An Evaluation

of the

Western Region Examining Board Dental Examination

Dr. Thomas M. Haladyna
Professor Emeritus
Arizona State University

June 10, 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 expert in testing provides information to candidates, dental schools, and the public that WREB's examination programs are fulfilling the promise that test scores can be validly interpreted and used for making pass/fail decisions for licensing in member and other participating states. The current examination program consists of four independent tests: *Operative*, *Endodontics*, *Periodonta*l, and *Comprehensive Treatment Planning*. Each candidate must pass each of the four tests.

Validity is the key concept in this evaluation. Test scores should be validly interpreted and validly used. The process of validation is an investigative procedure that beings 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 the judgment of validity.

A useful supporting mechanism is the *Standards for Educational and Psychological Testing* (2014). This publication contains specific standards that bear on test development and validation. Throughout this evaluation, standards were identified, organized, and used in conjunction 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) examiner training, (5) examination administration, (6) scaling and comparability, (7) standard setting, (8) 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.

Each of the four tests was individually evaluated. Thus, validity evidence was either collective for all four tests or unique for each test.

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. Threats to validity have been eliminated. No significant weaknesses have been identified. WREB is congratulated for developing and validating its examination program. WREB should be very proud of its achievement regarding this examination program.

Dr. Thomas M. Haladyna
Phoenix, Arizona 2017

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 helped me get correct how the programs should be described as they currently exist. For this I am very appreciative.

Dr. Thomas Haladyna
Phoenix, Arizona 2017

Table of Contents

INTRODUCTION

Examining boards like Western Region Examining Board (WREB) periodically undergo an external evaluation to determine 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 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. Covers the complexity of conjunctive scoring and how validity evidence was organized for this evaluation.
7. Presents validity evidence used for all four tests that comprise the *Dental Examination Program*.
8. Presents unique validity evidence for the *Operative* test.
9. Presents unique validity evidence for the *Endodontics* test.
10. Presents unique validity evidence for the *Periodontal* test.
11. Presents unique validity evidence for the *Comprehensive Treatment Planning* test.
12. Provides a summative evaluation.

For clarity of language, the term *examination or examination program* is used to refer to the entire testing program, which is the *Dental Examination Program*. This examination program consists of four tests.

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

WREB conducts clinical examinations in dentistry and dental hygiene. This organization was formally incorporated in 1976. WREB has provided 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 January 20, 2017. Another good source is any annual report (e.g., WREB, January 23, 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 Exam Review bodies, which allows state board members to focus on examination content, development and oversight. The former Exam 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, includes members selected by the Examination Review Boards and is responsible for strategic, administrative, legal, and financial decisions.

An Examination Review Board oversees the *Dental Examination Program*. WREB has a committee for each of the four tests. These committees meet regularly, review policies and procedures, and recommend changes intended to improve the examination program (see Appendix C). 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 examining boards that affects validity.

No examination program with its battery of tests 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 alone does not determine who receives a license. 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 in that state.

Although WREB is a testing agency that provides validated test scores, it is each state's responsibility that it validly interprets and uses test scores. WREB provides assurances through its documentation of validity that the test 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 tests comprising the examination program that validly measures competence in the profession, and
3. Validation of the interpretation and use of examination scores derived from administering that examination.

Testing specialists have developed a way of thinking about validation (e.g., Kane, 2006a, Kane, 2006b; 2016). One might think of validation as an investigation of validity. This investigation involves many related sequential steps. These 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 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 four independently scored tests, this report will present validity evidence and analysis for each of the four tests.

PART II: DESCRIPTION OF THE *DENTAL EXAMINATION PROGRAM*

Table 1 describes briefly the *Dental Examination Program.* More detailed descriptions of the four tests of this examination appear in the *Dental Exam Candidate Guide* (WREB, 2017d).

Table 1: Tests Comprising the *Dental Examination Program*	
Operative	Candidate must diagnose and treat two restorative procedures on a patient. The procedures are chosen from four options: 1. Direct posterior Class II amalgam restoration 2. Direct posterior Class II composite restoration 3. Direct anterior Class III composite restoration 4. Indirect posterior Class II cast gold restoration
Endodontics	Candidate must perform endodontic treatment of two extracted teeth, mounted in a segmented arch, which is then mounted in an articulated full arch of a manikin.
Periodontal	Candidate must diagnose a need for periodontal treatment and perform scaling and root planing on at least one quadrant of a patient's mouth.
Comprehensive Treatment Planning	The test consists of three patient cases of varying complexity, one of which is a pediatric patient. For each case, candidates assess patient history, photographs, radiographs, and clinical information, create and submit a treatment plan, and then compose case-dependent, constructed-response answers to clinical questions including tasks related to written communication to dental laboratories and pharmacies.

It is the responsibility of each candidate to identify and bring a qualifying patient to the examination site for the two clinical examination sections: *Operative* and *Endodontics*, and *Periodontal*. Each candidate will treat a patient. Candidates bring qualifying extracted teeth for the *Endodontics* test. Highly skilled and calibrated examiners evaluate the results of each candidate's performance. WREB has developed extensive criteria for appointing the examiners (WREB 2009a). Professional judgment is a key element in determining these scoring rules. All judgments are validated, as will be described later in great detail in this evaluation. Performance as determined from ratings is transformed into points using conversion charts that WREB has studied and approved by a committee of its Subject-matter experts through a consensus. These charts also appear in the *Dental Exam Candidate Guide* (WREB, 2017d). Scoring by examiners is mainly objective but rating scales are also used that require subjective judgment. Examiners are trained and calibrated to the rating scale criteria using multiple exemplars at each level of performance. These results are used to compute a total score for each of the four tests comprising this examination. The cut score (passing score) will be described in subsequent appropriate sections. Criteria for scoring and rater consistency and reliability are reviewed in appropriate sections of this report.

PART III: VALIDITY

A test score should represent a candidate's degree of professional competence. If a test 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 test quality are subsumed under validity.

An assessment of validity requires professional judgment about the reasonableness of an interpretation or use of a test score. The *Standards for Educational and Psychological Testing*–hereafter referred to as 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 a test score obtained from any of the *Dental Examination Program* battery of four tests mean? How valid is it for a state to make a pass/fail decision based on a score for each of these four tests? Validity focuses on the meaningfulness of an interpretation of a test 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 dentistry that is usually derived from a practice analysis (Raymond, 2016; Raymond & Neustel, 2006). Then to validate interpretations and uses of examination scores, we need certain elements in this validation:

1. An argument that describes what WREB plans to measure and how test scores will be validly interpreted and used;

2. a claim that the test 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 test 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 four tests.

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 *Dental Examination Program*	
Argument	The Joint Commission on National Dental Examinations administers a *National Board Dental Examination*. This examination measures the knowledge and skills thought to be necessary for safe and competent practice. This examination derives principally from a practice analysis of the profession of dentistry. The WREB *Dental Examination Program* is a battery of performance tests intended to measure clinical competence. It also is based on a practice analysis. These two examination programs represent complementary aspects of dentistry. WREB's *Dental Examination Program* is the capstone in this licensing process for dentists.
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 test standards. WREB's technical reports and other documents cited in this report offer validity evidence supporting this argument. Appendices A, B, and C 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 endeavor to 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 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

Parts VII to XI of this report provides validity evidence for each of the four tests. 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.

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 tests used to measure competence do not match very well the definition of dentistry. ***Fidelity*** is the technical term we use to assess the connection of the tasks on the examination to the definition of dentistry. 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 tests under-represent the construct of competence in dentistry. When we combine the results of the *National Board's Dental Examination* with WREB's *Dental Examination Program*, we have important complementary pieces of information that provide adequate representation of the construct of dentistry. 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 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 test 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 testing 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 competence, then formulate an argument about the validity of using WREB's *Dental Examination Program* test scores as unique, complementary measures of clinical competence. A claim is made by WREB for its member and nonmember states using these test 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 test. Participating states can use this judgment to guide them in deciding if the test score information they receive is adequate for their needs. As mentioned previously, 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* (2014) was published by three large organizations committed to the improvement of testing and in support of valid test score interpretations and uses: American Educational Research Association, American Psychological Association, and the National Council on Measurement in Education. This book contains a comprehensive set of guidelines that help test developers achieve a high degree of validity in the interpretation and use of test scores that result from the administration of their tests.

