An Evaluation of the

Western Regional Examining Board's

Dental Examination Program

This document consists of excerpts from the report of the evaluation completed by Dr. Tom Haladyna in January 2005. The report is 46 pages long and is very detailed. The report is quite positive in confirmation of the use of WREB candidate scores to inform licensing agencies so they may accurately determine that a candidate should or should not receive a dental license. It is our intent that this reduced report will inform interested persons without requiring psychometric knowledge for an interpretation.

Dr. Haladyna is highly qualified to accomplish evaluations of licensure testing programs. He has been a consultant for 13 certification/licensure programs, including the American Dental Association, and for 12 school testing programs. He has been a professor of educational psychology at Arizona State University at the Phoenix west campus since 1989 and was previously the Director of Health Programs at ACT testing in Iowa City. He has authored numerous books and journal articles in the field of testing.

Organizations who desire a complete version of Dr. Haladyna's report should request a copy of the report from the WREB Executive Director.

Table of Contents

BACKGROUND

Examining boards exist to provide one source of information to states who must decide who receives a license to practice a profession. These professions include dentists, dental hygienists, certified public accountants, physicians, teachers, and social workers among many other professions.

The main concern of any examining board is to increase the likelihood that the professionally licensed person will treat the patients safely. The tests sponsored by any examining board must validly identify those who may be a threat to patient safety in their professional practice. No test or battery of tests is adequate for this purpose, and no system of making pass/fail decisions is infallible. Nonetheless, all states and jurisdictions engage in testing to inform decision making about who receives a license to practice a profession. Testing specialists have developed a system of validation supporting this practice. This system begins with a logical argument, a claim for validity, and supporting evidence that using such test scores to make pass/fail decisions affecting licensure are legally defensible and fair. Of course, the test alone does not determine who receives a license, but in most states and jurisdictions, passing a test is one criterion for licensure that all candidates must achieve.

WREB is an organization that conducts clinical examinations in dentistry and dental hygiene. Its corporate office is in Phoenix, Arizona. Its bylaws were amended by the membership (WREB, January 11, 2003). It provides testing information to participating states on clinical performance for candidates for dentist and dental hygienist licenses. This system and organization appear to be working well. WREB is continually introducing improvements. Documents cited in this report and archived in WREB's office give ample testimony to the continual improvement of WREB's examination programs.

Testing experts have recommended that all examining boards undergo an intensive, regular evaluation (Downing & Haladyna, 1997; Madaus, 1992). The main purpose of this evaluation is to make a summative judgment about the quality of examination that is based on validity evidence that has been assembled and documented for this project. WREB has consistently validated its test score uses and improved its dental and dental hygienist examination programs. This evaluation reveals the long-term development and evolution of this organization to provide the most valid use of test scores as is possible given the resources and the challenges it faces.

DESCRIPTION OF THIS EXAMINATION PROGRAM

The *Dental Examination Program* provides test scores to states for use in making licensing decisions for dentists.

Highlights of the *Dental Examination Program*

The examination consists of four parts:

1. Operative–52 points
2. Periodontics–20 points
3. Endodontics–18 points
4. Prosthodontics–10 points

Total examination score is 100 points.

The cut score (passing score) is 75 points.

This 100-point scale is not a raw-score scale. Performance as determined from ratings is transformed into points using conversion charts that WREB has studied and approved.

Information about this examination program can be found in the *Dental Candidate Guide* (WREB, 2004). Another source of information is the WREB web page: http://www.wreb.org/

Information about validity can be obtained from technical reports done for each year. This report is one of the most comprehensive, well-documented sources of validity evidence. Other documents cited in this evaluation also provide validity evidence. WREB has many documents in an archive in its office that also document validity evidence.

The 2003 annual technical report shows the median and mean scores for the years 1997 through 2003. These scores range from 77 to 79 roughly.

The failure rate fluctuates between 9.8 and 16.3.

First time candidates have the highest probability of passing (86.8%). Candidates taking a retest have an increasingly lower probability of passing as a function of the number of times they have to retest.

Examiners seldom deviate more than one point on any rating scale when rating candidate performance. This fact implies that examiners are very consistent in their ratings.

