the review is conducted in an educational context to simultaneously facilitate continuous education of physicians and improve patient care and outcomes.

Documentation of surgical activities and patient outcomes in electronic medical records, and summarization of these data, comprises the assessment technique that has been used by a group of general surgeons in Britain to improve surgical practice.[37] Known as the Lothian Surgical Audit (LSA), the surgeons have convened weekly and annually to discuss practice activities, learn about approaches to improving practice, and develop standardized codes for surgical activities. Data have demonstrated decreased mortality rates for vascular and colorectal cancer surgeries and increased use of more acceptable approaches to treating anal fissure, breast cancer, and peptic ulcers.

This example indicates that medical record review may be used to assess the PBLI step of practice improvement. Because the surgeons developed a mechanism for documenting surgical activities and outcomes, and had a structured approach for summarizing and reviewing data, they also engaged in practice analysis. Using electronic medical records demonstrated selected information technology skills and the educational meetings about best practices implied skills in the appraisal and application of research literature.

Although not explicitly discussed, predictive validity may be inferred from the positive outcomes of the LSA. Similarly, the reliability of data yielded from the LSA was not discussed, however, the process includes characteristics known to enhance reliability. The use of standardized codes increases the reliability of data entry. Electronic data collection eliminates chart abstraction protocols, thus omitting a source of error in medical record review. In addition, ongoing data collection

and the large volume of cases per procedure and diagnosis provide more stable estimates of physician performance. With regard to feasibility, medical record review may be considered a reasonable approach to assessing PBLI because it can be built on an existing documentation system. In the previous example, feasibility was enhanced by using electronic medical records and by the systematic and structured approach used to identify and implement improvements.

Performance Ratings

Performance ratings have long been employed to assess patient care activities of residents and practicing physicians. The use of performance ratings to ascertain PBLI, however, is relatively new. Since 1999, the College of Physicians and Surgeons of Alberta in Canada has used information obtained via performance ratings to stimulate practice improvement (i.e., the Physician Achievement Review or PAR). Data are collected at least once every 5 years from the physician (i.e., self-evaluation), and the following people are identified by the physician: 6 physician peers, 6 consultants or referring physicians, 6 nonphysician coworkers, and 25 patients.[38] Forms are distributed by the physician, via mail or hand delivery, and are self-administered. The 40-item patient form addresses communication, professionalism, patient education, prevention, and office logistics; the form uses a 5-point scale ranging from 1 (*strongly disagree*) to 5 (*strongly agree*). The other forms have 31 items, address patient care, and use a 5-point scale ranging from 1 (*among the worst*) to 5 (*among the best*).

In the late 1980s, Ramsey and colleagues[39] (American Board of Internal Medicine or ABIM) began to explore the use of feedback about physician competence via performance rating forms voluntarily completed by physician and nonphysician coworkers. This approach was initially used to ascertain the predictive validity of ABIM certification examinations. Since then, however, the ABIM has begun to use peer, patient, and self-ratings as an elective part of recertification, to stimulate reflection and practice improvement. Using an automated system, the questionnaires are administered over the phone. The patient form has 10 items, which address communication and professionalism, and uses a 5-point rating scale from 1 (*poor*) to 5 (*excellent*). The peer- and self-completed forms have 11 and 21 items, respectively, use a 9-point scale anchored by scoring rubrics, and address patient care, communication, and professionalism. Based on feedback derived from the performance ratings, physicians are required to develop and submit a quality improvement plan (QUIP).[40]

The previous examples indicate that performance ratings may help assess two PBLI steps. Using performance ratings to gauge practice behaviors and outcomes may be considered evidence of practice monitoring. In addition, data obtained from repeated administration and collection of rating forms may be used to check the impact of changes in physician practice. Practice improvement, or lack thereof, could thus be inferred from rating form data obtained longitudinally.

With regard to validity, research on the impact of the PAR indicated that 66% of responding participants reported a change in at least one aspect of their practice.[41] These changes were initiated most frequently in the areas of patient communication and support, such as helping patients with their fears and worries. Feedback about the performance rating and QUIP components of ABIM recertification revealed that 61% of physician participants thought the module had provided them with a valuable learning experience, 82% said they would continue to seek feedback from patients and peers, and 42% reported they would change professionalism and communication strategies used with patients.[40] Because the physicians in this study were volunteers, the findings may not be representative of all board certified internists.

