

An Evaluation
of the
Western Region Examining Board Dental Hygiene Examination

Dr. Thomas M. Haladyna
Professor Emeritus
Arizona State University

October 2017

Executive Summary

Periodically, the Western Region Examining Board (WREB) undergoes an external evaluation of its dental and dental hygiene examination programs. An opinion by an external testing expert informs candidates, dental schools, and the public concerning whether WREB's examination programs are fulfilling the promise that test scores can be validly interpreted and used with other information about candidates for making pass/fail decisions for licensing in member and other participating states. The current *Dental Hygiene Examination Program* consists of three examinations: (1) *Local Anesthesia Examination*, (2) *Dental Hygiene Clinical Examination* and (3) the *Restorative Examination*.

Validity is the key idea in this evaluation. Test scores should be validly interpreted and used. The process of *validation* is an investigative procedure. It begins with a claim for validity, the forming of an argument supporting that claim, the collecting of evidence supporting the argument, and a judgment by the author of this evaluation regarding validity. Sometimes validity evidence is missing or weak. Sometimes, evidence is negative, and that weakens validity. This evaluation serves as a validation.

A useful supporting mechanism is the *Standards for Educational and Psychological Testing*. This publication contains specific standards that bear on test development and validation. Throughout this evaluation, standards were identified, organized, and used with validity evidence to evaluate the claim for validity.

Should WREB endure a legal challenge regarding validity, evaluations such as this one and technical reports are useful tools for defending against unjust legal action.

The largest portion of the evaluation contains validity evidence. This evidence consists of documents and data linked to these standards. The categories of validity evidence include (1) content, (2) reliability, (3) item quality, (4) examination administration, (5) setting the cut score, (6) examiner recruiting, training, and scoring, (7) scaling and comparability, (8) score reporting, (9) candidate and patient rights, (10) security, and (11) documentation. Concerning documentation, this evaluation contains an extensive listing of documents that provide attestation of validity evidence.

The summative evaluation at the end of this report supports the claim for validity. This examination program has been developed over many years with many improvements and refinements. WREB is congratulated for developing a high quality examination program that meets many testing industry standards and achieves a high degree of validity in test score interpretation and use.

Dr. Thomas M. Haladyna
Phoenix, Arizona

Acknowledgment

This report would not be possible without the assistance of Dr. Sharon Osborn Popp. She has supplied me with documents and data that were vital to the collection of validity evidence. She advised me of changes in the examination program and offered useful suggestions. The technical report she authored and cited in this report was a very important source of information and great aid in completing this evaluation. For this I am very appreciative.

Dr. Thomas Haladyna
Phoenix, Arizona

Table of Contents

Introduction	1
Part I: Why Is WREB’s <i>Dental Hygiene Examination Program</i> Being Evaluated?	2
Part II: Description of this <i>Dental Hygiene Examination Program</i>	4
Part III: Validity and Validation	5
Part IV: <i>Standards for Educational and Psychological Testing</i>	8
Part V: Legal Defensibility	9
Part VI: Compensatory and Conjunctive Scoring	10
Part VII: Validity Evidence	11
1. Content related	13
2. Reliability	16
3. Item Quality	18
4. Examination Administration	19
5. Setting the Cut Score for Pass/Fail Decisions	22
6. Examiner Recruiting, Training, and Scoring	23
7. Scaling and Comparability	26
8. Score Reporting	28
9. Candidate and Patient Rights	30
10. Security	31
11. Documentation	33
Part VIII: Summative Evaluation	34
References	35
Appendix A: Examination Review Board Meetings and Minutes (Documentation)	37
Appendix B: Committee Reports and Minutes	38

INTRODUCTION

Examining boards like Western Region Examining Board (WREB) periodically undergo an external evaluation to find out how validly test scores are interpreted and used by participating states for licensure decisions. The external evaluation is done by a highly qualified testing specialist, who provides an opinion regarding validity. The process of evaluation entails many steps explained in the section on validity in this report. The sections of the report are briefly summarized below to give the reader an overview of what follows.

1. Explains the reason for this evaluation.
2. Describes the *Dental Hygiene Examination Program*.
3. Explains *validity* and the investigative process known as *validation*.
4. Discusses national testing standards followed in this evaluation.
5. Describes the threat of legal challenge to a test score decision (pass/fail) and how to defend against it.
6. Presents validity evidence for the *Dental Hygiene Examination Program*.
7. Provides a summative evaluation.

For clarity of language, the term *examination program* is used to refer to the entire testing program, which is the *Dental Hygiene Examination Program*. This program has three independent examinations.

1. *Local Anesthesia Examination* (This examination has two parts, scored independently.)
2. *Dental Hygiene Clinical Examination*
3. *Restorative Examination*

The word *test* is synonymous with the word *examination*.

PART I: WHY IS THE *DENTAL HYGIENE EXAMINATION PROGRAM* BEING EVALUATED?

WREB administers clinical examinations in dentistry and dental hygiene. WREB, which was formally incorporated in 1976, has provided testing services to states and candidates in growing numbers. Its corporate office is in Phoenix, Arizona. Its bylaws were amended by its membership (WREB, January 11, 2003; January 7, 2006). A history of WREB is available on its website: <https://wreb.org>, retrieved August 17, 2017.

Examining boards provide important information to states. Each member state must decide who receives a license to practice a profession in that state's jurisdiction. These professions include dentistry, dental hygiene, accountancy, architecture, medicine, education, social work, law, and law enforcement, among many others. WREB provides this service to 41 states that accept WREB results, including 20 members and affiliate member states. Test scores are used with other information to decide licensure for each candidate in a state.

WREB has a Board of Directors (also known as the *Governing Board*). This board meets quarterly to discuss policy and oversee examination development and validation. Meeting minutes provide documentation of the process of governance restructuring that WREB accomplished between 2009 and 2012. The restructuring included the separation of the roles of the Board of Directors and the Examination Review bodies, which allows state board members to focus on examination content, development and oversight. The former Examination Review Committees were replaced with the Dental Examination Review Board and the Dental Hygiene Examination Review Board, consisting of representation from every active member state. The Governing Board replaced and expanded the Executive Committee. The Governing Board includes members selected by the Examination Review Boards and is responsible for strategic, administrative, legal, and financial decisions.

An Examination Review Board Committee oversees the *Dental Hygiene Examination Program*. WREB has a committee for each of the three examinations comprising this program. These committees meet regularly, review policies and procedures, and recommend changes intended to improve the examination program (See Appendix B for a comprehensive list of meeting minutes.). The structure of committees and the way staff serves WREB and the committees are clearly shown in any annual report to states (WREB, June 24, 2016).

Responsibilities of Examining Boards Like WREB

The main concern of any examining board is to increase the likelihood that a professionally licensed person will treat their patients safely. The content of these examinations is professional competence. This content usually consists of knowledge, skills, and abilities (KSAs). Defining KSAs is a very important task of these examination boards. The focus is validity.

No examination program with its battery of examinations is infallible in helping identify candidates who might jeopardize public safety. Nonetheless, all states and jurisdictions engage in licensing examinations to inform decision making about who receives a license to practice a profession. Of course, the examination program alone does not determine who receives a license. In most states and jurisdictions, passing the examinations in an examination program is one

important criterion for licensure that all candidates must achieve if they are to be allowed to practice in that state.

Although WREB is an examination board that provides validated test scores, it is each state's responsibility that it validly interprets and uses examination scores. WREB provides assurances through its documentation of validity that the examination scores and the cut score guidelines it provides are in the best interests of the states and the citizens of each state.

Evaluation of an Examination Program

An external evaluation of an examination program is highly recommended by testing experts (Buckendahl & Plake, 2006; Downing & Haladyna, 1996; Geisinger, 2016; Madaus, 1992). The benefit of such an evaluation is to verify that the examination program is providing valid information about the professional competency of its candidates. The external evaluation also provides constructive criticism intended to improve validity.

Every examination program consists of three important, logical, sequential, related elements:

1. Defining the profession as to KSAs needed to practice safely and competently,
2. Development of the examinations comprising the examination program that validly measures competence in the profession, and
3. Validation of the interpretation and use of examination scores.

Examination specialists have developed a way of thinking about validation (e.g., Kane, 2006a, 2006b; 2016). One might think of validation as an investigation of validity. This investigation involves many related, sequential steps. These steps include an argument about validity, a claim for validity, the gathering of evidence to support the argument and claim, and an evaluation of this evidence bearing on the logic of the argument and the claim.

Earlier evaluations of WREB's *Dental Hygiene Examination Program* provided validity evidence and opinions that were current to the date of each evaluation's publication (Haladyna, 1998, 2006, 2010). The organization and emphases in the current report differ from earlier evaluations to reflect changes in the concept of validity and validation (Kane, 2016). Greater emphasis is placed on reliability in this evaluation. Also, as the current examination consists of three independently scored examinations, this report will present validity evidence and analysis for each of the three examinations.

