

## The Evidence Base:

### A Compendium of the Data and Research Informing the Proposed Two-Part Recertification Model under Consideration by NCCPA

As an organization accredited by the National Commission on Certifying Agencies (NCCA), NCCPA is required to provide high quality examinations and to periodically review them for relevance based on a scientific analysis of the current state of PA practice. NCCA defines continuing competence as a demonstration of specified levels of knowledge, skills, or ability not only at the time of initial certification but throughout an individual's professional career. To that end, NCCPA continues to research and study what makes recertification a relevant, meaningful and reliable indicator of baseline PA qualifications.

Here NCCPA outlines the significant evidence base that informed the development of this model, including evidence gathered from the following sources:

- Feedback from certified PAs
- Data on PA practice patterns
- Published works

The report concludes with a description of data gathering efforts still underway as we compile the complete body of evidence that will inform the NCCPA Board's ultimate decision on whether to adopt, adapt or abandon this possible new approach to PA recertification.

### Feedback from Certified PAs

In August 2015, NCCPA convened a focus group of 29 participants, PAs who had no previous NCCPA service and whose demographic and practice profile was by design representative of the larger body of certified PAs. Over the course of three days, those PAs considered the recertification exam process from a variety of perspectives and then worked in small groups to develop models they deemed to be relevant and appropriate for PA recertification. The insights and preferences of the focus group provided a new body of qualitative evidence that was used to develop a survey of all PAs. That survey was issued through a link in the September issue of NCCPA News; approximately 10,000 PAs completed that survey, the results of which were instrumental in the selection of the model that was ultimately published for public comment. The following findings came from that survey.

- **71% of PAs were in favor of PANRE including both a specialty and general component**
  - The proposed model includes assessment of both core medical knowledge and specialty knowledge.

- **Less than half of PAs (42%) feel that PANRE is relevant, with ranges of 65% for primary care PAs and 25% for surgical PAs; even fewer (41%) feel that *preparing* for PANRE is relevant, with ranges of 65% for primary care PAs and 22% for surgical PAs; and only 39% of PAs feel PANRE is a meaningful experience, with the highest ratings (44%) coming from the PAs working in primary care and the lowest (34%) from those working in surgery.**
  - In the proposed model, the proctored, specialty-focused exam would be focused more narrowly on content that – for the vast majority of PAs – would be more closely related to their day-to-day practice than today’s broad-based PANRE. Preparation for that exam could also then be focused on content relevant to practice.
- **Only 36% feel the questions on today’s PANRE are at an appropriate difficulty level for their current practice.**
  - This informed the decision to replace the concept of a “general” medical knowledge exam with a “core medical knowledge” exam that focuses on the essential “walking around knowledge and skills” that all PAs should have to support their ability to successfully transfer from one specialty to another, thus preserving PA mobility.
- **92% favored an exam that provided learning opportunities with feedback for incorrect exam responses.**
  - Immediate feedback on missed items (including links to references with more information about the topic) is a key feature of the take-at-home, core medical knowledge exams in this model.
- **88% preferred an exam with opportunities for remediation instead of retesting when they don’t meet the required standard.**
  - The take-at-home, core medical knowledge exams and the specialty-focused exam include remediation through CME focused in areas of knowledge deficiency instead of retesting for PAs within a certain performance band.
- **When asked to rank various factors in terms of their overall importance, the top three factors for PAs, in rank order were: relevance to practice, patient safety, and ongoing opportunity to learn.**
  - Enhancing relevance to practice was a central concern in the conceptualization of the PA-choice approach to the proctored exam, through which PAs will choose from a menu of exams the one whose content is most relevant to their practice; patient safety concerns will be central to the effort to define what constitutes core medical knowledge; and the concept of ongoing learning is at the heart of the take-at-home core medical knowledge assessments.

## Data on PA Practice Patterns

- **PAs are trending into specialties outside of primary care.** In 1997, approximately 50% of PAs worked in primary care, defined as family/general medicine, general internal medicine, and general pediatrics (Hooker, Cawley, & Asprey, *Physician Assistants: Policy and Practice*, 2009). Based on the latest NCCPA data collected from over 90,000 certified PAs, as published in the [2015 Statistical Profile of Certified Physician Assistants](#), the percent of PAs in primary care practice has dropped to 28%. Of the approximately 13% of PAs who work in two or more clinical positions, only 21.1% of those PAs are working in a primary care specialty in that second position. Furthermore, the 2015

Practice Analysis (conducted April-June 2015, 17% response rate, n=15,771) provides data on what PAs do in clinical practice. Respondents rated knowledge and skill statements and evaluated the criticality and frequency of specialty specific knowledge in 11 specialties. The PAs completing these specialty sections generally rated the specialty-specific knowledge and skills statements highly, offering empirical support that PA practice in these specialty areas is different from that of other PAs.

- This influenced the design of the proposed recertification exam model to include an array of exams focused in 10-12 specialty areas from which PAs would be able to select the one best aligned with their practice.
- **One of the strengths of the profession has been the ability for PAs to move between practice areas to fill gaps to help meet healthcare needs.** NCCPA's Profile data of 69,685 certified PAs as of Dec. 31, 2013 shows that PAs changed specialties an average of 2.13 times throughout their career. According to a *Health Affairs* article that looked at specialty movement over the past four decades, 49% of PAs changed specialties at least once in their career. Of those, 24% changed to an unrelated specialty. Approximately 11% worked in three or more specialties throughout their career (Hooker, Cawley, & Leinweber, 2010).
  - The proposed model was developed with preservation of the generalist nature of the PA-C credential as a guiding principle; all PAs who complete the recertification process would continue to be issued the same generalist PA-C credential NCCPA has always conferred. The core medical knowledge assessment is designed to promote and ensure maintenance of the critical knowledge base PAs should maintain across medicine, regardless of what specialty area in which they practice.

## Published Works

NCCPA has reviewed relevant research literature related to the proposed PANRE model and best practices in assessment and certification. The citations provided in this document's appendix are but a sample of the significant body of research on these topics. These articles demonstrate the following converging lines of evidence and reasoning that informed the development of the recertification model under consideration:

1. Medical knowledge grows and changes at a rapid pace.
2. Performance on assessments of medical knowledge decline the further one is away from training.
3. Certainly some, but not all, CME is effective at improving practice and outcomes; the type of CME that is effective tends to involve more effort and input on the part of the participant.
4. Self-assessment of one's areas of strengths and weakness is poor without external input such as examinations, and it is worse for those lower in ability than high in ability.
5. When structured correctly, testing is a valuable learning tool that is more effective than study alone.
6. Certification and recertification examinations are associated with improved practice.
7. Feedback is critical to address misconceptions and can enhance learning.
8. An effective recertification model is a comprehensive approach involving multiple aspects of continuous learning and assessment among other things.

9. The public has high expectations of health care providers and relies on regulatory and certification agencies to put policies in place to provide assurance of the quality of preparation and maintenance of knowledge and skills.

**The annotated bibliography of these works begins on p. 5.**

### **Additional Data Gathering in Progress**

Since the publication of the proposed model, NCCPA launched a public comment period and solicited feedback on the model from PA organizations. Also, individual certified PAs were invited to participate in a survey to gather their input about various aspects of the proposed model and to gauge their preference between the current recertification exam and the proposed model.

In addition to that feedback from PAs, NCCPA is also surveying state licensing board executives and members, employers and credentialing specialists, and the public. The aim of those surveys is to understand how those stakeholder groups view the current recertification model, how they view the proposed model, and how adoption of that new model would be likely to impact how they perceive or what they require of certified PAs. NCCPA is partnering with an independent consumer advocacy organization on these surveys of external stakeholders to ensure an objective and unbiased approach to these data gathering efforts.

All of that information will weigh into the NCCPA Board's decision to adopt, adapt or abandon the changes under consideration.

### **Conclusion**

If NCCPA is to fulfill its responsibility to the public and to those we certify, the recertification model must align with PA practice and with current theory and evidence regarding best practices in psychometric and cognitive science. By analyzing the current state of PA practice, collecting and analyzing the opinions of multiple stakeholder groups, and engaging in a responsible and thorough review of the literature, we are working to identify improvements that enhance the relevance and meaningfulness of this important process without overburdening PAs in terms of cost or time.

## Appendix – Annotated Bibliography

### *Introduction*

Since late 2014, NCCPA has engaged in study and data gathering to inform the development and consideration of new approaches to the PA recertification exam. In November 2015, we published for public comment the potential new model that emerged from that year's discussions and evaluation as the one worthiest of further exploration based on: considerations of data on the current state of PA practice, expressed needs and preferences of the PA profession, NCCPA's responsibility to serve the public interest, the latest research and capabilities in assessment, and approaches made possible by today's technology.

Part of that study has involved an intensive review of literature related to testing and assessment, including studies that address the link between knowledge and clinically important practices and processes, and the link between practices and processes and patient safety and/or outcomes. This sort of study is part of the ongoing work of NCCPA. To date, in relation to the recertification exam model under consideration, we have reviewed well over 300 articles. The remaining articles will be added over time, and this document will be republished periodically as the list is expanded. The completed work will be available prior to the NCCPA Board's consideration of whether to pursue the proposed model, and this will be part of the evidence base considered in that decision-making process.

In our view, the preponderance of evidence supports the value of examinations and the changes represented in the proposed recertification exam model under consideration by the NCCPA Board of Directors; after all, this was part of the evidence base considered in its design. However, we encourage others to examine the evidence base for themselves. Often research of this nature does not represent a simple, clear picture. For example, there are many factors that affect patient safety and outcomes. Outcomes themselves can be difficult to measure, given the nature of team-based care and the way that coding and billing practices often obfuscate the role of the certified PAs in care at the patient level.

While PAs rely on medical literature to inform thinking about treatment options, they do not rely on the findings of one study for clinical decision making. Rather, they assess findings from a variety of sources and meta analyses, interpret the findings based on that evaluation of the medical literature, and assess those against patient populations and the individual patient to make decisions. The same thoughtful approach should be applied to the review of this literature as well. No single study and certainly no single statement will illuminate this issue.

### *The Format of this Literature Review*

In the pages that follow, citations are provided with NCCPA's very brief summation of the article's key message and the full-text of the work's abstract or introduction. We also tagged each article with the relevant subset of **six key content descriptors** so those looking for content that addresses particular aspects of this issue can focus their review. Those six descriptors relate to: studies conducted in health care or medical education settings, the testing effect and the value of exams, the importance or impact of feedback on exam performance, CME, self-assessment or the evaluation of areas of needed remediation, patient or public perception, and patient outcomes.

Agarwal, P. K., Karpicke, J. D., Kang, S. H., Roediger, H. L., & McDermott, K. B. (2008). Examining the testing effect with open-and closed-book tests. *Applied Cognitive Psychology, 22*(7), 861-876.

#### Key Message

Testing enhances learning, and it occurs in both open- and closed-book tests.

#### Abstract

Two experiments examined the testing effect with open-book tests, in which students view notes and textbooks while taking the test, and closed-book tests, in which students take the test without viewing notes or textbooks. Subjects studied prose passages and then restudied or took an open- or closed-book test. Taking either kind of test, with feedback, enhanced long-term retention relative to conditions in which subjects restudied material or took a test without feedback. Open-book testing led to better initial performance than closed-book testing, but this benefit did not persist and both types of testing produced equivalent retention on a delayed test. Subjects predicted they would recall more after repeated studying, even though testing enhanced long-term retention more than restudying. These experiments demonstrate that the testing effect occurs with both open- and closed-book tests, and that subjects fail to predict the effectiveness of testing relative to studying in enhancing later recall.

#### Key Content Descriptors

*Testing Effect and the Value of Examinations*

*Feedback*

*Self-Assessment or Evaluation of Areas of Needed Remediation*

**Agrawal, S., Norman, G. R., & Eva, K. W. (2012). Influences on medical students' self-regulated learning after test completion. *Medical Education*, 46(3), 326-335.**

### **Key Message**

The use of tests to enhance learning is further established by this paper. By structuring tests in such a way as to gather information about the test-takers' perceptions of accuracy and confidence in answers, it is possible to promote learning and retention of material with which they struggle. Testing may help to enhance metacognitive strategies.

### **Abstract**

**CONTEXT:** The inadequacy of self-assessment as a mechanism to guide performance improvements has placed greater emphasis on the value of testing as a pedagogic strategy. The mechanism whereby testing influences learning is incompletely understood. This study was performed to examine which aspects of a testing experience most influence self-regulated learning behaviour among medical students.

**METHODS:** Sixty-seven medical students participated in a computer-based, multiple-choice test. Initially, participants were instructed to attempt only items for which they felt confident of their response. They were then asked to indicate their best responses to deferred items. Students were then given an opportunity to review the items, with correct responses indicated. Accuracy, the attempt / defer decision and the time taken to reach this decision were recorded, along with participants' ratings of their confidence in each response and the time spent reviewing each item on completion of the test.

**RESULTS:** Students correctly answered a larger proportion of attempted items than deferred items (71% versus 40%;  $p < 0.001$ ), and indicated a higher mean confidence in responses to items they answered correctly compared with items they answered incorrectly (70 versus 46;  $p < 0.001$ ). They spent longer reviewing items they had answered incorrectly than correctly (8.3 versus 4.0 seconds;  $p < 0.001$ ), and paid particular attention to items for which the attempt / defer decision and accuracy were discordant ( $p < 0.01$ ). The amount of time required to make a decision on whether or not to answer a test question was also related to reviewing time.

**CONCLUSIONS:** Medical students showed a robust ability to accurately and consciously self-monitor their likelihood of success on multiple-choice test items. By focusing their subsequent self-regulated learning on areas in which performance and self-monitoring judgements were misaligned, participants reinforced the importance of providing learners with opportunities to discover the limits of their ability and further elucidated the mechanism through which test-enhanced learning might be derived.

### **Key Content Descriptors**

*Health Care/Medical Education Setting*

*Testing Effect and the Value of Examinations*

*Feedback*

*Self-Assessment or Evaluation of Areas of Needed Remediation*

Ahmed, K., Wang, T. T., Ashrafian, H., & Layer, G. T. (2013). The effectiveness of continuing medical education for specialist recertification. *Canadian Urological Association Journal*, 7 (7-8), 266-272.

### Key Message

This paper, in conjunction with others, points out that CME can be an effective tool for maintenance and improvement of knowledge and skills as part of recertification activities but that the guidelines for this should be set by both clinicians and health care authorities. CME is clearly a valuable part of continuous learning, but not all CME is equally useful, reliable, or valid. The more involved the participant is with the materials, the better the CME tends to be. It is also noted that hours of credit do not always align with changes in quality of performance.

### Abstract

Evolving professional, social and political pressures highlight the importance of lifelong learning for clinicians. Continuing medical education (CME) facilitates lifelong learning and is a fundamental factor in the maintenance of certification. The type of CME differs between surgical and non-surgical specialties. CME methods of teaching include lectures, workshops, conferences and simulation training. Interventions involving several modalities, instructional techniques and multiple exposures are more effective. The beneficial effects of CME can be maintained in the long term and can improve clinical outcome. However, quantitative evidence on validity, reliability, efficacy and cost-effectiveness of various methods is lacking. This is especially evident in urology. The effectiveness of CME interventions on maintenance of certification is also unknown. Currently, many specialists fulfill mandatory CME credit requirements opportunistically, therefore erroneously equating number of hours accumulated with competence. New CME interventions must emphasize actual performance and should correlate with clinical outcomes. Improved CME practice must in turn lead to continuing critical reflection, practice modification and implementation with a focus towards excellent patient care.

### Key Content Descriptors

*Health Care/Medical Education Setting*

*Feedback*

*Continuing Medical Education*

*Patient Outcomes*



**Alper, B. S. (2012, March). How much does practice-guiding medical knowledge change in one year? Medicine 2.0 Conference. Toronto, Canada: JMIR Publications Inc.**

**Key Message**

Much medical knowledge is in flux, and a great deal can change in a short time. As a responsible certifying body, NCCPA should move to more frequent assessments of changes in core medical knowledge.

**Abstract**

Background: Medical knowledge is estimated to double every two to nineteen years, with the most frequently reported estimates of doubling every seven to ten years. Sixteen percent of highly cited research is subsequently contradicted. However, it is unknown how often these changes occur in areas directly affecting clinical decision-making. We have created and successfully maintained a point-of-care clinical reference that is systematically derived and updated daily from the best available evidence. This provides the opportunity to determine in the current era how frequently the evidence and guidance informing clinical practice changes.

Objective: To measure the proportion of core practice-driving medical knowledge that changes in one year.

Methods: Two hundred DynaMed topics were selected for pilot testing of another project to evaluate regionalization of content for a country. These 200 topics were selected in collaboration with physicians from that country to represent common conditions and also uncommon conditions that are important to consider in clinical care, such as emergency management. For these topics management overview sections which synthesize the most valid and relevant evidence and guidance will be evaluated. For each topic the selected sections will be checked at a current date and compared to a date when they were revised for the aforementioned project. The range of time between assessment points for individual topics is expected to be 9 months to 2 years. For each topic we will analyze the number of lines that changed (added, deleted, or modified). For each change we will classify the change as: 1) Change due to publication of new evidence or guidance (the primary outcome for this analysis), 2) Change due to reviewer and user feedback, including reviewer feedback, 3) Change due to internal editorial continuous quality improvement activities.

