

An Evaluation of the  
Western Regional Examining Board's  
Dental Hygiene Examination Program

Dr. Thomas M. Haladyna  
Professor of Educational Psychology  
College of Teacher Education and Leadership  
Arizona State University

July 12, 2005

## Table of Contents

Part I: Why Is WREB’s Dental Hygienist Examination Program Being Evaluated? .....	1
Part II: Description of this Examination Program .....	3
Part III: Validity .....	5
Part IV: <i>Standards for Educational and Psychological Testing</i> .....	9
Part V: Legal Defensibility .....	12
Part VI: Validity Evidence .....	13
Content-related validity evidence .....	14
Items and rating scales .....	17
Reliability .....	19
Comparability .....	22
Standard setting .....	23
Administration .....	24
Scoring .....	26
Reporting .....	30
<i>Candidate Guide</i> and rights of candidates .....	31
Security .....	32
Part VII: Summative Evaluation .....	34
References .....	35

## **Part I: Why Is WREB's *Dental Hygienist Examination Program* Being Evaluated?**

Examining boards provide important information to states, who must decide who receives a license to practice a profession in that state's jurisdiction. These professions include dentistry, dental hygiene, accountancy, medicine, education, social work, law, and law enforcement among many others.

The main concern of any examining board is to increase the likelihood that the professionally licensed person will treat the public that they serve safely. The examination sponsored by any examining board must validly identify those who may be a threat to patient safety in their professional practice. No examination or battery of tests is infallible in this regard. Nonetheless, all states and jurisdictions engage in licensing 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 valid. Of course, the examination alone does not determine who receives a license, but in most states and jurisdictions, passing an examination is one important criterion for licensure that all candidates must achieve if they are to be allowed to practice.

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 member states on clinical performance for candidates for dentist and dental hygienist licensure. WREB also provides information on test performance for non member states that use this information in licensing decisions.

WREB has a Board of Directors (also known as the *governing board*) and a Examination Review Committee that oversee the Dental Hygiene Examination Program. WREB has a Dental Hygiene Subcommittee that meets regularly, reviews policies and procedures, and recommends changes intended to improve the examination program (WREB, July 18-19, 1998; July 10-11, 1999; October 15-17, 1999; June 30-July 2, 2000; November 3-5, 2000; July 6-8, 2001; September 28-30, 2001; July 5-7, 2002; September 7-8, 2002; July 7-8, 2003; September 26-28, 2003; July 9-11, 2004; September 24-26, 2004). These documents from WREB's archive provide substantial evidence that WREB has an annual review process that seeks to improve this testing program. The recommendations of this subcommittee go to the Examination Review Committee, which, in turn, recommends changes to the Board of Directors.

WREB's Examination Review Committee also issues annual reports to the Board of Directors (WREB, July 2000; July 7-8, 2002; July 5-7, 2002; July 14, 2004). Although these reports are brief, they contain substantive recommendations that are intended to improve the overall quality of the examination program.

Testing experts have recommended that all examining boards undergo an extensive,

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 a validity argument, a claim for validity, and supporting validity evidence that has been assembled and documented. A similar evaluation was done in 1998 for the Dental Hygiene Examination Program (Haladyna, May, 4, 1998). Thus, the current report incorporates changes in policy, procedures, and documents since the last review. Also, the previous report identified areas of concern.

WREB has consistently validated its test score uses and improved its dental and dental hygienist examination programs. Reports like this one and annual reports and other documents in WREB's archive provide substantial documentation of this continued effort to provide a high quality, validated examination program.

## Part II: Description of the *Dental Hygiene Examination Program*

The Dental Hygiene Examination Program provides test scores to states for use in making licensing decisions for dental hygienists. Table 1 provides highlights of this program. More detailed descriptions of these highlights appear in the *Dental Hygiene Examination 2005 Candidate Guide* (WREB, 2005a).

Table 1: Highlights of the Dental Hygiene Examination Program
The examination consists of three parts: 1. Probe Depths/Recession–15 points 2. Extra/Intraoral Exam–10 points 3. Calculus Removal and Tissue Trauma–75 points Total examination score is 100 points.
Possible Point Deductions x-ray penalty–4 points First patient unacceptable–4 points Second patient unacceptable–3 points No acceptable patient–failure Late penalties–1 point per minute that patient is late for checkout.
Cut (passing) score is 75.
Information about validity can be obtained from annual technical reports and other documents in the archive. This report provides references to many documents in WREB’s archive.
The 2003 annual technical report shows the median and mean total scores for the years 1997 through 2003. The median scores range between 88 % and 92 %. The fail rates range between 8.3% and 14.3%.
The number of candidates ranged between 1,120 and 1,278 per year.
Examiners receive training in the examination process and are validated using a performance test. These examiners seldom deviate more than one point on any rating scale when rating candidate performance.
Harshness and leniency in ratings of these examiners were very low. Data bearing on this threat to validity is presented in this report.
Information about this examination program can be found in the <i>Dental Hygiene Examination 2005 Candidate Guide</i> (WREB, 2005a). Another source of information is the WREB web page: <a href="http://www.wreb.org/">http://www.wreb.org/</a>

From Table 1, some comments and observations seem germane to the evaluation of validity. The *Dental Hygiene Examination 2005 Candidate Guide* (WREB, 2005a) was very helpful in providing information about this examination. *WREB Policy Guide* (WREB, 2005c) provides policy information about three independent examinations, including the Dental Hygiene Examination Program. Annual technical reports provide information about important validity evidence that includes content-related, item quality, reliability, examiner training, and examiner

consistency and bias. 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.

### Part III: 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 Robert Linn (2000), validity takes on more importance that requires that more attention be paid to validation. With any professional examination, such as this one, a dental hygienist's future depends on the outcome of this examination. Further, the examination 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 the Dental Hygiene Examination Program mean? How valid is it for a state to make a pass/fail decision based on this test score? 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, certain components are needed:

1. an argument that lays out what WREB plans 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 evidence relating to this argument and claim; and
4. a professional judgment that incorporates this argument, claim, and 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 the claim. Negative evidence should be inconsequential. Negative validity evidence usually leads to recommendations that eliminate or reduce any threats to validity.