A large, representative committee of testing experts and other qualified volunteers participated in developing this book. The American Association of Dental Examiners (2005) published *Guidance for Clinical Licensure Examinations in Dentistry*. Although not specifically cited, 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 test 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 test 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 January 23, 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, May 2015).

PART VI: CONJUNCTIVE SCORING AND ITS IMPLICATIONS FOR VALIDITY AND VALIDATION

Two scoring models exist: compensatory and conjunctive. The choice of a scoring model is a significant consideration in the design of any testing program. The choice affects all candidates who take a licensing examination (Haladyna & Hess, 1999). The following presents the two scoring models and what led WREB to choose the conjunctive model.

Compensatory Scoring

With a compensatory scoring model, the sum of the four tests constitutes a test score where a pass/fail decision is made. If a candidate scores are very low in one of these four tests, higher performance in other tests can compensate and result in an overall passing score. Thus, low performance is tolerated if the overall performance is adequate. The compensatory scoring model sets a lower standard for passing and failing. If a candidate scores very low in one of the four tests, does it pose a threat to patient safety? Because the compensatory model combines scores from related tests, reliability tends to be high. If competency is viewed as a holistic concept, the compensatory model is appropriate because the scores from each of the four tests contribute to the overall assessment of competency.

Conjunctive Scoring

Where the sponsors of a credentialing testing program believe that a low score in any test in the battery of tests may infer that the candidate's professional practice may harm a patient, a conjunctive scoring model is preferred to a compensatory scoring model. Each test in the examination program is subject to a pass/fail decision. A candidate cannot meet a state's standards until all tests in the tests are passed. Usually, the tests used in conjunctive scoring represent independent aspects of competence. Conjunctive scoring usually results in lower reliability. Development of more reliable conjunctively scored tests is more extensive and costly.

WREB's Position

WREB has determined that the licensing of dentists involves four important areas of competency and a pass/fail decision is justified for each of these areas. The history of this important transition is well documented in the Examination Review Committee Minutes (Appendix B). At the July 11, 2007 (Appendix B) meeting, the issue of conjunctive versus compensatory scoring was discussed. At the July 9, 2008 (Appendix B) meeting, the motion to use a conjunctive scoring system was passed. As a result, WREB began conjunctive scoring for the examination in 2009. The dental exam was completely compensatory among sections until 2005. Between 2005 and 2009, scoring was partially conjunctive, since sections were considered separately and there was a limit to the compensation possible. There was a minimum score below which a retake of the section was required. The minimum required was 55%. Several types of validity evidence contribute to the wisdom of such a decision. The most important of these considerations is that content areas can be identified and prioritized via a practice analysis. These four content areas must be clearly identified as uniquely important to patient safety. Another consideration for a conjunctive scoring model is that the reliability estimate of each of the test scores is sufficiently high to warrant an accurate pass/fail decision.

PART VII: VALIDITY EVIDENCE THAT APPLIES TO ALL FOUR TESTS

Table 3 lists standards related to validity. Standards appearing in italics are paraphrased for the sake of brevity. Those not appearing in italics are cited verbatim. From that table, standards 1.0 and 1.1 reflect the role that the *Dental Examination Program* results play in state's decision to license a dentist. Test results are clearly part of the criteria used to make a pass/fail decision for licensing in member states. Standard 1.2 mentions validity evidence, which is extant in this evaluation and found in the many documents cited in the appendix. Also, the technical report (WREB, May 2015) provides more evidence supporting 1.2. Standard 1.5 addresses the outcome of a test score use: pass/fail. This is clearly stated in the *Dental Exam Candidate Guide* (WREB, 2017d). The standard 1.7 reflects the tendency for coaching to disaffect a test score interpretation and threaten validity. Coaching for a practical performance test is fair. Thus, 1.7 is in conformance with the design of this testing program. Standard 3.0 presents the idea of construct-irrelevant variance (CIV)–also know as bias. Standard 4.13 also refers to CIV. Later in this technical report, this threat to validity is discussed in greater scope. Throughout this report, such threats are mentioned as they occur in various aspects of the testing program. As shown throughout this report, there are no substantive instances of CIV. Standard 3.1 addresses the design of the test battery to provide validly interpretable results for the widest range of examinees. Throughout this report, such evidence is well documented. Standard 3.4 refers to treating examinees comparably. By that, the battery of four tests comprising the *Dental Examination Program* is standardized. The test items are identical for each examination, and administration is done the same way at each test site. Standard 4.0 may be repetitious, but points to the emphasis on validity, which pervades this report and brings evidence to bear with the argument for validity. In the technical report (WREB, May 2015) and in the *Dental Exam Candidate Guide* (WREB, 2017d), the interpretation and use of test scores in the conjunctive model are very clearly presented.

Table 3: *Standards* Generally Related to Validity	
1.0	Clear articulation of each intended test 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 test developer should set forth clearly how test scores are intended to be interpreted and consequently used. The population(s) for which a test is intended should be delimited clearly, and the construct or constructs that the test is intended to assess should be described clearly.
1.2	A rationale should be presented for each intended interpretation of test 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 test 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 test performance, or a decision made therefrom, is claimed to be essentially unaffected by practice and coaching, then the propensity for test performance to change with these forms of instruction should be documented.
3.0	*Construct-irrelevant variance (CIV) should be avoided in all aspects of test development, administration, scoring, and reporting.*
3.1	Those responsible for test development, revision, and administration should design all steps of the testing

	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	Test developers are responsible for developing tests that measure the intended construct and for minimizing the potential for tests' being affected by construct-irrelevant characteristics, such as linguistic, communicative, cognitive, cultural, physical or other characteristics.
3.4	Test takers should receive comparable treatment during the test administration and scoring process.
4.0	Tests and testing programs should be designed and developed in a way that supports validity of interpretations of test scores for their intended uses.
4.13	When credible evidence indicates that irrelevant variance could affect scores from the test, then to the extent feasible, the test developer should investigate sources of irrelevant variance. Where possible, such sources of irrelevant variance should be removed or reduced by the test developer.
6.0	To support useful interpretation of score results, assessment instruments should have established procedures for test 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	*A clear statement of intended interpretation of a test score and the use to which it is intended should be made clear to test takers.*

Conclusion

Regarding the standards displayed in Table 3, WREB's *Dental Examination Program* appears to achieve all standards. There is ample validity evidence in the documents cited throughout this report to support this conclusion.

Content-related Validity Evidence

The most fundamental basis 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, 2007), which formed the basis for the new 2009 *Dental Examination Program.* A new practice analysis was done and discussed in a meeting (December 18, 2015). Table 4 presents standards bearing on content-related validity evidence.

Table 4: *Standards* Related to Content-related Validity Evidence	
1.11	*The basis for defining and identifying content should be clearly specified.*
1.13	If the rationale for a test score interpretation for a given use depends on premises about the relationships among test items or among parts of the test, evidence concerning the internal structure of the test 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	Test specifications should describe the purpose(s) of the test, 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 test results for the intended purpose(s).
4.2	*Test specifications should be very comprehensive regarding content, test length, item formats, ordering of items and sections, and administration time.*
4.3	*All test development activities should be documented.*
4.12	Test developers should document the extent to which the content domain of a test represents the domain defined in the test specifications.
5.1	Test 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 test content requires a thorough and explicit definition of the content domain of interest.
11.3	*When test content is a primary source of validity evidence, a close link between test content and the profession being assessed is required.*
11.13	*The content domain should be clearly described and justified in light of the professional competency being tested.*

Regarding the standards in Table 4, the *Dental Exam Candidate Guide* (WREB 2017) 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 (WREB, December 18, 2015).

