



# The Case Against One-Shot Testing for Initial Dental Licensure

David W. Chambers, EdM, MBA, PhD; Arthur A. Dugoni, DDS, MSD; Ian Paisley, DDS

## ABSTRACT

High-stakes testing are expected to meet standards for cost-effectiveness, fairness, transparency, high reliability, and high validity. It is questionable whether initial licensure examinations in dentistry meet such standards.

Decades of piecemeal adjustments in the system have resulted in limited improvement. The essential flaw in the system is reliance on a one-shot sample of a small segment of the skills, understanding, and supporting values needed for today's professional practice of dentistry. The "snapshot" approach to testing produces inherently substandard levels of reliability and validity. A three-step alternative is proposed: boards should (1) define the competencies required of beginning practitioners, (2) establish the psychometric standards needed to make defensible judgments about candidates, and (3) base licensure decisions only on portfolios of evidence that test for defined competencies at established levels of quality.

Much of the current licensing examination system is driven by tradition and efforts to keep it in operation have been piecemeal rather than grounded in a comprehensive understanding of its purpose or standards prevailing in other licensure communities. It is the result of a political process rather than based on psychometric or other rational principles.

Initial dental licensure examinations are examples of what is known in the assessment community as "high-stakes performance assessment."<sup>1,2</sup> For the public and the candidate, a great deal rides on a moment in time and a small sample of performance. Such testing is used regularly and effectively for the civil service, drivers' licenses, admission to college, and for professions such as medicine.<sup>3-6</sup>



**Authors** / David W. Chambers, EdM, MBA, PhD, is associate dean for academic affairs and scholarship at the University of the Pacific School of Dentistry.

Arthur A. Dugoni, DDS, MSD, is dean and professor of orthodontics at the University of the Pacific School of Dentistry.

Ian Paisley, DDS, is a 2003 graduate of the University of the Pacific, School of Dentistry. He currently is attending the General Practice Residency Program at Denver Veterans Affairs Medical Center in Denver, Colo. He was the ASDA National Consultant on Dental Education and Licensure.

Table 1

### Standards for High-Stakes Testing

Standard	Quality of Initial Licensure Examinations in Dentistry
<b>Cost effectiveness</b> <i>(Higher net contribution to society than alternatives)</i>	States lack resources, resources diverted from enforcement Potential high rate of false positives — incompetents licensed Unprofessional culture of hassle for recent graduates As much as a year's lost income for 15 percent of recent graduates
<b>Fairness</b> <i>(Decisions based only on competence)</i>	Good record of objectivity and protection against bias Different standards applied to candidates and practitioners
<b>Transparency</b> <i>(Information available to those who need to know)</i>	No published evidence that initial licensure protects public Partial and untimely reporting have prompted ADA intervention Psychometric data reported in peer-reviewed literature Incomplete disclosure of reliability and validity
<b>Reliability</b> <i>(Reduction of all sources of random variation)</i>	Dental licensure $r=.40$ compared to standards of $r=.80$ Focus on inter-rater calibration which is trivial source of error One-time testing cannot be improved to standard
<b>Validity</b> <i>(Decision based on all components of practice)</i>	No definitions of competency have been developed by boards Initial licensure samples very small subset of practice Patient management cannot be evaluated in one-shot format

As shown in **Table 1**, there are five standards traditionally applied to high-stakes testing. These include: (1) cost-effectiveness — the tests must have a higher return for investment than comparable alternatives; (2) fairness — licensure decisions should be influenced by no criteria other than competence; (3) transparency — those concerned should be able to receive timely, meaningful, and comprehensible information about how decisions are made; (4) reliable — all sources of randomness must be identified and reduced to an acceptable minimum; and (5) validity — the tests measure what practitioners do.

In this paper, we raise concerns

that the traditional examination for initial dental licensure that was developed to meet conditions prevalent 80 years ago now fail to reach conventional standards for high-stakes examinations, especially in the areas of reliability and validity. In particular, the one-shot nature of initial licensure examinations makes it virtually impossible to satisfy the five necessary standards for such tests. "Tweaking" or outsourcing in a similar format are not viable options in anything other than a political sense. It is possible, however, to use the steps traditionally taken in developing high-stakes testing to create a satisfactory alternative.

### Can We Meet the Standards With One-Shot Examinations?

#### *Cost-effectiveness*

Initial licensure examinations tend not to run smoothly. They can be characterized as cumbersome with regard to handling scoring and reporting and non-responsive in scheduling. Issues exist in the areas of the ethics of testing using live patients, difficulties for candidates to manage the logistics of testing, availability of remediation, and appropriate care for patients of those candidates who failed sections of the exams. The Dental Board of California is in a budget crisis. Staff shortages resulted in candidates from the spring 2003 testing being informed of their pass or fail status weeks after the announced date and after the deadline for candidates to apply for the next testing. A large backlog exists for the restorative technique (bench test) examination. The short written tests in areas such as endodontics have not been updated in years and are likely compromised.