Results: As of March 6, 2012 the management overviews of 24 topics have been evaluated for a period of 0.84 to 1.78 years. Eighteen topics (75%) were changed based on new evidence and eighteen (75%) were changed based on new guidelines. Twenty-two topics (92%) were changed based on new evidence or new guidelines. Among 1,328 lines in these topic overviews, 349 (26%) were modified and 166 (12.5%) were deleted; 269 new lines (20%) were added. Of the 784 changes, 393 (50%) were due to new evidence or guidelines. The annualized rate of change of overview content due to new evidence or guidelines was 20.8%. Interim results (about 100 topics) will be presented at the conference.

Conclusions: More than 20% of core information guiding clinical practice is changed within one year based on new evidence or guidelines. When completed,

**Key Content Descriptors**

*Health Care/Medical Education Setting*

this research will provide the first quantifiable representation of how frequently the core content guiding clinical practice changes. In an era of evidence-based medicine and electronic data management, using textbooks as a marker of standard practice is no longer warranted. Quantifying the frequency of change affecting core clinical practice has significant implications for the frequency of updating evidence-based clinical references, guidelines, medical education, and medical informatics applications.

**Alper, B. S., Hand, J. A., Elliott, S. G., Kinkade, S., Hauan, M. J., Onion, D. K., & Sklar, B. M. (2004). How much effort is needed to keep up with the literature relevant for primary care? *Journal of the Medical Library Association*, 92(4), 429.**

**Key Message**

Medical knowledge grows at an extremely fast rate, as indexed by the large amount of information published in relevant research journals. The proposed PANRE model offers an opportunity to provide feedback to PAs about gaps in knowledge to help determine which areas are most important to review.

**Abstract**

**Objectives:** Medicine must keep current with the research literature, and keeping current requires continuously updating the clinical knowledgebase (i.e., references that provide answers to clinical questions). The authors estimated the volume of medical literature potentially relevant to primary care published in a month and the time required for physicians trained in medical epidemiology to evaluate it for updating a clinical knowledge base.

**Methods:** We included journals listed in five primary care journal review services (ACP Journal Club, DynaMed, Evidence-Based Practice, Journal Watch, and QuickScan Reviews). Finding little overlap, we added the 2001 "Brandon/Hill Selected List of Print Books and Journals for the Small Medical Library." We counted articles (including letters, editorials, and other commentaries) published in March 2002, using bibliographic software where possible and hand counting when necessary. For journals not published in March 2002, we reviewed the nearest issue. Five primary care physicians independently evaluated fifty randomly selected articles and timed the process.

**Results:** The combined list contained 341 currently active journals with 8,265 articles. Adjusting for publication frequency, we estimate 7,287 articles are published monthly in this set of journals. Physicians trained in epidemiology would take an estimated 627.5 hours per month to evaluate these articles.

**Conclusions:** To provide practicing clinicians with the best current evidence, more comprehensive and systematic literature surveillance efforts are needed.

**Key Content**

**Descriptors**

*Health  
Care/Medical  
Education Setting  
Knowledge grows  
and changes*

**Archer, J. C. (2010). State of the science in health professional education: effective feedback. *Medical Education*, 44(1), 101-108.**

**Key Message**

There are many missed opportunities to provide feedback in health care settings. For feedback to be effective, it must be meaningful, systematic, and provided in a way that aids the recipient in changing his or her thinking or practice.

**Abstract**

**BACKGROUND** Effective feedback may be defined as feedback in which information about previous performance is used to promote positive and desirable development. This can be challenging as educators must acknowledge the psychosocial needs of the recipient while ensuring that feedback is both honest and accurate. Current feedback models remain reductionist in their approach. They are embedded in the hierarchical, diagnostic endeavours of the health professions. Even when it acknowledges the importance of two-way interactions, feedback often remains an educator-driven, one-way process.

**LESSONS FROM THE LITERATURE** An understanding of the various types of feedback and an ability to actively seek an appropriate approach may support feedback effectiveness. Facilitative rather than directive feedback enhances learning for high achievers. High achieving recipients undertaking complex tasks may benefit from delayed feedback. It is hypothesised that such learners are supported by reducing interruptions during the task. If we accept that medical students and doctors are high achievers, we can draw on some guiding principles from a complex and rarely conclusive literature. Feedback should focus on the task rather than the individual and should be specific. It should be directly linked to personal goals. Self-assessment as a means to identify personal learning requirements has no theoretical basis. Motivated recipients benefit from challenging facilitated feedback from external sources.

**A NEW MODEL:** To achieve truly effective feedback, the health professions must nurture recipient reflection-in-action. This builds on self-monitoring informed by external feedback. An integrated approach must be developed to support a feedback culture. Early training and experience such as peer feedback may over time support the required cultural change. Opportunities to provide feedback must not be missed, including those to impart potentially powerful feedback from high-stakes assessments. Feedback must be conceptualised as a supported sequential process rather than a series of unrelated events. Only this sustained approach will maximise any effect.

**Key Content Descriptors**

*Health Care/Medical Education Setting*

*Feedback*

*Self-Assessment or Evaluation of Areas of Needed Remediation*

**Baker, K. (2011). Determining resident clinical performance: getting beyond the noise. *Journal of the American Society of Anesthesiologists*, 115(4), 862-878.**

**Key Message**

The author describes a method of assessing clinical performance in residents to help adjust for positive bias in faculty ratings. As part of the evidence provided in the article, the author notes that this method is a valid way to represent clinical performance of residents in that the method shows a statistical relationship to standardized tests of medical knowledge.

**Abstract**

Reliable measures of clinical performance are needed to enhance and direct learning, determine which trainees are ready for advanced training, and identify which are in need of remediation.<sup>1,2</sup> Unfortunately, evaluations of resident clinical performance suffer from a number of limitations,<sup>3-5</sup> such as trainees not being directly observed,<sup>3</sup> faculty leniency and grade range restriction,<sup>6-8</sup> concerns about validity of what is being assessed,<sup>9-11</sup> and the finding that even highly valid tests of medical knowledge may not<sup>12,13</sup> or may only modestly<sup>14-17</sup> predict competence in patient care. There are also issues of generalizability because Objective Structured Clinical Examinations (OSCEs)<sup>18</sup> and simulation-based examinations<sup>19,20</sup> sample only a subset of the domain of interest, and performance may not generalize to different circumstances.<sup>10,21,22</sup> Furthermore, even when faculty members observe the same clinical performance, they may disagree about their observations<sup>23</sup> or what constitutes an acceptable performance<sup>24</sup> or response to a situation.<sup>25</sup> Lastly, and of considerable importance, is that physicians' scores on high-stakes OSCEs may not predict what they do in actual practice.<sup>26</sup> Thus, measures of competence (what a physician can do) may not relate to performance (what a physician actually does in everyday practice).<sup>27,28</sup>

**Key Content Descriptors**

*Health  
Care/Medical  
Education Setting  
Feedback*

**Bashook, P. G., & Parboosingh, J. (1998). Recertification and the maintenance of competence. *BMJ*, 316(7130), 545.**

**Key Message**

In the United States, recertification is taken as an index of the quality of care that a health care provider offers. As a result, competition for medical specialty boards has increased, with new organizations being established with similar names and granting similar certificates or accreditations. However, the rigor of certification can vary appreciably, which is of concern to the public in that they may not be sure of the differing quality of the competing certifications. Additionally, the authors argue for a more comprehensive assessment process that occurs on a continuous basis.

**Abstract**

Recertification should assess real performance in practice and competence to continue to learn. Recertification programmes in the United States use examinations and performance assessments as “snapshots” of competence taken every 710 years. In other countries most programmes evaluate documented participation in continuing education as evidence of continuing competence as a specialist. The proposed continuous recertification programme uses computer technology to document self-directed learning from practice and to monitor performance. Poor performers could be recognised early, given focused assistance and additional periodic examinations at testing centres, and if necessary their certificates could be rescinded.

**Key Content Descriptors**

*Health  
Care/Medical  
Education Setting  
Patient/Public  
Perception*

**Bjork, E. L., Little, J. L., & Storm, B. C. (2014). Multiple-choice testing as a desirable difficulty in the classroom. *Journal of Applied Research in Memory and Cognition*, 3(3), 165-170.**

**Key Message**

Repeated quizzing leads to better performance on material conceptually related to the content matter in the quizzes. This transfer of learning demonstrates that the use of repeated tests serves to enhance general knowledge about a subject area.

**Abstract**

We examined whether the power of tests as learning events, frequently demonstrated in the laboratory, would also occur in a large undergraduate course. Our goals were to determine: if learning of information tested on multiple-choice quizzes administered across the course would be enhanced compared to non-tested control information; and what the effects of quizzing would be for the learning of information conceptually related to the tested information but not itself tested on the quizzes. Given that retrieval practice can have positive (testing effect) and negative consequences (retrieval-induced forgetting), our concern was that the learning and later retention of non-tested conceptually related information might be impaired by the multiple-choice quizzes. Importantly, learning benefits were found for both types of information on the final exam, indicating that quizzing within a course can enhance not only the learning of specifically tested information, but the learning of non-tested conceptually related information as well.

**Key Content Descriptors**

*Testing Effect and  
the Value of  
Examinations*

Bloom, B. S. (2005). Effects of continuing medical education on improving physician clinical care and patient health: a review of systematic reviews. *International Journal of Technology Assessment in Health Care*, 21(3), 380-385.

### Key Message

Not all CME is helpful in changing health care providers' behavior for the better. Unfortunately, the less effective forms of CME are very popular and used very often.

### Abstract

**Objectives:** The objective of physician continuing medical education (CME) is to help them keep abreast of advances in patient care, to accept new more-beneficial care, and discontinue use of existing lower-benefit diagnostic and therapeutic interventions. The goal of this review was to examine effectiveness of current CME tools and techniques in changing physician clinical practices and improving patient health outcomes.

**Methods:** Results of published systematic reviews were examined to determine the spectrum from most- to least-effective CME techniques. We searched multiple databases, from 1 January 1984 to 30 October 2004, for English-language, peer-reviewed meta-analyses and other systematic reviews of CME programs that alter physician behavior and/or patient outcomes.

**Results:** Twenty-six reviews met inclusion criteria, that is, were either formal meta-analyses or other systematic reviews. Interactive techniques (audit/feedback, academic detailing/outreach, and reminders) are the most effective at simultaneously changing physician care and patient outcomes. Clinical practice guidelines and opinion leaders are less effective. Didactic presentations and distributing printed information only have little or no beneficial effect in changing physician practice.

**Conclusions:** Even though the most-effective CME techniques have been proven, use of least-effective ones predominates. Such use of ineffective CME likely reduces patient care quality and raises costs for all, the worst of both worlds.

### Key Content Descriptors

*Health Care/Medical Education Setting*

*Continuing Medical Education*

Brennan, T. A. (2005). Recertification for internists—one “Grandfather's” experience. *New England Journal of Medicine*, 353(19), 1989-1992.

**Key Message**

This is one internist's perspective of taking recertification examinations despite the fact that he is "grandfathered" into practice. He brings up several poignant aspects of recertification, such as grandfathering, self-assessment modules, standardized high-stakes examinations, exam material relevance, and patients' perceptions of health care providers.

**Abstract**

Practicing internal medicine can be isolating: lacking the clear-cut outcomes of one-time interventions like surgery, internists work from day to day essentially without signposts indicating success or failure. Without performance data, not only do I, as an internist, have little sense of where I stand, but the clinical leader of my practice group knows little about my fund of knowledge. She has some key evidence with regard to the group's younger doctors, for they are being recertified. If they fail their board examination, it probably indicates some deficiency. But most members of our practice, having become certified before 1990, are grandfathered into the old system of lifetime certification. Now, in the face of public demand for greater transparency, many are pushing for greater accountability through universal recertification — a worthy goal, as my recent experience with voluntary recertification convinced me.

**Key Content Descriptors**

*Health  
Care/Medical  
Education Setting  
Patient/Public  
Perception*

Brennan, T. A., Horwitz, R. I., Duffy, F. D., Cassel, C. K., Goode, L. D., & Lipner, R. S. (2004). The role of physician specialty board certification status in the quality movement. *Journal of the American Medical Association, 292*(9), 1038-1043.

#### Key Message

Couched in a discussion of the quality movement stemming from the publication of *To Err is Human: Building a Safer Health System*, this article makes a strong argument for the increasing focus on individual health care providers to achieve better quality in health care. Standardized examinations are useful to assess individual's levels of medical knowledge and skills because these examinations are reliable and valid; they are developed in line with industry best practice in measurement of cognitive skills, with extensive input from subject matter experts in the medical field. These examinations also correlate with other measures of competence of health care providers, such as type and amount of formal training as well as supervisor ratings of clinical skills. Also, there is a great deal of evidence that additional training and certification activities are associated with better patient outcomes and fewer disciplinary actions. Finally, the public expects that health care providers are held to high standards and are regularly assessed on their capabilities.

#### From the Paper

"The evolving requirements for certification and maintenance of certification are spurred by many leaders in the profession agreeing that physicians must do more to demonstrate to the public that they are skilled and knowledgeable. This momentum predates the IOM quality reports but is now given further impetus by the general activism surrounding quality."

"The changing scope of medical information, the increased concern of the public for the need to recredential physicians, and some evidence that knowledge and skills of practicing physicians decay over time motivated specialty boards to develop recertification programs and to limit the duration of certificates. Realizing that satisfactory performance on a single examination does not guarantee that physicians remain competent through their careers, the ABMS has taken on the challenge to insist that all member boards' maintenance of certification programs include the 6 certification competencies, organized into a 4-part framework, now referred to as "maintenance of certification."

"Certification and maintenance of certification evaluate a physician's evidence of possessing the requisite habits of practice (practice performance assessment) and robust knowledge base (cognitive examination) needed to prevent both types of errors. A physician who performs well on a certification examination and who maintains certification by routine review of the medical literature presumably has demonstrated ability to access a base of clinical knowledge and uses this same skill and knowledge when faced with a patient problem. Common sense suggests that the physician with a broad and readily manipulated knowledge base will be more likely to arrive at the correct answer to a clinical question, although no empirical studies are available on this point."

"Reasonable empirical evidence suggests that certification and maintenance of certification programs will improve quality, and more research is under way. That evidence is supported by the theory of error prevention and even by common sense assumptions about medical practice. Our polling data suggest the public is convinced that there is a connection, no doubt swayed by common sense."

#### Abstract

The Institute of Medicine's reports and discussions on quality of medical care have focused on a systems-based approach to quality improvement. Our objective is to summarize evidence and theory about the role of a physician's

#### Key Content Descriptors

Health



<p>current board certification status in quality improvement. The first body of evidence includes the validity of board certification demonstrated by the testing process, the relationship of examination scores with other measures of physician competence, and the relationship between certification status and clinical outcomes. The second body of evidence involves the adaptation of error prevention theory to medical care. Patient safety is enhanced when problem-solving uses readily accessed habits of behavior, the same behavior necessary to achieve board certification. The third body of evidence, obtained through a Gallup poll, demonstrates that certification and maintenance of certification are highly valued by the public. The majority of respondents thought it important for physicians to be reevaluated on their qualifications every few years and that physicians should do more to demonstrate ongoing competence than is currently required by the profession. We conclude that a physician's current certification status should be among the evidence-based measures used in the quality movement.</p>	<p><i>Care/Medical Education Setting</i></p> <p><i>Testing Effect and the Value of Examinations</i></p> <p><i>Continuing Medical Education</i></p> <p><i>Patient/Public Perception</i></p> <p><i>Patient Outcomes/Patient Safety</i></p>
--	--

Buscemi, D., Wang, H., Phy, M., & Nugent, K. (2013). Maintenance of certification in internal medicine: participation rates and patient outcomes. *Journal of Community Hospital Internal Medicine Perspectives, 2*(4).

#### Key Message

While dismissing the relationship between MOC and patient outcomes, this literature review points out multiple studies that show a positive association between the two (eg, 15% reduction in mortality for board-certified physicians in a sample of more than 40,000 patients). The recommendations from these authors are in line with the NCCPA's proposed changes to the recertification model.

#### Abstract

The clinical practice of internal medicine continues to evolve with the addition of new information and new technology. Most internists in practice will have erosion of their knowledge after they complete training unless life-long learning occurs. The American Board of Internal Medicine (ABIM) began to issue time-limited certification in 1990 and asserts that the Maintenance of Certification (MOC) program promotes the professional development of internists. However, the available medical literature does not provide strong support for the assumption that internists with certification or recertification have better patient outcomes. This relationship between recertification and patient outcomes needs more study. In addition, the participation in the Maintenance of Certification program by internists with lifetime certifications has been low, and recertification by leaders in internal medicine has also been relatively low. Some physicians in practice have concerns about the relevance of the program and the cost. Our review suggests that the ABIM needs to review its current Maintenance of Certification program and make changes to enhance its clinical relevance and educational value. We suggest that professional development should be based on focused reviews of the current literature, which is immediately relevant to clinical practice, and that recertification could be based on completion of modules and more frequent, less onerous testing.

