

CLEAR Exam Review

Volume XXIX, Number 1
Spring 2019

A Journal

CLEAR Exam Review

VOLUME XXIX, NUMBER 1

SPRING 2019

CLEAR Exam Review is a journal, published twice a year, reviewing issues affecting testing and credentialing. CER is published by the Council on Licensure, Enforcement, and Regulation, 108 Wind Haven Drive, Suite A, Nicholasville, KY 40356.

Design and composition of this journal have been underwritten by Prometric, which specializes in the design, development, and full-service operation of high-quality licensing, certification and other adult examination programs.

Subscriptions to CER are sent free of charge to all CLEAR members and are available for \$30 per year to others. Contact CLEAR at (859) 269-1289 or cer@clearhq.org for membership and subscription information.

Advertisements and Classified (e.g., position vacancies) for CER may be reserved by contacting CLEAR at the address or phone number noted above. Ads are limited in size to 1/4 or 1/2 page, and cost \$100 or \$200, respectively, per issue.

Editorial Board
Steven Nettles
Retired, Applied Measurement Professionals, Inc.

Jim Zukowski
Independent Consultant

Coeditor
Elizabeth Witt, Ph.D.
Witt Measurement Consulting
Laingsburg, MI
WittMeasure@aol.com

Coeditor
Sandra Greenberg, Ph.D.
ACT
New York, NY
Sandy.Greenberg@act.org

Contents

FROM THE EDITORS	1
<i>Sandra Greenberg, Ph.D.</i>	
<i>Elizabeth Witt, Ph.D.</i>	

COLUMNS

Abstracts and Updates	3
<i>George T. Gray, Ed.D.</i>	

Legal Beat	8
<i>Dale J. Atkinson, Esq.</i>	

Perspectives on Testing.....	11
<i>Grady Barnhill, M.Ed., Fae Mellichamp, M.S., Cyrus Mirza, Ph.D., Chuck Friedman, Ph.D., and Lawrence J. Fabrey, Ph.D.</i>	

2018 CLEAR Quick Poll Results	16
<i>Carla M. Caro, M.A.</i>	

ARTICLES

Are You Kidding Me? Keep Your Exams Legally Defensible	23
<i>Sarah Wennik, M.B.A.</i>	

The Future of Medical Continuing Certification Assessment: Relevant, Dynamic, and Frequent.....	27
<i>Sarah D. Schnabel, Ph.D.</i>	

The Future of Medical Continuing Certification Assessment: Relevant, Dynamic, and Frequent

SARAH D. SCHNABEL, Ph.D.
American Board of Ophthalmology

Editors' note: In this issue of the CLEAR Exam Review, we present an article on what may be perceived as a controversial issue in continuing competence assessment. The author advocates for the need for frequent, summative assessments for maintenance of certification (MOC) among American Board of Medical Specialties (ABMS) members, including the assessment of core knowledge for all physicians in the specialty. Other credentialing organizations (including some medical specialty boards) may have different perspectives regarding best practices for supporting continuing competence. We invite feedback from readers with both supporting and opposing views at cer@clearhq.org.

Elizabeth A. Witt, Ph.D.
& Sandy Greenberg, Ph.D., Eds.

Certification is the process by which a credential is provided to an individual who meets certain standards. The duty of a specialty certification board is to award a certificate to those individuals who have demonstrated mastery of the necessary knowledge, skills, and abilities needed to practice competently. Primarily, valid and reliable criterion-referenced, summative assessment(s) are used to determine which individuals earn the initial certificate (Buckendahl, 2017). But what happens next? What do certified professionals need to do in order to maintain the necessary knowledge, skills, and abilities and, therefore, their certification?