Although virtually all graduates of the five California dental schools pass the initial licensure test within one year and the national average licensure rate is about 97 percent,<sup>7</sup> the cost in lost income due to delays normally exceed the entire cost of dental education for initially failing candidates. Dental services denied to the public are also significant, especially as concerns over access to care grow. The growing negative attitude in organized dentistry<sup>8-24</sup> and among recent graduates<sup>25,26</sup> regarding initial licensure must also be regarded as a cost.

It is appropriate to note the costs associated with licensing incompetent practitioners. The low reliability of one-shot exams (discussed later) actually means that it is more likely to make false-positive decisions (granting a license to an incompetent candidate) than false-negative decisions (requiring

that a failing candidate be retested). Dental malpractice among licensed dentists is a serious issue and, nationally, each year about the same number of dentists who have passed an initial licensure examination have their licenses disciplined as the number who do not pass a board within a year.<sup>27-28</sup> (See side bar article, *Why Some of the Best Graduates Fail the Boards and Why Incompetent Graduates are Licensed.*)

### *Fairness*

States that manage their own initial licensure examinations and regional examining agencies have done a commendable job of removing opportunity for bias based on personal characteristics of candidates other than technical performance. Such “blind objectivity” is a welcome change from what is rumored to have occurred in the early days of testing. Such practices also tend to reduce judgments about competency to the same kind of “quality of product” appraisals used to authorize insurance claims. They impose constraints on validity.

Another side of fairness must also be weighed. Are candidates for licensure being held to the same standards that practitioners must observe? Would, for example, a practicing dentist lose his or her license for transposing numbers in a laboratory prescription, for getting a pulpal exposure, or for leaving an amalgam restoration too high? Typically, boards look at a pattern of incompetent performance in reaching decisions to discipline an existing license. To be fair, shouldn't boards also consider a pattern of performance in granting licenses in the first place?

### *Transparency*

Transparent systems are open to review by those who need to know. Dental boards have traditionally grounded their mandate in a legislated charge to protect the public. The “California Business and Professions Code,” popular-

ly known as the “practice act,” does not actually contain such language<sup>29</sup> and there is no published evidence that initial licensure examinations in dentistry have that effect. (It is traditional in other professions such as medicine to publish such data in peer-reviewed journals where it is subject to the scrutiny of experts. See<sup>30-33</sup> for a representative sample.) In California, results of initial board examinations are made available to dental schools in May of each

The current one-shot initial dental licensure system misclassifies at least 20 percent of candidates who must retake the tests, plus an unknown number of candidates who pass the tests by luck and should not have been granted a license.

year — approximately a year after such data would be useful for curricular review purposes. In the late 1990s, the House of Delegates of the American Dental Association called upon states and examining agencies to report their pass rate statistics on an annual basis. Some states have not been able to do this on a consistent basis and the report has not been published each year.<sup>7</sup> Although some examining agencies do report overall reliability results for inter-rater consistency, none report all sources of unreliability. It has been general practice in California that patients are not informed by the board that the care they received is inadequate and should be replaced or corrected when the candidate is given a failing score.

### *Reliability*

Standards have gradually emerged in high-stakes testing, such as licensure in various professions, admissions to advanced educational opportunities, or for highly selective jobs.<sup>1-2</sup> Reliability coefficients of .80 to .90 are usually expected, although, occasional r-values as low as .70 may be encountered.<sup>34-37</sup> With an anticipated high pass rate (as is the case in dental licensure examinations), a test system with reliability of .90 would fail about 1 percent of candidates who would pass if tested again immediately under the same circumstances. Even with reliability as low as  $r=.70$ , the error rate is still generally regarded as acceptable at about 3 percent. It is known that the National Board Dental Examinations have reliabilities above .90.<sup>38</sup>

Nationally, the reliability of initial licensure examinations in dentistry is lower than .40.<sup>39-41</sup> This number is calculated from the most recent available 1998 ADA data on all licensure jurisdictions<sup>7</sup> and includes all subtests taken in combination. It does not include an estimate of those candidates who pass the initial licensure tests but would not pass if given the identical test immediately (false positive results). This means that the current one-shot initial dental licensure system misclassifies at least 20 percent of candidates who must retake the tests, plus an unknown number of candidates who pass the tests by luck and should not have been granted a license. In California, the reliability is estimated to be slightly higher at  $r=.45$  until this year. Preliminary reports indicate a sharp drop in reliability in the spring 2003 tests.