With regard to reliability, the PAR pilot study indicated that the internal consistency of the ratings ranged from 0.91 to 0.95 across respondent groups. To yield stable estimates of performance, PAR required 6 to 10 forms from physician peers and associates and 25 to 30 forms from patients.[38] Similarly, research on the ABIM performance ratings found that 25 patient forms and 10 peer forms were sufficient to obtain reliable estimates of physician performance.[42] The nature of relationships between assessors and those being assessed may influence performance ratings. This comprises another reliability concern. Although this source of bias was not substantial with ABIM forms,[39] analysis of PAR data indicated that physician peers who did not know the physicians well, or not at all, were significantly more likely to give favorable ratings.[38] This group of raters, however, comprised a small percentage of all raters.

Two findings suggest that the PAR approach is feasible. First, 79% to 96% of those who were sent rating forms completed them. Second, in 1999, the estimated cost per physician was $200 (Canadian).[38] The feasibility of the ABIM approach is enhanced by low cost, the small number of items per form, which take about 8 min to complete, and by an effective phone system for data collection.[40]

A Suggested Approach to Assess Residents' PBLI Abilities

Portfolios may be a suitable approach to assessing the multicomponent nature of PBLI. First, portfolios are flexible; portfolio content can be tailored to specific residents, specialties, and programs. This flexibility also allows portfolios to capture assessment information relevant to all PBLI steps, as indicated in the

following list of possible portfolio items: (a) A critical incident log of challenging clinical tasks or near misses could provide evidence of the ability to monitor practice experience, (b) clinical questions derived from clinical challenges or chart stimulated recall could provide evidence of the ability to analyze practice experiences, and (c) brief narratives that describe specific examples of applying learning to practice or information gleaned from chart review could provide evidence of the ability to apply practice improvements. With regard to ascertaining skills and knowledge needed to engage in PBLI, results of a literature search could serve as an indicator of the ability to locate research and use information technology to access online medical information. The application of population-based health could be tracked by using a structured form to document case presentations that address precepts such as risk-factor assessment and epidemiology.[43]

Alone, projects, performance ratings, or medical record reviews are not comprehensive enough to address the scope of items that can be included in a portfolio. In addition, because projects have been completed in the final year of residency, they cannot sample the kinds of learning and improvement that occur over a 3-year period. Furthermore, project completion assumes that residents have the skills required to implement and monitor practice improvements. On the other hand, PBLI portfolio content can be expanded over time so that new residents begin with simple tasks, such as generating clinical questions from near misses or critical incidents. As residents gain experience, however, more complex requirements can be added. Another advantage of portfolios is that they can be constructed on the basis of residents' practice experience: assessment information generated from what residents' actually do rather than knowledge tests alone, or information elicited from simulated scenarios. Because portfolios are designed to consist of a set of tangible items that reflect ability in selected areas, construct validity can be addressed relatively easily. From the learner's perspective, these tangible items provide evidence of accomplishments, and when gathered over time, provide learners with examples of improvement from 1 year to the next. Responsibility for collecting and maintaining portfolio content can be assigned to the learner, thus mitigating the administrative burden that arises from data collection. On the other hand, it is critical that physician mentors are involved with portfolio development and rating; they should prompt collection of meaningful portfolio items and help learners to recognize such items. A final advantage of using portfolios to assess PBLI is that data collection could be integrated into residents' usual routines. This would certainly be true in cases where potential portfolio items are already being generated for other purposes, such as difficult clinical problems for case conference presentations, literature searches for rounds, or procedure logs for certification. Other approaches to assess PBLI may not be as easily integrated into residents' routines.

A Suggested Approach to Assess Practicing Physicians' PBLI Abilities

Practicing physicians are often far removed from the educational resources offered by the academic health centers where many residents work and learn. Traditionally, practicing physicians have participated in post-residency education through CME activities, specialty society memberships, and journal subscriptions. Although recent publications have advocated practice-based learning experiences,[2,10] most CME has focused on teaching medical knowledge in a lecture format. Recent advances in assessing practicing physicians' PBLI abilities include PCDiary©,[28] PAR,[38] and the LSA.[37] Although the PCDiary© may aid reflection and plans to change practice, it does not explicitly require documentation of practice improvement. The PAR and the ABIM approaches provide information about the perspectives of others via performance ratings. They can yield reliable data and are relatively inexpensive to administer; however, the forms may need to be administered at least annually to ascertain meaningful improvement, and feedback usefulness is limited by the nature and number of rating form items. In other words, if a physician receives generally positive feedback, he or she may be reluctant to improve his or her practice in areas not addressed by the form. The LSA is a specialty-specific, peer-driven, and peer-managed initiative that includes ongoing educational meetings among involved physicians. By designing a mechanism for efficiently documenting, extracting, and summarizing patient care activities and outcomes, the LSA has developed an assessment approach likely to yield valid and reliable data. Parallel activities have begun in the United States. The Physician Consortium for Performance Improvement, for example, involves physicians in multiple specialties who thus far have developed and implemented performance measurement sets for diabetes, prenatal care, and coronary artery disease.[36,44] Other organizations are also beginning to address practice performance measures that may be used with individual physicians (e.g., ABMS, Joint Commission on Accreditation of Healthcare, National Commission for Quality Assurance). Evidence of ongoing participation in such initiatives would be a reasonable approach to assess practicing physicians' PBLI activities.