The first question to ask is, what are the skills possessed by the certificate holder and are they subject to change over time? Although this article addresses the question in terms of certified medical doctors, the principles may extend to other types of professional certification. Several studies of physicians have shown that knowledge and skills are not constant and may in fact decrease over time (Ayanian et al., 1994; Choudhry, Fletcher, & Soumerai, 2005; Hawkins, Lipner, Ham, Wagner, & Holmboe, 2013; Salem-Schatz, Avorn, & Soumerai, 1990). Choudhry et al. (2005) reviewed 62 studies on U.S. physicians and found that “physicians who have been in practice for more years and older physicians possess less factual knowledge, are less likely to adhere to appropriate standards of care, and may also have poorer patient outcomes” (p. 269). These findings are likely affected by the rapid growth in technology, procedures, new treatments, and other changes to the landscape of providing care to patients.

When a certifying board has reason to believe that certificate holders' knowledge and skills required for safe and effective practice could decline over time, it is the board's duty to periodically verify whether the certificate holder still possesses the requisite competencies representative of board certification. While educational tools and formative assessments likely help certificate holders stay up to date, a summative assessment represents a direct measure of whether an individual has maintained the standards of the initial certification. (Note: The certificate holder's actual scope of practice may not be as broad as the areas covered by the initial credentialing examination, thus making the initial exam somewhat less relevant as a direct measure of current practice.)

A summative assessment is often called an assessment of learning. It is used to determine whether (or the extent to which) an individual has learned or mastered some knowledge or skill. In contrast, a formative assessment is used to provide feedback to the examinee to encourage ongoing learning (AERA, APA, & NCME, 2014) and is often called assessment *for* learning. An assessment can have both formative and summative components; for example, an assessment may be used to make a classification

decision and also to provide users with feedback on their performance. By definition, only a summative assessment can provide a classification decision related to an individual's mastery of a domain of knowledge.

The challenge with requiring periodic summative assessments for certificate holders is that they are sometimes considered irrelevant and burdensome and may not ultimately result in better patient care (Cook, Blachman, West, & Wittich, 2016; Drazen & Weinstein, 2010). These contentions have risen to the level of state legislatures ruling that maintenance of certification for specialty doctors (through the American Board of Medical Specialties) cannot be required for employment (Finkel, 2016; Sullivan, 2018). Summative assessment requirements for physician certification maintenance must address these issues while simultaneously maintaining the high standards for reliability and validity that will allow certification boards to feel confident that they are recertifying the right professionals. The remainder of this article is about one potential model of the future for maintaining certification. How do we balance the challenges to the status quo with the necessary attention to psychometric rigor? The medical recertification model of the future should feature assessments that are relevant, dynamic, and frequent.

Any assessment must be relevant to practice. Relevancy will be explored in two ways: fidelity to practice and specificity to practice. The first issue, fidelity to practice, is a measure of how well an assessment measures what a practitioner does as opposed to just what he or she knows. Miller's oft-cited clinical assessment framework distinguishes four levels: knows (knowledge), knows how (competence), shows how (performance), and does (action) (Miller, 1990). Fidelity and standardization are inversely related: on one end of the continuum is a fixed-form multiple choice test (high standardization, low fidelity), and on the other is observation of an individual's practice (high fidelity, low standardization) (Swygert & Williamson, 2016). It is no doubt important to select a high-fidelity assessment, so the question is ultimately, how standardized must the assessment tool be? Or in other words, how easily comparable must scores or performances of different individuals be to one another? How much equivalence in scores is needed? If the current standardized assessment for recertification isn't perceived as relevant to certificate holders, perhaps trading in some standardization for fidelity is required (climbing Miller's pyramid). The assessment tools that better measure "shows how" are things like oral examinations, simulations, and OSCEs. Assessment of "does" includes direct observation and case review. Reliable implementation of these types of assessment, however, is the "international challenge of the century for all involved in clinical competence-testing" (Wass et al., 2001).