### *Validity*

In 1990, in response to a request from the ADA, the licensure examination community conducted a systematic content analysis process to ensure that the tasks tested on one-shot licen-

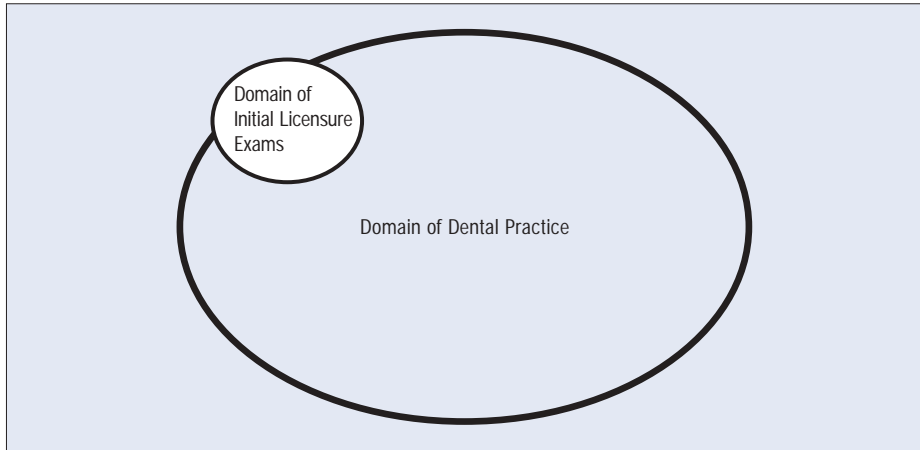


Figure 1. Schematic representation of overlap between skills, understanding, and supporting values required of beginning dental practitioners and those sampled in initial licensure examinations.

sure examinations were realistic representations of tasks performed by dentists.<sup>42</sup> Generally this was a well-conducted exercise, but it was a poorly conceived one. It failed to provide evidence that the tasks performed in dental offices were the same as those tested on the examinations. (A is part of B does not prove that B is part of A.) It was a “validation of the test” not a demonstration that the test is a valid measure of practice.<sup>43</sup> See **Figure 1**.

In particular, the one-shot evaluation format is limited to testing one-shot tasks. Increasingly, dentistry is about managing patients and their oral health over extended periods of time.<sup>21</sup> No matter how well designed, a test of one-shot performance will not be able to measure the large realm of dental practice behaviors that occur in context and over time.<sup>39-48</sup>

### Can the One-Shot Model Be Saved Through Adjustment?

State dental boards are answerable to the public regarding the competency of those dentists whom they license.<sup>29, 49</sup> Examining agencies or testing committees in states such as California that conduct their own initial licensure examinations, are answerable that the data they provide to the boards for mak-

ing licensure decision are of high quality. Currently, dental boards in America accept a lower standard for such data than do boards in other professions or other who use high-stakes evaluation data. In fact, most dental boards do not actually have formal standards for what constitutes an acceptable level of evaluation evidence. The process is not transparent to the public whom boards serve.

Discussions about improving initial licensure examination in dentistry in recent decades have been about tweaking the system.<sup>42,49-53</sup> Sufficient data are now available to demonstrate that the one-shot system cannot be adjusted enough to reach conventional standards approximating  $r=.80$ .<sup>5,39-41,48</sup> The hard work and good intentions of “tweakers” should not be applauded. The situation resembles what happens when patients seek shortcuts through alternative medicine. Even when no harm can be directly attributed to the unproven therapies, they preclude or delay a proper remedy.

Four arguments are often advanced in favor of tweaking the one-shot approach to initial dental licensure. They will be analyzed.

#### *Does Delegation to Examining Agencies Improve Decision-making?*

Regional testing agencies<sup>28</sup> offer

advantages such as centralization of expertise and staff resources; the need to accommodate state-to-state variations in standards — hence more generic criteria; relieving individual states of the cost of maintaining their own testing system; and larger data bases with their potential for deeper psychometric analysis. They also offer candidates the benefits of multiple test sites and dates. They are vulnerable, as is any monopolistic organization that must make its budget, to taking a defensive posture.<sup>49-54</sup>

The use of regional testing agencies moves the problems of low reliability and lack of accountability for testing standards; it does not solve them. The testing practices used by regional examining agencies are essentially the same as those used by states that do their own testing.<sup>7,28,42</sup> State boards can delegate information gathering, but they cannot delegate responsibility for making licensure decisions. When such delegation is made in business or public agencies, it is customary to require that suppliers demonstrate capacity to meet performance standards — in this case, measured in terms of predetermined criteria for reliability and validity.<sup>55</sup>

#### *Is It Possible to Improve Individual Exams?*

In the past 30 years, tremendous effort has gone into improving the selection of test situations, logistics of administering the examinations, selection and calibration of examiners, and scoring.<sup>49</sup> They embody high intentions. But effort alone is not enough to pass a candidate or qualify an examining group, and the reliability and validity of one-shot initial licensure examinations remains below acceptable levels despite half a century of numerous minor adjustments to the system.

In order to understand why asymptote has already been reached in the existing model of licensure testing, it is

*Continued on Page 248*

# Why Some of the Best Graduates Fail the Boards and Why Incompetent Graduates are Licensed

**I**f you were the worst tennis player at the club, would you want to play the pro “one-time, winner takes all” or “the best four out of seven?” Remember, the truth will come out — certainly given time.

On any given day, people perform above or below their true ability level, but the more performances are averaged together, the closer they will be to the true performance capability.