Structure of the Content of the Four Tests

The decision to use conjunctive scoring for each of the four tests depends on many factors. Foremost is the answer to this question: Does low performance on any of the four tests pose a threat to patient safety? If the four tests are highly correlated, then it is unlikely that a single test score would be low and other test scores high. In other words, the four tests measure complementary aspects of dental ability (proficiency). If the four tests are uncorrelated, then it is likely that one score among four for a candidate might be low and pose a threat to patient safety. In that instance, the conjunctive scoring model is highly desirable.

Using data from the 2016 examination, a simple matrix of product-moment correlations produced these results. Correlations among the four sets of test scores for 2,173 candidates ranged from 0.08 to 0.21. Whereas these coefficients are statistically significant, practically speaking these coefficients are very close to zero. If we combined scores from the four tests, a measure of internal consistency, Cronbach's alpha, was a very low 0.34. This coefficient summarizes the information from the correlation matrix of scores from the four tests. Finally, factor analysis is another dependable way to study structure. Using a principal components analysis with equamax rotation for a four-factor solution, the results were virtually perfect. Four equally loaded factors appears, one for each test. Thus, a reasonable conclusion is that these four tests are slightly related but, for the most part, independent. These findings address standards 1.13 and 1.14.

Fidelity

An issue facing all test developers for a clinically-based professional competency examination is whether each test has *fidelity* with its criterion. Fidelity is a judged characteristic of any test by which each test item should resemble or replicate what a practicing professional does in actual clinical practice. Testing agencies need to show how much each test is relevant to professional practice as shown in the practice analysis. For example, a multiple-choice test would have low fidelity for a test of clinical competence, because it measures knowledge not skill. The clinical examination in dentistry has extremely high fidelity because the tasks performed by candidates on the clinical examination resemble those done in actual practice. This statement applies to the *Operative* test, the *Endodontics* test, and the *Periodontal* test, and the *Comprehensive Treatment Planning* test. More information about these tests is found in the *Dental Exam Candidate Guide* (WREB 2017d). WREB has released a statement discussing the pros and cons of using manikins instead of actual patients (http://www.wreb.org/Files/Articles/pos-supt.pdf). WREB's conclusion was to maintain a high-fidelity testing situation and not use manikins on the *Operative* and *Periodontal* tests. The *Comprehensive Treatment Planning* test uses realistic patient materials in a simulated computer environment. The *Endodontics* test has always used manikins because the candidates treat extracted human teeth. WREB's position is that until simulations have high enough fidelity it will not be employed. Since then the fidelity of simulated teeth has improved considerably and WREB will likely be using simulated teeth for endodontics in the future as field tests are underway.

Conclusion

WREB has assembled a comprehensive and appropriate body of evidence supporting the content of the four tests comprising the *Dental Examination Program*. As the four tests in the current examination includes parts of the previous compensatory-scored examination, there is a historical continuity in the content of the tests that suggest stability with respect to content.

Reliability

Table 5 lists seven standards addressing reliability. This section generally addresses reliability and its importance. Subsequent sections address reliability for each of four tests.

Table 5: *Standards* 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 testing procedures, and with the intended interpretation for the use of test scores.
2.5	Reliability estimation procedures should be consistent with the structure of the test.
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	*Method of estimation of reliability should be documented.*
11.14	Estimates of the consistency of test-based credentialing decision should be provided in addition to other sources of reliability evidence.

Reliability coefficients are reported for each test (standard 2.0). The interpretation of reliability is appropriate in the following way (2.2). Test scores are very negatively skewed because the examinees are highly trained to perform in a criterion-referenced way. Thus, reliability estimates tend to be attenuated (weakened) because the statistical procedure used to estimate reliability depends on variance and is usually best when there is a normal distribution instead of a skewed distribution. However, estimates of random error are more important, and the conditional standard error of measurement is the statistic that matters. Inter-judge and intra-judge consistency ratings are studied and monitored (2.7). These are documented for each of the four tests in subsequent sections of this evaluation. Conditional standard error of measurement is reported in the technical report (2.13). The method of estimation of reliability is documented (2.19). The consistency of pass/fail decision is reported for each of the four tests (11.14). However, this index is largely dependent on the number of examinees who score at or near the cut score.

Conclusion

In subsequent sections of this evaluation more evidence is reported bearing on specific tests. Generally, this section addresses the standards for reliability. These standards appear to have been achieved in this testing program, but a more focused review appears subsequently in this evaluation.

Item Quality

Table 6 lists some standards addressing the quality of the performance test items that WREB uses. This section of the evaluation addresses in a general way item quality. In subsequent sections, each test is evaluated vis a vis item quality. As will be shown, there is ample documentation in the technical report (WREB, May 2015) and in this evaluation regarding the quality of items used in each of the four tests.

Table 6: S*tandards* 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 test 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	*Statistical properties of item scores should be studied in an appropriate theoretical context.*

Standards 4.7 and 4.8 are addressed in Appendix C extensively. All items have high fidelity with actual clinical practice. Statistical properties of test items are found in the technical report (WREB, May 2015–in Appendix A).

Conclusion

These standards have been met in the development of this examination program. There is substantial documentation in Appendix C 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 (December 18, 2015–Appendix A).

Examiner Training

Table 7 presents standards addressing examiner training. Examiners may contribute to random and systematic errors that undermine validity. WREB has taken steps to decrease random errors by having many tasks (test items) and three examiners.

Table 7: *standards* Related to Scoring	
1.9	*When candidate performance is judged, the process for identifying, recruiting, training, and monitoring judges should be documented.*
2.7	*Inter-judge and intra-judge consistency of ratings should be studied, monitored, and documented.*
4.18	Procedures for scoring and, if relevant, scoring criteria should be presented by the test 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	*Processes for identifying, training, and evaluating judges should be well developed and documented.*
4.21	*Rater consistency and rater effects should be studied, documented, and, if feasible, improved.*
5.0	Test scores should be derived in a way that supports the interpretations of test scores for the proposed uses of tests. Test developers and users should document evidence of fairness, reliability, and validity of test scores for their proposed uses.
6.8	Those responsible for test scoring should establish scoring protocols. Test scoring that involves human judgment should include rubrics, procedures, and criteria for scoring.
6.9	Those responsible for test 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.

As standard 1.9 requires, WREB has a system for recruiting examiners (WREB, 2009a). WREB has documentation showing how examiners are prepared for the examination (4.18 and 4.20). Each examiner receives an updated *Examiner Manual* (WREB, 2017a). This manual contains general and specific information. The general information addresses basic issues that apply to all clinical tests. The *Examiner Manual* has sections for each clinically-based test. Additional on-site calibration training and testing are accomplished on the day before the beginning of each test administration. The specific training information for each of the relevant tests and sub-test is provided in subsequent sections.

Rater consistency and bias are routinely evaluated. This process is described and reported in subsequent sections of this report. This evaluation bears on standards 2.7 and 4.21. Regarding standard 5.0, the derivation of test scores was based on many committee meetings over many years (see Appendix C). The same is true for 6.8, test scoring. Standard 6.9 is an omnibus standard reflected in all documentation in this section of the evaluation.

Conclusion

WREB has evidence supporting its examiner training activities. This system of training has evolved over many years, and Board-appointed Review Committees serve to improve it continuously.

Examination Administration

Table 8 lists standards related to test administration. McCallin (2006, 2016) provides an extensive analysis of ways that the validity of test score interpretations and uses can be weakened by poor administration practices. Following these standards is one way of contributing to the overall assessment of validity for this examination program.

Table 8: *Standards* Related to Test Administration	
4.16	The instruction presented to test takers should contain sufficient detail so that test takers can respond to a task in the manner that the test 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 test's classification or domain should be provided to the test taker prior to the administration of the test, or should be included in the testing material as part of the standard administration instructions.
6.1	Test administration should follow carefully the standardized procedures for administration and scoring specified by the test developer and any instruction from the test user.
6.4	The testing environment should furnish reasonable comfort with minimal distractions to avoid construct-irrelevant variance.
6.5	Test 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 test scores by eliminating opportunities for test takers to attain scores by fraudulent or deceptive means.
6.7	Test users have the responsibility of protecting the security of test material at all times.