Conclusion

This article describes approaches that have been used to assess PBLI steps. Methods include portfolios,

a project, medical record review, and performance ratings. Each assessment approach has strengths and weaknesses that have to be weighed against the circumstances in which a method will be used. At the same time, additional steps can mitigate weaknesses associated with any single assessment approach. Ultimately, the challenge is one of balancing requirements for valid and reliable data with the practical limitations posed by feasibility issues.

Although a relatively new assessment approach in medical education, portfolios are recommended as an approach to assess residents' PBLI abilities. This suggestion is based on the flexibility, comprehensiveness, and potential for integration afforded by portfolios. On the other hand, active participation in practice improvement initiatives that includes data collection and review is suggested as an approach to assess PBLI activities of practicing physicians. This suggestion is based on the apparent success and functional nature of these peer-driven initiatives.[37,44]

Apart from being a component of PAR and ABIM recertification activities, the use of self-assessment rating forms as a trigger for PBLI does not appear to have been addressed in the medical education literature although it is a key component of many specialty-specific practice improvement modules now available on the Internet. Whereas the validity of self-assessment data remains debatable,[45] questionnaire items presented in concrete behavioral terms, such as those used in the aforementioned modules, may be more likely to yield valid data.[46] Given the wide use and relative speed of self-assessment, it may be worthwhile to examine the extent to which these rating forms and the learning modules that accompany them stimulate new learning and practice improvement.

The rationale and impetus for teaching and assessing PBLI are clear. The next step is to tailor and apply existing approaches to assess the PBLI abilities of residents and practicing physicians in the United States. The feasibility of these assessment approaches should continue to be studied while simultaneously examining the validity and reliability of data yielded.

References

1. Institute of Medicine. *Crossing the quality chasm: A new health system for the 21st century.* Washington, DC: National Academy Press, 2001.
2. Davis D, Thompson O'Brien MA, Freemantle N, Wolf FM, Mazmanian P, Taylor-Vaisey A. Impact of formal continuing medical education: Do conferences, workshops, rounds, and other traditional continuing education activities change physician behavior or health care outcomes? *Journal of the American Medical Association* 1999;282:867–74.
3. Epstein RM. Mindful practice. *Journal of the American Medical Association* 1999;282:833–9.
4. Headrick LA, Richardson A, Priebe GP. Continuous improvement learning for residents. *Pediatrics* 1998 101:768–73.
5. Berwick DM. Continuous improvement as an ideal in health care. *New England Journal of Medicine* 1989;320:53–6.
6. Berwick DM. Developing and testing changes in delivery of care. *Annals of Internal Medicine* 1998;128:651–6.
7. Langley GJ, Nolan KM, Nolan TW. *The foundation of improvement.* Silver Springs, MD: API Publishing, 1992.
8. Staker LV. Practice-based learning for improvement: The pursuit of clinical excellence. *Texas Medicine* 2000;96:53–60.
9. Accreditation Council for Graduate Medical Education. ACGME Outcome Project. Retrieved October 9, 2001, from www.acgme.org/Outcome
10. Fox RD, Bennett NL. Learning and change: Implications for continuing medical education. *British Medical Journal* 1998;316:466–8.
11. Handfield-Jones RS, Mann KV, Challis ME, et al. Linking assessment to learning: A new route to quality assurance in medical practice. *Medical Education* 2002;36:949–58.
12. Slotnick HB. How doctors learn: Physician's self-directed learning episodes. *Academic Medicine* 1999;74:1106–17.
13. Grol R. Beliefs and evidence in changing clinical practice. *British Medical Journal* 1997;315:418–21.
14. American Educational Research Association, American Psychological Association, National Council on Measurement in Education. *Standards for educational and psychological testing.* Washington, DC: AERA, 1999.
15. Nunally J. *Psychometric theory* (2nd ed.). New Delhi: Tata McGraw-Hill, 1981.
16. Fletcher RH, Fletcher SW, Wagner EH. *Clinical epidemiology: The essentials.* Baltimore, MD: Williams & Wilkins, 1982.
17. Jenson GM, Saylor C. Portfolios and professional development in the health professions. *Evaluation and the Health Professions* 1994;17:344–57.
18. Wenzel LS, Briggs KL, Puryear BL. Portfolio: Authentic assessment in the age of the curriculum revolution. *Journal of Nursing Education* 1998;37:208–12.
19. Challis M. AMEE Medical Education Guide No. 11 (revised): Portfolio-based learning and assessment in medical education. *Medical Teacher* 1999;21:370–86.
20. Shapiro ES. Self-monitoring procedures. In *Child behavioral assessment* (pp. 148–165). New York: Pergamon, 1984.
21. Flanagan JC. The critical incident technique. *Psychological Bulletin* 1954;51:327–58.
22. Challis M, Mathers NJ, Howe AC, Field NJ. Portfolio-based learning: Continuing medical education for general practitioners: A mid-point evaluation. *Medical Education* 1997;31:22–6.
23. Mathers NJ, Challis MC, Howe AC, Field NJ. Portfolios in continuing medical education—Effective and efficient? *Medical Education* 1999;33:521–30.
24. Kelly DR, Murray TS. The development and evaluation of a personal learning log for senior house officers. *Medical Education* 1999;33:260–6.
25. Snadden D, Thomas ML. Portfolio learning in general practice vocational training—Does it work? *Medical Education* 1998;32:401–6.
26. Challis M. AMEE Medical Education Guide No. 19: Personal learning plans. *Medical Teacher* 2000;22:225–36.
27. Fung Kee Fung M, Walker M, Fung Kee Fung K, et al. An internet-based learning portfolio in resident education: The KOALA™ multicentre programme. *Medical Education* 2000;34:474–9.
28. Campbell CM, Parboosingh JT, Gondocz ST, et al. Study of physicians' use of a software program to create a portfolio of their self-directed learning. *Academic Medicine* 1996;71:S49–51.
29. Campbell C, Parboosingh J, Gondocz T, Babitskaya G, Pham B. A study of the factors that influence physicians' commitments to change their practices using learning diaries. *Academic Medicine* 1999;74:S34–6.