This understanding can be applied to performance in dental school and in the one-shot initial licensure examination. Imagine that there is a scale from 0 to 100 and that a score of 75 is required to practice dentistry for the rest of one’s life with no further demonstration of skill needed. We have two graduates, Star and Dud. Star’s ability is estimated based on a series of test cases, faculty ratings, and other data to be 80; and Dud is thought to be 70. The confidence of these estimates is  $r=.80$  based on review of the psychometric characteristics of the evaluation methods. Both graduates sit for a one-shot initial licensure examination with a known consistency of  $r=.40$ . The ability of both candidates remains the same from school to licensure examination and in subsequent licensure tests.

The results of this scenario are illustrated in **Figure 2**. The solid horizontal lines are 95 percent confidence intervals based on evaluation of performance in dental school. Ninety-five percent of the evaluations of Star’s ability will place his or her true ability between 75.6 and 84.4. Similarly, 95 percent of the evaluations of Dud will have his or her ability below the 75 threshold. These conclusions apply only to sets of ratings such as those conducted in dental school where a large amount of information can be aggregated to generate highly consistent conclusions. When less consistent methods of evaluation are used,

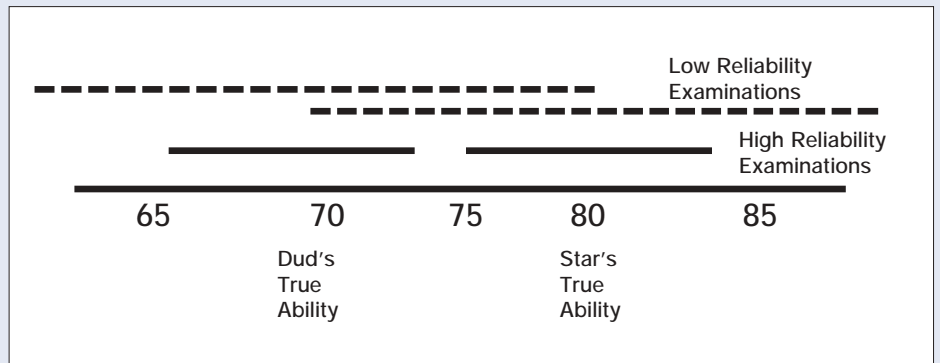


Figure 2. Ninety-five percent confidence intervals for performance of qualified and unqualified candidates under conditions of high and low reliability.

such as one-shot initial licensure examination, a wide berth must be reserved for estimating true ability — for any given true ability level. The dashed horizontal lines represent the 95 percent confidence intervals for the same candidates based on consistency typical of one-shot initial licensure situations. The range now extends almost 10 points on either side of the true ability. This follows simply from the reduced reliability of such examinations, but it can be estimated precisely using statistical methods.

Candidate Star had almost no chance of failing a comprehensive evaluation (one with acceptable reliability), but he or she has about a 17 percent chance of being a false negative — a failure on the board who should not have failed. Candidate Dud has a similar 17 percent chance of being a false positive — a passing candidate who lacks true ability.

Boards have recognized the unfair conditions resulting from one-shot tests with low reliability, so they make provision for candidates to be tested up to three times without prejudice. In our case of Star, there is less than 1 percent likelihood that a candidate of this ability level will fail to pass in three tries. The obvious penalty to such candidates comes from lost income and negative attitudes toward the profession. The

hidden cost of this system intended to compensate for a design flaw in one-shot initial licensure evaluation is the false positive decisions it creates. Candidate Dud, who certainly lacks ability, has a 70 percent chance of passing at least one in three of the tests at his or her level of competence.

The one-in-three rule raises some ethical issues. It is a means of lowering the standard of competence in the profession while appearing to hold to a high standard. No one would argue that a practicing dentist should be allowed to retain his or her license if only one out of every three crowns is functional or 33 percent of the diagnoses are correct. Arguing that this is the fault of the schools for sending forward unqualified candidates is an example of moral hazard. If only the boards or delegated examining agencies are qualified to determine who is competent and if schools and one-shot initial licensure tests often disagree, schools cannot be criticized for failing to do the board’s work. (If the proposals outlined in the accompanying paper were in effect, boards could and would be expected to withdraw the delegated authority to present licensure data given to schools or examining agencies if they failed to meet standards for consistency established by the board.) **CDA**

*Continued from Page 246*

necessary to have more than a superficial understanding of evaluation theory.<sup>2,34,44</sup> The variation observed in test scores results from multiple sources, such as differences in true ability across candidates, examiner differences, testing circumstances, patient variability, instructions given to examiners and candidates at different sites and their physical layout, the “culture” of professional assistants and patients, and a myriad of other factors that can be lumped under the heading of random chance. Reliability is defined as the square root of the ratio of true differences in competency between candidates to all sources of variation taken together.<sup>35,37</sup>

Some readers who are familiar with licensure testing know that examiner

calibration has been raised in some cases to  $r=.60$  or even  $.70$  (the exact numbers are not available because state boards do not require transparency in psychometric properties from testing agencies). It would be tempting to argue that the reliability of examinations must be higher than the  $r=.40$  quoted earlier in this paper if the reliability for examiners is  $r=.60$ . Unfortunately, that is not sound reasoning. The “reliability of the examiners” is determined by dividing differences in candidates by differences in candidate plus differences in examiners. The “reliability of the test” is determined by dividing differences in candidates by all the differences, including examiners and all other sources of variation.