#### Key Content Descriptors

*Health Care/Medical Education Setting Testing Effect and the Value of Examinations Patient Outcomes/ Patient Safety*

**Butler, A. C. (2010). Repeated testing produces superior transfer of learning relative to repeated studying. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 36(5), 1118-1133.**

**Key Message**

Repeated testing produced superior retention and transfer of knowledge on final testing compared with repeated studying. Mnemonic benefits of test-enhanced learning extend to the transfer of knowledge in a variety of contexts, possibly due to the enhanced encoding variability introduced during repeated testing, which increases the number of potential retrieval routes. This study shows through multiple experiments that the testing effect extends to transfer of learning within a content domain, indicating that it would be in line with aspects of the proposed PANRE model.

**Abstract**

The present research investigated whether test-enhanced learning can be used to promote transfer. More specifically, 4 experiments examined how repeated testing and repeated studying affected retention and transfer of facts and concepts. Subjects studied prose passages and then either repeatedly restudied or took tests on the material. One week later, they took a final test that had either the same questions (Experiment 1a), new inferential questions within the same knowledge domain (Experiments 1b and 2), or new inferential questions from different knowledge domains (Experiment 3). Repeated testing produced superior retention and transfer on the final test relative to repeated studying. This finding indicates that the mnemonic benefits of test-enhanced learning are not limited to the retention of the specific response tested during initial learning but rather extend to the transfer of knowledge in a variety of contexts.

**Key Content Descriptors**

*Testing Effect and the Value of Examinations*

Butler, A. C., Karpicke, J. D., & Roediger III, H. L. (2008). Correcting a metacognitive error: feedback increases retention of low-confidence correct responses. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 34(4), 918.

#### Key Message

Feedback for incorrect answers on an initial test improves performance on a later test. Additionally, feedback for correct answers on an initial test for which participants had a low level of confidence enhances performance on a later test. When structured appropriately, feedback can not only correct misconceptions but can also enhance memory of facts when a person is not confident in his or her knowledge. Also, feedback can be used to aid in metacognitive judgments about the relationship between perceived and actual performance.

#### Abstract

Previous studies investigating posttest feedback have generally conceptualized feedback as a method for correcting erroneous responses, giving virtually no consideration to how feedback might promote learning of correct responses. Here, the authors show that when correct responses are made with low confidence, feedback serves to correct this initial metacognitive error, enhancing retention of low confidence correct responses. In 2 experiments, subjects took an initial multiple-choice test on general knowledge facts and made a confidence judgment after each response. Feedback was provided for half of the questions, and retention was assessed by a final cued-recall test. Taking the initial test improved retention relative to not testing, and feedback further enhanced performance. Consistent with prior research, feedback improved retention by allowing subjects to correct initially erroneous responses. Of more importance, feedback also doubled the retention of correct low-confidence responses, relative to providing no feedback. The function of feedback is to correct both memory errors and metacognitive errors.

#### Key Content Descriptors

*Testing Effect and the Value of Examinations*

*Feedback*

Butler, A. C., & Roediger, H. L. (2008). Feedback enhances the positive effects and reduces the negative effects of multiple-choice testing. *Memory & Cognition*, 36(3), 604-616.

#### Key Message

This article makes a clear case for the provision of feedback when using multiple-choice questions. The proposed PANRE model includes feedback to enhance memory of core medical knowledge.

#### Abstract

Multiple-choice tests are used frequently in higher education without much consideration of the impact this form of assessment has on learning. Multiple-choice testing enhances retention of the material tested (the testing effect); however, unlike other tests, multiple-choice can also be detrimental because it exposes students to misinformation in the form of lures. The selection of lures can lead students to acquire false knowledge (Roediger & Marsh, 2005). The present research investigated whether feedback could be used to boost the positive effects and reduce the negative effects of multiple-choice testing. Subjects studied passages and then received a multiple-choice test with immediate feedback, delayed feedback, or no feedback. In comparison with the no-feedback condition, both immediate and delayed feedback increased the proportion of correct responses and reduced the proportion of intrusions (i.e., lure responses from the initial multiple-choice test) on a delayed cued recall test. Educators should provide feedback when using multiple-choice tests.

#### Key Content Descriptors

*Testing Effect and the Value of Examinations Feedback*

Cepeda, N. J., Vul, E., Rohrer, D., Wixted, J. T., & Pashler, H. (2008). Spacing effects in learning a temporal ridgeline of optimal retention. *Psychological Science, 19*(11), 1095-1102.

#### Key Message

This study implies that the spacing of testing can be adjusted to improve the efficiency of long-term retention. It is a very interesting study in that it gives specific guidance as how to best structure testing policy to best enhance long-term retention of factual information. Findings suggest the existence of large and nonmonotonic spacing effects that unfold over long periods of time. The relationship of gap (between studying events) with test delay implies that many educational practices are highly inefficient (i.e., a better arrangement of study gap and test delay can enhance retention). The ratio of gap to retention interval is highly critical. The timing of learning sessions can have powerful effects on retention accounting for study time. The optimally efficient gap between study sessions is not some absolute quantity that can be recommended but depends dramatically on retention interval. If you want to know the optimal distribution of study time, you need to decide how long you wish to remember something. For example, if a person wishes to retain information for several years, a delayed review of at least several months seems likely to produce a highly favorable return on the time investment.

#### Abstract

To achieve enduring retention, people must usually study information on multiple occasions. How does the timing of study events affect retention? Prior research has examined this issue only in a spotty fashion, usually with very short time intervals. To characterize spacing effects over significant durations, over 1350 individuals were taught a set of facts and – after a gap of up to 3.5 months – given a review on the same facts. A final test was administered at a further delay of up to 1 year. At any given retention interval, an increase in the inter-study gap at first increased, and then gradually reduced, test performance. The optimum gap value was about 20% of the test delay for delays of a few weeks, falling to about 5% when delay was one year. The interaction of gap and test delay implies that many educational practices are likely to be highly inefficient.

#### Key Content Descriptors

*Testing Effect and the Value of Examinations*

**Cervero, R. M., & Gaines, J. K. (2014). Effectiveness of continuing medical education: updated synthesis of systematic reviews. Report by Accreditation Council for Continuing Medical Education (ACCME), Chicago, IL.**

**Key Message**

The more involved and interactive the CME experience, the better it is for health care providers.

**Abstract**

We identified eight systematic reviews of CME effectiveness that were published since 2003, with the inclusion criteria: 1) primary research studies in CME were reviewed, 2) physicians' performance and/or patient health outcomes were included as outcome measures, and 3) the reports were published since 2003. Five of the eight systematic reviews asked the question: "Does CME improve physician performance and patient health outcomes?" The reviews consistently reached the same conclusion as the previous synthesis (Robertson, Umble, & Cervero, 2003) of the systematic review literature: CME has a positive impact on physician performance and patient health outcomes. Consistent with the previous synthesis, the five reviews also conclude that CME has a more reliably positive impact on physician performance than on patient health outcomes. These eight systematic reviews also asked the question: "What types of CME are effective?" The reviews buttress previous research showing that CME leads to improvement in physician performance and positive patient health outcomes if it is more interactive, uses more methods, involves multiple exposures, is longer, and is focused on outcomes that are considered important by physicians.

**Key Content Descriptors**

*Health  
Care/Medical  
Education Setting*

*Feedback*

*Continuing Medical  
Education*

*Patient Outcomes/  
Patient Safety*

Chang, J. T., Hays, R. D., Shekelle, P. G., MacLean, C. H., Solomon, D. H., Reuben, D. B., ... & Wenger, N. S. (2006). Patients' global ratings of their health care are not associated with the technical quality of their care. *Annals of Internal Medicine*, 144(9), 665-672.

#### Key Message

Although ratings of elderly patient satisfaction were associated with communication skills of physicians, the ratings were not indicative of the quality of technical care.

#### From the Text

"Patients' overall ratings of their care are not a reliable indicator of adherence to practice norms."

#### Abstract

Background: Patient global ratings of care are commonly used to assess health care. However, the extent to which these assessments of care are related to the technical quality of care received is not well understood.

Objective: To investigate the relationship between patient-reported global ratings of health care and the quality of providers' communication and technical quality of care.

Design: Observational cohort study.

Setting: 2 managed care organizations.

Patients: Vulnerable older patients identified by brief interviews of a random sample of community-dwelling adults 65 years of age or older who received care in 2 managed care organizations during a 13-month period.

Measurements: Survey questions from the second stage of the Consumer Assessment of Healthcare Providers and Systems program were used to determine patients' global rating of health care and provider communication. A set of 236 quality indicators, defined by the Assessing Care of Vulnerable Elders project, were used to measure technical quality of care given for 22 clinical conditions; 207 quality indicators were evaluated by using data from chart abstraction or patient interview.

Results: Data on the global rating item, communication scale, and technical quality of care score were available for 236 vulnerable older patients. In a multivariate logistic regression model that included patient and clinical factors, better communication was associated with higher global ratings of health care. Technical quality of care was not significantly associated with the global rating of care.

Limitations: Findings were limited to vulnerable elders who were enrolled in managed care organizations and may not be generalizable to other age groups or types of insurance coverage.

#### Key Content Descriptors

*Health Care/Medical Education Setting*

*Patient/Public Perception*

*Patient Outcomes*



<p>Conclusions: Vulnerable elders' global ratings of care should not be used as a marker of technical quality of care. Assessments of quality of care should include both patient evaluations and independent assessments of technical quality.</p>	
---	--

Choudhry, N. K., Fletcher, R. H., & Soumerai, S. B. (2005). Systematic review: the relationship between clinical experience and quality of health care. *Annals of Internal Medicine*, 142(4), 260-273.

### Key Message

Many studies show that there is a decline of medical knowledge and skills over time.

### Introduction

Quality assurance and performance evaluation have become central issues in medicine. Care is suboptimal in many different medical conditions and clinical settings (1–6). Although delivering high-quality care is important to all clinicians, this issue may be particularly relevant to certain subgroups, such as physicians with less specialized training and those who see a smaller volume of patients (7–10). Physicians who have been in practice for more years may also be less likely to deliver high-quality care (11–12). Medical advances occur frequently, and the explicit knowledge that physicians possess may easily become out of date. Therefore, although it is generally assumed that the tacit knowledge and skills accumulated by physicians during years of practice lead to superior clinical abilities (13), it is also plausible that physicians with more experience may paradoxically be less likely to provide technically appropriate care. Few existing studies have had the specific goal of evaluating the effects of experience on the quality of medical care (11). However, length of time in clinical practice has been included as part of a set of physician characteristics that might explain variations in quality or that may be confounders of the association between quality and other factors (13–18). The purpose of this paper is to assess the robustness of the relationship between clinical experience and quality of care by systematically reviewing empirical studies. Although we define experience as the number of years a physician has been in practice, physician age and time in practice are highly correlated (11, 19–20); therefore, for the purposes of this paper, we consider these variables to be interchangeable.

### Editors' Notes

Context: While “practice makes perfect” in some situations, physicians' knowledge and performance may decline with the passage of time.

Contribution: Of 62 published studies that measured physician knowledge or quality of care and described time since medical school graduation or age, more than half suggested that physician performance declined over time for all outcomes measured. Only 1 study showed improved performance for all outcomes measured.

Implications: This review should provoke careful study of the relationship of physician experience and the quality of care. It also raises concerns about the adequacy of continuing professional education in medicine.

### Key Content Descriptors

*Health Care/Medical Education Setting*

*Continuing Medical Education*

*Self-Assessment or Evaluation of Areas of Needed Remediation*

Colthart, I., Bagnall, G., Evans, A., Allbutt, H., Haig, A., Illing, J., & McKinstry, B. (2008). The effectiveness of self-assessment on the identification of learner needs, learner activity, and impact on clinical practice: BEME Guide no. 10. *Medical Teacher*, 30(2), 124-145.

### Key Message

Those who are least able in terms of medical knowledge tend to be the least able to accurately self-assess. There is no solid evidence base in the health professions' literature that establishes the effectiveness of self-assessment in identifying learner needs, influencing learning activity, and changing clinical practice. The accuracy of self-assessment in clinical training may be improved by increasing the learner's awareness of the standard to be achieved. There is some indication that practical skills in clinical training may be better self-assessed than knowledge-based activities. Self-assessment needs to be used as one tool among other sources of feedback to provide a more complete appraisal of competence in health care practice. Future research should address the role of self-assessment in the everyday practice of health care decision-making. "If self-assessment is to remain the cornerstone of continuing professional development and in determining how regulatory appraisal requirements are to be met, we need to have a greater understanding of what forms of self-assessment are useful in determining learning needs, and what impact these have on future learning activities."

### Abstract

**Background:** Health professionals are increasingly expected to identify their own learning needs through a process of ongoing self-assessment. Self-assessment is integral to many appraisal systems and has been espoused as an important aspect of personal professional behaviour by several regulatory bodies and those developing learning outcomes for clinical students. In this review we considered the evidence base on self-assessment since Gordon's comprehensive review in 1991. The overall aim of the present review was to determine whether specific methods of self-assessment lead to change in learning behaviour or clinical practice. Specific objectives sought evidence for effectiveness of self-assessment interventions to: a. improve perception of learning needs; b. promote change in learning activity; c. improve clinical practice; d. improve patient outcomes.

**Methods:** The methods for this review were developed and refined in a series of workshops with input from an expert BEME systematic reviewer, and followed BEME guidance. Databases searched included Medline, CINAHL, BNI, Embase, EBM Collection, Psychlit, HMIC, ERIC, BEI, TIMElit and RDRB. Papers addressing self-assessment in all professions in clinical practice were included, covering under- and post-graduate education, with outcomes classified using an extended version of Kirkpatrick's hierarchy. In addition we included outcome measures of accuracy of self-assessment and factors influencing it. 5,798 papers were retrieved, 194 abstracts were identified as potentially relevant and 103 papers coded independently by pairs using an electronic coding sheet adapted from the standard BEME form. This total included 12 papers identified by hand-searches, grey literature, cited references and updating. The identification of a further 12 papers during the writing-up process resulted in a total of 77 papers for final analysis.

**Results:** Although a large number of papers resulted from our original search only a small proportion of these were of sufficient academic rigour to be included in

### Key Content Descriptors

*Health Care/Medical Education Setting*

*Self-Assessment or Evaluation of Areas of Needed Remediation*

our review. The majority of these focused on judging the accuracy of self-assessment against some external standard, which raises questions about assumed reliability and validity of this 'gold standard'. No papers were found which satisfied Kirkpatrick's hierarchy above level 2, or which looked at the association between self-assessment and resulting changes in either clinical practice or patient outcomes. Thus our review was largely unable to answer the specific research questions and provide a solid evidence base for effective self-assessment. Despite this, there was some evidence that the accuracy of self-assessment can be enhanced by feedback, particularly video and verbal, and by providing explicit assessment criteria and benchmarking guidance. There was also some evidence that the least competent are also the least able to self-assess accurately. Our review recommends that these areas merit future systematic research to further our understanding of self-assessment.

Conclusion: As in other BEME reviews, the methodological issues emerging from this review indicate a need for more rigorous study designs. In addition, it highlights the need to consider the potential for combining qualitative and quantitative data to further our understanding of how self-assessment can improve learning and professional clinical practice.

Cook, D. A., Levinson, A. J., Garside, S., Dupras, D. M., Erwin, P. J., & Montori, V. M. (2010). Instructional design variations in Internet-based learning for health professions education: a systematic review and meta-analysis. *Academic Medicine, 85*(5), 909-922.

### Key Message

Internet-based learning, which is very similar in many ways to the proposed core medical knowledge assessments, emphasizes practice exercises, feedback, and repeated studying, which are linked to higher learning outcomes across many studies. The synthesized evidence suggests that interactivity, practice exercises, repetition, and feedback improve learning outcomes and that interactivity, online discussion, and audio improve satisfaction in Internet-based learning for health professionals.

### Abstract

**Purpose:** A recent systematic review (2008) described the effectiveness of Internet based learning (IBL) in health professions education. A comprehensive synthesis of research investigating how to improve IBL is needed. This systematic review sought to provide such a synthesis.

**Method:** The authors searched MEDLINE, CINAHL, EMBASE, Web of Science, Scopus, ERIC, TimeLit, and the University of Toronto Research and Development Resource Base for articles published from 1990 through November 2008. They included all studies quantifying the effect of IBL compared with another Internet-based or computer-assisted instructional intervention on practicing and student physicians, nurses, pharmacists, dentists, and other health professionals. Reviewers working independently and in duplicate abstracted information, coded study quality, and grouped studies according to inductively identified themes.

**Results:** From 2,705 articles, the authors identified 51 eligible studies, including 30 randomized trials. The pooled effect size (ES) or learning outcomes in 15 studies investigating high versus low interactivity was 0.27 (95% confidence interval, 0.08–0.46;  $P = .006$ ). Also associated with higher learning were practice exercises (ES 0.40 [0.08–0.71;  $P = .01$ ]; 10 studies), feedback (ES 0.68 [0.01–1.35;  $P = .047$ ]; 2 studies), and repetition of study material (ES 0.19 [0.09–0.30;  $P = .001$ ]; 2 studies). The ES was 0.26 (0.62 to 1.13;  $P = .57$ ) for three studies examining online discussion. Inconsistency was large ( $I^2 \geq 89\%$ ) in most analyses. Metaanalyses for other themes generally yielded imprecise results.