30. Pitts J, Coles C, Thomas P. Educational portfolios in the assessment of general practice trainers: Reliability of assessors. *Medical Education* 1999;33:515–20.
31. Pitts J, Coles C, Thomas P, Smith F. Enhancing reliability in portfolio assessment: Discussions between assessors. *Medical Teacher* 2002;24:197–201.
32. Lough JRM, Murray TS. Audit and summative assessment: A completed audit cycle. *Medical Education* 2001;35:357–63.
33. Lough JRM, McKay J, Murray TS. Audit and summative assessment: System development and testing. *Medical Education* 1997;31:219–24.
34. Englert J, Davis KM, Koch KE. Using clinical practice analysis to improve care. *The Joint Commission Journal on Quality Improvement* 2001;27:291–301.
35. Palmer RH, Hargreaves JL. The ambulatory care medical audit demonstration project: Research design. *Medical Care* 1996;34:SS12–28.
36. Sutherland JE, Hoehns JD, O'Donnell B, Wiblin RT. Diabetes management quality improvement in a family practice residency program. *Journal of the American Board of Family Practice* 2001;14:243–51.
37. Aitken RJ, Nixon SJ, Ruckley CV. Lothian Surgical Audit: A 15-year experience of improvement in surgical practice through regional computerised audit. *Lancet* 1997;350:800–4.
38. Hall W, Violata C, Lewkonia R, et al. Assessment of physician performance in Alberta: The Physician Achievement Review. *Canadian Medical Association Journal* 1999;161:52–7.
39. Ramsey PG, Wenrich MD, Carline JD, Inui TS, Larsen EB, LoGerfo JP. Use of peer ratings to evaluate physician performance. *Journal of the American Medical Association* 1993;269:1655–60.
40. Lipner RC, Blank LL, Leas BF, Fortna GS. The value of patient and peer ratings in certification. *Academic Medicine* 2002;77:S64–6.
41. Fidler H, Lockyer JM, Toews J, Violato C. Changing physicians' practices: The effect of individual feedback. *Academic Medicine* 1999;74:702–14.
42. Ramsey PG, Carline JD, Blank LL, Wenrich MD. Feasibility of hospital-based use of peer ratings to evaluate the performance of practicing physicians. *Academic Medicine* 1996;71:364–70.
43. Raik B, Fein O, Wachspress S. Measuring the use of the population perspective on internal medicine attending rounds. *Academic Medicine* 1995;70:1047–9.
44. Kmetik KS, Williams JR, Hammons T, Rosof BM. The American Medical Association and physician performance measures: Information for improving patient care. *Texas Medicine* 2000;96:80–3.
45. Ward M, Gruppen L, Regehr G. Measuring self-assessment: Current state of the art. *Advances in Health Science Education* 2002;7:63–80.
46. Fowler FJ. *Survey research methods.* Newbury Park, CA: Sage, 1993.

Received 17 December
Final revision received 1 May 2003