A reasonable question to ask would

be, “is the variance in examiners very large compared to other sources?” If it is a major factor, examiner calibration is very important; if it is a small contributor, examiner calibration is nice but not a significant issue in establishing the credibility of initial licensure examinations. This question was explored by Chambers and Loos<sup>48</sup> in a simulation of initial licensure testing. It was found that examiners contribute less than 1 percent of the variance. Research in other fields report similar findings.<sup>56-58</sup> Because examiner calibration has already been improved to near its practical limit and because it makes so little difference anyway, hoping to improve one-shot initial licensure examination by working with examiners is a fruitless strategy (other than its obvious political value).

It might be argued that research on testing has not yet ruled out all opportunities for tweaking, even if examiners are not a viable alternative. Generalizability research, of the type described in the paragraph above allows for partitioning of variation into multiple categories, including a leftover category of unexplained error. It has been shown in several disciplines,<sup>56-58</sup> including dentistry,<sup>48</sup> that the single category of candidate-patient-trial is the largest source of variation. It is larger than differences between candidates, much larger than examiners, and larger than the leftover category that includes all the potential tweaking items. Variability for each candidate from occasion-to-occasion and patient-to-patient swamps all other considerations when trying to get an accurate read on competence. One-shot examination systems are by definition blind to the major (overwhelmingly major) source of error in data that lead to licensure decisions.

The only reasonable method to improve licensure examinations is by increasing the number of patients candidates are tested on. Tests were performed with the Chambers and Loos data and it was confirmed that there are no possible adjustments to factors such as examiners or test logistics that would have as beneficial effect in improving reliability as testing the candidates twice. This conclusion remains true even when it is assumed that the system is tweaked to perfection.

### *Would Simulation Be An Improvement?*

The dental education community in particular has argued recently for the elimination of live patients from initial licensure examinations. Most often, this position is advanced on ethical grounds<sup>10,11,15,18,20</sup> — patients should not be treated by unlicensed practitioners or the logistics of testing create conditions of moral hazard. Others critics suggest that simulations (work per-

formed on typodont teeth) would be preferable as a means of eliminating patient variability.<sup>13,26</sup> Simulation is an attractive hope for those who would like to maintain the one-shot initial licensure system by finding some acceptable adjustment.

Efforts to develop computer simulations<sup>59</sup> have been on going for 15 years and do not seem to be making progress. Data gathered by the Central Regional Testing Agency show that variance is actually larger on typodonts than on patients' teeth<sup>52</sup> — a finding contrary to the belief about standardization of task.

The only reasonable method to improve licensure examinations is by increasing the number of patients candidates are tested on.

Most importantly, reliability should not be purchased at the cost of validity.<sup>3,39,43,47</sup> Dentists do not restore plastic teeth in practice. All dental schools currently use typodont simulations such as those being proposed by some “tweakers” as a screening mechanism to determine which students should be allowed to enter the clinic. Valid licensure examination requires testing on patients under realistic circumstances.

### *Is the Current System A “Test of Minimal Competence?”*

Some defenders of the current examination system agree that it may not be possible to make precise predictions of future competence, but they argue that a clinical failure is automatic evidence of inability to practice. “At least,” they say, “the current test weeds out the grossly incompetent” or the

“three percent who should not practice.” (How that quota of incompetent candidates has been determined has yet to be explained.)

This argument confuses performance with ability, and while ability must always be inferred from performance, they are not the same. It would not be argued, for example, that all candidates who perform acceptably on one occasion should never have their ability questioned at any time in the future. Neither should it be argued that experienced practitioners should immediately turn in their licenses if they obtain one exposure. This would be one-shot licensure examination taken to the extreme.

### **What Needs to Be Done Instead**

No endodontist would limit diagnosis to a single test to support a decision about the vitality of a tooth. Multiple tests are indicated, especially when the first results are equivocal. An initial finding of vitality would never be accepted as evidence that the tooth should never be tested at a later date.

There are two flaws in the current conception of initial licensure based on a one-shot approach to testing. The first misconception is to assume that the essence of dental practice can be observed in a few standardized tests. The second misconception is that single measures of performance are sufficiently reliable sources of data to support high-stakes decisions. Better one-shot test, or having someone else perform the one-shot tests have attractive advantages. Achieving standard for professional licensure decisions is not one of those advantages.

The fundamental issues — defining competence (realistically and comprehensively) and obtaining data to support decisions (reliably and validly) — can both be overcome. The three steps outlined below also hold promise of enhancing cost-effectiveness, fairness, and transparency.