**Conclusions:** Interactivity, practice exercises, repetition, and feedback seem to be associated with improved learning outcomes, although inconsistency across studies tempers conclusions. Evidence for other instructional variations remains inconclusive.

### Key Content Descriptors

*Health Care/Medical Education Setting*

*Testing Effect and the Value of Examinations*

*Continuing Medical Education*

*Feedback*

Crenshaw, K., Curry, W., Salanitro, A. H., Safford, M. M., Houston, T. K., Allison, J. J., & Estrada, C. A. (2010). Is physician engagement with Web-based CME associated with patients' baseline hemoglobin A1c levels? The Rural Diabetes Online Care study. *Academic Medicine*, 85(9), 1511-1517.

### Key Message

This study provides evidence that active engagement with CME is linked to better patient outcomes. Additionally, the highest-performing health care providers are most likely to be actively engaged with CME and lower-performing health care providers are less likely to take advantage of useful CME.

### Abstract

**Purpose:** To investigate the association between physician participants' levels of engagement in a Web-based educational intervention and their patients' baseline diabetes measures.

**Method:** The authors conducted a randomized trial of online CME activities designed to improve diabetes care provided by family, general, and internal medicine physicians in rural areas of 11 southeastern states between September 2006 and July 2008. Using incidence rate ratios derived from negative binomial models, the relationship between physicians' engagement with the study Web site and baseline proportion of their patients having controlled diabetes (hemoglobin A1c  $\leq 7\%$ ) was explored.

**Results:** One hundred thirty-three participants (intervention = 64; control = 69) provided information for 1,637 patients with diabetes. In the intervention group, physicians in practices in the worst quartiles of A1c control were least engaged with the study Web site in nearly all dimensions. Total number of pages viewed decreased as quartile of A1c control worsened (137, 73, 68, 57;  $P = .007$ ); similarly, for a given 10% increase in proportion of patients with controlled A1c, participants viewed 1.13 times more pages (95% CI: 1.02–1.26,  $P = .02$ ). In the control group, engagement was neither correlated with A1c control nor different across quartiles of A1c control.

**Conclusions:** Engagement in Web-based interventions is measurable and has important implications for research and education. Because physicians of patients with the greatest need for improvement in A1c control may not use online educational resources as intensely as others, other strategies may be necessary to engage these physicians in professional development activities.

### Key Content Descriptors

*Health Care/Medical Education Setting*

*Continuing Medical Education*

*Feedback*

Custers, E. J., & ten Cate, O. T. (2011). Very long-term retention of basic science knowledge in doctors after graduation. *Medical Education*, 45(4), 422-430.

### Key Message

This report supports the literature showing that there is degradation of knowledge learned in training programs over time. Additionally, performance on a test of basic science knowledge indicates that current medical students retain roughly 40% of that knowledge and physicians who are many years out of school retain roughly 15% to 30%.

### Abstract

**CONTEXT:** Despite frequent complaints that biomedical knowledge is quickly forgotten after it has been learned, few investigations of actual long-term retention of basic science knowledge have been conducted in the medical domain.

**OBJECTIVES:** Our aim was to illuminate the long-term retention of basic science knowledge, particularly of unrehearsed knowledge.

**METHODS:** Using a cross-sectional study design, medical students and doctors in the Netherlands were tested for retention of basic science knowledge. Relationships between retention interval and proportion of correct answers on a knowledge test were investigated.

**RESULTS:** The popular notion that most of basic science knowledge is forgotten shortly after graduation is not supported by our findings. With respect to the full test scores, which reflect a composite of unrehearsed and rehearsed knowledge, performance decreased from approximately 40% correct answers for students still in medical school, to 25–30% correct answers for doctors after many years of practice. When rehearsal during the retention interval is controlled for, it appears that little knowledge is lost for 1.5–2 years after it was last used; from then on, retention is best described by a negatively accelerated (logarithmic) forgetting curve. After  $\geq 25$  years, retention levels were in the range of 15–20%.

**CONCLUSIONS:** Conclusions about the forgetting of unrehearsed knowledge in this study are in line with findings reported in other domains: it proceeds in accordance with the Ebbinghaus curve for meaningful material, except that in our findings the 'downward' part appears to start later than in most other studies. The limitations of the study are discussed and possible ramifications for medical education are proposed.

### Key Content Descriptors

*Health  
Care/Medical  
Education Setting*

*Knowledge  
Degrades Over  
Time*

Davis, D. A., Mazmanian, P. E., Fordis, M., Van Harrison, R. T. K. E., Thorpe, K. E., & Perrier, L. (2006). Accuracy of physician self-assessment compared with observed measures of competence: a systematic review. *Journal of the American Medical Association, 296*(9), 1094-1102.

#### Key Message

A thorough review of the literature on the connection between self-assessment and observed measures of competence indicates that there is reason to be concerned about the ability of health care providers to self-assess accurately. Of particular concern, it was noted that those health care providers who were least skilled were inaccurate at self-assessment and were the most confident in their abilities.

#### Abstract

Context: Core physician activities of lifelong learning, continuing medical education credit, relicensure, specialty recertification, and clinical competence are linked to the abilities of physicians to assess their own learning needs and choose educational activities that meet these needs. Objective: To determine how accurately physicians self-assess compared with external observations of their competence.

Data Sources: The electronic databases MEDLINE (1966-July 2006), EMBASE (1980-July 2006), CINAHL (1982-July 2006), PsycINFO (1967-July 2006), the Research and Development Resource Base in CME (1978-July 2006), and proprietary search engines were searched using terms related to self-directed learning, self-assessment, and self-reflection.

Study Selection: Studies were included if they compared physicians' self-rated assessments with external observations, used quantifiable and replicable measures, included a study population of at least 50% practicing physicians, residents, or similar health professionals, and were conducted in the United Kingdom, Canada, United States, Australia, or New Zealand. Studies were excluded if they were comparisons of self-reports, studies of medical students, assessed physician beliefs about patient status, described the development of self-assessment measures, or were self-assessment programs of specialty societies. Studies conducted in the context of an educational or quality improvement intervention were included only if comparative data were obtained before the intervention.

Data Extraction: Study population, content area and self-assessment domain of the study, methods used to measure the self-assessment of study participants and those used to measure their competence or performance, existence and use of statistical tests, study outcomes, and explanatory comparative data were extracted.

Data Synthesis: The search yielded 725 articles, of which 17 met all inclusion criteria. The studies included a wide range of domains, comparisons, measures, and methodological rigor. Of the 20 comparisons between self- and external assessment, 13 demonstrated little, no, or an inverse relationship and 7 demonstrated positive associations. A number of studies found the worst accuracy in self-assessment among physicians who were the least skilled and those who were the most confident. These results are consistent with those found in other professions.

Conclusions: While suboptimal in quality, the preponderance of evidence suggests that physicians have a limited ability to accurately self-assess. The processes currently used to undertake professional development and evaluate competence may need to focus more on external assessment.

#### Key Content Descriptors

Health  
Care/Medical  
Education Setting  
Self-Assessment or  
Evaluation of Areas  
of Needed  
Remediation



Day, S. C., Norcini, J. J., Webster, G. D., Viner, E. D., & Chirico, A. M. (1987, December). The effect of changes in medical knowledge on examination performance at the time of recertification. In *Research in Medical Education: Proceedings of the... Annual Conference. Conference on Research in Medical Education* (Vol. 27, pp. 139-144).

#### **Key Message**

The longer physicians are away from formal training, the worse they perform on tests of medical knowledge; this is especially true for new knowledge and knowledge that changes.

#### **Abstract**

The purpose of this paper was to determine if the decline in performance with time since completion of training on the 1980 ABIM Recertification Examination can be explained by a difference in performance on items testing different types of knowledge. Results showed that candidates further out of training performed less well on items testing new or changing knowledge, while performance on items testing stable knowledge was relatively constant across age groups.

#### **Key Content Descriptors**

*Health  
Care/Medical  
Education Setting*

*Knowledge  
Degrades Over  
Time*

Ely, J. W., Osherooff, J. A., Ebell, M. H., Bergus, G. R., Levy, B. T., Chambliss, M. L., & Evans, E. R. (1999). Analysis of questions asked by family doctors regarding patient care. *BMJ*, 319(7206), 358-361.

**Key Message**

Even when resources are available, health care providers do not always seek answers to questions they have regarding patient care.

**Abstract**

**Objectives:** To characterise the information needs of family doctors by collecting the questions they asked about patient care during consultations and to classify these in ways that would be useful to developers of knowledge bases.

**Design:** Observational study in which investigators visited doctors for two half days and collected their questions. Taxonomies were developed to characterise the clinical topic and generic type of information sought for each question.

**Setting:** Eastern Iowa.

**Participants:** Random sample of 103 family doctors.

**Main outcome measures:** Number of questions posed, pursued, and answered; topic and generic type of information sought for each question; time spent pursuing answers; information resources used. **Results:** Participants asked a total of 1101 questions. Questions about drug prescribing, obstetrics and gynaecology, and adult infectious disease were most common and comprised 36% of all questions. The taxonomy of generic questions included 69 categories; the three most common types, comprising 24% of all questions, were "What is the cause of symptom X?" "What is the dose of drug X?" and "How should I manage disease or finding X?" Answers to most questions (702, 64%) were not immediately pursued, but, of those pursued, most (318, 80%) were answered. Doctors spent an average of less than 2 minutes pursuing an answer, and they used readily available print and human resources. Only two questions led to a formal literature search.

**Conclusions:** Family doctors in this study did not pursue answers to most of their questions. Questions about patient care can be organised into a limited number of generic types, which could help guide the efforts of knowledge base developers.

**Key Content Descriptors**

*Health Care/Medical Education Setting*

Eva, K. W. (2002). The aging physician: changes in cognitive processing and their impact on medical practice. *Academic Medicine*, 77(10), S1-S6.

### Key Message

This paper is a review of age-related effects on memory. In most of the papers cited, there is a decrease in knowledge and performance as physicians age. One paper discussed in detail shows the opposite effect (Hobus et al, 1987). In this paper, family physicians were better than medical students at formulating diagnoses on the basis of short histories. Although the physicians performed better than the students, their diagnosis was correct for only approximately 12 of the 32 histories. The overall message of this paper is that aging declines do exist, especially in effortful, analytical situations in which health care providers are required to incorporate multiple sources of sometimes conflicting information. The recommendations of the author to handle the negative effects of aging are to provide external support, engage in deliberate practice, and rely on education and testing.

### Abstract

Most of the medical education research that has used age as an independent variable has been performed in the context of physician-review programs. This literature suggests that aging induces cognitive changes in the way that diagnosticians approach clinical cases. There are discrepant findings, however, in terms of whether clinical performance improves or declines with aging. In this paper we systematically examine the small amount of evidence available in medical education that highlights the issue of aging and attempt to reconcile contradictory findings by drawing on the much larger psychological literature on pre-senile aging. Finally, we identify some of the specific implications for continuing education.

### Key Content Descriptors

*Health Care/Medical Education Setting*

*Knowledge Degrades Over Time*

*Testing Effect and the Value of Examinations*

*Feedback*

*Self-Assessment or Evaluation of Areas of Needed Remediation*

*Continuing Medical Education*

**Eva, K. W., & Regehr, G. (2005). Self-assessment in the health professions: a reformulation and research agenda. *Academic Medicine*, 80(10), S46-S54.**

### **Key Message**

This article describes a variety of theoretical explanations for why self-assessment is poor. The authors summarize some of the research on self-efficacy and self-concept, cognitive and metacognitive theory, social cognition, models of expert performance and the development of expertise, and the concept of reflective practice. Each of these theoretical approaches offer insight into why self-assessment is a challenge for professionals. The authors call for a change to the way in which self-assessment is conceptualized and specifically point to the need for external sources of information to address aspects such as continuing medical education.

### **From the Text**

"...the traditional model of self-regulated continuing professional development presumes that an individual will select ongoing learning activities that fill professional gaps, but this presumes that the professional can effectively self-assess."

"If issues raised by the literatures we have explored are to be taken seriously, they cast doubt on the extent to which data published in the health professions literature on self-assessment provide evidence relevant to the ability of health professionals to function as self-regulating professionals. The flaws in the way that self-assessment has been conceptualized and operationalized in the current literature are sufficiently fundamental that scale tweaking and refinement of the criterion variables will not correct them. At the core of these flaws has been our communal presumption of the importance of personally generated summary judgments of overall performance: a concern about whether individuals are able to rate themselves relative to their peers, or to rate their own strengths and weaknesses relative to one another, or to accurately estimate the percent of items correctly answered on a test. We now believe that placing the burden of personal self-regulation on this "personally generated summary judgment" form of self-assessment is inappropriate for two reasons. First, the literature from a variety of fields would suggest that our literature's findings are, in fact, correct: people cannot effectively engage in these actions in any regular and stable way. Thus, it is time to recognize that, when trying to identify and redress gaps in learning, seeking and incorporating external evaluations will be a better model for effecting self-awareness than any form of personally generated summative assessment. Second, and to us more important, the focus on self-assessment as "summary judgments" fails to capture the context to which self-assessment is, in fact, critical to self-regulation: the context of reflection in practice. Self-assessment as a mechanism of ongoing monitoring must take precedence over self-assessment as a mechanism for identifying and redressing gaps."

### **Abstract**

Many researchers and educators have identified self-assessment as a vital aspect of professional self-regulation. This rationale has been the expressed motivation for a large number of studies of self-assessment ability in medical education, health professional education, and professions education generally.

Unfortunately, the outcome of most studies would seem to cast doubt on the capacity for self-assessment, with the majority of authors concluding that self-assessment is, in fact, quite poor. In a recent article, Ward and colleagues suggested that this conclusion must be questioned because the methodologies used to evaluate self-assessment are fraught with methodological weaknesses. However, even studies that have attempted to address the weaknesses within the

### **Key Content Descriptors**

*Health Care/Medical Education Setting*

*Feedback*

*Self-Assessment or Evaluation of Areas*

<p>methodological paradigm have produced little evidence for effective self-assessment. Thus, the health professional education community is left with a conundrum that can only be resolved by deciding either that the conclusions of the studies are wrong, or that a critical premise underlying the concept of “self-regulation” in the professions is unsupportable.</p>	<p><i>of Needed Remediation</i></p> <p><i>Continuing Medical Education</i></p>
--	--

<p><b>Eva, K. W., &amp; Regehr, G. (2007). Knowing when to look it up: a new conception of self-assessment ability. <i>Academic Medicine, 82(10)</i>, S81-S84.</b></p>	
<p><b>Key Message</b></p> <p>In general, people are not very good at self-assessment. However, tests can be structured with more objective indicators than traditional self-assessment in such a way as to aid individuals in identifying areas of concern in their knowledge base. By encouraging reflection in the moment as people are testing their knowledge, it is possible to come up with better predictors of areas of strength and weakness than self-assessment.</p> <p><b>From the Text</b></p> <p>"Three broad findings seem to dominate the literature: (1) there is little or no correlation between a group of individuals' self-assessments and externally generated assessments of those individuals, (2) all but the highest performers tend to overestimate their performance and ability, and (3) the worst offenders are in the lowest quartile of performance, with most of these individuals nonetheless believing that they are above average in performance.<sup>5</sup> Such findings place the notion of self-directed learning and self-regulation in jeopardy."</p> <p>"When we used the traditional measures of self-assessment to interpret our data, our results suggested that the overall summative, “guess your grade” self-assessments of our participants were moderate at best. However, the new behavioral measures used in this study were aimed more directly at capturing self-assessment as an ongoing monitoring process, and these results support a more optimistic outlook for self-assessment in daily practice. Participants did seem to be sensitive (whether consciously or unconsciously), in the moment, to whether they were likely to make an error. It seems that they could prioritize tasks in an order consistent with their overall accuracy on those tasks. They tended to defer answering specific questions for which their responses were likely to be incorrect. And, looking at the time taken to decide whether to answer or not, it seems that on a question-by-question basis, respondents knew what they knew (quickly deciding to answer questions on which they were very likely to be correct) knew what they did not know (quickly deciding to defer questions on which they were very likely to be incorrect), and appropriately considered more slowly whether to answer questions on which their eventual level of accuracy proved to be less certain (in the middle ranges)."</p>	
<p><b>Abstract</b></p> <p>Background: Although self-assessment is widely acknowledged as a vital skill for members of self-regulating professions, a ubiquitous finding in the research literature is that self-ratings are quite poor when compared with externally generated measures of ability. Many researchers have identified this as a serious problem for the concept of self-regulation in the professions. However, we question the sufficiency of the operational definitions of self-assessment on which the previous research is based. This study examines the validity of a new</p>	<p><b>Key Content Descriptors</b></p> <p><i>Testing Effect and the Value of Examinations</i></p> <p><i>Feedback</i></p>

<p>conceptualization of self-assessment in practice and evaluates a series of measures for capturing self-assessment ability as defined by this new conceptualization.</p> <p>Method: Using a computer-delivered free response test, the authors generated three measures intended to capture situational awareness: (1) response times to questions, (2) the ability to avoid responding to questions for which the respondent is less likely to be correct, and (3) the ability to select questions from content areas in which respondents have greater ability. In addition, the traditional measures of self-assessment (e.g., predictions of how many questions one would answer correctly) were administered.</p> <p>Results: Participants showed behavioral indications of being aware of the limits of their ability. They took longer to respond when their eventual answer was incorrect relative to when it was correct, they were able to avoid answering questions on which they were likely to be incorrect, and they selected content based domains in an appropriate order given their accuracy.</p> <p>Discussion: These results provide evidence in favor of this new framework that should reorient the way in which self-assessment “skills” are conceptualized, taught, and evaluated in medical school and beyond.</p>	<p><i>Self-Assessment or Evaluation of Areas of Needed Remediation</i></p>
---	--

Eva, K. W., & Regehr, G. (2008). "I'll never play professional football" and other fallacies of self-assessment. *Journal of Continuing Education in the Health Professions*, 28(1), 14-19.