PART II: DESCRIPTION OF THE *DENTAL HYGIENE EXAMINATION PROGRAM*

Table 1 describes briefly the *Dental Hygiene Examination Program*. More detailed descriptions of the three examinations of this examination program appear in the dental hygiene candidate guides (WREB, 2017a, 2017i, 2017n).

Table 1: Description of the <i>Dental Hygiene Examination Program</i>	
Name	Description
<i>Local Anesthesia Examination</i>	This examination has two parts: written and clinical. Candidates must pass both parts; the written must be passed before candidates take the clinical part. With the written part, the candidate is required to respond to a series of discipline-based and case based selected-response examination items. The cut score for the written part is 75%. With the clinical part, the candidate must successfully perform two injections on a patient. To fail, two examiners must validate (agree) that a candidate did not perform the injection properly. This kind of critical error is described in the candidate’s guide (WREB, 2017a).
<i>Dental Hygiene Clinical Examination</i>	This examination is performance-based. The candidate must perform calculus removal and a series of periodontal assessments on a patient. The candidate is evaluated on the following: Patient selection, extraoral and intraoral examination, diagnostic quality of radiographs, calculus detection and removal, tissue management, accuracy of periodontal pocket measurement and recording, and accuracy of gingival recession assessment and recording. Grading is done by three independent examiners scoring performance as error/no error. The maximum high score is 100. Validated calculus remaining errors, probing depth errors or recession errors result in a loss of points. A final score of 75 points or higher is required for passing.
<i>Restorative Examination</i>	This examination is performance-based. The candidate must place, carve, and finish two restorative procedures on dentiform teeth (one maxillary and one mandibular). The procedures required are (1) Class II amalgam restoration, and (2) Class II composite restoration. Three independent examiners rate performance using a rating scale from 1 to 5. Candidates must score 3 or higher to pass.

PART III: VALIDITY

An examination score should represent a dental hygiene candidate's degree of professional competence. If an examination score is used as one criterion to advance or prevent advancement of a candidate to licensure, the decision to pass or fail must be highly valid. The focus of this evaluation is validity. All other ideas about examination quality are subsumed under validity.

An assessment of validity requires professional judgment about the reasonableness of an interpretation or use of an examination score. The *Standards for Educational and Psychological Testing*—hereafter called the *Standards* (American Educational Research Association-AERA, American Psychological Association-APA, & National Council on Measurement in Education-NCME, 2014) provides guidelines for evaluating validity. Additionally, the American Association of Dental Examiners-AADE (2005) issued guidelines for clinical performance examinations that include both dentistry and dental hygiene.

What does an examination score obtained from any of the *Dental Hygiene Examination Program* battery of three examinations mean? How valid is it for a state to make a pass/fail decision based on a score for each of these three examinations? Validity focuses on the meaningfulness of an interpretation of an examination score and the reasonableness of its use in making pass/fail decisions.

As noted previously, the investigative process for evaluating validity is validation (Kane, 2006a, 2006b, 2016). This process begins with a definition of dental hygiene that is usually derived from a practice analysis (Raymond, 2016; Raymond & Neustel, 2006). Then to validate interpretations and uses of examination scores, we need these elements in this validation:

1. An argument that describes what WREB plans to measure and how examination scores will be validly interpreted and used;
2. a claim that the examination scores are validly interpreted and used;
3. a collection of validity evidence related 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 stated interpretation and use of examination scores. Negative validity evidence or lack of evidence should be inconsequential.

No examination program reaches its ultimate in validity. The attainment of the highest degree of validity is a goal. All examination programs undergo improvement in an evolutionary path, but the road is steep and long. This evaluation report presents the argument and claim for validity, and it displays the evidence supplied by WREB. The author of this report has evaluated the argument and evidence to make a summative judgment about validity of each of the three examinations.

Table 2 shows the constituent elements in validation. This table also shows the reasoning process used in this validation.

Table 2: Validation of WREB’s <i>Dental Hygiene Examination Program</i>	
Argument	The Joint Commission on National Dental Examinations administers a <i>National Board Dental Examination</i> . This examination measures the knowledge, skills, and abilities thought to be necessary for safe and competent practice. This examination derives from a practice analysis of the profession of dental hygiene. The WREB <i>Dental Hygiene Examination Program</i> is a battery of three examinations intended to measure clinical competence. It also is based on a practice analysis. These two examination programs represent complementary aspects of dental hygiene professional competence.
Claim About Validity	WREB claims those examination scores obtained from candidates represent clinical competence and can be used with confidence by participating states, along with other criteria, to make licensing decisions.
Evidence Supporting the Argument	This evaluation report provides validity evidence of many types that are based on national examination standards. WREB’s technical report and other documents cited in this report offer validity evidence supporting this argument. Appendices A and B are part of this body of evidence.
Evidence Weakening the Argument	In this report, to the extent possible, evidence may be presented that weakens this argument. In the judgment of this evaluator, this kind of evidence as discussed in this report is inconsequential to validity. Nonetheless, WREB should consider threats to validity and act accordingly to diminish the threat. By that, WREB strengthens the evidence supporting the argument and the claim for validity.
Lack of Evidence	Gaps in evidence are noted in this report if it exists.
Summative Judgment	This evaluator considers the argument, claim, and evidence before making a judgment about validity of WREB’s <i>Dental Hygiene Examination</i> 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 VII of this report provides validity evidence for each of the three examinations. The sources of evidence are information found in documents and the results of statistical analysis. Validity evidence should never be noted in a checklist. Instead, the evaluator considers the body of evidence before making a summative evaluation. This evidence is used in the same manner that a jury considers evidence and decides that it supports either the prosecutor’s claim or the defense’s claim. Besides the evidence presented in this evaluation, WREB’s 2016 technical report (WREB, July, 12, 2017) provides a substantial amount of validity evidence used and cited often in this report.

Evidence Weakening the Argument

In any evaluation for validity, honest examination of evidence that undermines validity is seldom done by examination sponsors. According to Messick (1989), 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 KSAs that comprise dental competence. This part of the evaluation seeks to uncover evidence that may undermine validity.

CUR is present if the examinations used to measure competence do not match very well the definition of dental hygiene. *Fidelity* is the technical term we use to assess the connection of the tasks on the examination to the definition of dental hygiene. If we used a multiple-choice examination of scientific knowledge or a multiple-choice examination of professional knowledge, we would not be representing clinical dental competence adequately. That is why the *National Board Dental Examination* is a necessary licensing requirement but it is not sufficient. These multiple-choice examinations under-represent the construct of competence in dental hygiene. When we combine the results of the *National Board's Dental Hygiene Examination* with WREB's *Dental Hygiene Examination Program*, we have important complementary pieces of information that provide adequate representation of the construct of dental hygiene competence. Thus, participating states see the value of using both the National Board's and WREB's examination programs due to their complementary nature with respect to the KSAs that comprise professional competency.

CIV is systematic error that undermines validity (Haladyna & Downing, 2004). In WREB's *Dental Hygiene Examination Program*, a major threat is rater bias. Raters may be too severe or too lenient. Fortunately, WREB is very aware of this threat and deals with this possibility in every examination administration. Subsequent sections of this report deal with this threat using the many-faceted Rasch Model.

Naturally, WREB and its member 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. For most examination programs, a validity research agenda is useful for exploring problems and solving problems that bear on validity (Haladyna, 2006).

Summary

This section on validity is best summarized in Table 2. It shows that we start with a definition of dental hygiene competence, then formulates an argument about the validity of using WREB's *Dental Hygiene Examination Program* scores as unique, complementary measures of clinical competence. A claim is made by WREB for its member and nonmember states using these examination scores in that way is highly valid. Validity evidence is collected and displayed. After all evidence is assessed, a summative judgment is made about the validity of each examination. Participating states can use this judgment to guide them in deciding if the examination score information they receive is adequate for their needs. As mentioned previously, all licensing boards 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 EXAMINATION

The *Standards for Educational and Psychological Testing* (2014) was published by three large organizations committed to the improvement of examination programs and in support of valid examination score interpretations and uses: American Educational Research Association, American Psychological Association, and the National Council on Measurement in Education. This publication contains a comprehensive set of guidelines that help examination developers achieve a high degree of validity in the interpretation and use of examination scores.

A large, representative committee of testing experts and other highly qualified volunteers participated in developing this book. The American Association of Dental Examiners (2003) published *Guidance for Clinical Licensure Examinations in Dentistry*. Although not specifically cited in this evaluation, these guidelines also apply to this evaluation. The two sets of guidelines are very similar in terms of principles related to validity. Specific standards are cited and supported by validity evidence throughout this document.

PART V: LEGAL DEFENSIBILITY

No examining board wants to be challenged legally for adverse examination score decisions that might be considered invalid. Such challenges are expensive to defend and if successful may lead to loss of credibility that can ultimately weaken and destroy an examination program. Moreover, someone's career objectives can be thwarted by an unfair decision based on weak validity or significant threats to validity.