Some comments and observations seem germane to the evaluation of validity. The operative part of the examination has the greatest weight. The *Dental Candidate Guide* is very helpful in providing information about this examination. The *Examiner's Manual* also provides good information about examiner training, calibration, and scoring. The *Policy and Procedures Manual* also provides considerable information about the organization and administration of this examination. Technical reports can be very useful in informing readers about the validity of the examination. The pattern of first- time candidates having the highest pass rate is consistent with most credentialing examinations. The data used in analyses for this report appears adequate for intended purposes. Having this background information is useful for understanding the results of this evaluation report.

VALIDITY

The most important concern in any examination program is ***validity***. In a high-stakes examination program such as this one, according to leading test expert Bob Linn (2004), validity takes on more importance due to fact that a candidate's future as a practicing dentist depends on the outcome of this examination. Further, the test is intended as a gatekeeper, screening out those candidates who are most likely to have a negative influence on public's welfare and safety. Therefore, the focus of this evaluation is validity. All other ideas are subsumed under validity, such as examination content, item quality, reliability, standardized administration, fairness, bias, equity, comparability of scores and scales, and the pass/fail standard, among many other considerations and issues.

Validity applies to a process involving judgment of the reasonableness of an interpretation or use of a test score. What does a test score obtained from WREB's dental clinical examination mean? How valid is it for a state to make a pass/fail decision based on this test score? Thus, validity does not address a test, so the term *test validity* is inappropriate. Validity focuses on the meaningfulness of an interpretation and the reasonableness of its use in making pass/fail decisions.

To argue in favor of the validity of a test score interpretation or use, we need certain components:

1. an argument that lays out what we plan to measure and how the measure will be validly interpreted and used;

2. a claim that the measure is validly interpreted and used;

3. a collection of positive and negative evidence relating to this argument and claim; and

4. a professional judgment that incorporates this argument, claim, and positive and negative evidence into a summary judgment.

For a positive evaluation, the argument has to be sound and compelling, the claim just, and the preponderance of evidence in favor of validity. Negative evidence should be inconsequential.

No examination program reaches its ultimate in validity. Validity is a goal. All examination programs undergo transformation in an evolutionary path upwards, but the road is steep and long. This evaluation report presents the argument and claim for validity, and also displays the evidence. Its author evaluated the argument and evidence to make a summative judgment about validity.

Summary

We start with an argument about the validity of using WREB's dental examination scores as a measure of clinical competence. A claim is made by WREB on behalf of its client states that using these test scores in that way is valid. We collect and display evidence both supporting and weakening this claim for validity. We also identify missing evidence. Then a summative judgment is made about the validity of WREB's test score interpretation and use.

Participating states can use this judgment to guide them in deciding if the service they receive is adequate for their needs. All licensing authorities have a responsibility to the public to do this. WREB exists to help these states accomplish this mission.

STANDARDS FOR EDUCATIONAL AND PSYCHOLOGICAL TESTING

The *Standards for Educational and Psychological Testing (*subsequently referred to as the *Standards)* was published in 1999 by the American Educational Research Association (AERA), the American Psychological Association (APA) and the National Council on Measurement in Education (NCME). A large, representative committee of testing experts and other qualified volunteers participated in developing these guidelines. For the purpose of this evaluation, these guidelines are used and often cited throughout this document. All of the referenced guidelines bear on the overall judgment of validity.

Other Standards and Guidelines

Many concepts, principles, and procedures of test development and validation are used in this evaluation that are not only based on the preceding standards, but also draw from other important sources. A document that relates the *Standards* (AERA, et al., 1999) to dentistry, *Guidance for Clinical Licensure Examinations in Dentistry,* was published by the American Association of Dental Examiners (AADE) (2003). This document reflects many of the standards identified in Table 3 but more directly reflects the nature of clinical testing and the specific types of validity evidence needed to support WREB's claim for validity. These standards were incorporated into the evaluation, although less directly, as reflected in the *Standards*, which is more comprehensive.

LEGAL DEFENSIBILITY

In addition to providing the highest quality examination program possible, WREB has motivation to ensure that it will not be successfully challenged legally. Such challenges are expensive and may lead to loss of credibility that can ultimately weaken and destroy an examination program.

Validation is an effort to provide evidence that supports the examination program and its purpose. By undertaking a validation, WREB provides assurance to its participating states that the test score information can be used validly. Such validation efforts can also be used with various constituencies and the public to ward off threats that arise from litigation. When potential litigants know that validation has been done and the evidence is available, they are disarmed.