## Define Initial Competence

Currently, no dental board has defined what it means to be competent to begin the practice of dentistry. Competency statements describe the skills, understanding, and supporting values of those who are capable of treating a general population of patients and managing their own continued professional growth.<sup>60,61</sup> The American Dental Education Association has a general set of competencies for both beginning dental practitioners and beginning dental hygiene practitioners.<sup>62,63</sup> All U.S. dental schools have defined the competency of their graduates; it is an accreditation requirement.<sup>64</sup>

Examples of competency statements include: “Restore single teeth for therapeutic reasons”; “Determine differential, provisional, and definitive diagnoses”; “Administer and prescribe medications commonly used in dentistry, including local anesthesia, and manage their complications”; “Practice consistent with sound business principles and legal requirements and regulations”; “Diagnose and treat only within one’s competence”; “Assume active responsibility for one’s lifelong learning”; and “Participate in organized dentistry.”

National Dental Board Examinations and “add on” tests used by some boards or examining agencies, including so-called “ethics” tests, measure only knowledge (not skill or values). One-shot clinical examinations are partial assessments of a few of the skill components of what it means to be a competent beginning practitioner. Many of the vital competencies of the profession are not amenable to testing in the one-shot format. For example, the management of patients’ comprehensive oral health over an extended period of time, deciding when to refer, and prevention, can only be determined in a

realistic context through observation over time.

Developing a set of competencies is typically an easy task.<sup>65</sup> Normally, two days of meetings, with appropriate background work are sufficient. There is usually a good deal of consensus in this process. In 1996, national representatives of the practicing community, registrars (boards), and educators developed a set of competencies that has guided all three groups in reasonable harmony for about a decade.<sup>66</sup>

The quality of any single test will normally be limited because of inherent variability.

## Set Standards for Evidence

This step requires the assistance of a psychometrician and careful examination of the level of examination quality achieved in a variety of settings. The goal is to establish standards for the confidence required by boards in the sources of evidence they use to make licensure decisions.<sup>67</sup> Once determined, boards can allow any group who is able to demonstrate that they can meet the standards consistently to provide the evidence used by boards.

Currently no boards have ex ante standards for the quality of the evidence they use to make licensure decisions.

## Obtain Sufficient Data to Support Sound Decisions

Fortunately, measurement precision is a function of both the consistency of measures and their number. Limited consistency can always be compensated

for by increasing the number of observations.<sup>34,36,47</sup> Even more fortunate is the fact that this compensatory relationship is well understood by measurement practitioners. This is known as a “portfolio” approach to performance assessment.<sup>39,68-71</sup> It is governed by the Spearman-Brown formula<sup>45</sup> in restricted cases and Cronbach’s generalizability theory<sup>72,73</sup> in general.

The quality of any single test will normally be limited because of inherent variability. The traditional solution — fix the amount of data to a single test and take whatever reliability results — seems like an indefensible choice. The alternative is to fix the desired level of decision-making confidence (say reliability of  $r=.80$  over a representative or even exhaustive range of competencies) and vary the amount of data collected so as to guarantee that level of confidence.

It is unlikely that state boards or examining agencies could either create realistic testing conditions or gather sufficient replications of examinations to approach the requirements of valid and reliable initial licensure examination. By contrast, dental schools are very close to doing this currently. What would be required is a new partnership between boards and schools (something like the delegation of information gathering that exists now between boards and examining agencies). Boards would determine the competencies for beginning practice, the nature of testing conditions, and the level of confidence required in order to make a decision. Where schools can meet these standards, boards would accept the evidence and then make their own decision. Where the available evidence fails to meet the board’s standards, there is always a fallback position of using the current approach.

At the very least, there would be many benefits in shared learning and cooperation if the boards and schools began discussions along these lines.

## Conclusion

Dentistry is complex and highly dependent on responding to individual patient needs over extended periods of time. It would be surprising if examinations could capture the essence of dental competence in a single snapshot with consistency much greater than  $r=.40$ . In this paper it has been argued that one-shot initial licensure is indefensible. It has been attempted by well-meaning individuals and organizations for half a century and still remains conspicuously short of the standard achieved and expected by other professions. The critical limiting factor is inability to use data collected in realistic settings and to gather enough of this data to support necessary decisions.

An alternative — portfolio evaluation — has been developed in detail in a companion paper appearing in the *Journal of the American Dental Association*.<sup>39</sup> It is argued there that state boards of dentistry (not testing agencies) have a responsibility to define the set of competencies required to begin dental practice and the quality (reliability and validity standards) of the data needed to identify candidates who possess those competencies. If existing testing agencies can meet those standards they should provide the required data (and on-going evidence of meeting the psychometric standards). If dental schools can meet these standards, their services should be used. Perhaps a two-stage process would be most effective. **CDA**

**References** / 1. American Education Research Association, American Psychological Association, National Council on Measurement in Education. Standards for educational and psychological testing. Washington: American Psychological Association, 1985.

2. Linn RL, Baker EL, Dunbar SB, Complex, performance-based assessment: expectations and validation criteria. *Educ Researcher* 20(8):5-21, 1991.