Validation provides evidence that supports the examination program and its purpose. By undertaking a validation, WREB provides assurance to its participating states that the examination score information can be used validly. Therefore, validation can discourage unwarranted litigation. When potential litigants know that validation has been done and the validity evidence is publicly available, they are less likely to challenge the examining board's examination score interpretations and uses.

Any examining board should have legal counsel that examines threats that arise from legal actions and its position in thwarting these threats. By engaging in this evaluation where validity evidence is collected and organized, WREB very effectively reduces the threat of legal action. Mehrens and Popham (1992) provided a useful discussion of legal threats and validity. By paying particular attention to validity, the intent is often sufficient to ward off legal challenge.

WREB has made public its validity evidence in technical reports and evaluations, such as this one. WREB's website is very informative and represents a model for other examining boards. (See <https://www.wreb.org/Information/Articles.aspx> Retrieved August 17, 2017). WREB's annual reports provide useful overviews of its examination program and other sources of information about its programs. A primary source of technical information and support is the annual technical report (WREB, July 12, 2017).

VI. COMPENSATORY AND CONJUNCTIVE SCORING

Governing boards, like WREB, face a dilemma in how to score examination results. With the *Dental Hygiene Examination*, three independent examinations are employed. The *Local Anesthesia Examination* has two parts: written and performance. Shall WREB combine all score into a total score? Such a score is usually very reliable and represents the sum of all performance. However, is it permissible or allowed that a weak performance in one of these examinations or parts of one examination is low but is compensated by a higher performance in another examination or part of the examinations? That is, would a weak performance on one of the three examinations be acceptable, if the candidate has a strong performance on the other two examinations. This is known as *compensatory* scoring.

Conjunctive scoring requires that each examination stand alone and is important. Thus, a candidate must pass each examination. The limitation of conjunctive scoring is that each examination has lower reliability than compensatory scoring. Thus, greater effort is needed to ensure that each of the three examinations of the *Dental Hygiene Examination* program has adequate reliability to ensure a confident pass/fail decision. Pros and cons of compensatory and conjunctive scoring are reviewed in Haladyna and Hess (1999).

WREB imposed a conjunctive scoring condition on both parts of the *Local Anesthesia Examination* for the same reason—patient safety. Thus, WREB has imposed a very high standard in scoring and making pass/fail decisions for the sake of preventing low-scoring candidates the ability to practice until they display excellence in all three parts of the *Dental Hygiene Examination Program*.

PART VII: VALIDITY EVIDENCE

Table 3 lists standards addressing validity in a general way. Standards 1.0, 1.1, 1.2, 1.5, d 4.0, and 11.1 are supported by the candidate guides (WREB, 2017a, 2017i, 2017n), WREB’s website, the annual technical report, and this evaluation report. Standard 1.7 addresses a threat to validity—that practice might incorrectly boost performance on the examinations. In this examination program, practice and coaching are openly encouraged because the tasks performed resemble actual professional practice. Standards 3.0 and 4.13 bear on an important threat to validity discussed elsewhere in this report. Standard 3.1 discusses the importance of ensuring that candidates with special needs are accommodated so that is professionally responsible yet ensures patient safety. Standards 3.4 and 6.0 emphasize the importance of a standardized examination experience. WREB’s examinations are consistent no matter time or place because the tasks are always the same and the examiners are highly trained and regulated to provide a uniform examination experience.

Table 3: <i>Standards</i> Generally Related to Validity ¹	
1.0	Clear articulation of each intended examination score interpretation for a specified use should be set forth, and appropriate validity evidence in support of each intended interpretation should be provided.
1.1	The examination developer should set forth clearly how examination scores are intended to be interpreted and consequently used. The population(s) for which an examination is intended should be delimited clearly, and the construct or constructs that the examination is intended to assess should be described clearly.
1.2	A rationale should be presented for each intended interpretation of examination scores for a given use together with a summary of the evidence and theory bearing on the intended interpretation.
1.5	When it is clearly stated or implied that a recommended examination score interpretation for a given use will result in a specific outcome, the basis for expecting that outcome should be presented together with relevant evidence.
1.7	If examination performance, or a decision made therefrom, is claimed to be essentially unaffected by practice and coaching, then the propensity for examination performance to change with these forms of instruction should be documented.
3.0	<i>Construct-irrelevant variance (CIV) should be avoided in all aspects of examination development, administration, scoring, and reporting.</i>
3.1	Those responsible for examination development, revision, and administration should design all steps of the examination process to promote valid score interpretations for intended score uses for the widest possible range of individuals and relevant subgroups in the intended population.
3.2	Examination developers are responsible for developing examinations that measure the intended construct and for minimizing the potential for examinations’ being affected by construct-irrelevant characteristics, such as linguistic, communicative, cognitive, cultural, physical or other characteristics.
3.4	Examination takers should receive comparable treatment during the examination administration and scoring process.
4.0	Examinations and examination programs should be designed and developed in a way that supports validity of interpretations of examination scores for their intended uses.

4.13	When credible evidence indicates that irrelevant variance could affect scores from the examination, then to the extent feasible, the examination developer should investigate sources of irrelevant variance. Where possible, such sources of irrelevant variance should be removed or reduced by the examination developer.
6.0	To support useful interpretation of score results, assessment instruments should have established procedures for examination administration, scoring, reporting, and interpretation. Those responsible for administering, scoring, reporting, and interpreting should have sufficient training and supports to help them follow the established procedures. Adherence to the established procedures should be monitored, and any material errors should be documented and, if possible, corrected.
11.1	<i>A clear statement of intended interpretation of an examination score and the use to which it is intended should be made clear to examination takers.</i>

¹Standards appearing in italics have been paraphrased due to their great length.

Some of these standards may seem repetitious. This is true because different panels of testing experts worked on different chapters of the *Standards* yet maintained a strong focus on validity.

1. Content-related Validity Evidence

The most fundamental way for identifying the content of any professional credentialing examination such as this one is to conduct a practice analysis (Raymond & Neustel, 2006; Raymond, 2016). This survey of the profession provides information about the KSAs needed to practice competently and safely in WREB member states. A practice analysis was completed (WREB, May 15, 2007), which formed the basis for the 2009 *Dental Hygiene Examination Program*. A new practice analysis was completed, and the results were discussed in a meeting (December 18, 2015). Table 4 presents standards bearing on content-related validity evidence.

1.11	<i>The basis for defining and identifying content should be clearly specified.</i>
1.13	If the rationale for an examination score interpretation for a given use depends on premises about the relationships among examination items or among parts of the examination, evidence concerning the internal structure of the examination should be provided.
1.14	When interpretation of subscores, score differences, or profiles is suggested, the rationale and relevant evidence in support of such interpretation should be provided.
4.1	Examination specifications should describe the purpose(s) of the examination, the definition of the construct or domain measured, the intended examinee population, and interpretations for intended uses. The specifications should include a rationale supporting the interpretations and uses of examination results for the intended purpose(s).
4.2	<i>Examination specifications should be very comprehensive regarding content, examination length, item formats, ordering of items and sections, and administration time.</i>
4.3	<i>All examination development activities should be documented.</i>
4.12	Examination developers should document the extent to which the content domain of an examination represents the domain defined in the examination specifications.
5.1	Examination users should be provided with clear explanations of the characteristics, meaning, and intended interpretation of scale scores, as well as their limitations.
11.2	Evidence of validity based on examination content requires a thorough and explicit definition of the content domain of interest.
11.3	<i>When examination content is a primary source of validity evidence, a close link between examination content and the profession being assessed is required.</i>
11.13	<i>The content domain should be clearly described and justified in light of the professional competency being tested.</i>

Regarding the standards in Table 4, the dental hygiene candidate guides (WREB, 2017a, 2017i, 2017n) provides information addressing 1.13, 4.1, 4.2, 5.1. This evaluation provides documentation related to 4.3. The practice analysis provides a basis for 4.12, 11.2, 11.3, and 11.13. As stated in the WREB technical report (WREB, July 12, 2017):

“Subject matter experts and WREB staff develop and review examination content in accordance with current professional standards and occupational analyses in dentistry and dental hygiene, including the 2005-2006 Survey of Dental Services Rendered (ADA, 2007),

the Standards for Clinical Dental Hygiene Practice (ADHA, 2008), the WREB Practice Analysis for General Dentist (WREB, 2007), the WREB Dental Hygiene Practice Analysis Report (WREB, 2009) and the professional standards of practice within member states. A current Dental Hygiene practice analysis is in development.”

Structure of the Content of the Three Examinations

One way to validate whether the content of the three examinations represents a single unified dimension (professional competency) or the content represents three relatively independent dimensions is through the study of dimensionality. This finding is important because pass/fail decisions are made on each of the three examinations. Thus, WREB has concluded that professional competency is best represented by three independent, complementary abilities that comprise dental hygiene competence. This conclusion motivates the development of three examinations and justifies the use of pass/fail decisions on each of these examinations.