VALIDITY EVIDENCE

Introduction

This part contains a body of evidence intended to support WREB's claim for validity for the use of examination scores from the *Dental Examination* for licensing decisions. Toward that end, many references to documents are provided in this section. The importance of these references can be found in the *Standards* (AERA, et al., 1999) in chapter 6. This chapter argues that all validity evidence should be documented. At the end of each category, a brief summary is given and conclusions are drawn about the adequacy of the evidence and the adherence to standards.

The categories are as follows:

1. Content-related validity evidence
2. Items and rating scales
3. Reliability
4. Comparability
5. Standard setting
6. Administration
7. Scoring
8. Reporting
9. Candidate Guide and rights of test takers
10. Security

1. CONTENT-RELATED VALIDITY EVIDENCE

The most fundamental type of validity evidence for a credentialing examination is content. A dental clinical examination should focus on the skills and abilities needed to successfully practice dentistry. We consider this domain of skills and abilities to represent professional clinical competence. This category of validity evidence directly addresses WREB's claim for evidence supporting the validity of using this test as a measure of clinical competence. Thus, it is important for WREB to put most of its effort in validation in the area of content-related validity evidence.

Summary and Conclusion

Content-related validity evidence is the most important category of evidence for a credentialing examination like this one. The results and discussion presented here are by no means comprehensive or exhaustive. Nonetheless, both documented procedures and empirical findings support several conclusions.

1. WREB has formed and supported four committees of content specialists who oversee the refinement of content of this examination and the tasks that comprise each sub-test. These committees' sole purpose is to improve the examination. Evidence presented here shows that this is happening.

2. The four sections of this examination are based on earlier studies and deliberations of four subject-matter committees over many years. WREB's Board of Directors have reviewed and approved recommended changes to increase the content-related validity for this examination.

3. The four distinct subscales representing the four major content areas of the test are distinct and moderately correlated. The structural integrity of each of these subscales is very strong due to high internal consistency that is due to high inter-examiner consistency.

4, The number of observations and the weighting for each of the four parts of the examination appear to be derived from reasoned analysis by the content committees followed by recommendations to the Examination Review Committee and ultimately approved by the Board of Directors.

WREB's content-related validity evidence is commendable for an examination program that relies centrally on defining and examining professional clinical competence.

2. ITEMS AND RATING SCALES

WREB's *Dental Examination Program* is very complex. The examination requires performance items where the results are evaluated by professionally trained dental examiners using many rating scales. To obtain the kind of reliability required for a high-stakes test where pass/fail decisions are made, test items should be numerous, have moderate difficulty for the population tested, and have high discrimination. Further, these items should meet certain qualifications based on their content and a reasoned analysis of the kinds of tasks in each content area that are linked to clinical competence.

The development of these items has been a continual process for WREB. The current test contains the results of committees' efforts. In general, minutes and other documents show that the items were drafted and reviewed by each committee, and recommendations were offered to the Examination Review Committee and the Board of Directors in turn, which accepted, rejected, or revised each committee's recommendations.

Many important steps go into the development of high-quality test items (Downing, in press, Haladyna, 2004, Welch, in press). Working from test specifications that specify the content to be tested, each committee must determine from their knowledge of the content what tasks best represent clinical competence and how many of these tasks must be included to provide an accurate sample of the skills and abilities to be measured.

Summary and Conclusions

WREB's item development and validation activities have been conducted over a long time under the aegis of four content committees. Minutes and reports provide a good trail of evidence about the process and substance of their achievements. Data presented in this report provide information on item performance.

WREB's item development and validation appear to be very appropriate. Only one standard applies to item development (3.7), and WREB is in compliance with this standard. Standard 3.9 deals with documenting item analysis and using it to improve tests. WREB is in compliance with this standard.

3. RELIABILITY

A primary form of validity evidence is **reliability**. Theoretically, every test score has a random amount of error, which can be large or small, positive or negative. The size and direction of this error are always unknown. However, we can estimate reliability and by that estimate the margin of error that a true score might have. So we have a way to estimate the average degree of error in a set of test scores. By estimating reliability, we can obtain an indication of the degree to which a test score might range randomly due to this error.