#### Key Message

As a rule, self-assessment is poorly done by professionals, most of all by those who are least able in a given knowledge area. There are a host of psychological factors that help to explain why this is so; unfortunately, there is little or nothing one can do to overcome these factors. Ultimately, external input is needed regarding areas of concern in one's knowledge of medicine if improvement in practice is to be achieved.

#### From the Text

"Personal, unguided reflections on practice simply do not provide the information sufficient to guide performance improvements adequately. This has been shown in the context of the continuing education exercises one chooses to enact, in the electives students choose to take, and in the effort one exerts to develop specific skills like suturing techniques."

"The evidence is clear and overwhelming: self-assessment is not and will never be a generic skill that one can develop. Its 'accuracy' is at best dependent on context and, perhaps more importantly, on the individual's level of competence within that particular context."

"As a result, self-regulating professionals and educators have not yet institutionalized the need to draw on their community for external feedback rather than prioritizing self-generated personal opinions."

"From the perspective of self-directed assessment seeking, we must recognize that for maintenance of competence efforts to be in any way meaningful, external feedback is essential."

"Research addressing questions like "Does engaging in self-reflection result in improved performance" could parallel the emerging literature that reveals the pedagogical benefits of externally derived assessment strategies (eg, multiple-choice tests of knowledge)."

"...it is time to move beyond the rhetoric that self-assessment as a general, personal, unguided judgment of ability should be taught and developed as a valid basis on which to direct performance improvements because this belief, when all is said and done, is the fundamental fallacy of self-assessment."

#### Abstract

It is generally well accepted in health professional education that self-assessment is a key step in the continuing professional development cycle. While there has been increasing discussion in the community pertaining to whether or not professionals can indeed self-assess accurately, much of this discussion has been clouded by the fact that the term self-assessment has been used in an unfortunate and confusing variety of ways. In this article we will draw distinctions between self-assessment (an ability), self-directed assessment seeking and reflection (pedagogical strategies), and self-monitoring (immediate contextually relevant responses to environmental stimuli) in an attempt to clarify the rhetoric pertaining to each activity and provide some guidance regarding the implications that can be drawn from making these distinctions. We will further explore a

#### Key Content Descriptors

*Feedback*

*Self-Assessment or Evaluation of Areas of Needed Remediation*

<p>source of persistence in the community's efforts to improve self-assessment despite clear findings from a large body of research that we as humans do not (and, in fact, perhaps cannot) self-assess well by describing what we call a "they not we" phenomenon. Finally, we will use this phenomenon and the distinctions previously described to advocate for a variety of research projects aimed at shedding further light on the complicated relationship between self-assessment and other forms of self-regulating professional development activities.</p>	
<p><b>Eva, K. W., &amp; Regehr, G. (2011). Exploring the divergence between self-assessment and self-monitoring. <i>Advances in Health Sciences Education, 16(3), 311-329.</i></b></p>	
<p><b>Key Message</b> When structured appropriately, tests can elicit observable behavior (choosing to defer answering a question until later, amount of time to answer a question) that test-takers can use to improve their understanding of the limits of their own knowledge (self-monitoring) better than traditional self-assessment judgments.</p> <p><b>From the Text</b> "The fact that self-monitoring may indeed be effective should not be mistaken as evidence that overall self-assessments are accurate."</p>	
<p><b>Abstract</b> Many models of professional self-regulation call upon individual practitioners to take responsibility both for identifying the limits of their own skills and for redressing their identified limits through continuing professional development activities. Despite these expectations, a considerable literature in the domain of self-assessment has questioned the ability of the self-regulating professional to enact this process effectively. In response, authors have recently suggested that the construction of self-assessment as represented in the self-regulation literature is, itself, problematic. In this paper we report a pair of studies that examine the relationship between self-assessment (a global judgment of one's ability in a particular domain) and self-monitoring (a moment-by-moment awareness of the likelihood that one maintains the skill/knowledge to act in a particular situation). These studies reveal that, despite poor correlations between performance and self-assessments (consistent with what is typically seen in the self-assessment literature), participant performance was strongly related to several measures of self-monitoring including: the decision to answer or defer responding to a question, the amount of time required to make that decision to answer or defer, and the confidence expressed in an answer when provided. This apparent divergence between poor overall self-assessment and effective self-monitoring is considered in terms of how the findings might inform our understanding of the cognitive mechanisms yielding both self-monitoring judgments and self-assessments and how that understanding might be used to better direct education and learning efforts.</p>	<p><b>Key Content Descriptors</b></p> <p><i>Feedback</i></p> <p><i>Self-Assessment or Evaluation of Areas of Needed Remediation</i></p>



Evans, A. W., Aghabeigi, B., Leeson, R., O'Sullivan, C., & Eliahoo, J. (2002). Are we really as good as we think we are? *Annals of the Royal College of Surgeons of England*, 84(1), 54.

**Key Message**

Ratings of one's own competence in regard to surgical skills can be inaccurate even when provided with specific criteria against which to judge performance.

**Abstract**

Differences are examined in assessment and self-assessment scores, in oral and maxillofacial surgery trainees and MSc postgraduates, following the surgical removal of lower third molar teeth. This study found evidence of a surprising and worrying over-rating of their own surgical skills by many trainees and postgraduates.

**Key Content Descriptors**

*Health  
Care/Medical  
Education Setting  
Self-Assessment or  
Evaluation of Areas  
of Needed  
Remediation*

Evans, A. W., McKenna, C., & Oliver, M. (2002). Self-assessment in medical practice. *Journal of the Royal Society of Medicine*, 95(10), 511-513.

**Key Message**

Self-assessment is a skill that can be enhanced with appropriate input. Formal guidelines can benefit the ultimate goals of self-assessment, namely demonstration of competence and continued growth and development.

**Abstract**

Although assessment of performance will be central to revalidation, to clinical governance and to maintenance of quality in the National Health Service, formal appraisal by a third party will be no more than an annual event for most medical staff. To be fully effective doctors should assess their own clinical performance more frequently, but many lack training in how to do this. Self-assessment is also a valuable exercise in its own right. Benefits include increased morale and motivation as well as improvements in knowledge, communication and performance<sup>1</sup>. This paper sets out the case for devoting more time to the teaching of self-assessment skills.

**Key Content Descriptors**

*Health  
Care/Medical  
Education Setting  
Self-Assessment or  
Evaluation of Areas  
of Needed  
Remediation*

**Facts About the ABMS Consumer Survey: Lifelong Learning and Other Qualities in Choosing a Doctor.**  
[http://www.abms.org/News\\_and\\_Events/Media\\_Newsroom/pdf/ABMS\\_Survey\\_Fact\\_Sheet.pdf](http://www.abms.org/News_and_Events/Media_Newsroom/pdf/ABMS_Survey_Fact_Sheet.pdf)

**Key Message**

The results of a survey of the general population indicated that many people (60% of a representative sample) consider "testing at regular intervals to assess the doctor's medical knowledge" very important. When MOC was defined as "a process by which doctors who are Board Certified continue to participate in a continuous process of lifelong learning and self-assessment in their specialties," 95% of respondents said it was important, with 66% saying it was very important.

**Abstract**

The American Board of Medical Specialties (ABMS) commissioned a consumer survey that was conducted in December 2010 by Opinion Research Corporation, Princeton, New Jersey. More than 1,000 U.S. adults were randomly selected to participate in a telephone survey that focused on knowledge of doctors' qualifications and factors considered important in choosing a doctor, including Board Certification and participation the ABMS Program for Maintenance of Certification (ABMS MOC®).

**Key Content  
Descriptors**

*Health  
Care/Medical  
Education Setting  
Patient/Public  
Perception*

Fenton, J. J., Jerant, A. F., Bertakis, K. D., & Franks, P. (2012). The cost of satisfaction: a national study of patient satisfaction, health care utilization, expenditures, and mortality. *Archives of Internal Medicine*, 172(5), 405-411.

### Key Message

The relationship between patient satisfaction ratings and patient outcomes is not well understood. Patient perceptions are one of many factors that need to be included in any system of monitoring quality of care and patient outcomes.

### Abstract

**Background:** Patient satisfaction is a widely used health care quality metric. However, the relationship between patient satisfaction and health care utilization, expenditures, and outcomes remains ill defined.

**Methods:** We conducted a prospective cohort study of adult respondents (N = 51 946) to the 2000 through 2007 national Medical Expenditure Panel Survey, including 2 years of panel data for each patient and mortality follow-up data through December 31, 2006, for the 2000 through 2005 subsample (n = 36 428). Year 1 patient satisfaction was assessed using 5 items from the Consumer Assessment of Health Plans Survey. We estimated the adjusted associations between year 1 patient satisfaction and year 2 health care utilization (any emergency department visits and any inpatient admissions), year 2 health care expenditures (total and for prescription drugs), and mortality during a mean follow-up duration of 3.9 years.

**Results:** Adjusting for sociodemographics, insurance status, availability of a usual source of care, chronic disease burden, health status, and year 1 utilization and expenditures, respondents in the highest patient satisfaction quartile (relative to the lowest patient satisfaction quartile) had lower odds of any emergency department visit (adjusted odds ratio [aOR], 0.92; 95% CI, 0.84-1.00), higher odds of any inpatient admission (aOR, 1.12; 95% CI, 1.02-1.23), 8.8% (95% CI, 1.6%-16.6%) greater total expenditures, 9.1% (95% CI, 2.3%-16.4%) greater prescription drug expenditures, and higher mortality (adjusted hazard ratio, 1.26; 95% CI, 1.05-1.53).

**Conclusion:** In a nationally representative sample, higher patient satisfaction was associated with less emergency department use but with greater inpatient use, higher overall health care and prescription drug expenditures, and increased mortality.

### Key Content Descriptors

*Health Care/Medical Education Setting*

*Patient/Public Perception*

*Patient Outcomes*

Freed, G. L., Dunham, K. M., Clark, S. J., Davis, M. M., & Research Advisory Committee of the American Board of Pediatrics. (2010). Perspectives and preferences among the general public regarding physician selection and board certification. *Journal of Pediatrics*, 156(5), 841-845.

#### Key Message

Parents have very high expectations when it comes to certification and recertification of pediatricians, and 88% feel that it is important or very important for pediatricians to pass a written test of medical knowledge at regular intervals.

#### Abstract

**Objectives:** To characterize parental attitudes regarding board certification and other factors that influence selection of physicians to care for children.

**Study design:** A web-based survey administered in 2008 to a random sample of 3621 adults  $\geq 18$  years of age stratified by parents and non-parents. Proportion of respondents who view board certification and other measures of quality as important factors in selecting a physician to care for children.

**Results:** Survey completion rate was 62%. Almost all (95%) believe it is important or very important for doctors who care for children to be assessed on their quality of care, receive high ratings from patients (91%), and pass a written test at regular intervals (88%). Most reported that recommendations from friends or family (84%) and board certification (82%) were important or very important factors in choosing a physician for their child. Seventy-seven percent of parents stated that they would be likely to change their child's physician if he/she did not maintain board certification.

**Conclusion:** Parents report a preference for board-certified physicians and expect them to participate in Maintenance of Certification. Greater understanding of quality measures and the board certification process would empower consumers to make more informed decisions in selecting a physician for their children.

#### Key Content Descriptors

*Health Care/Medical Education Setting*

*Patient/Public Perception*

Hawkins, R. E., Lipner, R. S., Ham, H. P., Wagner, R., & Holmboe, E. S. (2013). American Board of Medical Specialties Maintenance of Certification: theory and evidence regarding the current framework. *Journal of Continuing Education in the Health Professions*, 33(S1), S7-S19.

### Key Message

This article provides both empirical data and a strong theoretical rationale for a comprehensive MOC program, part of which involves external assessment of medical knowledge.

### Abstract

The American Board of Medical Specialties Maintenance of Certification Program (ABMS MOC) is designed to provide a comprehensive approach to physician lifelong learning, self-assessment, and quality improvement (QI) through its 4-part framework and coverage of the 6 competencies previously adopted by the ABMS and the Accreditation Council for Graduate Medical Education (ACGME). In this article, the theoretical rationale and exemplary empiric data regarding the MOC program and its individual parts are reviewed. The value of each part is considered in relation to 4 criteria about the relationship of the competencies addressed within that part to (1) patient outcomes, (2) physician performance, (3) validity of the assessment or educational methods utilized, and (4) learning or improvement potential. Overall, a sound theoretical rationale and a respectable evidence base exists to support the current structure and elements of the MOC program. However, it is incumbent on the ABMS and ABMS member boards to continue to examine their programs moving forward to assure the public and the profession that they are meeting expectations, are clinically relevant, and provide value to patients and participating physicians, and to refine and improve them as ongoing research indicates.

### From the Text

"The goal of MOC Part II activities is to empower the physician to accept responsibility for their own learning strategy and provide them with access to tools to guide their learning and practice improvement activities. In MOC, the learning activities are embedded in a comprehensive assessment program that supports identification of physician learning and improvement needs. Indeed, a portion of the CME activities in Part II must be based on an external objective assessment. The requirement for externally guided self-assessment is important given concerns regarding the inaccuracy of physician self-assessment."

"Cognitive theory research suggests that physicians need both a sound content knowledge base and strong clinical skills to create an appropriate problem representation."

### Key Content Descriptors

*Health Care/Medical Education Setting*

*Testing Effect and the Value of Examinations*

*Feedback*

*Self-Assessment or Evaluation of Areas of Needed Remediation*

*Continuing Medical Education*

*Patient/Public Perception*

*Patient Outcomes*

Holmboe, E. S., Wang, Y., Meehan, T. P., Tate, J. P., Ho, S. Y., Starkey, K. S., & Lipner, R. S. (2008). Association between maintenance of certification examination scores and quality of care for Medicare beneficiaries. *Archives of Internal Medicine*, 168(13), 1396-1403.

#### Key Message

Recertification examination scores are predictive of some quality of care indicators, with higher performers on the examinations providing better care. Also, physicians who were more than 20 years out of school were more likely to score in the lowest quartile of their recertification examination.

#### Abstract

**Background:** The relationship between physicians' cognitive skill and the delivery of evidence-based processes of care is not well characterized. Therefore, we set out to determine associations between general internists' performance on the American Board of Internal Medicine maintenance of certification examination and the receipt of important processes of care by Medicare patients.

**Methods:** Physicians were grouped into quartiles based on their performance on the American Board of Internal Medicine examination. Hierarchical generalized linear models examined associations between examination scores and the receipt of processes of care by Medicare patients. The main outcome measures were the associations between diabetes care, using a composite measure of hemoglobin A1c, and lipid testing and retinal screening, mammography, and lipid testing in patients with cardiovascular disease and the physician's performance on the American Board of Internal Medicine examination, adjusted for the number of Medicare patients with diabetes and cardiovascular disease in a physician's practice panel; frequency of visits; patient comorbidity, age, and ethnicity; and physician training history and type of practice.

**Results:** Physicians scoring in the top quartile were more likely to perform processes of care for diabetes (composite measure odds ratio [OR], 1.17; 95% confidence interval [CI], 1.07-1.27) and mammography screening (OR, 1.14; 95% CI, 1.08-1.21) than physicians in the lowest physician quartile, even after adjustment for multiple factors. There was no significant difference among the groups in lipid testing of patients with cardiovascular disease (OR, 1.00; 95% CI, 0.91-1.10).

**Conclusion:** Our findings suggest that physician cognitive skills, as measured by a maintenance of certification examination, are associated with higher rates of processes of care for Medicare patients.

#### Key Content Descriptors

*Health Care/Medical Education Setting*

*Testing Effect and the Value of Examinations*

Juul, D., Vollmer, J., Shen, L., & Faulkner, L. R. (2016). Comparison of certification and recertification examinee performance on multiple-choice items in forensic psychiatry. *Journal of the American Academy of Psychiatry and the Law Online*, 44(1), 91-95.

**Key Message**

Despite the authors' claim that there is no clear evidence of age-related decline of knowledge in forensic psychiatry, more than half of the outcome measures showed a negative correlation between age and performance on a recertification examination, and three of these were statistically significant.