3. Kane M, The validity of licensure examinations. *Amer Psychol* 37:911-8, 1982.

4. LaDuca A, Validation of professional licensure examinations: professions theory, test design,

and construct validation. *Eval Health Prof* 17:178-97, 1994.

5. McGaghie W, Professional competence evaluation. *Educ Researcher* 20:3-9, 1991.

6. Sireci SG, Green PC III, Legal and psychometric criteria for evaluating teacher certification tests. *Educ Meas: Issues Prac* 19:22-34, 2000.

7. American Dental Association, 2000 survey of clinical testing agencies. Chicago: American Dental Association; 2002.

8. American Dental Association. Resolution 64H. ADA House of Delegates, October 2000.

9. Berry T, The board examination: a true test or only a rite of passage? *Operative Dent* 20:85, 1995.

10. Buchanan R, Problems related to the use of human subjects in clinical evaluation for follow-up care. *J Dent Educ* 55:797-8, 1991.

11. Chiodo GT, Tolle SW, An ethical perspective on licensure by state board examinations. *Gen Dent* 44:18-25, 1996.

12. Collins W, Conflicting interests of state boards and the public welfare. *J Dent Educ* 49:743-5, 1985.

13. Damiano PC, Shugars DA, Freed JR, Clinical board examinations: variation found in pass rates. *J Am Dent Assoc* 123(6):68-73, 1992.

14. Dugoni AA, Licensure — entry level examinations: strategies for the future. *J Dent Educ* 56:251-3, 1992.

15. Formicola AJ, Shub JL, Murphy FJ, Banning live patients as test subjects on licensure examinations. *J Dent Educ* 66:605-9, 2002.

16. Gaines JH, Licensure. *J Dent Educ* 57:221-2, 1993.

17. Hangorsky U, Clinical competency level of fourth-year dental students as determined by board examiners and faculty members. *J Am Dent Assoc* 102:35-7, 1981.

18. Hasegawa TK, Ethical issues of performing invasive/irreversible dental treatment for purposes of licensure. *J Amer Coll Dent* 69(2):43-46, 2002.

19. Howard W, What do state dental board examinations really accomplish? *J Acad Gen Dent* 39:310, 1991.

20. Jenson LE, Is it ethical to involve patients in state board examinations? *J Amer Coll Dent* 69(2):39-42, 2002.

21. Kalkwarf K, Commentary on Melnick, Patalochi, and Formicola papers. *J Dent Educ* 66:610-1, 2002.

22. Meskin L, Time for a dental board check-up. *J Am Dent Assoc* 125:1418-9, 1994.

23. Meskin L, Dental licensure revisited *J Am Dent Assoc* 127:292-4, 1996.

24. Swankin DA, Assuring the continued competence of health care professionals in the public interest. *J Amer Coll Dent* 64(4):35-8, 1997.

25. Kirkham D, A fresh look at how dental schools prepare dentists for today's practice. *J Amer Coll Dent* 68(3):16-9, 2001.

26. Meeske J, Clinical board examinations: feedback from graduates. *J Dent Educ* 59:351, 1995.

27. Division of Educational Policy and Research, 1999/2000 survey of dental educators. Chicago: American Dental Association; 2002.

28. American Association of Dental Examiners, Composite, Chicago: American Association of Dental Examiners; 1997.

29. (California) Board of Dental Examiners, Dental practice act with rules and regulations. Sacramento, CA: The Board, 1994.

30. Brennan RL, Misconceptions at the intersection of measurement theory and practice. *Educ Meas: Issue Prac* 17:5-9, 30, 1998.

31. Clouser BE, Clyman SG, Swanson DB, Components of rater error in a complex performance assessment. *J Educ Meas* 36:29-45, 1999.

32. Kane M, Validating the performance standards associated with passing scores. *J Educ Res* 64:425-61, 1994.

33. Wang T, Kolen MJ, Harris DJ, Psychometric properties of scale scores and performance as levels for performance assessments using polytomous IRT. *J Educ Meas* 37:141-162, 2000.

34. Crocker L, Algina J, Introduction to classical and modern test theory. Fort Worth, TX: Holt, Rinehart & Winston; 1986.

35. Feldt LS, Brennan RL, Reliability. In Linn RL, ed. Educational measurement. Washington: American Council on Education and the National Council on Measurement in Education; 1989.

36. Moss PA, Can there be validity without reliability? *Educ Researcher* 23:5-12, 1994.

37. Stanley JC, Reliability. In Thorndike R, ed. Educational measurement. Washington: American Council on Education; 1971.

38. Joint Commission on National Dental Board Examinations. Technical Report: the National Board Dental Examinations. Chicago, IL: American Dental Association, 2003.

39. Chambers DW, Portfolios for determining initial licensure competency. *J Am Dent Assoc* (Feb. 2004).

40. Chambers, DW, The roles of evidence and the baseline in dental decision making. *J Am Coll Dent* 66(2): 60-8, 1999.

41. Ranney RR, Wood M, Gunsolley JC, Works in progress: a comparison of dental school experiences between passing and failing NERB candidates. *J Dent Educ* 63(3):311-6, 2003.