Reliability Coefficient

There are many ways to estimate reliability, and some lead to spurious estimates of reliability. The most common method is coefficient alpha, which is actually a measure of internal consistency. Based on the sample of 1,292 candidates, the coefficient provided by Del Hammond using item analysis software is 0.903. Reliability was estimated for this report using different computer software and the result was 0.899. In both instances, the technique was coefficient alpha and the minuscule difference of 0.004 is due to rounding error in the computing formula. Thus, there are two independent sources to suggest that the internal consistency reliability of the total score is very high. Another technique was used that is described in Nunnally and Bernstein (1994) on page 269 that involves the reliability of a linear combination. The resulting coefficient was 0.916. The closeness of these coefficients for estimating reliability suggests that the reliability of the Dental Examination scores is very high. According to Nunnally and Bernstein, the linear combination method is more justifiable then internal consistency. Thus, I am inclined to accept the slightly higher coefficient as the most accurate.

Descriptive Statistics and the Reliability Estimates for Subscores

The table below provides basic descriptive statistics for the examination subscales. Most subscore means are on the high side of the scale, indicating high performance, as is expected from looking at the total score means. Most subscore means exhibit little variability, once again reflecting high, consistent performance of these candidates. The internal consistency reliability estimates for these subscales is very high considering the few numbers of items and the range of the rating scales. This finding indicates that the subscores have fairly high reliability.

Descriptive Statistics for Components of the Examination					
Scale Abbreviation and Name	Abbreviated Item Name	Mean	S. D.	Range of Points	Reliability Estimate (Alpha)
APRP – amalgam prep.	AOUTEX AINTFM AOPEN	4.5 4.7 2.7	0.9 2.7 0.4	0-6 0-6 0-3	0.696
CPRP-composite prep.	COUTEX CINTFM COPEN	4.4 4.6 2.6	1.3 1.3 0.7	0-6 0-6 0-3	0.893
GPRP-gold prep.	GOUTEX GINTFM GOPEN	0.2 0.2 0.1	1.0 1.0 0.6	0-6 0-6 0-3	0.968
AFIN-amalgam finish	AAFM AMARG AFFD	3.0 3.2 2.4	0.5 0.5 0.4	0-4 0-4 0-3	0.794
CFIN-composite finish	CAFM CMARG CFFD	2.8 3.0 2.3	0.9 0.8 0.6	0-4 0-4 0-3	0.928
GFIN-gold finish	GAFM GMARG GFFD	0.2 0.2 0.1	0.7 0.7 0.6	0-4 0-4 0-3	0.968
Deduction Points	DEDPTS	1.4	2.6	0-17	--
Endodontics	AACC ACON PACC PCON	3.2 4.1 3.0 3.8	0.6 1.0 0.8 1.1	0-4 0-5 0-4 0-5	0.680
Prosthodontics	PROSCOR	45.5	5.1	0-58	--
Periodontics	PERASST PERTRT	6.8 11.5	1.3 1.4	0-8 0-12	0.843
Total Points		80.0	8.9	10.9-94.9	0.916

Summary and Conclusion

WREB technical reports have separate sections devoted to reliability and examiner consistency. The large number of observations (84) and high inter-examiner consistency has led to high estimates of reliability from independent methods. Despite the dampening effect of a restriction of range of scores of these candidates, reliability is very high. It is hard to imagine how reliability could be improved in this examination program.

4. COMPARABILITY

This section addresses the important issue of scaling to achieve comparability of results. A standardized performance test should be consistent from site to site and over years that the examination is administered, so that the cut score is also consistently and accurately applied. The 100-point scale should retain the same meaning each time the examination is given as the difficulty of the examination is the same for every administration.

WREB's *Dental Examination* is standardized. The examination items are the same each time the examination is administered. The rating scales are the same. Although examiners who score at each administration may vary, all receive the same training and are calibrated at each examination site. The ratings reported in this evaluation show a high degree of accuracy and consistency.

Evidence has been presented in the annual technical report (WREB, 2003a) showing the consistency of results from 1997 through 2003, which is a sign of effective scaling to achieve comparability. The highly standardized nature of this examination, which is discussed in a subsequent section of this report, the examiner consistency discussed in the reliability section, examiner accuracy discussed in the scoring section of this report, and high reliability all provide sources of evidence supporting the integrity of the score scale used in this examination.

Summary and Conclusion

Scaling for comparability is a difficult challenge in any high-stakes examination, particularly where it is performance-oriented and the judgments are subjective. Evidence has been presented in this section and cited elsewhere attesting to the complex issue of scaling for comparability. The crux of the evidence appears to be located with reliability and examiner consistency and also the extensive training system for examiners. The standardized features of this examination also contribute.