**Abstract**

Research on the association between age and performance on tests of medical knowledge has generally shown an inverse relationship, which is of concern because of the positive association between measures of knowledge and measures of clinical performance. Because the certification and maintenance of certification (MOC) examinations in the subspecialty of forensic psychiatry draw on a common item bank, performance of the two groups of examinees on the same items could be compared. In addition, the relationship between age and test performance was analyzed. Performance on items administered to certification and MOC examinees did not differ significantly, and the mean amount of time spent on each item was similar for the two groups. Although the majority (five of eight) of the correlations between age and test score on the certification and MOC examinations were negative, only three were significant, and the amount of variance explained by age was small. In addition, examination performance for those younger than 50 was similar to those 60 and older, and diplomates recertifying for the second time outperformed those doing so for the first time. These results indicate that in this subspecialty, there is no clear evidence of an age-related decline in knowledge as assessed by multiple-choice items.

**Key Content Descriptors**

*Testing Effect and the Value of Examinations Knowledge Degrades Over Time*

Kerfoot, B. P., Lawler, E. V., Sokolovskaya, G., Gagnon, D., & Conlin, P. R. (2010). Durable improvements in prostate cancer screening from online spaced education: a randomized controlled trial. *American Journal of Preventive Medicine*, 39(5), 472-478.

#### Key Message

Testing can be a valid means of improving the practice of health care providers in terms of reducing inappropriate screening for prostate cancer. The use of only a few multiple-choice questions at a time can produce improved practice in terms of reduced unnecessary screening procedures over 108 weeks, when administered appropriately.

#### From the Text

"In the present RCT, spaced education significantly reduced clinicians' inappropriate screening for prostate cancer and aligned their clinical practice patterns more closely with CPG standards. In addition, the current study is the first to show that online education can produce demonstrable improvements in clinical practice that persist for more than 1 year after the intervention. The fact that this modest intervention (36 interactive e-mails over a 36-week period) generated such substantial results suggests that spaced education is a potent methodology for CME. With content tailored to meet specific needs, spaced education is the type of intervention that can be deployed across healthcare systems to improve the quality of patient care."

"However, overlearning (the repeated presentation of content after mastery has been achieved) significantly improves the long-term retention of learning."

#### Abstract

Context: Prostate cancer screening with prostate-specific antigen (PSA) is frequently performed, counter to clinical practice guidelines.

Background: It was hypothesized that an e-mail-based intervention termed "spaced education" could reduce clinicians' inappropriate screening for prostate cancer.

Design: The study was conducted as an RCT.

Setting/participants: The study involved 95 primary care clinicians in eight Veterans Affairs medical centers from January 2007 to February 2009.

Intervention: Participants were randomized into two cohorts: spaced education clinicians received four isomorphic cycles of nine e-mails over 36 weeks (zero to two e-mails per week), whereas control clinicians received no intervention. Each e-mail presented a clinical scenario and asked whether it was appropriate to obtain a PSA test. Participants received immediate feedback after submitting their answers.

Main outcome measures: The primary outcome was the number and percentage of inappropriate PSA screening tests ordered. Inappropriate testing was defined as use of PSA for prostate cancer screening in patients aged >76 or <40 years. Appropriateness of screening was dichotomized based on patient age at time of screening. Patients with PSA testing for non-screening reasons were excluded using a validated protocol. Logistic regression with adjustment for patient

#### Key Content Descriptors

*Health Care/Medical Education Setting*

*Testing Effect and the Value of Examinations*

*Feedback*



clustering by clinician was performed. Analyses were conducted in 2009.

Results: During the intervention period (Weeks 1–36), clinicians receiving spaced education e-mails ordered significantly fewer inappropriate PSA screening tests than control clinicians (10.5% vs 14.2%,  $p=0.041$ ). Over the 72-week period following the intervention (Weeks 37–108), spaced education clinicians continued to order fewer inappropriate tests compared to controls (7.8% vs 13.1%, respectively,  $p=0.011$ ), representing a 40% relative reduction in inappropriate screening.

Conclusions: Spaced education durably improves the prostate cancer screening behaviors of clinicians and represents a promising new methodology to improve patient care across healthcare systems.

**Larsen, D. P., Butler, A. C., & Roediger III, H. L. (2008). Test-enhanced learning in medical education. *Medical Education*, 42(10), 959-966.**

**Key Message**

Test-enhanced learning seems particularly useful for health care providers in that it facilitates improved retention of factual knowledge.

**Abstract**

**CONTEXT:** In education, tests are primarily used for assessment, thus permitting teachers to assess the efficacy of their curriculum and to assign grades. However, research in cognitive psychology has shown that tests can also directly affect learning by promoting better retention of information, a phenomenon known as the testing effect.

**COGNITIVE PSYCHOLOGY RESEARCH:** Cognitive psychology laboratory studies show that repeated testing of information produces superior retention relative to repeated study, especially when testing is spaced out over time. Tests that require effortful retrieval of information, such as short-answer tests, promote better retention than tests that require recognition, such as multiple-choice tests. The mnemonic benefits of testing are further enhanced by feedback, which helps students to correct errors and confirm correct answers.

**APPLICATION TO MEDICAL EDUCATION:** Medical educational research has focused extensively on assessment issues. Such assessment research permits the conclusion that clinical expertise is founded on a broad fund of knowledge and effective memory networks that allow easy access to that knowledge. Test enhanced learning can potentially strengthen clinical knowledge that will lead to improved expertise.

**CONCLUSIONS:** Tests should be given often and spaced out in time to promote better retention of information. Questions that require effortful recall produce the greatest gains in memory. Feedback is crucial to learning from tests. Test-enhanced learning may be an effective tool for medical educators to use in promoting retention of clinical knowledge.

**Key Content Descriptors**

*Health Care/Medical Education Setting*

*Testing Effect and the Value of Examinations*

*Feedback*

Larsen, D. P., Butler, A. C., & Roediger III, H. L. (2009). Repeated testing improves long-term retention relative to repeated study: a randomised controlled trial. *Medical Education*, 43(12), 1174-1181.

### Key Message

The benefits of repeated testing in a medical education setting persisted over six months in this study.

### Abstract

**CONTEXT:** Laboratory studies in cognitive psychology with relatively brief final recall intervals suggest that repeated retrieval in the form of tests may result in better retention of information compared with repeated study.

**OBJECTIVES:** Our study evaluates if repeated testing of material taught in a real-life educational setting (a didactic conference for paediatric and emergency medicine residents) replicates these findings when measured at a more educationally relevant final recall interval of 6 months.

**METHODS:** Residents participated in an interactive teaching session on two topics: (i) status epilepticus, and (ii) myasthenia gravis. Residents were randomised to two counter-balanced groups which either took tests on status epilepticus and studied a review sheet on myasthenia gravis (SE-T/MG-S group) or took tests on myasthenia gravis and studied a review sheet on status epilepticus (MG-T/SE-S group). Testing and studying occurred immediately after teaching and then at two additional times at intervals of about 2 weeks. Residents received feedback after each test. Tests consisted of short-answer questions and the review sheets consisted of information identical to that on the answer sheets for the tests. At about 6 months residents took a final test on both topics.

**RESULTS:** Nineteen residents in the SE-T/MG-S group and 21 residents in the MG-T/SE-S group completed the study. Collapsing across groups, repeated testing produced final test scores that were an average of 13% higher than those produced by repeated study (39% versus 26%) at > 6 months after the initial teaching session ( $t[78] = 3.93$ , standard error of the difference = 0.03,  $P < 0.001$ ,  $d = 0.91$ ).

**CONCLUSIONS:** Repeated testing with feedback appears to result in significantly greater long-term retention of information taught in a didactic conference than repeated, spaced study. Testing should be considered for its potential impact on learning and not only as an assessment device.

### Key Content Descriptors

*Health Care/Medical Education Setting Testing Effect and the Value of Examinations*

Larsen, D. P., Butler, A. C., Lawson, A. L., & Roediger, H. L. (2013). The importance of seeing the patient: test-enhanced learning with standardized patients and written tests improves clinical application of knowledge. *Advances in Health Sciences Education, 18*(3), 409-425.

#### Key Message

Test-enhanced learning was demonstrated in a medical education setting and can be a powerful and efficient way to improve knowledge of medical procedures.

#### Abstract

Previous research has shown that repeated retrieval with written tests produces superior long-term retention compared to repeated study. However, the degree to which this increased retention transfers to clinical application has not been investigated. In addition, increased retention obtained through written testing has not been compared to other forms of testing, such as simulation testing with a standardized patient (SP). In our study, 41 medical students learned three clinical topics through three different learning activities: testing with SPs, testing using written tests, and studying a review sheet. Students were randomized in a counter-balanced fashion to engage in one learning activity per topic. They participated in four weekly testing/studying sessions to learn the material, engaging in the same activity for a given topic in each session. Six months after initial learning, they returned to take an SP test on each topic, followed by a written test on each topic 1 week later. On both forms of final testing, we found that learning through SP testing and written testing generally produced superior long-term retention compared to studying a review sheet. SP testing led to significantly better performance on the final SP test relative to written testing, but there was no significant difference between the two testing conditions on the final written test. Overall, our study shows that repeated retrieval practice with both SPs and written testing enhances long-term retention and transfer of knowledge to a simulated clinical application.

#### Key Content Descriptors

*Health Care/Medical Education Setting*

*Testing Effect and the Value of Examinations*

*Feedback*

**Larsen, D. P., Butler, A. C., & Roediger III, H. L. (2013). Comparative effects of test-enhanced learning and self-explanation on long-term retention. *Medical Education*, 47(7), 674-682.**

### Key Message

Testing is better at enhancing retention of medical knowledge than participants generating explanations of the material to be learned. The opportunity at retrieval practice that testing affords is a more effective mechanism than studying at enhancing retention of medical knowledge.

### Abstract

**Context:** Educators often encourage students to engage in active learning by generating explanations for the material being learned, a method called self-explanation. Studies have also demonstrated that repeated testing improves retention. However, no studies have directly compared the two learning methods.

**Methods:** Forty-seven Year 1 medical students completed the study. All students participated in a teaching session that covered four clinical topics and was followed by four weekly learning sessions. In the learning sessions, students were randomised to perform one of four learning activities for each topic: testing with self-generated explanations (TE); testing without explanations (T); studying a review sheet with self-generated explanations (SE), and studying a review sheet without explanations (S). Students repeated the same activity for each topic in all four sessions. Six months later, they took a free-recall clinical application test on all four topics.

**Results:** Repeated testing led to better long-term retention and application than repeatedly studying the material ( $p < 0.0001$ ,  $\eta^2 = 0.33$ ). Repeated generation of self-explanations also improved long-term retention and application, but the effect was smaller ( $p < 0.0001$ ,  $\eta^2 = 0.08$ ). When data were collapsed across topics, both testing conditions produced better final test performance than studying with self-explanation (TE = 40% > SE = 29% [ $p = 0.001$ ,  $d = 0.70$ ]; T = 36% > SE = 29% [ $p = 0.02$ ,  $d = 0.48$ ]). Studying with self-explanation led to better retention and application than studying without self-explanation (SE = 29% > S = 20%;  $p = 0.001$ ,  $d = 0.68$ ). Our analyses showed significant interaction by topic ( $p = 0.001$ ,  $\eta^2 = 0.06$ ), indicating some variation in the effectiveness of the interventions among topics.

**Conclusions:** Testing and generating self-explanations are both learning activities that can be used to produce superior long-term retention and application of knowledge, but testing is generally more effective than self-explanation alone.

### Key Content Descriptors

*Health Care/Medical Education Setting*

*Testing Effect and the Value of Examinations*

Leape, L. L., Berwick, D. M., & Bates, D. W. (2002). What practices will most improve safety? Evidence-based medicine meets patient safety. *Journal of the American Medical Association*, 288(4), 501-507.

#### Key Message

Quite often, process improvement and reduction of error is not easily pinned on any one change, but rather on a host of common sense changes that together provide an effect of importance. Additionally, trying to perform randomized controlled trials on every single aspect is impractical and even nonsensical. This speaks to the idea that not every aspect of certification maintenance needs to be a stand-alone mechanism for improving quality, but that in concert, assessment and continuing education can improve patient safety through complementary means.

#### From the Text

"Aviation safety, for example, was not built on evidence that certain practices reduced the frequency of crashes. Instead, it relied on the widespread implementation of hundreds of small changes in procedures, equipment, training, and organization that aggregated to establish an incredibly strong safety culture and amazingly effective practices. These changes made sense; were usually based on sound principles, technical theory, or experience; and addressed real-life problems, but few were subjected to controlled experiments. In health care, the progress in anesthesia safety is a comparable example. Everyone, including Shojania et al, agrees that the current practice of anesthesia provides an outstanding example of how a high level of safety can be achieved in health care. Anesthesia is the only system in health care that begins to approach the vaunted "six sigma" level of perfection that other industries strive for. Mortality from elective anesthesia has declined 10-fold in the past several decades as the result of a concerted effort to improve safety. This outstanding achievement is attributable not to any single practice or development of new anesthetic agents or even any type of improvement (such as technological advances) but to application of a broad array of changes in process, equipment, organization, supervision, training, and teamwork. However, no single one of these changes has ever been proven to have a clear-cut impact on mortality. Rather, anesthesia safety was achieved by applying a whole host of changes that made sense, were based on an understanding of human factors principles, and had been demonstrated to be effective in other settings. Safety, they showed, is doing a lot of little things that, in the aggregate, make a big difference. In addition, the anesthesia community has measured its progress over time, accumulating a "time series" track record whose signal is virtually incontrovertible. To say that convincing evidence of progress and effect is lacking because randomized trials of all safe anesthesia practices have not been conducted would be Luddite."

#### Abstract

The Institute of Medicine (IOM) report *To Err is Human*<sup>1</sup> converted an issue of growing professional awareness to one of substantial public concern in a manner and pace unprecedented in modern experience with matters of health care quality. The epidemiologic finding that more than 1 million injuries and nearly 100,000 deaths occur in the United States annually as a result of mistakes in medical care came from studies nearly a decade old, but it was new information for the public, and it resonated strongly. In short order, the US Congress initiated hearings and the president ordered a government-wide feasibility study, which led to a subsequent directive to governmental agencies to implement the recommendations of the IOM report. The IOM called on all parties to make improving patient safety a national priority. In response, physicians, hospitals, and health care organizations have been searching for safe practices and asking what they should do to make health care safer.

#### Key Content Descriptors

*Health Care/Medical Education Setting*

*Continuing Medical Education*

*Patient Outcomes/ Patient Safety*

Lee, D. S., Tu, J. V., Chong, A., & Alter, D. A. (2008). Patient satisfaction and its relationship with quality and outcomes of care after acute myocardial infarction. *Circulation*, 118(19), 1938-1945.

### Key Message

In this study, patient satisfaction with medical care was not associated with quality of care or survival related to myocardial infarction. Whether or not patients in this study were satisfied with the care they received had no relationship with quality indicators or survival for survivors of acute myocardial infarction.

### Abstract

**Background:** Patient satisfaction is a widely used measurement for the evaluation of medical care. We examined the extent to which quality of care received after acute myocardial infarction predicted subsequent patient satisfaction with care and whether patient satisfaction itself was associated with long-term survival after acute myocardial infarction.

**Methods and Results:** In a longitudinal cohort of acute myocardial infarction, we examined the associations of patient-reported satisfaction with care with clinical characteristics, physical and psychological function measures, quality indicators of myocardial infarction care, and outcomes. Among 1933 eligible patients (mean age 62.9±12.8 years, 70.5% men), 1866 survey respondents were analyzed. Of the study cohort, 1711 (91.7%) reported that they were satisfied with their overall care. Patients who reported satisfaction with care were older (mean age 63.1±12.7 versus 60.1±13.3 years, P=0.005), had improved physical function, and were less likely to be depressed. Better physical function, measured by the Specific Activity Scale, predicted higher satisfaction, with an OR of 1.75 (95% CI 1.17 to 2.68, P=0.008) for intermediate versus poor function and 2.96 (1.39 to 7.34, P=0.009) for high versus poor function, after adjustment for age, sex, income tertile, and ethnicity. Depression was the major predictor of dissatisfaction with overall care, with an OR of 0.44 (95% CI 0.29 to 0.67, P<0.001). Quality indicators for myocardial infarction care and clinical outcomes were not associated with patient satisfaction.

**Conclusions:** Satisfaction with care was more likely in patients who were older, in those without depression, and in those with better functional capacity, but it was not associated with the quality of myocardial infarction care or survival.

### Key Content Descriptors

*Health Care/Medical Education Setting*

*Patient/Public Perception*

*Patient Outcomes*

Lipner, R., Song, H., Biester, T., & Rhodes, R. (2011). Factors that influence general internists' and surgeons' performance on maintenance of certification exams. *Academic Medicine, 86*(1), 53-58.

#### Key Message

Increased activity in CME is associated with increased scores on MOC examinations. A finding such as this emphasizes the value of MOC examinations as indicative of the types of change that CME is designed to bring about. In other words, MOC examinations are also a valuable indicator of knowledge.

#### Abstract

**Purpose:** Good clinical judgment is important to providing high-quality patient care. Keeping current in one's field is challenged by rapid advances in health care and demanding practices. Understanding the collective factors that influence a practicing physician's clinical judgment could help medical educators design improvement programs that target specific audiences.