As the *Dental Examination Program* was revised for 2009 administrations, evidence was presented about how dental schools and their deans were informed of this fact (Ramos, October 30, 2008; October 14, 2009a; October 14, 2009b; October 14, 2009c). The most comprehensive and authoritative source of information about examination administration is the *Policy and Procedures Manual* (WREB, 2009b; 2015). This manual describes the application process, how examination results are prepared, travel and shipping, and administrative issues and procedures.

Overall planning for administration is documented showing how test sites and its coordinators need to prepare for examination (Appendix A; WREB, 2017a; 2017b; 2017c; 2017d). Every candidate receives a letter informing the candidate about the orientation day and the three clinical testing days and, also, the computer simulation test and subtest. The *Dental Exam Candidate Guide* (WREB, 2017d) is the most important document 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 four tests in this examination program.

Conclusion

As with examiner training, WREB's extensive use of appropriate subject-matter expert committees has served continuously to improve test administration. Considerable documentation is available to support the conclusion that examination administration is among the strengths of this examination program.

Scaling and Comparability

Table 9 presents standards addressing the test score scale of each of the four tests and why interpretations of test scores must be consistent from administration to administration. If one test were harder or easier than another or if the administration were somehow different resulting in an easier or harder degree of performance that would introduce bias into the testing process that undermines validity.

Table 9: *Standards* 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	Testing programs that attempt to maintain a common scale over time should conduct periodic checks of the stability of scale on which scores are reported.

The substance of standard 5.2 is discussed in the *Dental Exam Candidate Guide* (WREB, 2017d). Standard 5.5 involves the cut score, which is discussed in another section of this evaluation. Nonetheless, the above-cited publication is clear about the pass/fail decision based on a total score for each test. Standard 5.6 is satisfied because the performance test has a natural common scale that is unaltered from time to time. As the items are the same on all test occasions, the scale is consistent no matter when the test 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 four tests have high fidelity with tasks performed by dentists. That assertion is supported by the practice analysis (WREB, December 18, 2015). Candidates are familiar with the performance test items and have opportunities to practice with these tasks in clinical setting in dental schools. Thus, there is transparency among three sequential activities: test preparation, the test, and actual dental practice. Also, there is a strong alignment among these three activities.

Comparability in this examination is considered test by test. Clinical performance tests have standardized administration and scoring using a standardized scale with the *Operative*, *Endodontics*, and *Comprehensive Treatment Planning* tests cut score set at three on a five-point scale. The Periodontal test cut score is set at 75 out of 100 points possible. All four tests are scored by trained, validated examiners. Thus, scores from these tests are be considered as having comparable score scales from test site to test site across years due to the standardization that is evident in all aspects.

Appendix C contains an extensive list of documents showing the development of these tests and the high degree of standardization achieved.

Conclusion

The *Operative*, *Endodontics*, and *Comprehensive Treatment Planning* tests have a common scale of one to five. The *Periodontal* test has a scale ranging from zero to 100 with each error resulting in a 12.5 point deduction. The cut scores have been defined to represent performance that reflects a minimally competent professional that may safely begin at entry-level practice. Thus, equating is not recommended or possible. That is, the test score scale is constant across administrations. The *Comprehensive Treatment Planning* test does employ equating weights across patient cases to ensure test form comparability.

Standard Setting

Setting the cut score is a very important activity, because it determines who will pass or fail. Table 10 lists some standards to guide in evaluating how the cut score was set for each test.

Table 10: *Standards* 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 rational for recommended score interpretations should be explained clearly.
5.21	When proposed test 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 test performance to the relevant criteria.
11.16	The level of performance required for passing a credentialing test 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 test.

Conclusion

WREB has reviewed evidence on score interpretations around cut scores for each test, and its subject-matter experts have endorsed the current cut scores as appropriate.

Reporting

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

Table 11: *Standards* Related to Score Reporting	
6.10	When test score information is released, those responsible for testing programs should provide interpretations appropriate to the audience. The interpretations should describe in simple language what the test covers, what the scores represent, the precision/reliability of the scores, and how scores are intended to be used.
6.14	*Test organizations should maintain confidentiality and protect the rights of test takers.*
6.15	When individual test data are retained, both the test 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 test content and purposes that is available to any test taker prior to testing should be made available to all test takers.
8.5	*Policies for release of test scores should be carefully considered and clearly recommended. Release of scores should be consistent with the purpose of the test and in consideration of the test takers and informed consent.*

A candidate score report should present test 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. An inspection of a typical school score report and the individual score report shows clear comprehensive candidate performance that includes previous attempts, point deductions, and performance points for each test. Passing candidates receive a report of passing. Failing candidates receive additional details of performance. Passing candidates may receive additional details upon request.

The *Dental Exam Candidate Guide* (WREB, 2017d, p.4) describes the scoring and the score report. The score report is designed to reveal candidate performance in all aspects of the examination on a point basis against possible points to be earned. Confidentiality of candidates' results is ensured. Candidates graduating from dental schools have the option of withholding their score report from their school. No other parties 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. Standards 6.10, 6.14, 6.15, 6.16, 8.1, and 8.5 appear to be satisfied.

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.

Candidate and Patient Guide and Rights

Table 12 presents standards related to candidates' rights. The sponsors of 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: *Standards* Related to the Rights of Test Takers	
8.2	Test takers should be provided in advance with as much information about the test, the testing process, the intended use, test scoring criteria, testing policy, availability of accommodations, and confidentiality protection as is consistent with obtaining valid responses and making appropriate interpretation of test scores.
8.6	*Transmission of test taker scores should be protected from improper use.*
8.8	*When test scores are used to make decisions, the test taker should have access to that information.*
8.9	*Test takers should be aware of the consequence of cheating.*
8.10	*In the instance of an irregularity, a test taker should be informed of any delay in score reporting.*
8.11	*In the instance where a test result is invalidated, the test taker must have access to all information bearing on that decision. Ample opportunity should be available for appeal and claims.*
8.12	*Test 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. Test takers should have a means for recourse of any dispute regarding the rejection of a test score for a decision.*

The *Dental Exam Candidate Guide* (WREB, 2017d) has been cited often in this report. This booklet contains essential information for candidates. The table of contents for this 86-page booklet provides general information, performance evaluation information, patient criteria, and examination procedures. The booklet is published each year and is updated as needed. Besides this guide, WREB's web page is helpful, and if candidates prefer can contact the WREB office by phone or by email for more information.

WREB also issues regular newsletters of general interest to dental students (WREB, Fall, 2015; 2016). A review of these newsletters will reveal that candidates are informed about various issues they might encounter in their attempt to pass this important examination. Some of these issues are appeals process, score information, characteristics of successful candidates, advice on test preparation, choosing patients, application process, scoring procedures, and examination calendars.

If candidates have special needs as provided in the Americans with Disabilities Act, WREB provides reasonable and appropriate accommodations (see *Policy and Procedures Manual*, WREB, 2015 and WREB, 2017d, p. 5). Patients are part of the examination process. Considerable attention is given to patient rights and care (WREB, 2017, 13-16). An annual survey is conducted of patients inquiring about their treatment and satisfaction (http://www.wreb.org/Files/Articles/PtSurveyart.pdf)

Conclusion

Information and references to information about candidates' rights have been presented here and appear in the archive. The *Standards* (AERA, *et al.*, 2014, pp. 131-135) are clear in chapter 8 about the rights and responsibilities of test takers. WREB meets these standards fully. WREB is commended for its *Dental Exam Candidate Guide* (WREB, 2017d) and its frequent and informative newsletters. 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 testing program.

Security

For tests where items are not disclosed, security is paramount. For WREB, security is less of a threat, because the items are tasks that candidates can practice. Also, candidates are fully informed about how the tasks are presented and how performance is scored. Nonetheless, cheating and other threats to validity are omnipresent in all high-stakes tests. Table 13 contains four standards addressing security.