WREB has met standards for scaling for comparability by creating an examination that is standardized in every aspect, by providing uniform effective training to examiners, and by ensuring that testing conditions are standardized each time the examination is administered. The rating scales provide a scale that is consistently used by examiners. The pass/fail point is built into these scales, and the pass/fail decision is accurate across all administrations of the test.

Standards 4.10 to 4.13 address issues that apply to the prosthodontics examination. The report of scale comparability for that examination was not presented in this report. Given that this examination is being revised, future forms of this test should be subject to greater scrutiny for comparability. The other three parts of this examination appear to have a strong basis for comparability given the standardized nature of its single test form.

5. STANDARD SETTING

WREB sets its passing score at 75 and recommends to participating states that its pass/fail recommendations based on performance on the clinical examination be accepted. Whether a state has a passing standard of 70 or 75 does not matter. States set arbitrary cut scores as part of their statutes for credentialing examinations. Testing agencies still have the responsibility of setting a cut score that meets standards and fairly determines who is recommended for a passing or failing decision. Hammond explains this in an article in the Fall 2003 issue of the *WREB Dental Student Newsletter*.

Passing Score Studies

WREB has periodically conducted passing score studies for the written portion of the prosthodontics examination that are consistent with established standards in the testing industry. For instance, for the prosthodontics test, a passing score study was done using the Ebel method (WREB, June 16-17, 2002). The passing standard for the examiner-rated portions of the examination were incorporated into the rating scales when they were developed or revised. As in all other aspects of the testing, the criteria in the rating scales were developed by subject matter experts (with testing specialist consultation) on the subcommittees and were reviewed and approved by the Examination Review Committee and Board of Directors.

Summary and Conclusion

The technology and methodology for setting cut scores on performance tests, particularly when states have a fixed scale pont in mind, is very new and complex. The cut score is mainly based on the examiner judgment as expressed in rating scales. With effective examiner training, a high degree of accuracy and consistency can be achieved which contributes to scoring yielding more validly interpretable scores. Thus, the pass/fail decision is more likely to be accurate under such conditions. Evidence has been presented here and in other sections that address issues like reliability, examiner consistency, examiner accuracy, training of examiners, a calibration of examiners, and the creation of many tasks to be rated by examiners. All of these factors have impact on the validity of this cut score.

Based on the evidence appearing in this report and archived, WREB appears to have substantial support for the validity of making pass/fail decisions using the current cut score. The rationale for the cut score is well made and found in the *Dental Candidate Guide*, the *Examiner Manual*, and any annual technical report. Standards 4.19 and 4.21 are related to WREB's *Dental Examination*. 4.19 asks that the procedures for setting the cut scores are well described and documented. 4.21 states that the manner in which the cut scores were created use the expertise of the content committees. WREB meets these standards.

6. ADMINISTRATION

Given that the *Dental Examination* is standardized, the administration of the test must meet certain conditions to provide an equivalent opportunity for success for all candidates. Also, the content of the test must remain exactly the same each time the test is given. And WREB's cut score must be consistently at 75. The *Dental Candidate Guide* gives a very good account of the many standardized features of this examination. Another important document that provides extensive discussion and information about administration is the *Policy and Procedures Manual* (WREB, 2005b). WREB has a differentiated staff with complementary abilities that work together to achieve a smoothly run examination. The *Dental Examiner Manual* provides much detail to examiners about how the test is administered and scored. This manual is very detailed, and it has evolved over many years. Inspection of this manual reveals many quality control checks in all aspects of the examination.

As documented in its *Dental Candidate Guide*, the *Examiner's Manual,* and in other documents in WREB's archive, WREB addresses many issues of administration that affect validity. These issues include training of administrators of the examination, advance information that is available in the *Dental Candidate Guide*, clarity of directions in this guide, conditions of testing, patient consent forms, avoiding disruptions in the examination process, test security, monitoring candidates during the examination, responding to questions of candidates, administration instructions, and time limits.

Having a differentiated staff with clear functions is an important aspect of administration. As evidenced in the *Policy and Procedures Manual* (WREB, 2005b), WREB has hired and trained staff members who provide valuable service to the administration of the examination. The duties include planning, preparation, administration, and post-test activities. The cycle of activities for each administration is well documented in this manual.