**Method:** Data from two medical specialty boards, the American Board of Internal Medicine and American Board of Surgery, were used. Multiple regression analyses were conducted relating first-attempt performance on the maintenance of certification (MOC) exam with physician age, amount of continuing medical education (CME) undertaken, number of physicians in the practice, medical school type, and prior exam performance. Data were based on demographics and exam scores of 18,447 general internists and 4,961 general surgeons who took the MOC exam for the first time between 2003 and 2007.

**Results:** Similar findings were obtained for general internists and surgeons. Younger physicians, those with higher scores on initial certification, physicians in group not solo practice, and U.S. medical graduates were significantly more likely to pass the MOC exam ( $P < .001$ ). Effect sizes were small except for performance on the initial certification exam. General internists with higher internal medicine program directors' ratings and more CME activities were also significantly more likely to pass ( $P < .001$ ).

**Conclusions:** Medical educators may target improvement programs for those who practice in isolation, are older, are international medical graduates, and performed poorly on their initial certification exam. Practicing without sustaining requisite clinical judgment has serious implications for patient care.

#### Key Content Descriptors

*Health Care/Medical Education Setting*

*Knowledge Degrades Over Time*

*Testing Effect and the Value of Examinations*

*Continuing Medical Education*



Logan, J. M., Thompson, A. J., & Marshak, D. W. (2011). Testing to enhance retention in human anatomy. *Anatomical Sciences Education*, 4(5), 243-248.

### Key Message

Repeated tests with feedback can be used to enhance learning and retention.

### Abstract

Recent work in cognitive psychology has shown that repeatedly testing one's knowledge is a powerful learning aid and provides substantial benefits for retention of the material. To apply this in a human anatomy course for medical students, 39 fill-in-the-blank quizzes of about 50 questions each, one for each region of the body and four about the nervous system, were developed. The quizzes were optional, and no credit was awarded. They were posted online using Blackboard, which provided feedback, and they were very popular. To determine whether the quizzes had any effect on retention, they were given in a controlled setting to 21 future medical and dental students. The weekly quizzes included questions on regional anatomy and an expanded set of questions on the nervous system. Each question about the nervous system was given three times, in a slightly different form each time. The second quiz was given approximately half an hour after the first one, and the third was given one week after the second to assess retention. The quizzes were unpopular, but students showed robust improvement on the questions about the nervous system. The scores increased by almost 9% on the second quiz, with no intervention except viewing the correct answers. The scores were 29% higher on the third quiz than on the first, and there was also a positive correlation between the grades on the quizzes and the final examination. Thus, repeated testing is an effective strategy for learning and retaining information about human anatomy.

### Key Content Descriptors

*Health  
Care/Medical  
Education Setting  
Testing Effect and  
the Value of  
Examinations  
Feedback*

McConnell, M. M., St-Onge, C., & Young, M. E. (2015). The benefits of testing for learning on later performance. *Advances in Health Sciences Education, 20*(2), 305-320.

#### Key Message

This study demonstrates the testing effect in a medical education setting and shows that context-rich questions promote better retention of factual information.

#### Abstract

Testing has been shown to enhance retention of learned information beyond simple studying, a phenomena known as test-enhanced learning (TEL). Research has shown that TEL effects are greater for tests that require the production of responses [e.g., short-answer questions (SAQs)] relative to tests that require the recognition of correct answers [e.g., multiple-choice questions (MCQs)]. High stakes licensure examinations have recently differentiated MCQs that require the application of clinical knowledge (context-rich MCQs) from MCQs that rely on the recognition of “facts” (context-free MCQs). The present study investigated the influence of different types of educational activities (including studying, SAQs, context-rich MCQs and context-free MCQs) on later performance on a mock licensure examination. Fourth-year medical students (n = 224) from four Quebec universities completed four educational activities: one reading-based activity and three quiz-based activities (SAQs, context-rich MCQs, and context-free MCQs). We assessed the influence of the type of educational activity on students’ subsequent performance in a mock licensure examination, which consisted of two types of context-rich MCQs: (1) verbatim replications of previous items and (2) items that tested the same learning objective but were new. Mean accuracy scores on the mock licensure exam were higher when intervening educational activities contained either context-rich MCQs (Mean z-score = 0.40) or SAQs (M = 0.39) compared to context-free MCQs (M = -0.38) or study only items (M = -0.42; all  $p < 0.001$ ). Higher mean scores were only present for verbatim items ( $p < 0.001$ ). The benefit of testing was observed when intervening educational activities required either the generation of a response (SAQs) or the application of knowledge (context-rich MCQs); however, this effect was only observed for verbatim test items. These data provide evidence that context-rich MCQs and SAQs enhance learning through testing compared to context-free MCQs or studying alone. The extent to which these findings generalize beyond verbatim questions remains to be seen.

#### Key Content Descriptors

*Health Care/Medical Education Setting*

*Testing Effect and the Value of Examinations*

*Feedback*

McDaniel, M. A., Roediger, H. L., & McDermott, K. B. (2007). Generalizing test-enhanced learning from the laboratory to the classroom. *Psychonomic Bulletin & Review*, 14(2), 200-206.

### Key Message

Testing effects are robust. An initial short-answer test produces greater gains in final testing than an initial multiple-choice test. Testing is not a neutral event in terms of learning but instead improves retention and later test performance. Production tests (short answer or essay) and feedback soon after learning increase learning and retention. Frequent testing has indirect positive effects of keeping students motivated and leading them to space out periods of study.

### Abstract

Test-enhanced learning refers to the fact that taking an initial test on studied material enhances its later retention relative to simply studying the material and then taking a final test. Most research on the testing effect has been done with materials such as word lists, and the general finding has been that the benefits of testing are greater when the initial test is a recall (production) test rather than a recognition test. We briefly summarize three experiments that extend these results to educationally relevant materials, namely brief articles, lectures, and materials in a college course. All three experiments demonstrated a robust testing effect and also revealed that an initial short-answer test produced greater gains on a final test than did an initial multiple-choice test. Furthermore, one experiment revealed a positive effect of immediate feedback given with the initial test. The educational implications are that production tests (short answer or essay) and feedback soon after learning increase learning and retention. In addition, frequent testing probably has the indirect positive effects of keeping students motivated and leading them to space out periods of study.

### Key Content Descriptors

*Testing Effect and the Value of Examinations*

*Feedback*

Nkenke, E., Vairaktaris, E., Bauersachs, A., Eitner, S., Budach, A., Knipfer, C., & Stelzle, F. (2012). Spaced education activates students in a theoretical radiological science course: a pilot study. *BMC Medical Education*, 12(1), 1.

### Key Message

Spaced education, in the form of e-mailed questions every two weeks with feedback provided after answers were given by learners, had a positive impact on dental students in a radiological science course in terms of time engaged with learning, perceptions of the quality of teaching, and fulfillment of learners' needs.

### Abstract

**Background:** The present study aimed at determining if the addition of spaced education to traditional face-to-face lectures increased the time students kept busy with the learning content of a theoretical radiological science course.

**Methods:** The study comprised two groups of 21 third-year dental students. The students were randomly assigned to a "traditional group" and a "spaced education group". Both groups followed a traditional face-to-face course. The intervention in the spaced education group was performed in way that these students received e-mails with a delay of 14 days to each face-to-face lecture. These e-mails contained multiple choice questions on the learning content of the lectures. The students returned their answers to the questions also by e-mail. On return they received an additional e-mail that included the correct answers and additional explanatory material. All students of both groups documented the time they worked on the learning content of the different lectures before a multiple choice exam was held after the completion of the course. All students of both groups completed the TRIL questionnaire (Trierer Inventar zur Lehrevaluation) for the evaluation of courses at university after the completion of the course. The results for the time invested in the learning content and the results of the questionnaire for the two groups were compared using the Mann-Whitney-U test.

**Results:** The spaced education group spent significantly more time ( $216.2 \pm 123.9$  min) on keeping busy with the learning content compared to the traditional group ( $58.4 \pm 94.8$  min,  $p < .0005$ ). The spaced education group rated the didactics of the course significantly better than the traditional group ( $p = .034$ ). The students of the spaced education group also felt that their needs were fulfilled significantly better compared to the traditional group as far as communication with the teacher was concerned ( $p = .022$ ).

**Conclusions:** Adding spaced education to a face-to-face theoretical radiological science course activates students in a way that they spend significantly more time on keeping busy with the learning content.

### Key Content Descriptors

*Health Care/Medical Education Setting*

*Testing Effect and the Value of Examinations*

*Feedback*

Norcini, J. J., Webster, G. D., Grosso, L. J., Blank, L. L., & Benson Jr, J. A. (1987). Ratings of residents' clinical competence and performance on certification examination. *Academic Medicine*, 62(6), 457-462.

#### Key Message

Moderate correlations were observed between examination scores and human ratings of clinical competence, indicating that the examinations and ratings are measuring similar aspects of the performance of health care providers.

#### Abstract

The American Board of Internal Medicine (ABIM) requires directors of internal medicine residency programs to rate their residents in overall clinical competence and its essential components. In the study reported here, the authors investigated the relationships among these ratings and compared them with the residents' performance on the ABIM's certification examination in the years 1980 through 1985. The ratings of the individual components of clinical competence were correlated moderately with examination performance and very highly with each other. The individual components that were less dependent on medical knowledge (for example, interpersonal skills or humanism) had slightly smaller correlations with examination performance. The ratings of overall clinical competence were also related moderately to examination performance. Changes in the pattern of the ratings over time indicated that fewer candidates were receiving lower ratings while more were receiving higher ratings. The pass rates for each rating level were the same or lower; for example, the pass rate for candidates rated 5 was 68 percent in 1980 and 56 percent in 1985. The similar ranking of examinees by the program directors and the examination provides evidence for the validity of the examination.

#### Key Content Descriptors

*Health  
Care/Medical  
Education Setting  
Testing Effect and  
the Value of  
Examinations*

Norman, G., Neville, A., Blake, J. M., & Mueller, B. (2010). Assessment steers learning down the right road: impact of progress testing on licensing examination performance. *Medical Teacher*, 32(6), 496-499.

### Key Message

Implementation of an objective, formative progress test can have large and lasting positive consequences for performance on a national licensing examination.

### Abstract

Although it is generally accepted that assessment steers learning, this is generally viewed as an undesirable side effect. Recent evidence suggests otherwise. Experimental studies have shown that periodic formative assessments can enhance learning over equivalent time spent in study (Roediger & Karpicke 2006). However, positive effects of assessment at a curriculum level have not been demonstrated. Progress tests are a periodic formative assessment designed to enhance learning by providing objective and cumulative feedback, and by identifying a subgroup of students who require additional remediation. McMaster adopted the progress test methods in 1992–1993, as a consequence of poor performance on a national licensing examination. This article shows the positive effect of this innovation, which amounts to an immediate increase of about one-half standard deviation in examination scores, and a consistent upward trend in performance. The immediate effect of introducing objective tests was a reduction in failure rate on the licensing examination from 19% to 4.5%. Various reasons for this improvement in performance are discussed.

### Key Content Descriptors

*Health Care/Medical Education Setting*

*Testing Effect and the Value of Examinations*

*Self-Assessment or Evaluation of Areas of Needed Remediation*

Peterson, L. E., Blackburn, B., Peabody, M., & O'Neill, T. R. (2015). Family physicians' scope of practice and American Board of Family Medicine recertification examination performance. *Journal of the American Board of Family Medicine*, 28(2), 265-270.

### Key Message

The relationship between what health care providers do on a daily basis and what a recertification examination measures is related to performance. In the case of this study, those whose practice was broad in scope did better on an examination that was similarly broad in scope. This finding speaks to the nature of the PA profession (increasing specialization) and the nature of PANRE (a generalist examination).

### Abstract

**Purpose:** Previous research indicated that rural family physicians were more likely to pass the American Board of Family Medicine (ABFM) Maintenance of Certification for Family Physicians (MC-FP) examination. One possible explanation is that rural family physicians may have a broader scope of practice.

**Method:** This was a cross-sectional study of family physicians taking the ABFM MC-FP examination in 2013. Examination results were linked with the Scope of Practice for Primary Care (SP4PC) scale. Linear and logistic regression models, with and without SP4PC score, determined associations between scope of practice and examination results.

**Results:** Among 10,978 examinees, rural physicians had a higher passing rate (90.7% vs 86.8%,  $P < .05$ ) and higher SP4PC score (16.1 vs 14.3  $P < .05$ ) compared with urban physicians. Regression models without SP4PC score confirmed that urban physicians were less likely to pass (OR = 0.73; 95% CI, 0.62–0.87) and scored lower, –15.6 points, compared with rural physicians. Including SP4PC score completely attenuated the relationship between practice location and passing (OR = 0.86; 95% CI, 0.73–1.02) and decreased the relationship between score and practice location (–5.8 points). Each point increase on the SP4PC score was associated with 9% higher odds of passing (OR = 1.09; 95% CI, 1.07–1.11) and 4.9 more points.

**Conclusion:** A broader scope of practice rather than rural or urban practice location, was associated with increased likelihood of passing the MC-FP examination. If higher board scores are associated with providing higher quality of care, then maintaining a broad scope of practice may enable the delivery of higher quality primary care.

### Key Content Descriptors

*Health Care/Medical Education Setting*

*Testing Effect and the Value of Examinations*

<p><b>Puffer, J. C., Bazemore, A. W., Jaén, C. R., Xierali, I. M., Phillips, R. L., &amp; Jones, S. M. (2012). Engagement of family physicians in maintenance of certification remains high. <i>Journal of the American Board of Family Medicine</i>, 25(6), 761-762.</b></p>	
<p><b>Key Message</b> Participation in maintenance of certification is higher than predicted in family medicine specialists despite the increased demands it places on participants.</p>	
<p><b>Abstract</b> Maintenance of Certification for Family Physicians was created to enhance the quality of care delivered by family physicians but risked decreasing their engagement due to the increased burden of meeting additional requirements to remain board-certified. Participation by family physicians in Maintenance of Certification remains higher than predicted.</p>	<p><b>Key Content Descriptors</b></p> <p><i>Health Care/Medical Education Setting</i></p> <p><i>Participation in MOC</i></p>

<p><b>Puffer, J. C., Bazemore, A. W., Newton, W. P., Makaroff, L., Xierali, I. M., &amp; Green, L. A. (2011). Engagement of family physicians seven years into maintenance of certification. <i>Journal of the American Board of Family Medicine</i>, 24(5), 483-484.</b></p>	
<p><b>Key Message</b> Nearly 90% of family physicians are actively engaged in maintenance of certification activities (MOC), which represents an increase of those who engaged in mandatory recertification prior to the adoption of ABMS's new MOC guidelines.</p>	
<p><b>Abstract</b> Transitioning Family Physicians to continuous Maintenance of Certification (MOC) was intended to support the quality improvement movement nationally, but it also risked decreasing their engagement due to the increased requirements for retaining certification. However, after completing the transition of all family physicians into MOC in 2010, participation appears higher than ever.</p>	<p><b>Key Content Descriptors</b></p> <p><i>Health Care/Medical Education Setting</i></p> <p><i>Participation in MOC</i></p>



Ramsey, P. G., Carline, J. D., Inui, T. S., Larson, E. B., LoGerfo, J. P., Norcini, J. J., & Wenrich, M. D. (1991). Changes over time in the knowledge base of practicing internists. *Journal of the American Medical Association*, 266(8), 1103-1107.

#### Key Message

This study found that as years in practice increase, scores on a test of general medical knowledge decrease. Also, this study found that practice within a specialty or subspecialty area was associated with worse performance on a test of general medical knowledge.

#### Abstract

**Objective:** To determine factors affecting the knowledge base of practicing internists.

**Design:** An 82-item multiple-choice examination with questions from the 1988 American Board of Internal Medicine (ABIM) certifying examination was used to assess the knowledge base of 289 internists.

**Setting and Participants:** Participants were selected from among practicing internists in New York, New Jersey, and Pennsylvania who had received ABIM certification 5 to 15 years previously.

**Results:** A significant inverse correlation ( $r = -.30$ ) was found between examination scores and the number of years elapsed since certification. Knowledge declined sharply within 15 years of certification. In addition, procedure-oriented subspecialists (cardiologists and gastroenterologists) had lower scores than other internists in this examination of general medical knowledge. Multivariate analyses showed that independent variables that predicted test performance were initial ABIM certifying examination score, time elapsed since certification, subspecialty classification, medical school type, and residency type.

**Conclusions:** These results support the recent decision for time-limited certification of internists and raise questions related to content and standard setting for recertification examinations.

#### Key Content Descriptors

*Health Care/Medical Education Setting*

*Knowledge Degrades Over Time*

Rao, M., Clarke, A., Sanderson, C., & Hammersley, R. (2006). Patients' own assessments of quality of primary care compared with objective records based measures of technical quality of care: cross sectional study. *BMJ*, 333(7557), 19.

#### Key Message

Older patients' assessments of quality of care are related to communication skills and trust in their physician but not to the technical quality of care. Other, more objective indicators are needed to assess quality of care.

#### Abstract

**Objective:** To investigate the relation between older patients' assessments of the quality of their primary care and measures of good clinical practice on the basis of data from administrative and clinical records.