To maintain the ongoing value of a physician's professional certification, a periodic summative recertification assessment must test knowledge across the breadth of the field listed on that certificate. However, allowing individuals to select certain clinical conditions or populations that they serve and tailoring parts of the assessment to those areas would improve the relevancy of the assessment to an individual's specific practice. Even more practice-specific features would include allowing certificate holders to select from their own cases as the basis for the assessment. What is no doubt a challenge about this format is developing standardized metrics for success. Many boards have adopted a hybrid model, wherein both core knowledge and subspecialty knowledge are assessed for continuing certification (American Board of Ophthalmology, 2018). As long as the core knowledge component is designed in a way that truly captures knowledge about the most critical aspects of the necessary field of knowledge, this hybrid approach can advance recertification assessment. Another model that addresses both practice fidelity and specificity is outcomes assessment. Patient outcomes are often considered the best measure of health care quality (Norcini, 2005). But outcomes are hard to measure. Additionally, while outcomes represent an important measure of overall health care quality, outcomes are influenced by non-provider factors (e.g., compliance, population) and may not tell us enough about individual provider competence. And while the medical assessment community is working to develop meaningful outcomes measurement, a proxy assessment including core knowledge needed by all certificate holders must remain a requirement.

Another feature of a future assessment is that it is dynamic: it assesses the most current, important information in the field. The standard of practice in standardized assessments is to develop a test blueprint (which outlines the content of each topic area on the examination and the relative percentage allocation of each topic on the examination) and to keep it constant for several years so that all tests have the same distribution of content. Unlike the historical design of most tests with a static test blueprint, a dynamic assessment requires a dynamic blueprint that highlights the most important topics and recent advances. As new research is published, important findings can be added to the blueprint and, thus, the assessment. Again, a trade-off exists between standardization/score equivalency (static blueprint) and fidelity (dynamic blueprint). A validity argument can be made in favor of the latter: if a certificate's meaning is that certificate holders are knowledgeable about current best practices and important recent advances, then it would be critical to develop a design in which new knowledge is assessed.

When it comes to burden, we have more to unpack. "Burden" is a combination of lack of perceived value, lack

of convenience, and amount of time spent. The value issue is clearly one that needs solving; but there also must be an understanding by certificate holders that reliable, valid assessment takes time. The duty of a certification board is to certify only those individuals that have mastery of an area of knowledge or skill. Short tests that lack reliability may be perceived as more convenient, but they ultimately do a disservice to the public if they do not correctly identify who should hold (and maintain) that certificate. There are two considerations for moving forward. First, all time spent by certificate holders needs to be purposeful. Boards need to critically evaluate their assessment practices and focus on those assessments that are defensible measures of the proposed construct(s) of interest. Second, innovations in remote assessment delivery and other technology to allow quicker and easier access to certification assessments should be explored.

Additionally, boards need to explain why periodic summative assessment is important and how it benefits certificate holders. Literature exists that links reliable summative assessment performance to improved outcomes and processes of care. Holmboe et al. (2008) found that physicians certified by the American Board of Internal Medicine who received higher scores on a maintenance of certification examination exhibited higher rates of performing processes of care for diabetes and mammography screening for Medicare patients. Norcini, Lipner, and Kimball (2002) found that “successful examination performance (i.e., certification in internal medicine or cardiology) was associated with a 19% reduction in mortality” (p. 853). Additional studies have identified links between American Board of Medical Specialties (ABMS) board certification status or summative assessment performance and decreased incidence of licensure disciplinary actions (Jones, Kopp, & Malangoni, 2018; Lipner, Young, Chaudhry, Duhigg, & Papadakis, 2016; Zhou et al., 2017). However, additional validity evidence needs to be developed and disseminated to certificate holders. Per the *Standards of Educational and Psychological Testing* (AERA, APA, & NCME, 2014), the burden of proof for the validity argument of an examination rests with the test developer (certification board). Critics of the validity evidence to date are right about one thing: more definitive evidence is needed.