No examination program reaches its ultimate in validity. Validation is an ongoing process to achieve perfection. All examination programs undergo transformation in an evolutionary path upwards toward a higher degree of validity, but the road is steep and long. This evaluation report presents the argument and claim for validity, and displays the evidence. Its author evaluated the argument and evidence to make a summative judgment about validity.

Table 2 shows the constituent elements in validation, which is the process of obtaining evidence supporting the claim about validity. This table also shows the reasoning process used in this validation.

Table 2: Validation of WREB's Dental Hygiene Examination	
Argument	The WREB <i>Dental Hygiene Examination</i> is a performance test that is intended to directly measure clinical competence of dental hygienists. WREB's clinical performance test is the capstone in the licensing process for dental hygienists. This performance test is believed to capture the essential skills and combination of skills that represent dental hygiene competence.
Claim About Validity	WREB claims that test scores obtained from WREB examinations represent dental hygienist clinical competence and using the scores for licensing decisions is highly valid. The resulting score can be used with a high degree of confidence by participating states, along with other criteria, to make licensing decisions.
Evidence Supporting the Argument	This evaluation report provides validity evidence. This evidence is linked to national testing standards. WREB's technical reports and other documents cited in this report offer validity evidence supporting this argument.
Evidence Weakening the Argument	If justified, this report might contain negative validity evidence. In fact, no negative evidence was found. Nonetheless, WREB should endeavor to consider threats to validity and take appropriate action to diminish the threat and, by that, strengthen the evidence supporting the argument and by that strengthen the claim for validity.
Lack of Evidence	An evaluation report of this type should look for evidence of all relevant types. One finding in this report is that there are no gaps in the validity evidence sought.
Summative Judgment	This evaluator considers argument for the claim, evidence supporting and weakening the argument about the validity of WREB scores as (1) a measure of professional clinical competence, and (2) for use by participating states in making pass/fail decisions.



## **Validity Evidence Used in This Evaluation**

Part IV of this report provides details about the kinds of validity evidence used in this report. The sources of evidence are many and come in different forms. Recommended procedures, documentation, and empirical results including statistics comprise the majority of the evidence found in this report. We should not consider this validation process as a type of check listing of evidence but instead consider the evidence as an accumulation that supports a judgment about this claim and the related argument. This evidence is used in the same manner that a jury weighs evidence and makes a decision that supports either the prosecutor's or the defense's claim about some litigation.

## **Evidence Weakening the Argument**

In any evaluation, honest assessment of evidence that undermines validity is seldom done by test sponsors (Cronbach, 1971). Two kinds of evidence that weaken validity are construct under-representation (CUR) and construct-irrelevant variance (CIV). The construct is another name for the domain of knowledge, skills, and abilities that comprise dental hygiene competence. This part of the evaluation seeks to uncover evidence that may work against validity. Naturally, WREB and its client states do not want such evidence to be strong, but its detection and eventual treatment are important steps in strengthening the overall validity argument and related claim. Every examination program is only as strong as its weakest link.

CUR is present if the definition of the construct (dental hygiene clinical competence) is not synchronous with what the actual test measures. If we used a multiple-choice test of scientific knowledge or a multiple-choice test of professional knowledge, we would not be representing dental hygiene clinical competence adequately. Thus, participating states see the value of using the Dental Hygiene Examination.

## **Missing Evidence**

An outcome of the current evaluation is that there is no missing evidence. Future evaluations should endeavor to find all relevant evidence and report missing evidence and recommend that such evidence be gathered for validation and future evaluations of validity.

## **Summary**

This section on validity is best summarized in Table 2. It shows that we start with an argument about the validity of using WREB's Dental Hygiene 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.

## **Part IV: Standards for Educational and Psychological Testing**

The *Standards for Educational and Psychological Testing* (heretofore 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.

Table 3 on the next page summarizes some of the more important guidelines that will be cited and used in this document. Of the many categories that appear in that table and throughout this report, several notable omissions exist that deserve special treatment here.

**Chapter 6: Documentation.** This evaluation report contains *all* documentation made available by WREB that is relevant to the validity claim stated in this evaluation. This chapter has many categories of validity evidence. The annual technical report is one source of documentation. This report is another. WREB keeps an archive of other documents that bear on validity. Chapter 6 of the *Standards* should be used as a guide for documenting its validity evidence. This documentation should be viewed as a kind of insurance that can be used to defend against criticism, legal challenges, and inquiries about the quality of WREB's examinations.

**Chapter 7: Fairness.** As this test is used in licensing dental hygienists, the issue of fairness is an important one. The design and administration of the *Dental Hygiene Examination* do not in any way violate any standard of fairness discussed in chapter 7. Examiners have no contact with candidates, and only see their patients. As this test is based on performance and is purporting to measure professional competency, there is no threat caused due to the gender, ethnicity, race, disability or other characteristics of candidates. Standard 7.12 is the most general of these and requires that all candidates be treated fairly and equitably in the examination process. Evidence presented throughout this report bears on the judgment of fairness of WREB's Dental Hygiene Examination.

**Chapter 9: Linguistic background.** As this performance test involves patient treatment under simulated natural conditions involving patient-dental hygienist, no threat due to inadequate linguistic background is perceived. These candidates should also be treated fairly. Noting the fail rates of this population and exploring factors that may contribute to their failing status is a good thing. All test sponsors should always be alert to any threat arising from a lack of understanding of the recommended procedures for this examination or other factors that may jeopardize a candidate whose primary language is not English.