A threat to validity may arise where some test sites are easier or harder than others. Hammond (WREB Fall 2003) discusses this threat and dismisses it with data showing that sites are immaterial as providing an advantage or disadvantage to a candidate. There is no reason or rational hypothesis supporting such a threat. Given the highly standardized nature of this examination, it is unlikely that this threat to validity is real.

Summary and Conclusion

The validity evidence addressing administration is described briefly above and well documented in the above references. These documents are substantial in scope. The *Standards* (AERA et al., 1999) contain 46 specific statements regarding administration. No attempt was made here to assess WREB's meeting these standards. WREB is likely to meet these standards. Via interviews and reviewing the *Dental Candidate Guide*, the *Policy and Procedures Manual* (WREB, 2005b) it is clear that WREB's administration protocols are excellent.

7. SCORING

As this test entails clinical performance by candidates for licensing as dentists, there are many threats to validity that arise from subjective judgments by examiners. Fortunately for WREB, excellent evidence was presented in the reliability section of this report to attest to high inter-judge consistency in ratings. However, validity evidence and threats are numerous. This section addresses many important issues related to scoring. Evidence is presented, and documentation is provided to attest to these types of validity evidence and threats to validity.

Selection of Examiners

Examiners are selected and prepared for training as part of the overall administration process. Examiners should meet qualifications as experts. Credentials, experience, a vita or resume, and other documents should be assembled for each examiner to attest to their expertise. The body of information about examiners along with performance data shows the participating states, other constituencies, and the public that WREB has the highest standard regarding who serves as an examiner.

Training of Examiners

WREB has a training system that has been refined through the years. This training system is well described in the *Dental Examiner Manual* (2005). As specified in the *Dental Examiner Manual*, training is an extended process that begins with a pre-training session followed by a formal training session just before the examination. Examiners are sent a letter and materials for training that include a CD ROM and the *Dental Examiner Manual*. These materials are very detailed and provide all examiners with background information to help them prepare for examining.

Examiners are expected to perform adequately. As stated in the *Dental Examiner Manual* (2005), examiners' performance is evaluated. Examiners get feedback on their performance and the degree to which they vary from their peer examiners. Examiners have to meet an 80% criterion, which means that they can vary from their counterpart examiners less than 20% of the time. If the error rate exceeds 20%, they need to be re-calibrated. WREB has an examiner dismissal clause for examiners (WREB, 2005, p. 8).

Indication of the extensiveness of this training is also found in content committees' reports. For example, the Operative Committee Minutes (June 21, 2003) provides discussion of pre-assessment of examiners and details of changes to improve administration. Such discussions are typical of committee proceedings. All meeting minutes of these committees are archived by WREB, and only a sampling of these is cited in this report to provide background and some documentation.

Scoring

Standard 3.22 addresses procedures for scoring and scoring criteria. Each year, Hammond (December 3, 2004) provides a report to dental examiners regarding their consistency and degree of agreement with other examiners. WREB fully meets this standard and supplies much relevant documentation in the *Dental Candidate Guide* and the *Examiner's Manual.* Also, in the annual technical report (WREB, 2003), examiner accuracy and consistency data are presented. These data have been presented and discussed in the reliability section and were very supportive of reliability.

Examiner Agreement (Consistency)

As reported in the reliability section of this report, high inter-examiner agreement was presented for the many tasks that comprise this examination. Thus, there is no further discussion here except to note that this is a strength in reliability and in scoring. This fact also supports a conclusion about the quality of training of examiners, where feedback to examiners and calibration provide high standards for examining, which appear to be achieved.

Quality Control

One problem that seems to be growing in standardized testing is scoring error. We have witnessed an epidemic of errors in scoring that have large consequences on candidates/students (see www.Fairtest.org). The fact of these numerous incidents reminds us that all testing agencies should have a policy for quality control that has checks and double-checks and verification that scores are accurate. WREB has many quality control procedures, scoring checks, and security measures in test administration procedures and employee job descriptions. Telling the participating states, dental schools, candidates, and the public that this is so is also reassuring.

Summary and Conclusion

Evidence was presented regarding the accuracy of data that is used to form a total score for each candidate. Some of this evidence came from the section on reliability. Data and documentation exist to support the validity of scoring. WREB appears to have a very thorough, refined system of scoring that is clearly stated in the *Examiner Manual* and also in the *Dental Candidate Guide*.