There hasn't been a large outcry from certificate holders for more frequent testing; nevertheless, this is an opportunity for improvement for medical certification boards. The ABMS specialty boards' recertification assessments of the past were often decennial (ABMS, 2018). Tamblyn, Abrahamowicz, and Dauphinee (2002) found “that [Canadian] family physicians' scores on an initial licensing examination had predictive validity for future performance on a number of quality measures up to 6 years after the examination”

(Holmboe et al., 2008, p. 1396). While the relationship between the competencies demonstrated on summative assessments and time is still not perfectly clear, rapid changes in medicine and the findings about individual physician changes in competence over time require certification boards to look closely at the frequency of summative assessments. Passing a decennial examination may not be the best way to determine whether someone is keeping up with rapid changes in patient care.

Additionally, more frequent assessment aids in retention (the “testing effect”) (Larsen et al., 2008). This is common knowledge: cramming for a single test is not likely to result in long-term knowledge retention. The more frequently individuals engage with the assessment, the more they learn and retain the knowledge. Developing more frequent assessments may not be perceived as less burdensome to practitioners (although there is some evidence of favorable reception [Lauer, 2018]), but this is an area of evolution for maintenance of certification that promotes stronger claims about individual continued competence.

Ultimately a valid assessment hinges on assessment purpose: what is the assessment designed to measure? If we assume that certification is meant to measure all aspects of what it is to be a “good” practitioner, it is likely there are several constructs or traits to measure. “No single assessment method can provide all the data required for judgment of anything so complex as the delivery of professional services by a successful physician” (Miller, 1990, p. 63). Different skills require different assessment tools: measuring knowledge of basic science and measuring communication skills would not be optimized with the same assessment tool. Each intended construct of measurement (skill) will need its own relevant, dynamic, and frequent assessment. To move forward, boards need to listen to and address the concerns of the certificate holders. At the same time, implementing reliable assessments that are more frequent and measure relevant knowledge, including some component of core knowledge, is essential to uphold the value of the certificate. Both sides are right: continuing certification assessments require some improvements, but abolishing them altogether is not the answer. A periodic summative assessment must be maintained indefinitely or until research is unequivocal in the ability of the initial summative assessment to maintain its predictive validity across the career span of a certificate holder. It is important that the validity arguments of both initial and recertification summative assessments are regularly reviewed and that alternative assessments are reviewed with equal rigor before adoption.