Table 3: Categories of Standards Used in This Evaluation	
Chapter 1: Validity. This chapter identifies fundamental concepts and types of validity evidence that appear throughout this evaluation report.	1.1, 1.2, 1.5, 1.6, 1.7, 1.11, 1.12, 1.15,
Chapter 2: Reliability. As a primary type of validity evidence, evidence is sought	2.1, 2.2, 2.10, 2.13, 2.14, 14.15
Chapter 3: Test Development. Performance testing is recognized as having special challenges in validation.	3.1, 3.2, 3.3, 3.4, 3.5, 3.6, 3.11, 3.13, 3.14, 3.15, 3.17, 3.19, 3.22, 3.23, 3.24
Chapter 4: Scales, Norms, and Score Comparability including Standard Setting.	4.1, 4.2, 4.9, 4.10, 4.19, 4.21, 14.16, 14.17
Chapter 5: Test Administration, Scoring and Reporting	5.1, 5.2, 5.3, 5.4, 5.5, 5.6, 5.8, 5.9, 5.10, 5.13, 5.15 , 5.16
Chapter 8: The Rights and Responsibilities of Test Takers	8.1, 8.2, 8.7, 8.11
Chapter 14.8: Testing in Employment and Credentialing	14.8, 14.9, 14.10, 14.11, 14.13, 14.14,

**Chapter 10: Testing individuals with disabilities.** A key issue with WREB's candidates is that each person is individually assessed with regard to disability and then any accommodation in the administration of the test is done in a way that does not alter the competence being measured. The *WREB Policy Guide for dental hygiene, anesthesia and restorative examinations* has a section bearing on examining individuals with disabilities.

**Chapter 11. The responsibilities of test users.** This category of standards applies to WREB's participating states who use test information. In general, states should have access to all information bearing on the validity of using test scorers for making pass/fail decisions. This is a state's responsibility; it is not WREB's responsibility. However, WREB should provide all participating states information that supports the validity of participating states' uses of test scores. WREB's 25-page *Dental Hygiene 2005 Candidate Guide* (WREB, 2005a) is published every year. WREB's technical reports are also good sources of information. This evaluation report constitutes another source of information that should be available to states. WREB also makes available much information on its web site: [www.wreb.org](http://www.wreb.org), and provides email addresses for individual queries: [hygieneinfo@wreb.org](mailto:hygieneinfo@wreb.org) and [generalinfo@wreb.org](mailto:generalinfo@wreb.org).

## **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. *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. Testing agencies like WREB are urged to follow these guidelines.

## **Part V: Legal Defensibility**

WREB has taken steps to discourage legal challenges by conducting periodic external evaluation of its examination programs and regularly validating its examination programs. Legal challenges are expensive to defend and may lead to loss of credibility of the examination program that can ultimately weaken and destroy it. Validation provides evidence that supports the examination program and its purpose. By undertaking validation, WREB provides assurance to its participating states that the test score information can be used validly. Validation can also be used with various constituencies and the public to repel litigation. When potential litigants know that validation has been done and the evidence is available, they are disarmed. By engaging in this evaluation where validity evidence is collected and organized, WREB very effectively reduces the threat of legal action. In all circumstances, any examining board should have continued legal counsel that examines threats that arise from legal actions and its position in thwarting these threats. Mehrens and Popham (1992) discussed legal threats and validity. In most instances, an active program of validation is the best defense against legal challenge.

## PART VI: VALIDITY EVIDENCE

### Introduction

Part VI of this report is very extensive. This part contains an organized body of evidence intended to support WREB's claim for validity for the use of examination scores from the Dental Hygiene 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. As noted previously in this report, this chapter argues that all validity evidence should be documented. In this part of the report, each category of validity evidence is presented. 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. Item quality
3. Reliability
4. Comparability
5. Standard setting
6. Administration
7. Scoring
8. Reporting
9. Candidate Guide and rights of test takers
10. Security

## **Content-related Validity Evidence**

A domain of skills and abilities comprise what is known as professional clinical competence in dental hygiene. Therefore, a dental hygienist clinical examination should focus on how well candidates perform in this domain of skills and abilities. The performance of each candidate is compared to the passing standard (cut score) to make a decision of the candidate's competence to successfully practice dental hygiene.

The most fundamental type of validity evidence for a credentialing examination is content. Content-related 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. A good source of guidance in this area is a recent chapter by Raymond and Neustel (in press). A thorough and competent analysis of content is primary validity evidence. The procedure for identifying this evidence should be defensible. The focus of content-related validity evidence as discussed in the *Standards* (AERA, et al., 1999, p. 156) can be summarized in this way:

*Panels of respected experts in the field often work in collaboration with qualified specialists in testing to define test specifications, including knowledge and skills needed for safe, effective performance, and an appropriate way of assessing that performance* (AERA, et al., 1999, p. 156).

Chapter 14 of the *Standards* (AERA, et al., 1999) is devoted exclusively to standards affecting licensure examinations, such as WREB's. As stated in that source on page 157 and in this report, content-related validity evidence is the most important. Not only is a testing agency like WREB expected to define clinical competence, but is also expected to show the constituent parts of competency as determined from a scientific study. Further, WREB is expected to show that its test specifications and clinical competency examination are based on the analysis of results. Standards 14.8, 14.9, 14.10, 14.11, and 14.14 all address slightly different but complementary aspects of practice analysis as leading to the creation of test specifications.

### **Practice Analysis**

A practice survey was conducted (WREB, September 3, 1996). The result of this survey was used to identify the categories of most frequent use and most importance as judged by a survey of dental hygienists. The ten WREB member states were represented, with the number of dental hygienists ranging from 29 to 41. The sample of 381 respondents included those with bachelor's and master's degrees. The majority (216) has a certificate in dental hygiene or an associate degree. This sample included respondents with varying degrees of experience. The survey contained 65 items. This study led to the examination specifications discussed in the next section (WREB, undated).



## Examination Specifications

WREB's examination specifications contain two components. Component A addresses patient assessment in three areas: (1) patient qualification, (2) extra/intraoral examination, and (3) periodontal measurements. Component B addresses patient treatment in two areas: (1) calculus removal and (2) tissue management. The examination specifications are complementary to the test specifications produced and published by the Joint Commission on National Dental Examinations. *The National Board Dental Hygiene Examination* is a comprehensive test consisting of 350 items that measure knowledge and skills that are fundamental for assessing the competence of a dental hygienist. WREB's clinical performance examination is intended to help assess the clinical proficiency of a candidate for licensure. The two tests are complementary in what they measure, and both are important in aiding the licensing decision.