Table 13: *Standards* Related to Security	
6.7	*Test users have the responsibility of protecting the security of test materials at all times.*
6.14	*Testing organizations should have a safe, secure system to store test information.*
6.15	When individual test data are retained, both the test protocol and any written report should also be preserved in some form.
6.16	Transmission of individually identified test 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 PCs. 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.

With any high-stakes examination, the temptation to cheat is great. WREB requires each candidate to be clearly and accurately identified and monitored during the examination process.

Each candidate must have their patient qualified. Failure to do this well or do it at all puts candidates in jeopardy of failing.

It is extremely unlikely that a single examiner could undermine the validity of any examination. 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 trained and monitored during the administration to reduce this possibility. The integrity of examiners is discussed in the *Dental Examiners Manual* (WREB 2017a). Conflicts of interest with candidates are monitored, and examiners are asked to recuse themselves if potential conflicts exist. Finally, as reported previously, WREB has high standards for selecting examiners (WREB, 2009a). Finally, the *Dental Exam Candidate Guide* (WREB, 2017d, pp 12) also provides information to candidates about security.

Conclusion

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

Documentation

Table 14 displays standards addressing documentation. The *Standards* has an entire chapter devoted to this important topic. The two editions of the *Handbook of Test Development* have chapters devoted to documentation (Becker & Pomplum 2006; Ferrera & Lai, 2016). These authors provide several reasons why documentation is so important. Having well-documented validity evidence informs users of the test, candidates, and the public that the testing program is properly developed and the interpretation and use of test scores are highly valid (Haladyna, 2002a). Also, legal challenges to test score decisions may be avoided with good documentation. Finally, documentation is a source of validity evidence.

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

Standard 7.0 is clearly addressed in the *Dental Exam Candidate Guide* (2017d). Standard 7.1 is well understood by test developers, dental schools, and candidates: the test is used with other information to license dentists 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 *Dental Exam Candidate Guide* and the documents appearing in Appendix A, B, and C. WREB has a large repository of documents that have been made available for this evaluation as the appendices attest.

Conclusion

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

PART VIII : *OPERATIVE* TEST VALIDITY EVIDENCE

Each candidate completes two restorative procedures on a patient. The procedures are chosen from four options:

1. direct posterior Class II amalgam restoration,
2. direct posterior Class II composite restoration,
3. direct anterior class III composite restoration, and
4. indirect posterior class II cast gold restoration.

As of 2016, a direct posterior Class II composite restoration is required and the second procedure may be chosen from all four restorative procedures (including a second direct posterior Class II composite restoration).

For each restorative procedure, the candidate does a preparation and a finish. For each preparation and finish, three examiners provide three ratings of performance.

Test Development

The transition of the *Operative* test to conjunctive scoring is well documented in committee minutes (See Appendix C). The *Operative* test is continually being evaluated, and these minutes reveal many fine-tuning adjustments to the examination with the intent of improving this test over these years.

Item Quality

The tasks to be performance and the descriptive rating scales are presented in the *Dental Exam Candidate Guide* (WREB, 2017d, pp. 24-56). As noted in the content-related validity section, the tasks are those done by dentists in practice, so each task has high fidelity with the criterion behavior it is supposed to represent. The five-point descriptive rating scales are well written and have the added feature of having the median rating represent the cut score for making pass/fail decision. Thus, each examiner determines if the minimum passing has been met and then assigns a rating of three, four, or five according to the perceived level of performance exhibited by the candidate. These rating scales appear on pages 42-45 and 53-55.

Training of Scorers and Scoring

The method of scoring for the *Dental Examination Program* was presented to the Examination Review Committee on July 8, 2008. This document provides information about the complex scoring arrived by the examination committee. The training materials for *Operative* test are extensive entailing documents that address a tutorial overview, a self-test, a self-test key, a preparation tutorial and a preparation self-test, and a preparation self-test key, and an operative finish tutorial, self-test, and key. The *Examiner Manual* (WREB 2017a, pp. 37-76) provides extensive information about what examiners need to know and do for the *Operative* test.

Three calibrated grading examiners grade the preparation and finish for each procedure. The medians are weighted and summed. Where applicable, points are deducted for penalties incurred. The chart that follows shows a record of point deductions for candidates for the past three years. With a point deduction, there is consensus among three examiners that such a deduction is valid.

Table 15: Point Deductions for Candidates						
Category	2014 N = 2,257		2015 N = 2,295		2016 N = 2,275	
Modification Request Penalty	10.9%	246	10.8%	249	10.5%	229
Submission Rejection Penalty	6.2%	141	7.2%	166	4.9%	108
Pulp Exposure	2.5%	56	2.0%	45	2.2%	49
Caries Remaining	1.3%	30	1.0%	22	1.6%	36
Late Penalties	1.2%	27	0.8%	19	0.6%	15
Wrong Material	0.6%	14	0.1%	3	0.1%	4
Wrong Surface	0.1%	3	0.2%	4	0.1%	5
Wrong Tooth	0.1%	2	<0.1%	1	0.1%	1

Is There Bias in Choice?

As candidates can choose among four procedures, is there bias in the choices? In the *WREB 2015 Technical Report* (WREB, May 2015, p. 25), analysis is presented showing that choice of a procedure did not materially affect performance.

Examiner Consistency

Examiner consistency is reported in Table 11 of the *WREB 2015 Technical Report* (WREB, May 2015). For the *Operative* test, there was a mean 89.6% agreement. The degree of agreement ranged from 80.4% to 94.8%.

Examiner Bias (Lenience and Harshness/Severity)

As noted earlier in this report, some examiners have a tendency to be too lenient or harsh in their ratings. Such tendencies threaten validity. On pages 26 and 27 of the technical report (WREB, May 2015), examiner leniency and harshness were assessed. The few instances of a very lenient or harsh rating were not consequential.

Reliability Estimate and the Conditional Standard Error of Measurement

In the *WREB 2015 Technical Report* (May 2015, p. 28) reports that the reliability estimate

for the total score was 0.87, and the conditional standard error of measurement (CSEM) was 0.018. The objective of estimating the CSEM is to observe the number of candidates who scores were in a range of uncertainty around the cut score due to random error. In this instance, the CSEM was 0.018, which is very small. In the data file of 2,275 candidates 15 had scores ranging from 2.99 to 3.01. Thus, due to random error some of these students might be misclassified. It would be very difficult and expensive to increase reliability and reduce the CSEM. As most candidates scored well above this zone of uncertainty, the few candidates whose scores fall in this zone of uncertainty near the cut score need to improve their performance if they are to receive a more confident pass decision.

Conclusion

The *Operative* test is very complex. The content of this test is based on the results of the practice analysis as interpreted by a committee of subject-matter experts. The procedures for this test are very clearly described in the *Dental Exam Candidate Guide* (WREB, 2017d) and the *Examiner Manual* (WREB, 2017a). Scoring is consistent. The reliability estimate is very high.

PART IX: *ENDODONTICS* TEST VALIDITY EVIDENCE

The *Endodontics* test provides for the treatment of two extracted teeth: one anterior and one multi-canal posterior. The *Dental Exam Candidate Guide* (WREB, 2017d, pp 57-69) describes the testing process. Scoring is described on pages 70 to 72.
Descriptive rating scales appear on page 71. If a tooth is rejected on the first or second trial, there is no point deduction. However, on the third rejection, no additional submissions are allowed and no points are awarded. Criteria for rejection are provided on page 70. The *Examiner Manual* (WREB, 2017a, pp. 89-98) also provides much information about this test.

Test Development

The Endodontic Committee meets regularly to discuss issues and policy related to the test (See Appendix C). Included in these minutes are discussions and a consensus reached on scoring criteria and administration procedures ensuring high reliability for the conjunctive scoring.

Item Quality

The *Dental Exam Candidate Guide* (WREB, 2017d, pp. 48-69) provides a comprehensive description of the performance tasks and how they are scored. This section provides a test overview, which teeth to select, which kinds of teeth are not acceptable, how to prepare teeth, radiograph submission, treatment, and after treatment. The test is well supported by previous test administration procedures. As reported in the technical report, test items/tasks are the same for all examinations (WREB, May, 2015). Item discrimination indexes are very high. Another source on item quality is the *Examiner Manual* (WREB, 2017a) and the WREB 2015 Technical Report Dental Examination (WREB, May 2015).