Regarding the standards that apply to scoring, Standard 5.6 addresses integrity that relates to fraud in obtaining a score. WREB has procedures for safeguarding against fraud. There are eight standards regarding examiners. These standards address such issues as selecting examiners, qualifications of examiners, training, recalibration of examiners, feedback to examiners, and dismissal of examiners. Four scoring criteria standards exist. WREB meets these standards and provides good documentation for these standards in the *Examiner Manual,* and the *Dental Candidate Guide*. Scoring is a very complex and important aspect of any testing program, particularly one with high-stakes consequences for candidates, such as this one. WREB stands very well with respect to the many standards that apply to scoring.

8. REPORTING

A candidate score report should be clearly presented and easily interpretable. The score report should help candidates understand the scoring procedure and the meaning of scores on the report that comprise the total score.

Summary and Conclusion

The *Candidate Guide* provides descriptions about scoring. Score reports are designed to reveal candidate performance in all aspects of the examination in a point basis against possible points to be earned. Confidentiality of candidates' results are ensured. Candidates graduating from dental schools have the option of withholding their score report from the dental school.

The *Standards* (AERA et al., 1999) provides more than 40 standards existing for reporting. Many of these standards are not relevant to a credentialing examination. WREB's score reports are clear and insightful. Candidates' rights regarding confidentiality are respected. WREB appears to satisfactorily observe these standards.

9. CANDIDATE GUIDE AND RIGHTS OF TEST TAKERS

As mentioned previously and often in this evaluation, WREB annually publishes a *Dental Candidate Guide*. This document provides extensive guidance to candidates, including general guidelines, infection control guidelines, testing candidates with disabilities, anonymity of candidates taking the examination, scoring information, procedures and criteria for dismissal of candidates, providing examination results, equipment and materials needed for an examination, patient selection, and late penalties. The largest part of this guide provides specific information about the content of the examination, the specific procedures and criteria to be assessed, and the rating scales used. This *Candidate Guide* is also presented in a web page: www.wreb.org. WREB has a well-documented history of its content committees revising this guide to tell candidates about examination changes. The citations offered previously for each content area provide some of the documentation for this history. The rest is contained in WREB's archives.

WREB (Fall 2004a) publishes a newsletter three times a year for the public. This publication contains information about the examination. It is a valuable communication tool for candidates and others, including those from the dental schools. This newsletter is another instance of evidence supporting standards about candidates' rights.

WREB (Fall 2003; Fall 2004b) also provides a publication to student candidates keeping them informed of issues in its examination and provides ethical test preparation advice and other useful information. This newsletter complements the information presented in the *Candidate Guide*. Published once a year, it is an excellent tool for updating candidates on the most important test they will ever take. Also, this publication provides many instances of validity evidence that support making pass/fail decisions using these test scores. Finally, WREB uses an email address (dentalinfo@WREB.org), which provides service to candidates to keep them informed and answer queries. To augment this, phone queries are also encouraged during the business day.

Summary and Conclusion

Information about candidates' rights, which is an important part of any credentialing examination process, has been presented here and also appears in the archive. The *Standards* (AERA et al., 1999) are very clear in chapter 8 about the rights and responsibilities of test takers. Standard 8.1 speaks to keeping candidates informed about the test. Standard 8.2 contains advice about keeping candidates informed about the intricacies of the examination process. WREB meets these standards fully. WREB is commended for its *Dental Candidate Guide*. It is exemplary as a communication tool for candidates, and it also provides a wide variety of well-documented validity evidence that assures the candidates and others about the quality of this testing program. Also, the WREB *Dental School Newsletter* performs an invaluable service in keeping dental students apprized of the examination that so greatly affects their future practice in the dental profession. Moreover, this newsletter provides considerable evidence supporting the validity of the examination.

10 SECURITY

The *Policy and Procedure Manual* (WREB 2005a) discusses security. Threats to validity arising from security breaches are increasing throughout the world. WREB does many things to safeguard against cheating during the examination. This aspect of its security policy is well in place.

WREB Office

As the WREB office is a testing agency, no candidates or any other persons not affiliated with the examination are allowed in the office without recognition or permission. It is customary to admit such persons by signing in, and scrutiny should be maintained so that test materials are not subject to exposure, tampering, or theft. WREB satisfies this requirement.

Computer Security

WREB's computer system has been evaluated by Braincore, a company specializing in computer security. In a letter dated December 21, 2002 in the WREB archive, the results of this network and security audit were presented.