WREB reviewed these examination specifications (WREB, October 15-17, 1999). Embedded in this review was a point deduction that is part of the standard setting process for this examination. These examination specifications were adopted. WREB reviewed these specifications in 2001 and considered a conjunctive standard setting strategy, which is discussed elsewhere in this report (WREB, 2001). The decision was made not to implement a conjunctive strategy due to logistical issues and program and substantive validity issues, including lower reliability for any subtest that would be used for a pass/fail decision. WREB (June 2004) issued the current version of the examination specifications. That document shows the point allocations and the deductions for unacceptable patient, radiography penalty, and late checkout. This is found in the *Dental Hygiene Examination 2005 Candidate Guide* (WREB 2005a).

## Structural Evidence

***Probe depths and recession.*** No analysis of structure was done due to the unique nature of these data. Each candidate's patient undergoes 72 separate evaluations. The number of errors detected is very small. Thus, an analysis of structure using factor analysis would be uninformative.

***Extra/intraoral examination.*** An analysis of the structure of these data using principal components factor analysis with equamax rotation revealed that the nine dimensions of this examination were consistent with the intended interpretation of nine distinct factors. This finding is augmented by high intra-factor examiner consistency (reported in the reliability section of this report). Correlations among these nine factors were very low (ranging from 0.001 to 0.224). One interpretation of these results is that these factors are independent events. Another interpretation is that these correlations are likely to be attenuated due to the fact that the range of performance is very restricted.

***Calculus removal.*** An analysis of the structure of these data again using principal components factor analysis with equamax rotation reveal 12 unique surface removal sets of ratings that are essentially uncorrelated. Intersurface correlation for combined ratings of three examiners were low. The correlations ranged from 0.043 to 0.308.

### **Summary, Conclusion, and Recommendations**

The findings show that most observations within each part of the examination have a strong internal structure and little association. These appear to be independent observations. The high performance levels observed would indicate a high degree of competence and little opportunity for these items to discriminate. If the candidate pool included less competent candidates, stronger evidence would be present showing discrimination.

The data and arguments presented in this section show that WREB follows guidelines from the *Standards* regarding content-related validity evidence.

WREB periodically updates its practice analysis for the content of this examination by convening a Dental Hygiene Examination Development Committee that reviews the examination. Changes in dental hygiene practice and any perceived inadequacies are addressed in recommendations to the Dental Hygiene Examination Review Committee and the WREB Board of Directors (WREB, June, 2005).

## **Items and Rating Scales**

### **Peridodontal Measurements and Probing/Gingival Recession**

In a conventional analysis of test item performance, difficulty and discrimination are often computed for each scorable unit (test item). In this performance test, candidates make observations as described on page 12 of the *Dental Hygiene Examination 2005 Candidate Guide* (WREB, 2005a). The test items, in this instance, the candidate's skill for observing probe depths and recession depths. Validity evidence for this scorable part of the test at the item (observation) level is based on the judgment of examiners that the quadrant selected for these observations is representative of the entire dentition.

As reported elsewhere in this report, for probe evaluation, there were 72 evaluations per patient for a total of 84,168 evaluations. Of these, errors were detected in only 1,410 examiner's observations, and 495 were validated. Also reported elsewhere in this report, in recession evaluation, there are 36 evaluations per patient making a total of 42,084 individual evaluations. A total of 2,225 examiner observations detected errors with 1,760 validated. The respective error rates among examiners for these tasks is very small relative to the large number of evaluations made ( 1.1% and 1.1%). In other words, these are very small error rates. Validated error detection is a goal in this kind of analysis. In general, the performance of candidates is very high, and the few unvalidated errors appear immaterial considering the high percentage of correct evaluations across the observation opportunities.

### **Extra/intraoral**

These nine categories were identified in the practice analysis as essential aspects of dental hygiene competence. The choice of these nine categories is based on the practice analysis reported elsewhere in this evaluation and is also based on the recommendations of the Dental Hygiene Examination Subcommittee (WREB, December 18-19, 1998).

Correlations of the total score for each of these categories with the total extra/intraoral score were consistently high. These coefficients ranged from 0.237 to 0.734 with a median value of 0.363. In the traditional sense of item analysis, these coefficients represent discrimination indexes and have a direct relation to the reliability estimate reported elsewhere. Despite the high degree of skewness in these scores, these coefficients are very high, which is also reflected in the high reliability estimate reported in the next section.

### **Calculus Removal**

The 12 surfaces observed for calculus removal and tissue trauma all had high, uniform correlations with the criterion total score. These coefficients ranged from 0.385 to 0.522, and like the previous part of the examination had high reliability largely due to these surface discrimination indexes. The judged quality of these surfaces as observation opportunities for

scoring calculus removal and tissue trauma is offered by the Dental Hygiene Examination subcommittee (WREB, July 2000; July 18-19, 2001; July 5-7, 2002).

### **Summary, Conclusions, and Recommendations**

The data presented here pertained to statistical performance of items used to create scores for extra/intraoral examination and for calculus removal. The choices of these items and the underlying rationale are based upon the *Dental hygiene practice survey*, WREB (September 3, 1996) and many discussions by subject matter experts in committees and board meetings referenced at the end of this report. The documentation of the rationales for the observations used to score is a very important source of evidence for item quality. There is strong evidence for consistency in evaluating performance and the documentation of item development and of the rationale for item selection is extensive.

## Reliability

A primary form of validity evidence is **reliability**. Examining boards, like WREB, should give considerable attention to reliability to ensure its member states that the candidates' scores released are as accurate as possible.

Theoretically, every test score has random error, which can be large or small, positive or negative. The size and sign 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. By estimating reliability, we can obtain an indication of the degree to which a test score might range randomly due to this error. The indicator of this degree of error is the **standard error of measurement (SEM)**. This standard error should be small relative to the distribution of scores found at the cut score (75 for this examination).