Training of Scorers and Scoring

The training materials for these examiners include an orientation and calibration session and another session devoted to basic principles of endodontic access followed by a self-test and a self-test key. The *Examiner Manual* (WREB, 2017a, pp. 77-98) provides general information, approval criteria and procedures, scoring, definitions, reference material, and a floor examiner job description. A total score is the median and not the mean of three examiner ratings. This procedure is defensible because for small samples, the median is the most representative measure of central tendency. On a five-point scale, means vary between 3.65 and 3.88.

Examiner Consistency and Bias

According to the technical report (WREB, May 2015), examiner consistency had a mean of 90.3% with a range from 78.9% to 100%. Harshness had a mean of 4.8% with a range of 0.0 to 16.3%. Lenience also had a percentage of 4.8 with a range of 0.0 to 21.1.

Reliability and the Conditional Standard Error of Measurement

On page 28 of the technical report (WREB, May 2015), a reliability estimate for *Endodontics* was reported as 0.81. The CSEM was 0.013. As noted previously, the CSEM is the important, consequential statistic because when applied to the cut score shows how many candidates are trapped in this zone of uncertainty regarding their pass/fail status. As with the *Operative* test, the number of candidates whose scores reside in the zone of uncertainty due to random error was counted. Sixty candidates had scores that may be been misclassified due to random error. Of the 2,277 total number, 60 is a very low number. As with the *Operative* test, reducing the CSEM is difficult. All these candidates scores were 3.0 so they were classified as passing.

Conclusion

The *Endodontics* test is well developed, and it has adequate validity evidence to support its use for making pass/fail decisions. The fact that 60 candidates in this data set scored 3.0 and were observed in this zone of uncertainty might give pause for concern, because all of them passed. Most candidates scored much higher on this test.

PART X: *PERIODONTAL* TEST VALIDITY EVIDENCE

The practice analysis provided the content basis for this test (WREB, May 18, 2015). The *Periodontal* test requires the candidate to provide a qualifying patient for periodontal treatment. The candidate will perform scaling and root planing in one or two quadrants. The test is described on pages 73-77 of the *Dental Exam Candidate Guide* (WREB, 2017d). Extensive documentation of test development can be found in Appendix C. Scoring is described on pages 78 to 80.

Item Quality

The test items (tasks) are scaling and root planing of teeth in one or two quadrants. These are performance tasks that are very familiar to all candidates. Performance on these tasks is expected to be perfect or nearly perfect. The instructions for the test are very detailed and clearly presented. The items are identical from one test administration to another. The only variation is the patient, who the candidate must select in accordance with the acceptance criteria listed in the *Dental Exam Candidate Guide* (WREB, 2017d). However, the procedures are not patient-dependent, so no bias is introduced.

Training of Scorers and Scoring

The *Examiners Manual* (WREB, 2017a, pp. 99-106) provides general information, treatment criteria, check-in criteria, reasons for patient rejection, and scoring information. The *Dental Exam Candidate Guide* (2017d, pp. 78-80) also provides information about scoring.

Scoring is based on deductions. A patient rejection results in a 10% deduction from the total possible points. Major tissue damage that is validated by two or three examiners results in a loss of points. There are also late penalties.

Rater Consistency and Bias

Descriptive statistics are reported in the *WREB 2015 Technical Report Dental Examination* (WREB, May 2015, pp. 25-29). Three examiners score eight tooth surfaces. As shown in the technical report, rater agreement is 83.5% with the range being 72.6% to 88.8% (WREB, December 18, 2015). Harshness and leniency were also evaluated. Harshness was 4.4% and lenience was 0.8%.

Reliability

Because performance on this test is extremely high, the estimate of reliability is 0.10. The conditional standard error of measurement is 6.69 referring to a 100-point scale. Of 2,275 candidates, 62 were trapped in this zone of uncertainty due to random error. These candidates received passing scores. As with the other two tests, these candidates had very low performance that placed them in jeopardy of being incorrectly placed as passing or failing.

Conclusion

This test is very well developed and provides validly interpretable results. The restriction in the range of performance and the high uniform performance results in a low reliability estimate, but the CSEM is the statistic that matters. There is little WREB can do to increase reliability. Most candidates score in the high 90%s. Those scoring close to 75% barely passed. Is there concern about the ability of these borderline candidates on this test? If so, what remedies exist to improve this situation? The *Periodontal* test has a very high passing rate, so data on low-scoring candidates is limited. However, results show that while most candidates, regardless of final score upon initial unsuccessful attempt, pass upon retake. Those with the lowest scores (i.e., 50 or below) are less likely to pass. Passing percentages tracked over multiple years indicate that a small percentage of candidates remain unsuccessful on the *Periodontal* test after multiple retakes (that is, 0.2% over six years (2011 to 2016)).

PART XI: *COMPREHENSIVE TREATMENT PLANNING* TEST (CTP) VALIDITY EVIDENCE

The development of this test is well documented (see Appendix C). *Comprehensive Treatment Planning Examiner Manual* (WREB, 2017b) gives description of this test. The test has been under development for some time (see Appendix C). The *Comprehensive Treatment Planning Exam Candidate Guide* (WREB, 2017c) describes the test. The CTP presents three patient cases, one of which is pediatric. The candidate must assess patient history, photos, radiographs, and other clinical information. Then the candidate must create and submit a treatment plan. After that, the candidate must compose case-dependent constructed-response answers to clinical questions including tasks related to written communication to dental laboratories and pharmacies. All candidates receive access to examples and an online tutorial and complete a 15-minute tutorial before the test begins.

Test Development

Test development is well documented (see Appendix C). The test consists of performance items and some short-answer essay items. All performance and responses to test items are evaluated by three examiners using descriptive rating scales. All test items are based on test specifications that derive from the practice analysis.

Item Quality

The best source of information about the test items can be found in the *2017 Comprehensive Treatment Planning Exam Candidate Guide* (WREB, 2017c). As with the other three tests, all test items are open-source, so candidates can practice the professional skills needed to treat patients safely. Specifications for treatment plans developed by candidates are stated on pages 2 and 3. Scoring is described on pages 3 and 4. Advice to candidates is provided on subsequent pages. Scoring guides are provided on pages 8 and 9. Sample treatment plans are provided on pages 11 to 13. Additional test items are provided on page 14.

Training of Scorers and Scoring

The best source of information about the training of scorers and scoring can be found in the *Comprehensive Treatment Planning Examiner Manual* (WREB, 2017b). A review of this booklet provides information about candidate material, treatment plans, supplemental questions, scoring, and details concerning administration.

From the *WREB 2015 Dental Exam Technical Report* (WREB, May 2015), mean scores for this test range from 2.38 to 4.90 on a five-point scale. Raw score standard deviations range from 0.49 to 1.70. In this technical report, 7,257 procedures/cases were scored (May 2015).

Examiner Consistency and Bias

As shown in the technical report (May, 2015), agreement percentage was 83.5% ranging from 72.6% to 88.8%. Harshness had the highest degree reported for the four tests, 8.6%. Lenience also had the highest degree reported among the four tests, 7.9%.

Reliability and the Conditional Standard Error of Measurement

The reliability estimate was 0.87 and the conditional standard error of measurement was 0.019. Those candidates whose scores were higher than 2.981 and less than 3.019 fall in this zone of uncertainty. Of the 2,275 candidates in the data file, just five candidates had a score in this zone of uncertainty. These results foreshadow very high confidence in making pass/fail decisions.

Conclusion

This test elicits comprehensive problem solving and patient treatment. The test has shown remarkable improvement over recent years. It stands out as an exemplary test of clinical competence.

XII: SUMMATIVE EVALUATION

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

Operative Test

Based on the validity evidence reviewed, this test appears well validated for test score interpretations and pass/fail decisions. The content basis for the test, item quality, administration procedures, examiner training and validation, standard setting and comparability, and examiner consistency and reliability all are very good.