The score file for all candidates was subjected to an analysis to reveal the dimensionality of scores from the three examinations. With the statistical procedure known as *factor analysis* three different rotations were tried, each based on the assumption that there is one factor (varimax) or several factors equally represented (equamax, quartimax). The sample size was 243 candidates, due to the fact that not all candidates took all three examinations. Thus, this limitations is mentioned and may bear on the results. The results show the following:

All three examinations showed low correlations with one another. Two reasons for this kind of result are (1) scores are very negatively skewed—as candidates performed very highly as expected and (2) the three examinations represent independent dimensions.

Local Anesthesia Examination. Although this examination consists of a written and performance component, this factor did not emerge. Evidently the result of the written part of this examination has little bearing on whether the candidate passes on the clinical part of the examination. WREB has determined that a candidate must pass both the knowledge and skill parts of this examination.

Dental Hygiene Clinical Examination. This examination produced a factor that is strong enough to stand independently from the other two. The written part of the Local Anesthesia Examination was weakly connected to this examination, but this result is artifactual because the correlations among these scores and subscores were very small.

Restorative Examination. This examination had the strongest factor because it is composed for four scores ranging from zero to five. It is independent of the other two examinations.

The findings reported here confirm WREB’s strategy for forming three independently evaluated examinations instead of combining all scores to form a total score representing professional dental hygiene competency.

Fidelity

An issue facing all examination developers for a clinically-based professional competency examination is whether each examination has *fidelity* with its criterion. Fidelity is a judged characteristic of any examination by which each examination item should resemble or replicate what a hygienist does in actual clinical practice. Examining boards need to show how much each examination is about professional practice as shown in the results of the practice analysis. For example, a selected-response (multiple-choice) examination would have low fidelity for an examination of clinical competence, because it measures knowledge not skill. The clinical examination in dental hygiene has extremely high fidelity because the tasks performed by candidates resemble those done in actual practice. The written examination in the *Local Anesthesia Examination* is presented in a selected-response format but uses high-fidelity discipline-based and case-based problems. This kind of testing is highly recommended by testing experts (Haladyna and Rodriguez, 2015).

As noted in the 2016 technical report (WREB, July 12, 2017), the performance tasks that candidates perform are sampled from the domain of professional practice. No examination can sample all tasks, so WREB's objective is to have its subject-matter experts select those tasks that best represents this domain of performance tasks. Subject-matter experts review the domain of tasks observed in practice and decide upon the sample of tasks and evaluation criteria for measurement that reflect performance of a minimally competent entry-level professional or better. In addition, subject-matter experts meet often to discuss fine-tuning and changes that increase the representation of this sampling of tasks. Appendix B contains a comprehensive list of committee meetings and their dates.

Conclusion

WREB has assembled a comprehensive and appropriate body of evidence supporting the content of the three examinations comprising the *Dental Hygiene Examination Program*.

2. Reliability

Table 5 lists seven standards addressing reliability in a general way. Subsequent sections address reliability for each of three examinations.

Table 5: <i>Standards</i> Related to Reliability	
2.0	Appropriate evidence of reliability/precision should be provided for the interpretation and use for each intended score use.
2.2	The evidence provided for the reliability/precision of the scores should be consistent with the domain of replications associated with the examination procedures, and with the intended interpretation for the use of examination scores.
2.5	Reliability estimation procedures should be consistent with the structure of the examination.
2.7	Inter-judge and intra-judge consistency of ratings should be studied, monitored, and documented.
2.13	The standard errors of measurement, both overall and conditional (if reported), should be provided in units of each reported score.
2.19	<i>Method of estimation of reliability should be documented.</i>
11.14	Estimates of the consistency of examination-based credentialing decision should be provided in addition to other sources of reliability evidence.

Reliability coefficients are reported for each examination (standard 2.0). The interpretation of reliability is appropriate in the following way (2.2). Examination scores are very negatively skewed because the examinees are highly trained. Thus, reliability estimates are attenuated (weakened) because the statistical procedure used to estimate reliability depends on variation of candidate scores. Reliability is best estimated when there is a normal distribution instead of a skewed test score distribution.

However, estimates of random error are more important. The conditional standard error of measurement is the statistic that matters. Inter-judge and intra-judge consistency ratings are also studied and monitored (2.7). These are documented for each of the three examinations in subsequent sections of this evaluation and in the annual technical report (WREB, July 12, 2017). Conditional standard errors of measurement are also presented in the technical report. The consistency of pass/fail decision is reported for each of the three examinations (11.14). However, this index is largely dependent on the number of examinees who score at or near the cut score.

Reliability and the Conditional Standard Error of the Local Anesthesia Examination

For the selected-response test, the technical report (WREB, July 2017, p. 26) provides reliability estimates for three test forms: 0.64, 0.67, and 0.67. As passing rates were very high (87%, 88%, and 85%), there is little risk of misclassifying candidates whose scores may fall close to the pass/fail cut score (75%). The conditional standard errors were 2.90, 2.92, and 2.90. An index of classification consistency was 0.89, 0.89, and 0.88 for the three test forms.

For the clinical test, the technical report (WREB, July 2017, p. 29) reports that conventional reliability and conditional standard error analysis were not applicable. A high degree of examiner agreement was stated. Also, those candidates who failed this clinical examination tended to repeat poor performances.

Reliability and the Conditional Standard Error of the Clinical Dental Hygiene Examination

From the technical report (July 12, 2017, p. 25), reliability is reported to be 0.65 and the conditional standard error is 3.57.

Reliability and the Conditional Standard Error of the Restorative Examination

The reliability and conditional standard error of this examination is reported in the technical report (July 12, 2017, pp. 29-32). Reliability estimate was reported as 0.87, which is very high. The conditional standard error was reported as 0.075.

Conclusion

WREB's technical report provides a comprehensive analysis of technical qualities to the 2016 *Dental Hygiene Examination Program* (July 12, 2017). Although reliability estimates are low, the conditional standard error of measurement shows that very few candidates are trapped in the zone of uncertainty due to random error in test scores. Lengthening the test is one remedy, but given that the distribution of test scores is very skewed, making the test longer would not be productive. The problem is that the few candidates scoring at or near the cut score of 75 should be advised to remediate and perform at a much higher level on the next test occasion before assurance is given that they have confidently passed a test.

3. Item Quality

Table 6 lists some standards addressing the quality of the performance examination items that WREB uses. This section of the evaluation addresses in a general way item quality. As will be shown, there is ample documentation in the technical report (WREB, July 12, 2017) and in this evaluation regarding the quality of items used in each of the three examinations.

Table 6: <i>Standards</i> Related to Item Quality	
4.7	The procedures used to develop, review, and try out items and to select items from the item pool should be documented.
4.8	The examination review process should include empirical analyses and/or the use of expert judges to review items and scoring criteria. When expert judges are used, their qualifications, relevant experiences, and demographic characteristics should be documented, along with the instructions and training in the item review process that the judges receive.
4.10	<i>Statistical properties of item scores should be studied in an appropriate theoretical context.</i>

Many sources and advice on performance on these items can be found in the three candidate guides (2017a, 2017i, and 2017n) and on WREB's website. The items are disclosed; that is, all candidates are aware of the tasks to be performed and have the opportunity to practice before actually taking the three examinations.

Standards 4.7 and 4.8 are addressed in meeting minutes, which are documented in Appendix B. The practice analysis is the basis for the high fidelity between tasks on each of the three examinations and what ideally is the domain of dental hygiene practice. In other words, each test consists of a sample of tasks taken from practice as determined by subject-matter experts committees. Statistical properties of examination items are found in the technical report (WREB, July 17, 2017, pp. 23-4, 25-8, 29-30).

Conclusion

The standards for item quality have been met in the development of this examination program. There is substantial documentation in the technical report (WREB, July 17, 2017) supporting this conclusion. Moreover, the practice analysis also provides a basis for the item development as derived from the definition of content that arises from the practice analysis.

4. Examination Administration

Table 7 lists standards related to examination administration. McCallin (2006, 2016) provides an extensive analysis of ways that the validity of examination score interpretations and uses can be weakened by poor administration practices. Following these standards is one way of contributing to the improvement of validity for this examination program.

4.16	The instruction presented to examination takers should contain sufficient detail so that examination takers can respond to a task in the manner that the examination developer intended. When appropriate, sample materials, practice or sample questions, criteria for scoring, and a representative item identified with each format or major area in the examination's classification or domain should be provided to the examination taker prior to the administration of the examination, or should be included in the examination material as part of the standard administration instructions.
6.1	Examination administration should follow carefully the standardized procedures for administration and scoring specified by the examination developer and any instruction from the examination user.
6.4	The examination environment should furnish reasonable comfort with minimal distractions to avoid construct-irrelevant variance.
6.5	Examination takers should be provided appropriate instructions, practice, and other support necessary to reduce construct-irrelevant variance.
6.6	Reasonable efforts should be made to ensure the integrity of examination scores by eliminating opportunities for examination takers to attain scores by fraudulent or deceptive means.
6.7	Examination users have the responsibility of protecting the security of examination material at all times.