Candidates

With any high-stakes examination, the temptation to cheat is great, particularly in an examination like this one, where the stakes are arguably highest. WREB does an excellent job of ensuring that each candidate is clearly identified and monitored during the examination process. WREB has a photo identification procedure and other safeguards that are stated in the *Dental Candidate Guide*.

Examiners

It is extremely unlikely that a single examiner or team of examiners could undermine the validity of any examination. Examiners are subjected to high standards of performance and scrutiny. Self-interest or other factors may contribute to unwarranted ratings. Although this kind of behavior is unlikely to be considered cheating, it is undesirable and a threat to validity. This problem is unlikely in the WREB environment where examining team assignments are carefully controlled and monitored to minimize this possibility. The integrity of examiners is discussed in the *Examiners Manual*. Conflicts of interest with candidates are monitored, and examiners are asked to recuse themselves if potential conflicts exist.

Written Examination

Written examinations are subject to threats to validity arising from cheating. Examinations can be compromised, disclosed, or tampered with in various ways. An administrator guide to the written examination is one step to increase security. Other measures include checking candidates' identification and counting test booklets when passed out and collected.

Summary and Conclusion

This section has given brief information about security for WREB. The procedures for security that has been established over many years are well documented.

WREB has provided evidence security is not a serious threat to validity. It appears to meet the single standard pertaining to security (standard 5.9). However, the *Standards* (AERA et al., 1999) appears inadequate in this regard.

SUMMATIVE EVALUATION

The argument, claim for validity, and evidence presented in this document, in WREB's technical reports, and other documents, strongly support the validity of using test scores for making pass/fail decisions that affect licensing of dentists in WREB's participating states. WREB is commended for developing an excellent examination program that has many strengths in terms of the categories of validity evidence presented here and no apparent weaknesses.

The greatest strength is the overall commitment to excellence that permeates all aspects of the program. This includes the Board of Directors, the Examination Review Committee, the staff who plans and administers the program and the participation of states, dental schools, and other constituencies that support such testing programs, such as the American Association of Dental Examiners, and the guidelines they recently published with WREB's help and support.

The content committees appear to operate smoothly and effectively. They promote growth and improvement in a very active way, and their minutes and recommendations give ample evidence of this commitment for improvement.

The evidence presented in this document and other evidence that is in WREB's archive is very strongly in support of WREB's participating states using these test scores for making pass/fail decisions for licensure in dentistry.

References

American Association of Dental Examiners (2003). *Guidance for clinical licensure examinations in dentistry*. Chicago: Author.

American Educational Research Association, American Psychological Association. National Council on Measurement in Education. (1999). *Standards for Educational and Psychological Testing*. Washington, DC: American Educational Research Association.

Downing, S.M. (In press). Selected-response item formats in test development. In S.M.

Downing and T.M. Haladyna (Eds). (1997). *Handbook of Test Development*. Mahwah, N.J.: Lawrence Erlbaum Associates.

Haladyna, T. M. (2004). *Developing and validating multiple-choice test items* (3rd ed.). Mahwah, NJ: Lawrence Erlbaum Associates.

Madaus, G. F. (1992). An independent auditing mechanism for testing. *Educational Measurement: Issues and Practice, 11(1)*, 26-31.

Nunnally, J. C., & Bernstein, I. H. (1994). *Psychometric theory* (3rd edition). New York: McGraw-Hill.

Welch, C. (In press). Item/prompt development in performance testing. In S. M. Downing and T. M. Haladyna (Eds). *Handbook of Test Development*. Mahwah, N.J.: Lawrence Erlbaum Associates.

Western Regional Examining Board (June 16-17, 2002). *Prosthodontics Committee Meeting*. Phoenix: Author.

Western Regional Examining Board (2003a) *WREB Annual Examining Report 2003*. Phoenix: Author.

Western Regional Examining Board (2003b). *WREB Dental Student Newsletter*. Phoenix: Author.

Western Regional Examining Board (January 12, 2002). *Western Regional Examining Board By Laws* (As amended by the Membership. Phoenix: Author.

Western Regional Examining Board (June 21, 2003). *Operative Committee Minutes*. Phoenix: Author.

Western Regional Examining Board (2004). *Dental Candidate Guide*. Phoenix: Author.

Western Regional Examining Board (2005a). *Dental Examiner Manual*. Phoenix: Author.

Western Regional Examining Board (2005b). *Policy and Procedures Manual*. Phoenix: Author.