Periodontal Test

This test consisted of two clinical periodontal sub-tests and a written periodontal test that measured independent, uncorrelated procedures. The written periodontal test was developed and administered by a joint venture among three testing agencies, including WREB. Since 2015, the *Periodontal* test consists of the periodontal clinical procedures. Periodontal assessment is now tested within the scope of the *Comprehensive Treatment Planning* test. Validity evidence supports the use of the test scores for making pass/fail decisions.

Endodontics Test

Based on the validity evidence reviewed, this test appears well validated for test score interpretations and pass/fail decisions. The content basis for the test, item quality, administration procedures, examiner training and validation, standard setting and comparability, and examiner consistency and reliability all are very good.

Comprehensive Treatment Planning Test

This test represents a high degree of sophistication needed in professional licensing. It emphasizes patient problem solving and treatment in a complex way using performance tasks and short-answer essays scored by three examiners. Like the other three tests, it is well validated.

Overall Evaluation of WREB's *Dental Examination*

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

All testing 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 Testing*. Washington, DC: American Educational Research Association.

Becker, D. F., & Pomplum, M. R. (2006). Technical reporting and documentation. In S. M. Downing and T. M. Haladyna (Eds). *Handbook of Test 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 Test 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, & Lai, (2016).

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. (2002a). 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. (2002). Research to improve large scale testing. pp. 483-497. In J. Tindal & T. M. Haladyna (Eds.) *Large scale assessment programs for all students*. Mahwah, NJ: Lawrence Erlbaum Associates.

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 test development,* 2nd ed. (pp. 64-80). NY: Routledge.

Lane, S., Raymond, M., & Haladyna, T. M. (2016). *Handbook of Test Development*, 2nd ed. NY: Routledge.

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

McCallin, R. (2006). Test administration. In S. M. Downing & T. M. Haladyna (Eds.) *Handbook of test development* (pp. 625-652). Mahwah, NJ: Lawrence Erlbaum Associates.

McCallin, R. (2016). Test 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

tests. *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.

Ramos, D. (October 30, 2008). *Changes for the examination year 2009.* Phoenix: WREB.

Ramos, D. (June 23, 2009). *Description of scoring and standard setting*. Phoenix: WREB.

Ramos, D. (October 14, 2009a). *School coordinator preparation instructions: 2010 WREB Dental Examination*. Phoenix: WREB.

Ramos, D. (October 14, 2009b). *Examination year preparation*. Phoenix: WREB.

Ramos, D. (October 14, 2009c). *2010 school information packets*. Phoenix: WREB.

WREB (January 11, 2003). *Western Regional Examining Board By Laws* (As amended by the Membership. Phoenix: Author.

WREB (2007). *WREB Practice Analysis for General Dentist.* Phoenix, AZ: Author

WREB (January 7, 2007). *Amendment and restatement of bylaws of WREB*. Phoenix: Author.

WREB (May 15, 2007). *Practice analysis for general dentist*. Phoenix: Author.

WREB (July 8, 2008). *Proposed scoring detail*. Phoenix: Author.

WREB (2009a). *Criteria to become a WREB examiner*. Phoenix: Author.

WREB (2009b). *Policy and procedures manual*. Phoenix: Author.

WREB (2009c). *WREB Annual report*. Phoenix: Author.

WREB (2010a). *Examiner standardization*. Phoenix: Author.

WREB (2010b). *Annual report for 2009*. Phoenix: Author.

WREB (2015). *Dental Policies and Procedures.* Phoenix, AZ: Author.

WREB (Fall 2015). *Dental student newsletter*. Phoenix: Author.

WREB (May 2015). *WREB 2015 Technical Report Dental Examination*. Phoenix, AZ: Author.

WREB (December 18, 2015). *Dental Practice Analysis Committee Meeting*. Phoenix, AZ:

WREB (Fall 2016). *Dental student newsletter*. Phoenix: Author.

WREB (2017a). *Examiner Manual*. Phoenix, AZ: Author.

WREB (2017b). *Comprehensive Treatment Planning Examiner Manual*. Phoenix, AZ: Author.

WREB (2017c). *Comprehensive Treatment Planning Exam Candidate Guide*. Phoenix, AZ: Author.

WREB (2017d). *Dental Examination Candidate Guide*. Phoenix, AZ: Author.

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

WREB (July 5, 2005). *WREB Board of Directors meeting minutes*. Phoenix: Author.
WREB (January 7, 2006). *WREB Board of Directors meeting minutes*. Phoenix: Author
WREB (July 19, 2006). *Examination Review Committee meeting minutes*. Phoenix: Author.
WREB (July 20, 2006). *WREB Board of Directors meeting minutes*. Phoenix: Author.
WREB (January 7, 2007). *WREB Board of Directors meeting minutes*. Phoenix: Author
WREB (July 11, 2007). *Examination Review Committee Meeting minutes*. Phoenix: Author.
WREB (July 12, 2007). *WREB Board of Directors meeting minutes*. Phoenix: Author.
WREB (July 9, 2008). *Examination Review Committee meeting minutes.* Phoenix: Author.
WREB (January 5, 2008). *WREB Board of Directors meeting minutes*. Phoenix: Author.
WREB (July 10, 2008). *WREB Board of Directors meeting minutes*. Phoenix: Author.
WREB (July, 15, 2009). *Examination Review Committee meeting minutes.* Phoenix: Author.
WREB July 16, 2009). *WREB Board of Directors meeting minutes*. Phoenix: Author.
WREB (January 30, 2010). *WREB Board of Directors meeting minutes*. Phoenix: Author.
WREB (July 21, 2010). *Dental Exam Review Board–Minutes*. Phoenix, AZ: Author.
WREB (July 20, 2011). *Dental Exam Review Board–Minutes*. Phoenix, AZ: Author.
WREB (July 18, 2012). *Dental Exam Review Board–Minutes*. Phoenix, AZ: Author.
WREB (July 19, 2013). *Dental Exam Review Board–Minutes*. Phoenix, AZ: Author.
WREB (February 21, 2014). *Dental Exam Review Board–Minutes*. Phoenix, AZ: Author.
WREB (July 25, 2014). *Dental Exam Review Board–Minutes*. Phoenix, AZ: Author.
WREB (October 17-19, 2014). *Dental Exam Review Board–Minutes*. Phoenix, AZ: Author
WREB (December 13-14, 2014). *Dental Exam Review Board–Minutes*. Phoenix, AZ: Author
WREB (February 13, 2015). *Dental Exam Review Board–Minutes*. Phoenix, AZ: Author.
WREB (June 24, 2016). *Dental Exam Review Board–Summary to Member States*. Phoenix, AZ: Author.
WREB (January 23, 2017). *Dental Exam Review Board–Executive Summary*. Phoenix, AZ: Author.

Baker, J. (March 14, 2003). *Dentistry interactive computerized examination system for prosthetics. Test specifications.* Victoria, Australia: Zoomorphix.

Baker, J. (November 11, 2004). *Specifications for data exchange between CSW exam builders and testing agency software* (CRDTS, SRTA, WREB data bases. Victoria, Australia: Zoomorphix Systems.

Baker, J. (June 24, 2005). *Periodontal report*. Victoria, Australia: Zoomorphix.

CSW (January 29, 2005). *CSW Joint Prostodontic Committee minutes*. Phoenix

CSW (September 9-10, 2005). *Periodontal Committee Meeting Minutes*. Phoenix: Author.

CSW (October 27, 2005). *Periodontal Committee Meeting Minutes*. Phoenix: Author.

CRDTS/SRTA/WREB (November 14-15, 2003). *Joint Periodontal Committee Minutes*. Phoenix, Author.

CRDTS/SRTA/WREB (February 14-15, 2004). *Joint Periodontal Committee Minutes*. Phoenix: Author.

CRDTS/SRTA/WREB (March 27-28, 2004). *Joint Periodontal Committee Minutes*. Phoenix: Author.

CRDTS/SRTA/WREB (July 10-11, 2004) *Joint Periodontal Committee Minutes*. Phoenix: Author.