The best source of information about examination administration can be found in the annual technical report (July 16, 2017, pp. 11-14). Additional information is presented in the in the candidate guides (WREB, 2016a, 2016i, WREB, 2016n), on the WREB website (<http://www.wreb.org>), and in the *WREB 2016 Policy Guide* (WREB, 2017r). Additional information can be found in the examiner manuals (See Appendix A.).

Overall planning for administration is documented showing how examination sites and its coordinators need to prepare for examination (WREB, July 12, 2017). Every candidate receives a letter informing the candidate about the orientation day and the three clinical examination days and, also, the examination. The three candidate guides (WREB, 2017a; 2017i, 2017n) are the most important documents related to standard 4.16. Standard 6.1 is also in evidence in the guide. Standards 6.4, 6.5, 6.6, and 6.7 are addressed specifically for each of the three examinations in this examination program.

The objective of standardized examination administration is to ensure that all candidates have an equal opportunity to perform to the best of their ability. WREB reviews regularly examination administration policies and procedures (WREB, July 12, 2017). Examination committee meetings are listed in Appendix A. Some concerns of the examination committee regard administration include timing, accommodations, patient safety and comfort, infection control, and site assignments of examiners. Having regular meetings to discuss refinements in administration is a very positive, constructive action that WREB maintains annually.

Examination Timing

The *Dental Hygiene Examination* allows two hours to complete treatment. The computer-based *Local Anesthesia Written Examination* is administered by Pearson VUE at examination centers. Candidates are allowed one hour to complete this examination. No strict time-limit is enforced on the *Local Anesthesia Clinical Examination*. Candidates are scheduled at times that provide approximately 20 minutes to complete the required injections, but the time to complete the examination may be shorter or longer. The *Restorative Examination* is administered in pre-assigned morning or afternoon groups. Once a group enters the clinic, candidates have one and one-half hours to complete the two procedures.

Accommodations

WREB makes every reasonable effort to offer examinations in a manner that ensures the comparability of scores for all candidates. This policy is consistent with the *Standards for Educational and Psychological Testing* (AERA, et al., 2014) and the Americans with Disabilities Act (1990). In appropriate instances, more time is allowed for some candidates. If an examination accommodation is requested and documented by the appropriate professional, WREB makes necessary provisions for an accommodation. One exception is if the accommodation alters the measurement of dental hygiene competency.

Patient Safety and Comfort

Guidelines and requirements regarding patient safety and comfort are addressed throughout candidate guides (2017a, 2017i, 2017n). For each patient, a candidate must complete and submit a patient consent form, a patient medical history form, and a follow-up care agreement. The candidate guide lists medical conditions and other factors to consider when selecting a patient to participate in the examination and describes expectations for candidates regarding patient care and comfort during the examination, such as nourishment, breaks, and administration of appropriate local anesthesia as needed. Candidate guides, examiner training materials, and staff training emphasize patient safety. The three candidate guides also describe situations where the health of the patient may require additional treatment or follow-up care. Instruction includes the review of WREB's zero-tolerance policy regarding actions or conduct that could be viewed as sexual harassment and sexual misconduct (Title VII of the Civil Rights Act of 1964).

Infection Control

The candidate guides (WREB 2017a; 2017i, 2017n), examiner training materials, and staff training emphasize adherence to published clinical treatment guidelines and standards for infection control procedures. Procedures are addressed regarding (1) proper infection control protocol, (2) compliance with OSHA guidelines for proper clinic attire, (3) protection from contaminated instruments and (4) proper disposal of biohazardous and pharmaceutical materials and sharps. Failure to maintain acceptable standards of infection control may result in failure or dismissal.

Site Assignments of Examiners

Examiners are trained and calibrated. However, WREB chooses examiner teams with great care to avoid the appearance of bias. Bias is regularly evaluated, and this kind of analysis is discussed in another section of this evaluation. WREB requires member states to be involved in all aspects of examination administration, development, and review. Experienced examiners are chosen for positions of team captain and chief examiner.

Conclusion

Considerable documentation is available to support the conclusion that examination administration is among the strengths of this examination program.

5. Setting the Cut Score for Pass/Fail Decisions

Setting the cut score is a very important activity, because it determines who will pass or fail each of the three examinations. The process of setting the passing standard must be credible, legally defensible, and well-informed to protect the public and also the rights of candidates. Table 8 lists standards to guide in evaluating how the cut score was set for each of the three examinations.

Table 8: <i>Standards</i> Related to Setting the Cut Score	
5.5	When raw scores or scale scores are designed for criterion-reference interpretation, including the classification of examinees into separate categories, the rationale for recommended score interpretations should be explained clearly.
5.21	When proposed examination score interpretations involve one or more cut scores, the rationale and procedures used for establishing cut scores should be documented clearly.
5.23	When feasible and appropriate, cut scores defining categories with distinct substantive interpretation should be informed by sound empirical data concerning the relations of examination performance to the relevant criteria.
11.16	The level of performance required for passing a credentialing examination should depend on knowledge and skills necessary for credential worthy performance in the occupation or profession and should not be adjusted to control the number or proportion of persons passing the examination.

The *Standards for Educational and Psychological Testing* (AERA, et al., 2014, p. 175) states that passing standards should be high, to protect the public and the profession by excluding unqualified individuals, but not so high as to “unduly restrain the right of qualified individuals to offer their services to the public.” The passing standards set by WREB examination committees are set high to prevent unqualified candidates from getting a license. Most candidates are extremely well prepared before taking any of the three examinations. Consequently, most candidates score well above the cut score for pass/fail decisions. Thus, the risk of misclassifying candidates due to random error is small.

WREB's examination committees determine passing scores based on professional standards of content and practice, even when arbitrary cut scores have been legislated, such as 75%. The standard-setting process for selected-response *s Dental Hygiene Local Anesthesia Clinical Examination* involves committee judgments of each item on the exam, according to Ebel's method (Ebel, 1972; Zieky, Perie, and Livingston, 2008).

Conclusion

WREB has met the standards in Table 10 regarding setting the cut score.

6. Examiner Recruiting, Training and Scoring

Table 9 presents standards addressing examiner training. Examiners may contribute to random and systematic errors that undermine validity. These types of errors are usually referred to as *rater effects*. WREB has taken steps to decrease random errors by having many tasks (examination items) and three examiners. The annual technical report provides extensive information about examination scoring (WREB, July 12, 2017, pp. 15-18).

1.9	<i>When candidate performance is judged, the process for identifying, recruiting, training, and monitoring judges should be documented.</i>
2.7	<i>Inter-judge and intra-judge consistency of ratings should be studied, monitored, and documented.</i>
4.18	Procedures for scoring and, if relevant, scoring criteria should be presented by the examination developer with sufficient detail and clarity to maximize the accuracy of scoring. Instructions for using rating scales or for deriving scores obtained by coding, scaling, or classifying constructed-responses should be clear. This is especially critical for extended-response items such as performance tasks, portfolios, and essays.
4.20	<i>Processes for identifying, training, and evaluating judges should be well developed and documented.</i>
4.21	<i>Rater consistency and rater effects should be studied, documented, and, if feasible, improved.</i>
5.0	Examination scores should be derived in a way that supports the interpretations of examination scores for the proposed uses of examinations. Examination developers and users should document evidence of fairness, reliability, and validity of examination scores for their proposed uses.
6.8	Those responsible for examination scoring should establish scoring protocols. Examination scoring that involves human judgment should include rubrics, procedures, and criteria for scoring.
6.9	Those responsible for examination scoring should establish and document quality control processes and criteria. Adequate training should be provided. The quality of scoring should be monitored and documented. Any systematic errors should be documented and corrected.

Most examination judgments in WREB examinations are made by three independent examiners. Examiners must be consistent with other examiners in rating performance. Also, examiners must not be lenient or harsh in their ratings. The median of the three grades is used. Using the median is fair, because it moderates an examiner who might be unduly harsh or lenient. Examination judgments for the *Local Anesthesia Clinical Examination* are made by two independent examiners. Where the two examiners are involved in a decision that affects the candidate's score, the two examiners must validate on the same rationale for rejection or penalization.

Examiner Recruiting

As standard 1.9 requires, WREB has a system for recruiting examiners (WREB, 2009). Most examiners are members or designees of their state licensing boards. Approximately 15% of examiners are educators. The proportion of educators is limited to prevent conflict of interest. All examiners must be actively licensed and in good standing, with no license restrictions, submitting proof of license renewal annually. Most examiners participate directly in scoring, while some

highly experienced examiners participate in leadership roles, such as chief floor examiner.

Examiner Training

Because of the importance of having examiner consensuses, all examiners are trained and calibrated to an acceptable level of agreement with respect to the scoring criteria for the examinations in which they participate.

WREB has documentation showing how examiners are prepared for the examination (4.18 and 4.20). Each examiner receives an updated examiner manual and other materials (WREB, 2017b, 2017c, 2017d, 2017e, 2017f, 2017j, 2017k, 2017r, 2017s).