Because reliability is a primary type of validity evidence, the *Standards* has an entire chapter devoted to the topic of reliability. These standards are 2.1, 2.2, 2.10, 2.13, 2.14, 14.15. The first, 2.1 calls for estimates of reliability and standard errors of measurement. Standard 2.10 calls attention to the importance of examiners and the consistency of their ratings. 2.13 and 2.14 refer to special conditions and the adequacy of reliability in these conditions. Standard 14.15 suggests that reliability be applied to the cut score and the risk that a candidate may be misclassified incorrectly as a pass or fail due to random error.

Reliability of performance tests is seldom high, due to many factors. One of these factors is the use of too few items. Another factor is lack of high rater consistency. Another factor is restriction in the range of performance of candidates. Licensing boards tend to examine high-performing candidates, thus have this difficulty in properly estimating reliability. Another factor is that the skill being evaluated may not be highly correlated with other skills.

WREB engages in studies of examiner consistency and reliability for its annual examination. The results of these studies are summarized in its annual reports. For this report, statistical studies of examiner consistency and reliability are reported here.

### Reliability Coefficient and the Standard Error of Measurement

Table 4 summarizes the reliability coefficients for each of the three scales. Note that each coefficient is attenuated by the fact that performances on these scales is mostly high, and this restriction in range tends to produce lower estimates of reliability than desired. If the level of competency were broader in this sample, reliability estimates would be much higher.

**Probe depths/recession.** The reliability estimate for this part cannot be estimated due to the unique scoring formula for this part of the examination. We know that the error rate among examiners is very small, and we know the mean of the distribution of scores for this part of the examination is very high. From these data, we can predict that the reliability might be very high.

The scoring formula might be changed to accommodate the estimation of reliability for this part of the examination, but the derivation of this formula based on subject matter expertise is the overriding factor. Thus, no change in the scoring formula is recommended.

**Extra-/intraoral examination.** The reliability estimate for this part of the examination is very high, and the standard error of measurement is very small. Despite the restriction in the range of competence of these candidates, high inter-rater consistency contributed importantly to this very positive result.

**Calculus removal and tissue trauma.** As with the previous part of the examination, reliability is very high considering the skewness in these scores. The standard error is very small relative to the high number of points possible on this examination (75).

**Reliability of the Total Score.** Because the data for each part of the examination is negatively skewed and because the three parts are essentially uncorrelated, the internal consistency type reliability estimate (coefficient alpha) is inappropriate. An appropriate technique is to create three test scores, each based on one-third of the examination and examine the internal consistency of these three part scores. However, as noted in the scoring section of this report, examiners don't provide scores for candidates. Scores are a combination of deductions and performance where, in some instances, examiners validate an observation. Thus, a candidate point score might be the product of a pair of examiners' observations. If the scoring were modified for the probe depths/recession, reliability could be estimated. However, such action is not recommended due to the fact that subject-matter experts chose the scoring method to fairly measure the skills assessed here.

Table 4 Reliability Estimate of Scores of Each Part of the Examination and Their Associated Standard Errors of Measurement			
Scale	Number of Points	Reliability Estimate	Standard Error of Measurement
Probe Depths/ Recession	15	Not estimated	Not estimated
Extra-/Intraoral Examination	10	0.891	0.48
Calculus Removal & Tissue Trauma	75	0.821	1.13
Total Score	100	Not estimated	Not estimated

## **Inter-examiner Consistency**

**Probe depths/recession.** On the probe evaluation, there were 84,168 observations or 72 per candidate. Of this number, 1,140 (1.7%) noted possible candidate errors, and 496 (0.6%) were validated. Thus, examiner error rate was 0.8%, which is very small. On the recession examination, each examiner codes whether an error occurred. For an error to be validated, one of the other two examiners must agree. Of the 42,048 recession observations, 2,225 noted errors and 1760 (79.1%) were validated. Of these 465 were not validated. Thus, the examiner error rate, which is a form of inconsistency, is very small, 1.1%. As most of the recessions observations amounted to no candidate error, about 96%, the detection of candidate error occurs only 4.2% of the time, and in these instances, 79.1% of these are validated.

**Extra-/intraoral examination.** Total scores for this examination were computed for each examiner, and correlations were computed among the three examiners. The correlations were very high: 0.749, 0.725, and 0.723 for 1,165 cases. These correlations are impressively high when the fact that the data is negatively skewed is considered. Skewness tends to attenuate correlation, but in this case, the high degree of interrater consistency was sustained. In addition to calculating overall consistency using the total score, examiner consistency was computed for each item in this part of the examination.

**Calculus removal and tissue trauma.** Observations for each of the three examiners were correlated with the ratings by other examiners. The three correlation coefficients were: 0.622, 0.608, and 0.587. These coefficients are moderate, but like reliability, there is a tendency for these coefficients to be attenuated due to the fact that most scores tend to reach the ceiling of the scale.

## **Summary, Conclusion, and Recommendations**

High inter-rater consistency contributes importantly to keeping reliability of scores high and the standard error of measurement small. Performance tests that are subjectively scored by trained examiners tend to yield low estimates of reliability. With relatively high reliability, WREB's Dental Hygiene Examination is an exception. WREB should continue to monitor reliability and examiner consistency on an annual basis and take any actions that may improve rater consistency that will improve reliability.

WREB is commended for an examiner training program that produces high degrees of inter-rater consistency. This is particularly impressive given the fact that candidates tend to score high on these parts of the examination making the discrimination of differential performance very challenging.

## **Scaling & 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. That is, the difficulty of the examination should be the same for every administration.

WREB's *Dental Hygiene Examination* is standardized in its origin, scale for interpreting results, administration and scoring. The examination items are the same each time the examination is administered. The rating scales are the same. Although examiners at each administration may vary, all receive the same training and are also calibrated before each examination. The ratings reported in this evaluation show a high degree of examiner accuracy and consistency.

### **Threats to Validity**

Major threats to validity for this examination are examiner consistency and bias in ratings, which is discussed elsewhere in this report. As rater consistency is very high, and WREB's training combats bias in examiner ratings, these threats are not material. Although such threats are omnipresent, no evidence exists thus far that suggests that these threats to validity are substantive.