CRDTS/SRTA/WREB (January 28-29, 2005) *Joint Periodontal Committee Minutes.* Phoenix: Author.

CRDTS/SRTA/WREB (June 24-25, 2005). *Joint Periodontal Committee Minutes*. Phoenix: Author.

CRDTS/SRTA/WREB (December 10, 2005). *Joint Periodontal Committee Minutes*.

Hammond, D., & Buckendahl, C. (March 2007). *Technical report for CSW computer simulations LLC.* Phoenix: CSW.

Hammond, D., & Buckendahl, C., (March 2008). *Technical report for CSW computer simulations LLC*. Phoenix: CSW

Hammond, D. (June 2009). *Technical report for CSW computer simulations* LLC. Phoenix: CSW.

Hammond, D. (February 2010). *Technical report for CSW computer simulations LLC.* Phoenix: CSW.

WREB (August 15-16, 1998). *Prosthetics committee meeting*. Phoenix, Author.

WREB (January 9-10, 1999). *Prosthetics committee meeting*. Phoenix, Author.

WREB (April 17, 1999). *Prosthetics committee meeting*. Phoenix, Author.

WREB (April 23, 1999). *Draft test specifications–prosthetics*. Phoenix, Author.

WREB (September 25-26, 1999). *Prosthetics committee meeting*. Phoenix: Author.

WREB (September 25-26, 1999). *Prosthetics test specifications*. Phoenix: Author.

WREB (October 8, 1999. *Prosthetics committee meeting*. Phoenix, Author.

WREB (December 11-12, 1999). *Prosthetics committee meeting*. Phoenix, Author.

WREB (August 18-19, 2001). *Prosthetics committee meeting*. Phoenix: Author.

WREB (November, 10-11, 2001). *Prosthetics committee meeting*. Phoenix: Author.

WREB (February 15, 2002). *Prosthetics committee task force meeting*. Phoenix: Author

WREB (August 2-3, 2002). *Prosthetics committee meeting minutes*. Phoenix: Author.

WREB (October 19, 2002). *Prosthetics committee meeting minutes*. Phoenix: Author.

WREB (February 1-2, 2003) *Periodontal committee meeting minutes*. Phoenix: Author.

WREB (February 14, 2003). *Prosthetics committee meeting minutes*. Phoenix: Author.
WREB (July 25-27, 2003). *Prosthetics committee meeting minutes*. Phoenix: Author.
WREB (May 19-20, 2006). *Operative test committee meeting minutes*. Phoenix: Author.
WREB (September 8-9, 2006). *Operative test committee meeting minutes*. Phoenix: Author.
WREB (February 2-3, 2007). *Operative test committee meeting minutes*. Phoenix: Author.
WREB (September 7-9, 2007). *Operative test committee meeting minutes*. Phoenix: Author.
WREB (November 30-December 1, 2007). *Operative test committee meeting minutes*. Phoenix: Author.
WREB (February 29-March 1, 2008). *Operative test committee meeting minutes*. Phoenix: Author.
WREB (May 9-10, 2008). *Operative test committee meeting minutes*. Phoenix: Author.

WREB (March 6, 2009). *Operative test committee meeting minutes*. Phoenix: Author.
WREB (April 9, 2009). *Periodontal Treatment test committee meeting minutes*. Phoenix: Author.
WREB (May 15-16, 2009). *Endodontics committee meeting minutes*. Phoenix: Author.
WREB (September 11-12, 2009). *Operative test committee meeting minutes*. Phoenix: Author.
WREB (February 19-20, 2010). *Operative Committee Meeting–Minutes*. Phoenix, AZ: Author.
WREB (May 7-8, 2010). *Operative Committee Meeting–Minutes*. Phoenix, AZ: Author.
WREB (August 6-7, 2010). *Endodontics Committee Meeting–Minutes*. Phoenix, AZ: Author.
WREB (September 10-11, 2011). *Operative Committee Meeting–Minutes*. Phoenix, AZ: Author.
WREB (May 13, 2011). *Endodontics Committee Meeting–Minutes*. Phoenix, AZ: Author.
WREB (May 20-21, 2011). *Operative Committee Meeting–Minutes*. Phoenix, AZ: Author.
WREB (June 22, 2011). *ICE Conference Call*. Phoenix, AZ: Author.
WREB (August 5-6, 2011). *Operative Committee Meeting–Minutes*. Phoenix, AZ: Author.
WREB (September 16, 2011). *Integrated Computer Exam Committee–Minutes*. Phoenix, AZ: Author.
WREB (December 6, 2011). *PATP Committee Meeting Minutes*. Phoenix, AZ: Author.
WREB (December 9-11, 2011). ICE Committee Meeting Minutes. Phoenix, AZ: Author.
WREB (March 16-17, 2012). *ICE/Comprehensive Treatment Planning Committee Meeting Minutes*. Phoenix, AZ: Author.
WREB (June 8-9, 2012). *Endodontics Committee Meeting–Minutes*. Phoenix, AZ: Author.
WREB (June 9-10, 2012). *ICE/Comprehensive Treatment Planning Committee Meeting Minutes*. Phoenix, AZ: Author
WREB (September 21, 2012). *PATP Committee Meeting Minutes*. Phoenix, AZ: Author.
WREB (May 11, 2012). *Operative Committee Meeting–Minutes*. Phoenix, AZ: Author.
WREB (May 11-12, 2012). *PATP Committee Meeting Minutes*. Phoenix, AZ: Author.
WREB (August 7, 2012). *Operative Committee Meeting–Minutes*. Phoenix, AZ: Author.
WREB (September 28-29, 2012). *ICE/Comprehensive Treatment Planning Committee Meeting Minutes*. Phoenix, AZ: Author.
WREB (December 14-16, 2012). *ICE/Comprehensive Treatment Planning Committee Meeting Minutes*. Phoenix, AZ: Author
WREB (February 2-3, 2013). *PATP Committee Meeting Minutes*. Phoenix, AZ: Author.
WREB (February 22-24, 2013). *Comprehensive Treatment Planning Committee Meeting–Minutes*. Phoenix, AZ: Author.
WREB (May 3, 2013). *Operative Committee Meeting–Minutes*. Phoenix, AZ: Author.

WREB (May 3-4, 2013). *Endodontics Committee Meeting–Minutes*. Phoenix, AZ: Author.
WREB (February 23-23, 2014). *Comprehensive Treatment Planning Committee Meeting–Minutes*. Phoenix, AZ: Author.
WREB (May 9, 2014). *Operative Committee Meeting–Minutes*. Phoenix, AZ: Author
WREB (May 16, 2014). *Endodontics Committee Meeting–Minutes*. Phoenix, AZ: Author.
WREB (February 20-21, 2015). *Comprehensive Treatment Planning Committee Meeting–Minutes*. Phoenix, AZ: Author.
WREB (April 10, 2015). *Periodontics Committee Meeting–Minutes*. Phoenix, AZ: Author.
WREB (July 24, 2015). *Dental Exam Review Board–Executive Summary*. Phoenix, AZ: Author.
WREB (October 9, 2015). *Operative Committee Meeting–Minutes*. Phoenix, AZ: Author.
WREB (October 23, 2015). *Endodontics Committee Meeting–Minutes*. Phoenix, AZ: Author.
WREB (November 6-7, 2015). *Comprehensive Treatment Planning Committee Meeting–Minutes*. Phoenix, AZ: Author.
WREB (February 26-27, 2016). *Comprehensive Treatment Planning Committee Meeting–Minutes*. Phoenix, AZ: Author.
WREB (September 16, 2016). *Comprehensive Treatment Planning Committee Meeting–Minutes*. Phoenix, AZ: Author.
WREB (October 7, 2016). *Endodontics Committee Meeting–Minutes*. Phoenix, AZ: Author.
WREB (October 7, 2016). *Periodontics Committee Meeting*. Phoenix, AZ: Author.
WREB (October 8, 2016). *Operative Committee Meeting–Minutes*. Phoenix, AZ: Author.