All examiners are required to complete a series of tutorials and self-assessments before each examination. Examiners spend approximately eight to 10 preparing at home with WREB-secure online training materials. Examiners must also attend orientation and calibration sessions that take place before every examination. New examiners participate in an additional, earlier session to discuss their preparation with the team captain. During calibration, examiners evaluate performance according to the grading criteria. Their judgments are compared with scores that have been previously selected by the examination committees as representative of the defined levels in the criteria. The examiner team completes calibration examinations until they have all reached an acceptable level of agreement. All calibration examinations are reviewed regularly for content and consistency.

Examiner Consistency and Bias

WREB evaluates consistency and bias for all examiners. This is part of the training and also a policy for retaining or dismissing examiners who fail to evaluate consistently or with bias. The annual technical report (July 12, 2017, p. 32-3) reports a variety of analyses concerning examiner agreement and bias. To put this kind of evaluation in perspective, most candidates score well above the cut score, so examiner inconsistency and bias cannot result in a candidate failing when the candidate true performance is well above the cut score. Nonetheless, in those rare instances, where a candidate's score is close to the cut score, inconsistency and bias are serious threats to validity and potential threats to preventing a competent candidate from practice or allowing a candidate to pass who may practice unsafely.

Examiner Scoring

Examination scores are dependent upon the judgments of examiners. Examiners receive regular feedback on their performance. Examiners with low percentages of agreement, high percentages of harshness or lenience, or erratic grading patterns are remediated and monitored to ensure increased understanding of criteria definitions. Continued lack of agreement may result in dismissal from the examination pool.

Conclusion

WREB has abundant evidence supporting its examiner recruiting, training, and scoring activities. The control and reduction of inconsistent scoring and bias are crucial to ensuring that each candidate's performance is treated fairly. WREB does an excellent job in this regard. This system of training has evolved over many years. Board-appointed review committees serve to improve it continuously.

7. Scaling and Comparability

The pass or fail decision regarding candidate performance on most WREB examinations is based on the final score.

1. Final scores for the *Dental Hygiene Examination* are calculated by applying point deductions from the total points possible for any examiner-validated errors or penalties.
2. Final scores for the *Local Anesthesia Written Examination* are calculated by re-scaling the sum of correct responses to a percentage-like scale of zero to 100.
3. The *Local Anesthesia Clinical Examination* does not produce a final score. Instead, each injection is graded as passing or failing. Both injections must be passing to pass the examination.
4. Final scores for the *Restorative Examination* are calculated by summing the weighted median ratings assigned by the examiners on each scoring criterion and then averaging the scores of the two preparations treated.

With the written examination, equating is necessary because several equivalent test forms are used. With the other three examinations above, no equating is necessary as the tasks (items) are identical, and the administration is standardized. The scores obtained from performance on the simulated preparations assigned for treatment in the *Restorative Examination* are assessed regularly to confirm that different teeth do not vary in how much challenge, and therefore do not require equating.

Table 10 presents standards addressing the examination score scale of each of the three examinations and why interpretations of examination scores must be consistent from administration to administration.

Table 10: <i>Standards</i> Related to Scaling and Comparability	
5.2	The procedures for constructing scales used for reporting scores and the rationale for these procedures should be clearly described in detail.
5.5	When raw scores or scale scores are designed for criterion-referenced interpretation, including the classification of examinees into separate categories, the rationale for recommended score interpretations should be explained clearly.
5.6	Examination programs that attempt to maintain a common scale over time should conduct periodic checks of the stability of scale on which scores are reported.

Standard 5.2 is discussed in the technical report (WREB, July 12, 2017). Standard 5.5 involves the cut score, which was discussed in another section of this evaluation. Standard 5.6 is satisfied because the performance examination has a natural common scale that is unaltered from time to time. As the items are the same on all examination occasions, the scale is consistent no matter when the examination is administered. This statement is mitigated by unusual circumstances, such as when administration is altered, ended prematurely, power failure, or

another external event of that nature.

The three examinations have high fidelity with tasks performed by dental hygienists. That assertion is supported by the practice analysis (WREB, December 18, 2015). Candidates are familiar with the performance examination items and have opportunities to practice with these tasks in clinical setting in dental schools. Thus, there is transparency among three sequential activities: examination preparation, the examination, and actual dental practice. Also, there is a strong alignment among these three activities.

Appendix B contains an extensive list of meetings where examination improvements and refinements are regularly discussed and implemented.

Conclusion

Examination scales have been appropriately constructed and used consistently from one administration to another. Test administration is standardized. Standards have been satisfied.

8. Score Reporting

Table 11 below shows six *standards* addressing score reporting.

Table 11: <i>Standards</i> Related to Score Reporting	
6.10	When examination score information is released, those responsible for examination programs should provide interpretations appropriate to the audience. The interpretations should describe in simple language what the examination covers, what the scores represent, the precision/reliability of the scores, and how scores are intended to be used.
6.14	<i>Examination organizations should maintain confidentiality and protect the rights of examination takers.</i>
6.15	When individual examination data are retained, both the examination protocol and any written report should also be preserved in some form.
6.16	Transmission of individually identified scores to authorized individuals or institutions should be done in a manner that protects the confidential nature of the scores and pertinent ancillary information.
8.1	Information about examination content and purposes that is available to any examination taker prior to examination should be made available to all examination takers.
8.5	<i>Policies for release of examination scores should be carefully considered and clearly recommended. Release of scores should be consistent with the purpose of the examination and in consideration of the examination takers and informed consent.</i>

A candidate score report should present examination results clearly and effectively. The score report should help candidates understand the scoring procedure and the meaning of scores on the report that comprise the total score. Score reports are confidential and are not public documents.

WREB ensures that examination results are available to candidates as soon as possible. *Dental Hygiene Examination* and *Local Anesthesia Clinical Examination* candidates receive their provisional results onsite, after completing the examination. All candidates are notified via electronic mail when they can find their official results at their secure WREB login online. Candidates receive their results within one week of the examination. Failing candidates receive additional information about their performance. They are counseled to consider all performance information in their preparation for re-take. Detailed score reports are available to successful candidates upon request.

Some errors may result in point deductions when performance is judged as inadequate, unsafe or harmful. One example is tissue trauma. If performance is improper, the candidate may be dismissed. The candidate then must obtain permission from the WREB Board of Directors to become eligible for reexamination. Penalty details, definitions, possible point deductions, and examples of improper performance and unethical conduct can be found in the candidate guides (WREB, 2017a, 2017i, 2017n).

WREB supplied 13 score reports of differing formats depending on whether a candidate passed or fails. Those passing received a simple score report informing them of a passing score. Failing candidates receive additional details of performance. Passing candidates may receive

additional details upon request.

WREB contractually provides reports to member states. Non member state boards also receive results upon request. Program directors of candidate schools also receive candidate score reports for the current year.

Conclusion

An inspection of score reports shows that information is provided to all candidates that is both summative and diagnostic. Confidentiality is ensured. Transmission of scores is done responsibly. All standards addressing score reports appear to have been met. Standards 6.10, 6.14, 6.15, 6.16, 8.1, and 8.5 appear to be satisfied.

9. Candidate and Patient Rights

Table 12 presents standards related to candidates' rights. The standards take very seriously how candidates in high-stakes credentialing examinations are treated. Fairness is paramount. The previous section also presents standards and validity evidence addressing this important topic.

Table 12: <i>Standards Related to the Rights of Examination Takers</i>	
8.2	Examination takers should be provided in advance with as much information about the examination, the examination process, the intended use, examination scoring criteria, examination policy, availability of accommodations, and confidentiality protection as is consistent with obtaining valid responses and making appropriate interpretation of examination scores.
8.6	<i>Transmission of examination taker scores should be protected from improper use.</i>
8.8	<i>When examination scores are used to make decisions, the examination taker should have access to that information.</i>
8.9	<i>Examination takers should be aware of the consequence of cheating.</i>
8.10	<i>In the instance of an irregularity, a examination taker should be informed of any delay in score reporting.</i>
8.11	<i>In the instance where an examination result is invalidated, the examination taker must have access to all information bearing on that decision. Ample opportunity should be available for appeal and claims.</i>
8.12	<i>Examination takers are entitled for fair treatment in the event of an irregularity that prevents a score from being reported or if a score is invalidated. Examination takers should have a means for recourse of any dispute regarding the rejection of a examination score for a decision.</i>

The three candidate guides have been cited often in this evaluation (WREB, 2017a; 2017i, 2017n). Each guide provides specific information about each examination in this examination program. The three guides are published annually. WREB provides a wealth of information to dental hygiene candidates on its website (<https://wreb.org/>). Candidates can contact the WREB office by phone or by email for more information.

As mentioned previously, if candidates have special needs as provided in the Americans with Disabilities Act, WREB provides reasonable and appropriate accommodations. Patients are part of the examination process policies regarding special needs.