**Design:** Cross sectional population based study using the general practice assessment survey.

**Setting:** 18 general practices in the Basildon primary care trust area, south east England.

**Participants:** 3487 people aged 65 or more.

**Main Outcome Measures:** Correlations between mean practice scores on the general practice assessment survey and three evidence based measures on survey of case records (monitoring for, and control of, hypertension, and vaccination against influenza).

**Results:** 76% of people (3487/4563) responded to the general practice assessment survey. Correlations between patient assessed survey scores for technical quality and the objective records based measures of good clinical practice were 0.22 (95% confidence interval – 0.28 to 0.62) for hypertension monitored, 0.30 (– 0.19 to 0.67) for hypertension controlled, and –0.05 (–0.50 to 0.43) for influenza vaccination.

**Conclusions:** Older patients' assessments are not a sufficient basis for assessing the technical quality of their primary care. For an overall assessment both patient based and records based measures are required.

#### Key Content Descriptors

*Health Care/Medical Education Setting*

*Patient/Public Perception*

*Patient Outcomes*

**Rawson, K. A., & Dunlosky, J. (2011). Optimizing schedules of retrieval practice for durable and efficient learning: how much is enough? *Journal of Experimental Psychology: General*, 140(3), 283-302.**

**Key Message**

The relationship between initial learning situations and long-term retention is complex, but the evidence suggests that the timing and spacing of initial learning can be optimized to promote long-term retention across years.

**Abstract**

The literature on testing effects is vast but supports surprisingly few prescriptive conclusions for how to schedule practice to achieve both durable and efficient learning. Key limitations are that few studies have examined the effects of initial learning criterion or the effects of relearning, and no prior research has examined the combined effects of these 2 factors. Across 3 experiments, 533 students learned conceptual material via retrieval practice with restudy. Items were practiced until they were correctly recalled from 1 to 4 times during an initial learning session and were then practiced again to 1 correct recall in 1–5 subsequent relearning sessions (across experiments, more than 100,000 short-answer recall responses were collected and hand-scored). Durability was measured by cued recall and rate of relearning 1–4 months after practice, and efficiency was measured by total practice trials across sessions. A consistent qualitative pattern emerged: The effects of initial learning criterion and relearning were subadditive, such that the effects of initial learning criterion were strong prior to relearning but then diminished as relearning increased. Relearning had pronounced effects on long-term retention with a relatively minimal cost in terms of additional practice trials. On the basis of the overall patterns of durability and efficiency, our prescriptive conclusion for students is to practice recalling concepts to an initial criterion of 3 correct recalls and then to relearn them 3 times at widely spaced intervals.

**Key Content Descriptors**

*Testing Effect and the Value of Examinations*

*Feedback*

Roediger, H. L., & Karpicke, J. D. (2006). Test-enhanced learning taking memory tests improves long-term retention. *Psychological Science, 17*(3), 249-255.

#### Key Message

Immediate testing after reading a prose passage promoted better long-term retention than repeatedly studying the passage. This occurred even though no feedback was provided during or after testing. Clearly, the testing effect is not simply a result of students gaining re-exposure to the material during testing, because restudying allowed students to re-experience 100% of the material but produced poor long-term retention (also cf. Wheeler et al, 2003). Testing is a valuable mechanism to enhance learning.

#### Abstract

Taking a memory test not only assesses what one knows, but also enhances later retention, a phenomenon known as the testing effect. We studied this effect with educationally relevant materials and investigated whether testing facilitates learning only because tests offer an opportunity to restudy material. In two experiments, students studied prose passages and took one or three immediate free-recall tests, without feedback, or restudied the material the same number of times as the students who received tests. Students then took a final retention test 5 min, 2 days, or 1 week later. When the final test was given after 5 min, repeated studying improved recall relative to repeated testing. However, on the delayed tests, prior testing produced substantially greater retention than studying, even though repeated studying increased students' confidence in their ability to remember the material. Testing is a powerful means of improving learning, not just assessing it.

#### Key Content Descriptors

*Testing Effect and the Value of Examinations*

Swankin, D., LeBuhn, R. A., & Morrison, R. (2006). *Implementing continuing competency requirements for health care practitioners*. AARP, Public Policy Institute.

#### Key Message

Routine, periodic assessment of medical knowledge and skills is something that the public expects of health care providers as part of their demonstration of fitness to practice.

#### From the Text

“A five-step competency assessment and demonstration model is most promising. After evaluating many of the existing competence-maintaining models, the Citizen Advocacy Center recommended a five-step framework for assessing and demonstrating continuing professional competence:

1. Routine Periodic Assessment
2. Development of a Personal Improvement Plan
3. Implementation of the Improvement Plan
4. Documentation
5. Demonstration of Competence, based on steps 1 through 4 above

Steps 1 through 4 constitute *quality improvement*; step 5 is the *quality assurance* component, without which the process is incomplete. The critical first step is routine periodic assessment, the key to pinpointing knowledge deficiencies needing correction and to tailoring lifelong learning choices to the needs of individual health care professionals. Assessment also reveals whether a practitioner *applies* his or her knowledge and skills competently in clinical situations.”

#### Abstract

Consumers rely on their personal physicians to ensure that they get good care. Regrettably, abundant evidence demonstrates that such confidence often may be misplaced. The Institute of Medicine explains that quality problems occur for many reasons, including (1) the growing complexity of science and knowledge; (2) an increase in chronic conditions; (3) poorly organized health delivery systems; and (4) not adopting health information technologies that could foster quality improvement.

In addition to the need for system redesign, experts also advise that “training and ongoing licensure and certification reflect the need for lifelong learning and evaluation of competencies.”<sup>1</sup> In the current environment, responsibility for assessment and assurance of continuing competency is scattered and inconsistent and, in the minds of many, ineffective. AARP commissioned this study from the Citizen Advocacy Center, an organization that has studied clinical licensure and competence extensively, to recommend how to address regular assessment of clinicians to ensure continuing competency. Although the authors identify state licensure boards as the logical entity to shoulder this responsibility, they do acknowledge the challenges of implementing valid and reliable programs to accomplish this objective and offer numerous recommendations on how to reach the goal of state-based programs that assure the public of the ongoing competency of their clinicians and other health professionals.

Public and private purchasers have begun to recognize the importance of

#### Key Content Descriptors

*Health Care/Medical Education Setting*

*Testing Effect and the Value of Examinations*

*Self-Assessment or Evaluation of Areas of Needed Remediation*

*Continuing Medical Education*

assessing physician performance to improve quality. It is also important to recognize that several professional organizations have already begun to address ways to advance programs to ensure continuing competency. The 24-member boards of the American Board of Medical Specialties (a private, nonprofit organization whose members issue 37 general and 92 subspecialty certificates) have all agreed to issue time-limited certificates that require recertification within specified time frames and to maintain certification programs that involve continuous processes of assessing competence.<sup>2</sup> These efforts may help to accelerate progress and should certainly inform the actions and activities of state licensure boards as the boards move to strengthen and improve licensing requirements.

**Tamblyn, R., Abrahamowicz, M., Brailovsky, C., Grand'Maison, P., Lescop, J., Norcini, J., ... & Haggerty, J. (1998). Association between licensing examination scores and resource use and quality of care in primary care practice. *Journal of the American Medical Association*, 280(11), 989-996.**

#### **Key Message**

Higher scorers on licensing examinations are more likely to refer patients to specialists, prescribe drugs that treat causes of diseases and disorders rather than symptoms alone, and use mammography screening for women. Higher scorers on the licensing exam were less likely to prescribe inappropriate medications.

#### **Abstract**

**Context:** Clinical competence is a determinant of the quality of care delivered, and may be associated with use of health care resources by primary care physicians. Clinical competence is assumed to be assessed by licensing examinations, yet there is a paucity of information on whether scores achieved predict subsequent practice.

**Objective:** To determine if licensing examination scores were associated with selected aspects of quality of care and resource use in initial primary care practice.

**Design:** Prospective cohort study of recently licensed family physicians, followed up for the first 18 months of practice.

**Setting:** The Québec health care system.

**Participants:** A total of 614 family physicians who passed the licensing examination between 1991 and 1993 and entered fee-for-service practice in Québec.

**Main Outcome Measures:** All patients seen by physicians were identified by the universal health insurance board and all health services provided to these patients were retrieved for the 18 months prior to (baseline) and after (follow-up)

#### **Key Content Descriptors**

*Health Care/Medical Education Setting*

*Testing Effect and the Value of Examinations*

<p>the physicians' entry into practice. Medical service and prescription claims files were used to measure rates of resource use (specialty consultation, symptom-relief prescribing compared with disease-specific prescribing) and quality of care (inappropriate prescribing, mammography screening). Baseline data were used to adjust for differences in practice population.</p> <p>Results: Study physicians saw a total of 1116389 patients, of whom 113535 (10.2%) were elderly and 83391 (7.5%) were women aged 50 to 69 years. Physicians with higher licensing examination scores referred more of their patients for consultation (3.8/1000 patients per SD increase in score; 95% confidence interval [CI], 1.2-7.0; P=.005), prescribed to elderly patients fewer inappropriate medications (-2.7/1000 patients per SD increase in score; 95% CI, -4.8 to -0.7; P=.009) and more disease-specific medications relative to symptom-relief medications (3.9/1000 patients per SD increase in score; 95% CI, 0.3 to 7.4; P=.03), and referred more women aged 50 to 69 years (6.6/1000 patients per SD increase in score; 95% CI, 1.2-11.9; P=.02) for mammography screening. If patients of physicians with the lowest scores had experienced the same rates of consultation, prescribing, and screening as patients of physicians with the highest scores, an additional 3027 patients would have been referred, 179 fewer elderly patients would have been prescribed symptom-relief medication, 912 more elderly patients would have been prescribed disease-specific medication, 189 fewer patients would have received inappropriate medication, and 121 more women would have received mammography screening.</p> <p>Conclusions: Licensing examination scores are significant predictors of consultation, prescribing, and mammography screening rates in initial primary care practice.</p>	
<p><b>Tamblyn, R., Abrahamowicz, M., Dauphinee, W. D., Hanley, J. A., Norcini, J., Girard, N., ... &amp; Brailovsky, C. (2002). Association between licensure examination scores and practice in primary care. <i>Journal of the American Medical Association</i>, 288(23), 3019-3026.</b></p>	
<p><b>Key Message</b> This study was a follow-up to Tamblyn et al (1998) and investigated whether the link between certification examination scores persisted over four to seven years. The same results were found; higher scorers were more likely to refer patients to specialists, prescribe disease-specific drugs rather than symptom-specific drugs, and order mammography screening and were less likely to prescribe drugs with contraindications (although this was not statistically significant in this study).</p>	
<p><b>Abstract</b> Context: Standards for licensure are designed to provide assurance to the public of a physician's competence to practice. However, there has been little assessment of the relationship between examination scores and subsequent practice performance.</p> <p>Objective: To determine if there is a sustained relationship between certification examination scores and practice performance and if licensing examinations taken at the end of medical school are predictive of future practice in primary care.</p>	<p><b>Key Content Descriptors</b></p> <p><i>Health Care/Medical Education Setting</i></p> <p><i>Testing Effect and the Value of</i></p>

<p>Design, Setting, and Participants: A total of 912 family physicians, who passed the Québec family medicine certification examination (QLEX) between 1990 and 1993 and entered practice. Linked databases were used to assess physicians' practice performance for 3.4 million patients in the universal health care system in Québec, Canada. Patients were seen during the follow-up period for the first 4 years (1993 cohort of physicians) to 7 years (1990 cohort of physicians) of practice from July 1 of the certification examination to December 31, 1996.</p> <p>Main Outcome Measures: Mammography screening rate, continuity of care index, disease-specific and symptom-relief prescribing rate, contraindicated prescribing rate, and consultation rate.</p> <p>Results: Physicians achieving higher scores on both examinations had higher rates (rate increase per SD increase in score per 1000 persons per year) of mammography screening (<math>\beta</math> for QLEX, 16.8 [95% confidence interval {CI}, 8.7-24.9]; <math>\beta</math> for Medical Council of Canada Qualifying Examination [MCCQE], 17.4 [95% CI, 10.6-24.1]) and consultation (<math>\beta</math> for QLEX, 4.9 [95% CI, 2.1-7.8]; <math>\beta</math> for MCCQE, 2.9 [95% CI, 0.4-5.4]). Higher sub-scores in diagnosis were predictive of higher rates in the difference between disease-specific and symptom-relief prescribing (<math>\beta</math> for QLEX, 3.9 [95% CI, 0.9-7.0]; <math>\beta</math> for MCCQE, 3.8 [95% CI, 0.3-7.3]). Higher scores of drug knowledge were predictive of a lower rate (relative risk per SD increase in score) of contraindicated prescribing for MCCQE (relative risk, 0.88; 95% CI, 0.77-1.00). Relationships between examination scores and practice performance were sustained through the first 4 to 7 years in practice.</p> <p>Conclusion: Scores achieved on certification examinations and licensure examinations taken at the end of medical school show a sustained relationship, over 4 to 7 years, with indices of preventive care and acute and chronic disease management in primary care practice.</p>	<p><i>Examinations</i></p>
--	----------------------------



Wenghofer, E., Klass, D., Abrahamowicz, M., Dauphinee, D., Jacques, A., Smee, S., ... & Tamblyn, R. (2009). Doctor scores on national qualifying examinations predict quality of care in future practice. *Medical Education*, 43(12), 1166-1173.

### Key Message

Physicians scoring in the lowest quartile of the examination were as much as 3.4 times more likely to be assessed as providing unacceptable quality of care.

### Abstract

**OBJECTIVES:** This study aimed to determine if national licensing examinations that measure medical knowledge (QE1) and clinical skills (QE2) predict the quality of care delivered by doctors in future practice.

**METHODS:** Cohorts of doctors who took the Medical Council of Canada Qualifying Examinations Part I (QE1) and Part II (QE2) between 1993 and 1996 and subsequently entered practice in Ontario, Canada (n = 2420) were followed for their first 7–10 years in practice. The 208 of these doctors who were randomly selected for peer assessment of quality of care were studied. Main outcome measures included quality of care (acceptable / unacceptable) as assessed by doctor peer examiners using a structured chart review and interview. Multivariate logistic regression was used to determine if qualifying examination scores predicted the outcome of the peer assessments while controlling for age, sex, training and specialty, and if the addition of the QE2 scores provided additional prediction of quality of care.

**RESULTS:** Fifteen (7.2%) of the 208 doctors assessed were considered to provide unacceptable quality of care. Doctors in the bottom quartile of QE1 scores had a greater than threefold increase in the risk of an unacceptable quality-of-care assessment outcome (odds ratio [OR] 3.41, 95% confidence interval [CI] 1.14–10.22). Doctors in the bottom quartile of QE2 scores were also at higher risk of being assessed as providing unacceptable quality of care (OR 4.24, 95% CI 1.32–13.61). However, QE2 results provided no significant improvement in predicting peer assessment results over QE1 results (likelihood ratio test:  $\chi^2 = 3.21$ , P-value (1 d.f.) = 0.07).

**CONCLUSIONS:** Doctor scores on qualifying examinations are significant predictors of quality-of-care problems based on regulatory, practice-based peer assessment.

### Key Content Descriptors

*Health Care/Medical Education Setting*

*Testing Effect and the Value of Examinations*

**Williams, B. W. (2006). The prevalence and special educational requirements of dyscompetent physicians. *Journal of Continuing Education in the Health Professions*, 26(3), 173-191.**

**Key Message**

This article estimates that the rate of dyscompetence (i.e., failure to maintain acceptable standards of professional practice) of practicing physicians is between 6% and 12%. The author points out that a comprehensive system of assessment that includes written examinations is important for the identification and remediation of dyscompetence.

**Abstract**

Underperformance among physicians is not well studied or defined; yet, the identification and remediation of physicians who are not performing up to acceptable standards is central to quality care and patient safety. Methods for estimating the prevalence of dyscompetence include evaluating available data on medical errors, malpractice claims, disciplinary actions, quality control studies, medical record review studies, and in-stream assessments of physician performance. These data provide a range of estimates from 0.6% to 50%, depending on the method. A reasonable estimate of dyscompetence appears to be 6% to 12%. Age related cognitive decline, impairment due to substance use disorders, and other psychiatric illness can contribute to underperformance, diminishing physicians' insight into their level of performance as well as their ability to benefit from an educational experience. Currently, dyscompetent physicians in the United States are identified through either the legal system or peer review. The primary method of resolving issues of underperformance in physicians is through continuing medical education (CME). Although a number of specialized assessment and education programs exist in the United States, these programs are largely underutilized. Similar programs exist in Canada and have provided evidence of the efficacy of a more specialized and individualized educational approach for underperforming physicians. Current specialty programs focused on this population employ individual assessments of knowledge and performance, individually designed educational programs, long-term plans for maintenance of educational activity, and repeated assessment of performance level. Noting that few CME programs offer these requirements, a number of changes to current medical quality assurance programs that might foster such educational requirements for underperforming physicians are provided.

**Key Content Descriptors**

*Health Care/Medical Education Setting Testing Effect and the Value of Examinations*