References

- American Board of Medical Specialties. (2014). *Standards for the ABMS program for maintenance of certification (MOC)* [PDF File]. Retrieved from <https://www.abms.org/media/1109/standards-for-the-abms-program-for-moc-final.pdf>
- American Board of Ophthalmology. (2018). *Quarterly questions*. Retrieved from <https://abop.org/maintain-certification/quarterly-questions/>
- American Educational Research Association [AERA], American Psychological Association [APA], & National Council on Measurement in Education [NCME]. (2014). *Standards for educational and psychological testing*. Washington, DC: American Educational Research Association.
- Ayanian, J., Hauptman, P., Guadagnoli, E., Antman, E., Pashos, C., & McNeil, B. (1994). Knowledge and practices of generalist and specialist physicians regarding drug-therapy for acute myocardial infarction. *New England Journal of Medicine*, *331*(17), 1136-1142. doi:10.1056/NEJM199410273311707
- Buckendahl, C. W. (2017). Credentialing: A continuum of measurement theories, policies, and practices. In C. Buckendahl & S. Davis-Becker (Eds.), *Testing in the professions* (pp. 1-20). New York, NY: Routledge.
- Choudhry, N., Fletcher, R., & Soumerai, S. (2005). Systematic review: The relationship between clinical experience and quality of health care. *Annals of Internal Medicine*, *142*(4), 260-273. doi:10.7326/0003-4819-142-4-200502150-00008
- Cook, D. A., Blachman, M. J., West, C. P., & Wittich, C. M. (2016). Physician attitudes about maintenance of certification: A cross-specialty national survey. *Mayo Clinic Proceedings*, *91*(10), 1336-1345. doi:10.1016/j.mayocp.2016.07.004
- Drazen, J. M., & Weinstein, D. F. (2010). Considering recertification. *New England Journal of Medicine*, *362*(10), 946-947. doi:10.1056/NEJMe1000174
- Finkel, E. (2016). Physicians take MOC fight to state level. *Medical Economics*, *93*(16), 44-47. Retrieved from <http://proxy.cc.uic.edu/login?url=https://search-proquest-com.proxy.cc.uic.edu/docview/1814345234?accountid=14552>
- 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*, S7-S19. doi:10.1002/chp.21201
- 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. doi:10.1001/archinte.168.13.1396
- Jones, A. T., Kopp, J. P., & Malangoni, M. A. (2018). Association between maintaining certification in general surgery and loss-of-license actions. *JAMA-Journal of the American Medical Association*, *320*(11), 1195-1196. doi:10.1001/jama.2018.9550
- Larsen, D. P., Butler, A. C., & Roediger, H. L. (2008). Test-enhanced learning in medical education. *Medical Education*, *42*(10), 959-966. doi:10.1111/j.1365-2923.2008.03124.x
- Lauer, A. K. (2018). The new face of board certification. *Retina Today: Business Matters*, *1*(2), 9-11. http://retinatoday.com/viewer.asp?f=pdfs/0618_supp.pdf
- Lipner, R. S., Young, A., Chaudhry, H. J., Duhigg, L. M., & Papadakis, M. A. (2016). Specialty certification status, performance ratings, and disciplinary actions of internal medicine residents. *Academic Medicine*, *91*(3), 376-381. doi:10.1097/ACM.0000000000001055
- Miller, G. (1990). The assessment of clinical skills competence performance. *Academic Medicine*, *65*(9), S63-S67. doi:10.1097/00001888-199009000-00045
- Norcini, J. (2005). Current perspectives in assessment: The assessment of performance at work. *Medical Education*, *39*(9), 880-889. doi:10.1111/j.1365-2929.2005.02182.x
- Norcini, J., Lipner, R., & Kimball, H. (2002). Certifying examination performance and patient outcomes following acute myocardial infarction. *Medical Education*, *36*(9), 853-859. doi:10.1046/j.1365-2923.2002.01293.x
- Salem-Schatz, S., Avorn, J., & Soumerai, S. (1990). Influence of clinical knowledge, organizational context, and practice style on transfusion decision making: Implications for practice change strategies. *JAMA-Journal of the American Medical Association*, *264*(4), 476-483. doi:10.1001/jama.264.4.476
- Sullivan, T. (2018, May 4). Anti-MOC laws picking up steam across the United States. *Policy & Medicine*. Retrieved from <https://www.policymed.com/2017/06/anti-moc-laws-picking-up-steam-across-the-united-states.html>
- Swygert, K. A., & Williamson, D. M. (2016). Using performance tasks in credentialing tests. In S. Lane, M. Raymond, & T. Haladyna (Eds.) *Handbook of test development* (pp. 294-312). New York, New York: Routledge.
- Tamblyn, R., Abrahamowicz, M., Dauphinee, W. D., Hanley, J. A., Norcini, J., Girard, N., . . . Grand'Maison, P. (2002). Association between licensure examination scores and practice in primary care. *JAMA-Journal of the American Medical Association*, *288*(23), 3019-3026. doi:10.1001/jama.288.23.3019
- Wass, V., Van der Vleuten, C., Shatzer, J., & Jones, R. (2001). Assessment of clinical competence. *Lancet*, *357*(9260), 945-949. doi:10.1016/S0140-6736(00)04221-5
- Zhou, Y., Sun, H., Culley, D. J., Young, A., Harman, A. E., & Warner, D. O. (2017). Effectiveness of written and oral specialty certification examinations to predict actions against the medical licenses of anesthesiologists. *Anesthesiology*, *126*(6), 1171-1179. doi:10.1097/ALN.0000000000001623