Conclusion

The *Standards* (AERA, *et al.*, 2015, pp. 131-135) are clear in chapter 8 about the rights and responsibilities of examination takers. WREB meets these standards fully. WREB is commended for its three candidate guides (WREB, 2017a, 2017i, 2017n). These documents are exemplary as communication tools for candidates, and these documents also provides a variety of well-documented validity evidence that assures the candidates and others about the quality of this examination program. The website also provides information useful to candidates about their rights.

10. Security

Table 13 contains four standards addressing security.

Table 13: <i>Standards Related to Security</i>	
6.7	<i>Examination users have the responsibility of protecting the security of examination materials at all times.</i>
6.14	<i>Examination organizations should have a safe, secure system to store examination information.</i>
6.15	When individual examination data are retained, both the examination protocol and any written report should also be preserved in some form.
6.16	Transmission of individually identified examination scores to authorized individuals or institutions should be done in a manner that protects the confidential nature of the scores and pertinent ancillary information.

The *WREB Policy and Procedures Manual* (WREB, 2015) discusses security. WREB has security processes and policies for both technology hardware and software. Organization data is stored and processed on servers, which run from locked rooms. The server rooms are secured using keypad entry locks, limited to executive and information technology team access. The WREB office suite is locked after normal business hours and only accessible after hours with key card access. Key cards are monitored by building security system. Data regarding office access and video surveillance of building entry ways is monitored and saved by building management company. Besides server security, electronic scoring system hardware is also stored in locked limited keypad access rooms.

As far as organization data is concerned, because data is stored and processed from central servers, critical files are not stored on individual personal computers. A data backup process runs several times per week locally, and once per week offsite. Access in and out of the WREB internal network is guarded by hardware and software fire walls. In case of travel or emergency, WREB staff may have access to office data files remotely. However, access is restricted to specific user roles, only available as needed and facilitated by WREB information technology team.

Offsite critical data is also copied for redundancy and secure. The WREB website is hosted offsite. Candidate data collected through the website is encrypted and verified with licensed SSL certificate. Credit card information from online candidate registrations is not available to WREB staff or saved in a database. Candidate-specific information is available on the website using candidates' individual login accounts. A secured section of the website is also available for examiners who have been approved for access by WREB staff after verifying their access rights to the information.

A primary concern for the local anesthetic computer-based examination is exposure and disclosure of examination items. WREB continually develops and field-examinations new examination items to support multiple examination forms. In addition, all examiners, staff, and observers at examinations, and subject-matter experts must sign a non-disclosure agreement regarding all secure examination material and information.

A primary concern for clinical examinations is candidate identification. Candidates must confirm that all school credentials, personal identification documents, and photographs submitted in support of the examination application are authentic and unaltered, as well as agree to not disclose examination items or other examination-related materials.

WREB reviews security practices regularly from several perspectives: administrative, technological, legal, and psychometric. Potential threats to examination security are identified and prevention and response strategies are discussed.

Conclusion

The procedures for security established over many years are well documented (WREB, 2015; July 12, 2017). WREB has provided excellent validity evidence bearing on security.

11. Documentation

The *Standards* has an entire chapter devoted to this important topic. The two editions of the *Handbook of Examination Development* have chapters devoted to documentation (Becker & Pomplun 2006; Ferrera & Lai, 2016). Having well-documented validity evidence informs users of the examination, candidates, and the public that the examination program concerning validity (Haladyna, 2002). Also, legal challenges to examination score decisions may be avoided with good documentation. Finally, documentation is a source of validity evidence. This evaluation also provides important documentation. Table 14 lists standards pertaining to documentation.

Table 14: <i>Standards</i> Related to Documentation	
7.0	Information relating to examinations should be clearly documented so that those who use examinations can make informed decisions regarding which examination to use for a specific purpose, how to administer the chosen examination, and how to interpret examination scores.
7.1	The rationale for an examination, recommended uses of the examination, support for such uses, and information that assists in score interpretation should be documented. When particular misuse of an examination can be reasonably anticipated, cautions against such misuses should be specified.
7.3	When the information is available and appropriately shared, examination documents should cite a representative set of studies pertaining to general and specific uses of an examination.
7.4	Examination documentation should summarize examination development procedures, including descriptions and the results of the statistical analyses that were used in the development of the examination, evidence of the reliability/precision of scores and the validity of their recommended interpretations, and the methods for establishing performance cut scores.
7.8	Examination documentation should include detailed instructions on how an examination is to be administered and scored.
7.10	Examinations designed to be scored and interpreted by examination takers should be accompanied by scoring instructions and interpretive materials written in a language the examination takers can understand and that assist them in understanding the examination scores.
7.13	<i>Supporting documents should be made available to the appropriate people in a timely manner.</i>

Standard 7.0 is addressed in the three candidate guides (2017a, 2017i, 2017n). Standard 7.1 is well understood by examination developers, dental hygiene schools, and candidates. The examination is used with other information to license dental hygienists in states and other jurisdictions. Standard 7.3 is represented by the annual technical report and periodic evaluations, as shown in this document. Standards 7.4, 7.8, and 7.10 are satisfied through the publication of the three dental hygiene candidate guides. WREB has a large repository of documents made available for this evaluation in Appendix A and B.

Conclusion

WREB is commended for having a very large and comprehensive collection of documents describing this examination program and supplying validity evidence. The appendices provide ample support for this conclusion.

XII: SUMMATIVE EVALUATION

The *Dental Hygiene Examination Program* consists of three independent examinations, each of which is subject to validation. This final section summarizes findings and recommendations of this evaluation.

Local Anesthesia Examination

Having two parts, this examination is subject to extra scrutiny. The two parts of this examination are not statistically related. However, WREB has determined that the information tested in the written test is relevant and important to this aspect of dental hygienist competency.

Dental Hygiene Clinical Examination

This examination is well designed to measure actual practice. Many preliminary conditions must be met before a patient's treatment is evaluated. WREB also has an extensive point deduction aspect that is well validated. This examination appears very well designed and validated.

Restorative Examination

This examination is the best of the three well-designed examinations. Procedures are clearly described, and candidates have ample time to practice. This examination is also very challenging to candidates.

Overall Evaluation of WREB's *Dental Hygiene Examination*

Over many years, WREB has responsibly valued validity and worked to achieve a high degree of validity in interpreting and using examination scores for pass fail decisions. As this evaluation shows, there is a long history of examination development and validation. The appendices give ample support regarding the high degree of documentation of activities and products that contribute to validity.

All examination programs, especially ones where the stakes are high, are in need of further improvement along the road to perfection. WREB has achieved a high degree of refinement. Its organization of committees and annual reviews gives ample testimony to its excellence.

References

- American Association of Dental Examiners (2005). *Guidance for clinical licensure examinations in dentistry*. Chicago: Author.
- American Educational Research Association, American Psychological Association, & National Council on Measurement in Education. (2014). *Standards for Educational and Psychological Examination*. Washington, DC: American Educational Research Association.
- Becker, D. F., & Pomplun, M. R. (2006). Technical reporting and documentation. In S. M. Downing and T. M. Haladyna (Eds). *Handbook of Examination Development*, pp. 711-724. Mahwah, NJ: Lawrence Erlbaum Associates.
- Buckendahl, C., & Plake, B. S. (2006). Evaluating tests. In S. M. Downing and T. M. Haladyna (Eds). *Handbook of Examination Development*, pp. 725-738. Mahwah, NJ: Lawrence Erlbaum Associates.
- Downing, S. M., & Haladyna, T. M. (1996). A model for evaluating high stakes testing programs: Why the fox should not guard the chicken coop. *Educational Measurement: Issues and Practice*, 15(1), 5-12.
- Ferrera, S. & Lai, E. (2016). Documentation to support examination score interpretation and use. In S. Lane, M. R. Raymond, & T. M. Haladyna (Eds). *Handbook of examination development*, 2nd ed. (pp. 603-621). NY: Routledge.
- Geisinger, K. (2016). Test evaluation. In S. Lane, M. R. Raymond, & T. M. Haladyna (Eds). *Handbook of test development*, 2nd ed. (pp. 624-638). NY: Routledge.
- Haladyna, T. M. (2002). Supporting documentation: Assuring more valid test score interpretations (pp. 89-108). In J. Tindal & T. M. Haladyna (Eds.) *Large scale assessment for all students*. Mahwah, NJ: Lawrence Erlbaum Associates.
- Haladyna, T. M., & Rodriguez, M. R. (2013). *Developing and validating test items*. NY: Routledge.
- Haladyna, T. M., & Downing, S. M. (2004). Construct-irrelevant variance in high-stakes testing. *Educational Measurement: Issues and Practice*, 23(1), 17-27.
- Haladyna, T. M., & Hess, R. K. (1999). Conjunctive and compensatory standard setting models in high-stakes testing. *Educational Assessment*, 6(2) 129-153 .
- Kane, M. T. (2006a). Content-related validity evidence. In S. M. Downing & T. M. Haladyna (Eds.) *Handbook of test development* (pp. 131-154). Mahwah, NJ: Lawrence Erlbaum Associates.
- Kane, M. T. (2006b). Validation. In R. L. Brennan (Ed.), *Educational Measurement* (4th ed., pp. 17-64). Westport, CT: American Council on Education, Praeger.
- Kane, M. T. (2016). Validation strategies: Delineating and validating proposed interpretations and uses of test scores. In S. Lane, M. R. Raymond, & T. M. Haladyna (Eds). *Handbook of examination development*, 2nd ed. (pp. 64-80). NY: Routledge.
- Madaus, G. F. (1992). An independent auditing mechanism for testing. *Educational Measurement: Issues and Practice*, 11(1), 26-31.
- McCallin, R. (2006). Examination administration. In S. M. Downing & T. M. Haladyna (Eds.) *Handbook of test development* (pp. 625-652). Mahwah, NJ: Lawrence Erlbaum Associates.
- McCallin, R. (2016). Examination administration. In S. Lane, M. Raymond, & T. M. Haladyna (Eds). *Handbook of test development*, 2nd ed. (pp.567-584). NY: Routledge.
- Mehrens, W. A., & Popham, W. J. (1992) How to evaluate the legal defensibility of high-stakes