42. American Dental Association / American Association of Dental Examiners Licensure Committee. Guidelines for valid and reliable dental licensure clinical examinations. Chicago: The Associations; 1992.

43. Smith IL, Hambelton RK, Content validity studies of licensing examinations. *Educ Meas Issues Prac* 9(4):7-10, 1990.

44. Cronbach LJ, Construct validation after 30 years. In Linn RL, ed. Intelligence: measurement, theory and public policy. Urbana, IL: University of Illinois Press; 1989.

45. Cronbach LJ, Essentials of psychological testing (5th ed.). New York, NY: Harper & Row; 1990.

46. Messick S, The interplay of evidence and consequences in the validation of performance assessments. *Educ Researcher* 23(2):13-23, 1994.

47. Ebel RL, Measuring educational achievement. Englewood Cliffs, NJ: Prentice-Hall; 1963.

48. Chambers DW, Loos L, Analyzing the sources of unreliability in fixed prosthodontics mock board examinations. *J Dent Educ* 61:346-53, 1997.

49. Cole JR II, Legal status of dentistry and licensure. *J Amer Coll Dent* 69(2):6-12, 2002.

50. Cartwright CB, NERB's methods in developing and administering examinations and evaluating results. Paper presented at the NERB symposium sponsored by the Ohio Dental Association, Columbus, OH: 1994.

51. Cole JR II, Maitland RI, A response from the American Association of Dental Examiners. *J Amer Coll Dent* 69(2):47-9, 2002.

52. Patalochi RE, Patients on clinical board examinations: an examiner's perspective. *J Dent Educ* 66:600-604, 2002.

53. Yaple N, Metzler J, Wallace W, Results of the Ohio non-patient dental board examinations



for 1990-1991. *J Dent Educ* 56:248-50,21, 1992.

54. Maitland RI, The New York State post-graduate fifth-year dental residency as a new licensure path: concerns for public protection. *J Dent Educ* 63(3):301-10, 2003.

55. Health care criteria for performance excellence. Baldrige National Quality Program. Gaithersburg, Md: 2003.

56. Baxter GP, Glaser R, Investigating the cognitive complexity of science assessments. *Educ Meas Issues Prac* 17:37-45, 1998.

57. Koretz D, Stechner B, Klein S, McCaffrey D, The Vermont portfolio assessment program. *Educ Meas Issues Prac* 13:5-18, 1994.

58. Shavelson RJ, Baxter GP, Gao X, Sampling variability of performance assessments. *J Educ Meas* 30:215-32, 1993.

59. Finkelstein MW, Johnson LA, Lilly GE, Interactive videodisc patient simulation of oral diseases. *J Dent Educ* 52(4):217-20, 1998.

60. Chambers, DW, Toward a competency-based curriculum. *J Dent Educ* 57:790-3, 1993.

61. Chambers DW, Competencies: A new view of becoming a dentist. *J Dent Educ* 58:342-5, 1994.

62. Commission on Dental Education. Accreditation standards for dental education programs. Chicago: American Dental Association; 1998.

63. American Dental Education Association. Competencies for the new dentist. *J Dent Educ* 61(7):556-8, 1997.

64. American Dental Education Association. Competencies for entry into the profession of dental hygiene. *J Dent Educ* 64(7):512-6, 2000.

65. Chambers DW, Gerrow JD, Manual for developing and formatting competency statements. *J Dent Educ* 58:361-6, 1994.

66. Boyd MA, Gerrow JD, Chambers DW, Henderson BJ, Competencies for licensure in Canada. *J Dent Educ* 60:842-6, 1996.

67. Delandshere G and Petrosky AR, Assessment of complex performances: limitations of key assumptions. *Educ Researcher* 27:14-24, 1998.

68. Arter JA, Spandel V, Using portfolios of student work in instruction and assessment. *Educ Meas Issues Prac* 11(1):36-44, 1992.

69. Jensen GM, Saylor C, Portfolios and professional development in the health professions. *Eval Health Prof* 17:344-57, 1994.

70. Moss PA, Beck JS, Matson B, Muchmore J, Steele D, Taylor C, Portfolios, accountability, and an interpretative approach to validity. *Educ Meas Issues Prac* 11:12-21, 1992.

71. Reckase MD, Portfolio assessment: a theoretical estimate of score reliability. *Educ Meas Issues Prac* 14:12-14, 31, 1995.

72. Brennan RL, Elements of generalizability theory. Iowa City, IA: American College Testing Program; 1983.

73. Cronbach LJ, Gleser GC, Nanda H, Rajaratnam N, The dependability of behavioral measurements: theory of generalizability for scores and profiles. New York: John Wiley & Sons; 1972.

**To request a printed copy of this article, please contact / David W. Chambers, EdM, MBA, PhD, UOP School of Dentistry, 2155 Webster St., San Francisco, Calif., 94115.**