### **Summary and Conclusion**

Given the highly standardized procedures followed in the design, administration, and scoring of the examination, there is no evidence to suggest problems with scaling. In fact, the evidence supports a conclusion that the examination provides an equivalent experience each time it is administered, and that differences in performance from site to site are a function of the candidates taking the examination.



## **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 (WREB, July 17-18, 2002; July 9, 2003; July 10, 2003; Fall 2003). Whether a state has a passing standard of 70 or 75 that is set by legal statute does not matter. States normally 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. In one subcommittee report (WREB, September 26-28, 2003), the validity of rescaling to achieve agreement with each state's statutes regarding the passing score was discussed and resolved.

### **Passing Score Studies**

WREB has periodically conducted passing score studies. The first of these was done for the 1997 dental hygiene written examination. For the extra/intra oral calibration exercise, (WREB, October 29, 2004) Dixon reported a procedure used by her subcommittee to recommend a standard. WREB provides extensive documentation for issues related to setting the cut score (WREB, July 7-8, 2003).

### **Conjunctive Versus Compensatory Standard-Setting Strategies**

As discussed on page 14 of this report, a conjunctive standard is desirable, but many logistical and validity issues exist that argue against using a conjunctive standard. Thus, WREB continues to use a compensatory strategy for pass/fail decisions.

### **Summary and Conclusion**

Documentation of procedures used to set the standard was provided. WREB appears to be in conformance with guidelines regarding the setting of a cut score and its documentation.

## Administration

The administration of the Dental Hygiene Examination is standardized. That means that certain conditions must be met that provide an equivalent examination experience to 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 Hygiene Examination 2005 Candidate Guide* (WREB, 2005a) gives a very good account of the many standardized features of the administration of this examination.

Another important document that provides extensive discussion and information about administration is the *Dental Hygiene 2005 Examiner Manual* (WREB, 2005b). WREB has a differentiated staff with complementary abilities that work together to achieve a smoothly run examination. The *Dental Hygiene 2005 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 Hygiene Examination 2005 Candidate Guide* (WREB 2005a), the *Dental Hygiene 2005 Examiner Manual* (WREB, 2005b), 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 Hygiene Examination 2005 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 *Dental Hygiene 2005 Examiner 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.

Documentation exists for annual attention to issues affecting the administration of this examination (e.g., WREB, July 7-8, 2003; July 14, 2004).

## **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 Hygiene Examination 2005 Candidate Guide* (WREB, 2005a), the *WREB Policy Guide* (WREB, 2005c) it is clear that WREB's administration protocols are excellent.

## Scoring

The scoring of the dental hygiene examination is very complex. The three parts are added to obtain a total score, and each part is weighted according to the judgments of the subject matter experts who serve on the advisory committee for this examination. The scoring procedure for each part of the examination varies from each other. None is done in a simple, linear fashion. The scoring procedures make the estimation of reliability unlikely, although there are indications that if reliability could be estimated, it might be very high.

## Selection of Examiners

WREB has an archive that documents the selection and qualifications of dental hygienists and educator dental hygienists who participate in test development and standard setting. A cursory review of these files suggests that all examiners are very well qualified and experienced to examine. Examiner selection for examination scoring is made by the WREB member state boards. Examiners are trained and qualified by WREB prior to scoring candidates. An Examiner Review Committee meets annually to review examiner performance and takes action to correct any problems in examiner scoring.

## Training and Evaluation of Examiners

WREB has extensive examiner training and a process for qualifying examiners for evaluating candidate's patient's treatments (see WREB 2005b). The *Dental Hygiene 2005 Examiner Manual* provides general information, procedures to follow in the candidate check-in the patient qualification, and the check out procedure. Later, after the examination is given, WREB analyzes rater performance and evaluates the examiner consistency (WREB, July 10-11, 1999; July 2000; July 7-8, 2003; July 9-11, 2004; September 24-26, 2004). A memorandum from committee chair Barbara Dixon (WREB, October 29, 2004) described the standard-setting procedure used in the training and qualifying of examiners. Team captains coordinate and conduct the training exercises (WREB, 2000; 2005d). Examiner training is constantly evolving with new, revised procedures (e.g., WREB, 2005e).

## Scoring

Scoring is complex. For each of the three parts of the examination, specific rules govern scoring. In addition, candidates may lose points if the patient they selected is unsuitable for examination. Each part of the examination is described here and evaluated. As noted on page 4 of the *Dental Hygiene Examination 2005 Candidate Guide* (WREB 2005a), WREB has rules governing both scoring and deductions for various types of undesirable actions. These scoring rules are also shown in WREB's test specifications (WREB, June 2004). These rules are subject to regular scrutiny and evaluation. For example, WREB, August 20-21, 1999) instituted the use of median scores instead of biased mean scores as a basis for candidate scores across three examiners' scoring.

***Selection of Patients.*** The selection of patients is a very important aspect of the examination. A candidate could fail the test by selecting patients and alternative patients who fail to qualify. The *Dental Hygiene Examination 2005 Candidate Guide* (WREB, 2005a) provides extensive discussion on pages 6-8 on patient selection. Candidates may have deductions in their score if their first and second patients are not qualified. WREB's *Student Newsletter* (WREB, Fall 2003) discusses patient selection.

**Probe depths/recession (15 points).** Scoring is based on a point deduction method, where 2.5 points are deducted for each error up to a maximum of 15 for periodontal measurements and probing. For gingival recession, 2.5 points are deducted. The sum of deductions for this category cannot exceed 15 points. Of the total sample for this analysis, 66% of the scores were perfect, and 28% met the minimum criterion. Few candidates lost more than 3 points on this part of the examination, as the competency level of all candidates appears to be very high.

**Extra/Intraoral (10 points).** For each of the first eight categories, one point is the maximum score; for the ninth category (periodontal assessment) two points is the maximum score. A six-point rating scale (0-5) is used to rate performance on the first seven categories. Category 8 (Occlusion) is scored 0-1. Category 9 is scored 0-2. A conversion table changes ratings to fractions of a point for the first seven categories (see page 4 of the candidate guide). This conversion is non linear, and 3 on the rating scale approximates the overall cut score for the examination. Thus, a rating of 3 could be interpreted as minimally competent on this part of the examination. Table 5 contains descriptive statistics for this part of the examination.