- examinations. *Applied Measurement in Education*, 5(3), 265-283.
- Messick, S. (1989). Validity. In R. L. Linn (Ed.), *Educational measurement* (3rd ed., pp. 13–104). New York: American Council on Education and Macmillan.
- Raymond, M. R. (2016). Job Analysis, practice analysis, and the content of credentialing examinations. In S. Lane, M. R. Raymond, & T. M. Haladyna (Eds). *Handbook of test development*, 2nd ed. (pp. 144-164). NY: Routledge.
- Raymond, M. R., & Neustel, S. (2006). Determining the content of credentialing examinations. In S. M. Downing and T. M. Haladyna (Eds). *Handbook of Test Development*, pp. 181-224. Mahwah, NJ: Lawrence Erlbaum Associates.

Appendix A: Archive of Cited Documents Providing Validity Evidence

- American Association of Dental Examiners (2005). *Guidance for clinical licensure examinations in dentistry*. Chicago: Author.
- Haladyna, T. M. (1998). *An Evaluation of the Western Regional Examining Board's 1998 Dental and Dental Hygiene Examination Program*. Phoenix: Author.
- Haladyna, T. M. (2006). *An Evaluation of the Western Regional Examining Board's 2005 Dental and Dental Hygiene Examination Program*. Phoenix: Author.
- Haladyna, T. M. (2010). *An evaluation of the Western Regional Examining Board's 2008 Dental and Dental Hygiene Examination Program*. Phoenix: Author.
- Hammond, D. (December 2009). *WREB Dental Hygiene Practice Analysis Report*. Phoenix, AZ: Author.
- WREB (2009). *Criteria to become a WREB examiner*. Phoenix: Author.
- WREB (2015). *Dental Policies and Procedures*. Phoenix, AZ: Author.
- WREB (August 2016). *WREB Dental Hygiene Process of Care Examination, 2011-2014. Psychometric Overview*. Phoenix, AZ: Author.
- WREB (July 12, 2017). *2016 Technical Report Dental Hygiene Examinations*. Phoenix, AZ: Author
- WREB (2017a). *Dental Hygiene Examination Candidate Guide*. Phoenix, AZ: Author.
- WREB (2017b). *Dental Hygiene Examiner Manual*. Phoenix: Author
- WREB (2017c). *Dental Hygiene Examiner Manual Examiner Forms*. Phoenix: Author.
- WREB (2017d). *Dental Hygiene Examiner Manual General Information*. Phoenix: Author.
- WREB (2017e). *Dental Hygiene Examiner Manual Procedures Criteria-grading*. Phoenix: Author.
- WREB (2017f). *Dental Hygiene Examiner Manual References*. Phoenix: Author.
- WREB (2017g). *Dental Hygiene Required Instruments*. Phoenix: Author.
- WREB (2017h). *Local Anesthesia Chief Examiner and Flow Examiner General Information*. Phoenix; Author.
- WREB (2017i). *Local Anesthesia Examination Candidate Guide*. Phoenix, AZ: Author.
- WREB (2017j). *Local Anesthesia Examiner Manual Reference/tables*. Phoenix: Author.
- WREB (2017k). *Local Anesthesia Examiner Outline of Malamed's Techniques*. Phoenix: Author.
- WREB (2017l). *Local Anesthesia Forms*. Phoenix: Author.
- WREB (2017m). *Restorative Examination: Amalgam and Composite Grading*. Phoenix: Author
- WREB (2017n). *Restorative Examination Candidate Guide*. Phoenix, AZ: Author.
- WREB (2017o). *Restorative Examiner Manual*. Phoenix: Author.
- WREB (2017p). *Restorative Examiner Manual: Criteria*. Phoenix: Author.
- WREB (2017q). *Restorative Examiner Manual: General Information*. Phoenix: Author.
- WREB (2017r). *WREB 2016 Policy Guide* . Phoenix, AZ: Author.
- WREB (2017s). *Local Anesthesia Examiner Manual: Grading Criteria and Procedures*. Phoenix: Author.

Appendix B: Examination Review Board Meetings and Minutes (Documentation)

Local Anesthesia Examination Committee

WREB. (June 13, 2010). *Local anesthesia committee meeting*. Phoenix: Author.
WREB. (June 4-5, 2011). *Local anesthesia committee meeting*. Phoenix: Author
WREB. (June 17-18, 2012). *Local anesthesia committee meeting*. Phoenix: Author.
WREB. (November 3, 2012). *Local anesthesia committee meeting*. Phoenix: Author.
WREB. (June 15-16, 2013). *Local anesthesia committee meeting*. Phoenix: Author.
WREB. (October 10-11, 2013). *Local anesthesia committee meeting*. Phoenix: Author.
WREB. (March 26, 2014). *Local anesthesia committee meeting*. Phoenix: Author.
WREB. (June 14-15, 2014). *Local anesthesia committee meeting*. Phoenix: Author.
WREB. (March 22-23, 2014). *Local Anesthesia Committee Meeting*. Phoenix: Author.
WREB. (June 6-7, 2015). *Local anesthesia committee meeting*. Phoenix: Author.
WREB. (October 1-2, 2015). *Local anesthesia committee meeting*. Phoenix: Author.
WREB. (October 15-16, 2016). *Local anesthesia committee meeting*. Phoenix: Author.

Dental Hygiene Examination Committee

WREB (June 10-12, 2010). *Dental hygiene committee meeting*. Phoenix: Author.
WREB (June 9-11, 2011). *Dental hygiene committee meeting*. Phoenix: Author.
WREB (May 31-June 2, 2012). *Dental hygiene committee meeting*. Phoenix: Author.
WREB (May 30-June 1, 2013). *Dental hygiene committee meeting*. Phoenix: Author.
WREB (February 20, 2014). *Dental hygiene committee meeting*. Phoenix: Author.
WREB (May 29-31, 2014). *Dental hygiene committee meeting*. Phoenix: Author.
WREB (May 28-30, 2015). *Dental hygiene committee meeting*. Phoenix: Author.
WREB (October 3-5, 2015). *Dental hygiene committee meeting*. Phoenix: Author.
WREB (September 22-24, 2016). *Dental hygiene committee meeting*. Phoenix: Author.

Dental Hygiene Examination Review Board

WREB (July 20, 2010). *Dental hygiene examination review board*. Phoenix: Author.
WREB (July 19, 2011). *Dental hygiene examination review board*. Phoenix: Author.
WREB (July 17, 2012). *Dental hygiene examination review board*. Phoenix: Author.
WREB (June 18, 2013). *Dental hygiene examination review board*. Phoenix: Author.
WREB (March 21, 2014). *Dental hygiene examination review board*. Phoenix: Author.
WREB (July 24, 2014). *Dental hygiene examination review board*. Phoenix: Author.
WREB (March 13, 2015). *Dental hygiene examination review board*. Phoenix: Author.
WREB (July 23, 2015). *Dental hygiene examination review board*. Phoenix: Author.
WREB (March, 11, 2016). *Dental hygiene examination review board*. Phoenix: Author.
WREB (June 23, 2016). *Dental hygiene examination review board*. Phoenix: Author.

Restorative Examination

WREB (June 4-5, 2010). *Restorative committee meeting*. Phoenix: Author.
WREB (June 3, 2011). *Restorative committee meeting*. Phoenix: Author.

WREB (June 15-16, 2012). *Restorative committee meeting*. Phoenix: Author.
WREB (June 14, 2013). *Restorative committee meeting*. Phoenix: Author.
WREB (February 20, 2014). *Restorative committee meeting*. Phoenix: Author.
WREB (April 17-19, 2015). *Restorative committee meeting*. Phoenix: Author.
WREB (November 20-21, 2015). *Restorative committee meeting*. Phoenix: Author.
WREB (October 1-2, 2016). *Restorative committee meeting*. Phoenix: Author.