<p style="text-align: center;">Table 5 Descriptive Statistics for Results of the Extra/Intraoral Part of the Examination for 1,165 Candidates</p>				
Category	Mean	Stan. Dev.	Highest	Lowest
Head & neck	0.941	0.224	1.00	0.00
Lymph nodes	0.955	0.190	1.00	0.00
TMJ	0.925	0.193	1.00	0.00
Floor of mouth	0.972	0.158	1.00	0.00
Oral mucosa/alveolar ridge & lips	0.798	0.373	1.00	0.00
Palate & oral pharynx	0.927	0.232	1.00	0.00
Tongue	0.933	0.239	1.00	0.00
Occlusion	0.761	0.426	1.00	0.00
Periodontal assessment	0.483	0.500	1.00	0.00

**Calculus removal (75 points).** This is the most consequential part of the examination. Any candidate loses 6 points for every validated error. A validated error is one where two or three examiners agree that the candidate has erred in removing calculus or there is an instance of tissue trauma. For the 1,165 candidates in this sample, the range of scores was 21 to 75. The mean score was 69.6, which is well above the cut score standard of 75% (equivalent to 56.25 for a 75-point subtest). Thus, the typical scores for calculus removal are very high. More than 94% exceeded the implied cut score of 75% for this part of the examination.

### **Examiner Agreement (Consistency)**

As reported in the reliability section of this report, rater consistency was very high for extra/intraoral and the calculus removal parts of the examination.

### **Examiner Accuracy**

A serious threat to validity, which is often overlooked in performance examinations of this type, is any of a set of rater effects. One of these is severity/leniency, where an examiner may be overly harsh or lenient. Fortunately, WREB uses a scoring method that defends against this kind of unfair rating. Instead of using the mean of three ratings, WREB uses the median. The analyses were done to seek information about severity/leniency in these ratings and to also try to identify two other tendencies that weaken the validity of these judgments of candidate competence: restriction of range and central tendency.

## 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](http://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 hygiene schools, candidates, and the public that this is so is also reassuring.

Another aspect of quality control is that any candidate whose scores are close to the cut score is in jeopardy of being misclassified due to their proximity to this cut score. As a service to participating states and the candidates, scoring procedures should ensure candidates that each one is receiving accurate scoring. Any candidate whose total score falls within one standard error of the cut score should have their results checked for accuracy to ensure that the pass/fail decision is correct.

In the *Examiner Manual (WREB, 2005b)*, specific guidelines are shown and used to ensure that examiners avoid factors or conditions that may question the integrity of the examination process, such as grading candidates who may be related to the examiner in some way. WREB handles this threat to validity very well by having a photo identification procedure and other safeguards that are stated in the *Dental Hygiene Examination 2005 Candidate Guide (WREB, 2005a)*.

## Summary and Conclusion

Evidence was presented about the validity of scoring. From the data presented on reliability, inter-examiner consistency is very high. The use of validated judgments and medians instead of means is a strategy that improves the accuracy of scores. Effective training of examiners is also very important. WREB has very thorough documentation of its training of examiners and scoring procedures. The *Standards* have one standard, 5.6, that addresses the integrity of scores and fraudulently obtaining a score. WREB has procedures safeguarding against that type of fraud. Eight other standards bear on 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 *Dental Hygiene Examination 2005 Candidate Guide (WREB 2005a)* and the *Dental Hygiene 2005 Examiner Manual (WREB, 2005b)*

## **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.

The score report consists of a single page containing a WREB logo, the candidate's name and identification code, and a point report for each test and aspects of each test (WREB, 2005f). At the bottom of the score report appears total score in points and the total possible points, 100. The score report provides a detailed table showing points earned and possible points for the nine categories of the extra/intra oral examination, the points earned and possible points for probe and recession, and the points and possible points for the calculus and tissue trauma. The bottom of the score report shows any point deductions for poor patient qualification or x-ray penalty.

Candidates are entitled to confidential score reports. Standards 5.13, 8.5, and 11.14 from the *Standards* (AERA, et al., 1999) are very clear about this need. Examination results are sent to participating states. No other parties should have access to these scores, unless expressly designated by the candidate. WREB contractually provides reports to member states. Schools are sent reports unless students do not wish to have the schools receive their scores.

## **Summary and Conclusion**

The *Dental Hygiene Examination 2005 Candidate Guide* (WREB, 2005a) 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 hygiene schools have the option of withholding their score report from their school.

The *Standards* (AERA et al., 1999) provides more than 40 standards existing for reporting. However, 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 observe these standards.



## **Candidate Guide and Rights**

The *Dental Hygiene Examination 2005 Candidate Guide* (WREB, 2005a) has been cited often in this report. This booklet contains essential information for candidates. The table of contents for this 25-page booklet provides general information, performance evaluation information, patient criteria, criteria for teeth and for calculus detection and removal, and examination procedures. The booklet is published each year and is updated as needed. In addition to this guide, WREB's web page is helpful, and if candidates prefer can contact the WREB office by phone or by email.

## **Summary and Conclusion**

Information and references to 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 Hygiene Program 2005 Candidate Guide* (WREB, 2005a). 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's newsletters perform an invaluable service in keeping dental hygiene students and others apprised of the examination that so greatly affects their future practice in the profession. Moreover, these newsletters provide considerable evidence supporting the validity of the examination.

## **Security**

The *WREB Policy and Procedures Manual* (WREB 2005g) discusses security. Threats to validity arising from security breaches are increasing throughout the world. WREB does many things to safeguard against cheating and other threats to validity 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. One feature of this issue with candidates is that of patient qualification. Part of the definition of competence involves the candidate bringing in patients who qualify for treatment. Failure to do this well or do it at all puts candidates in jeopardy of failing. WREB has a patient qualification process and point deductions for candidates who try to advance a patient who does not qualify for treatment.

### **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* (WREB 2005b). Conflicts of interest with candidates are monitored, and examiners are asked to recuse themselves if potential conflicts exist.

## **Summary and Conclusion**

This section has given brief information about security for WREB. The procedures for security that have 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).

## **Part VII: 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 dental hygienists 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 Dental Hygiene Examination Committee, and the staff who plan and administer the program and the participation of states, dental hygiene 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 Dental Hygiene Examination subcommittee operates effectively. They are advocates for improvement, and their minutes and recommendations give ample evidence of this commitment.

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 dental hygiene. All indicators point to the presence of a very high quality examination.

Another salient observation is that this examination program was evaluated in 1998 (Haladyna, 1998), and the findings of that evaluation were reviewed and recommendations were made and implemented subsequently (see WREB, July 18-19, 1998). Extensive documentation cited throughout this report gives testimony to the idea that WREB has continuously improved this examination program.

## 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.
- Cronbach, L. J. (1971). Test validation. In R. L. Thorndike (Ed.), *Educational Measurement* (2nd ed., pp. 443–507). Washington, DC: American Council on Education.
- Dixon, B. (October 29, 2005). Memorandum to WREB: Passing level exercise.
- Downing, S. M., & Haladyna, T. M. (1996). Model for evaluating high-stakes testing programs: Why the fox should not guard the chicken coop. *Educational Measurement: Issues and Practice*, 15, 5-12.
- Haladyna, T. M. (May 4, 1998). *Evaluation of the Dental Hygienist Examination Program of the WREB*. Phoenix, Author.
- Joint Commission on National Dental Examinations (March 1998). *Dental Hygiene Examination Specifications*. Chicago, Author.
- Linn, R. L. (2000). Assessment and accountability. *Educational Researcher*, 29(2), 4-16.
- Madaus, G. F. (1992). An independent auditing mechanism for testing. *Educational Measurement: Issues and Practice*, 11(1), 26-31.
- Mehrens, W. A., & Popham, W. J. (1992). How to evaluate the legal defensibility of high-stakes tests. *Applied Measurement in Education*, 5(3), 265-283.
- Raymond, M. .R., & Neustel, S. (In press). Determining test content for credentialing examinations. In S.M. Downing and T.M. Haladyna (Eds). *Handbook of Test Development*. Mahwah, N.J.: Lawrence Erlbaum Associates.
- WREB (June, 2005) *Report of the Dental Hygiene Examination Review Committee*. Phoenix: Author.
- WREB (2005a). *Dental Hygiene Examination 2005 Candidate guide*. Phoenix: Author.
- WREB (2005b). *Dental Hygiene 2005 Examiner Manual* . Phoenix: Author.
- WREB (2005c). *WREB Policy Guide: Dental Hygiene, Anesthesia, Restorative Exams*. Phoenix: Author.
- WREB (2005d). *Team captain script*. Phoenix: Author.
- WREB (2005e). *New examiner exercises*. Phoenix: Author.
- WREB (2005f). *Score report*. Phoenix: Author.
- WREB (2005g). *WREB Policy and Procedures Manual*. Phoenix: Author.
- WREB (October 29, 2004). *Memorandum on examiner training*. Phoenix: Author.
- WREB (September 24-26, 2004). *Dental hygiene subcommittee meeting*. Phoenix: Author.
- WREB (July 14, 2004). *Dental hygiene exam review committee*. Phoenix: Author.
- WREB (July 9-11, 2004). *Meeting of the dental hygiene subcommittee*. Phoenix: Author.
- WREB (June 2004). *WREB dental hygiene examination test specifications*. Phoenix: Author.
- WREB (Fall 2003). *Dental hygiene student newsletter*. Phoenix: Author.
- WREB (September 26-28, 2003). *Dental hygiene subcommittee meeting*. Phoenix: Author.
- WREB (July 10, 2003). *Board of directors meeting*. Phoenix: Author.

WREB (July 9, 2003). *Dental hygiene exam review committee*. Phoenix: Author.

WREB (July 7-8, 2003). *Meeting of the dental hygiene subcommittee*. Phoenix: Author.

WREB (January 11, 2003). *Dental hygiene subcommittee meeting*. Phoenix: Author.

WREB (September 7-8, 2002). *Dental hygiene exam review committee*. Phoenix: Author.

WREB (July 17-18, 2002). *Dental hygiene exam review committee*. Phoenix: Author.

WREB (July 5-7, 2002). *Report of the dental hygiene sub-committee meeting*. Phoenix: Author.

WREB (2001). *Notes on 2001 test specifications*. Phoenix: Author.

WREB (September 28-30, 2001). *Dental hygiene subcommittee meeting*. Phoenix: Author.

WREB (July 2001). *Dental hygiene clinical subcommittee recommendations and justifications*.  
Phoenix: Author

WREB (July 18-19, 2001). *Dental hygiene sub-committee meeting*. Phoenix: Author.

WREB (December 2000). *Training examiners*. Phoenix: Author.

WREB (November 3-5, 2000). *Dental hygiene subcommittee meeting*. Phoenix: Author.

WREB (July 2000). *Dental hygiene clinical subcommittee recommendations and justifications*.  
Phoenix: Author.

WREB (July 5-7, 2000). *Dental hygiene subcommittee meeting*. Phoenix: Author.

WREB (June 30, 2000). *Dental hygiene subcommittee meeting*. Phoenix: Author.

WREB (October 15-17, 1999). *Dental hygiene sub-committee meeting*. Phoenix: Author.

WREB (August 20-21, 1999). *Dental hygiene exam review committee*. Phoenix: Author.

WREB (December 18-19, 2002). *Dental hygiene exam review committee*. Phoenix: Author.

WREB (July 10-11, 1999). *Report of the dental hygiene subcommittee meeting*. Phoenix: Author.

WREB (July 18-19, 1998). *Clinical sub-committee meeting*. Phoenix: Author.

WREB (September 3, 1996). *Dental hygiene practice Survey*. Phoenix, Author.

WREB (undated). *Dental hygiene examination specifications*. Phoenix, Author.