CRDTS' National Dental Hygiene Examination

Technical Report

for the

Year Ending 2016

ACKNOWLEDGMENT

Dr. Thomas Haladyna, Professor Emeritus at Arizona State University, is a well-recognized measurement specialist and educator. Author of many articles and books, Dr. Haladyna has been associated with many projects related to clinical evaluation in the professions. In particular, he was a valuable contributor to the 2004 *AADE Guidance for Clinical Licensure Examinations in Dentistry.* His comprehensive knowledge of the 2015 *Standards for Educational and Psychological Testing* enabled him to cross-reference those *standards* relevant to the principles and guidelines in the AADE Guidance document. In recent years, Dr. Haladyna has begun serving as a consultant to the Equal Employment Opportunity Commission, applying his knowledge of measurement principles to the review of those examinations that may be used as prerequisites for employment.

Dr. Haladyna authored CRDTS' *2010 Technical Report on Clinical Examination in Dentistry,* followed by the *2014 Technical Report on CRDTS National Dental Examination* upon the publication of the revised *Standards for Educational and Psychological Testing* in 2015. For each of these technical reports, including this *2016 Technical Report on CRDTS National Dental Examination,* Dr. Haladyna has reviewed all CRDTS' reports, analyses, and examination documents and identified those current *standards* relevant to clinical evaluation and correlated them with the appropriate aspects of CRDTS' *National Dental Examination.*

CRDTS is very pleased that Dr. Haladyna's expertise has documented the validity evidence accumulated in the development, scoring and administration of CRDTS' clinical dental examination. He is a gifted measurement specialist with a knowledge and understanding of clinical evaluation that is outstanding among his professional colleagues.

Lynn M. Ray, RDH, BS
CRDTS Director of Analysis
February 2017

Table of Contents

Introduction

The *National Dental Hygiene Examination (NDHE)* is developed by Central Regional Dental Testing Service (CRDTS). The purpose of this testing program is to provide validly interpretable test score information to states and jurisdictions to help each make a licensing decision for those wanting to practice dental hygiene.

A technical report has the important responsibility of displaying the qualities of a testing program that support validity. A technical report summarizes the argument for validity and the body of evidence supporting that argument. Thus, this technical report contains information useful in evaluating the validity of *CRDTS' NDHE* test score interpretation and use.

This technical report is organized in the following way:

1. Validity is defined and discussed as it applies to the *CRDTS' NDHE.*
2. The *Standards for Educational and Psychological Testing* (American Educational Research Association, American Psychological Association, and the National Council on Measurement in Education, 2015) are applied systematically to support validity.
3. The *CRDTS' NDHE* is described.
4. The largest section of this report presents validity evidence in support of using test scores as part of the information used to license dental hygienists.

Toward improving clarity in this technical report, some terms are defined here.

Standards (2015) refers to the above publication. The *Standards* is a highly respected and well-used set of guidelines for test planning, development, and validation. When *standards* appears in lowercase, this term refers to specific statements in the above publication.

A *test* is a device containing many tasks developed for obtaining responses scored to form a test score. Often the word *examination* or *exam* is used to mean test.

A *testing program* is an organization devoted to designing and developing a test, and validating test score interpretations and any uses. Sometimes the term *examination program* is used as a synonym for testing program.

Validity is the most important consideration in any testing program. *Validity* refers to the reasonableness of interpreting a test score as an indication of a candidate's professional competence. Validity is defined more adequately in a subsequent section of this technical report.

Construct is a technical term that refers to the domain of tasks performed by a dental hygienist. A more recognizable term is *content*. The content of the *CRDTS' NDHE* is the construct of professional competence in dental hygiene. Often we think of content as a domain of tasks, which is called the *target domain*. The target domain is a critical idea in the development of the construct of professional competence in dental hygiene and the validation of using test scores for licensing decisions.

Validity

Validity refers to the judged degree to which an argument and evidence support a specific interpretation and a specific use of a test score. In dental hygiene, the intended interpretation of the *CRDTS' NDHE* test score is how the candidate stands concerning a domain of tasks performed by dental hygienists in practice. The test is a representative sample from this domain. This domain is limited to those of normal, everyday practice and does not include tasks that are rarely performed. The intended use of these test scores is for states and other jurisdictions. The test scores are part of the body of evidence used to decide to license a dental hygienist in that state or jurisdiction.

Validation is an investigative process by which the claim for validity, a supporting argument, and validity evidence can be judged by a competent observer. The judgment is as to degrees of validity. Generally, the body of evidence is considered in totality but weaknesses in this body of evidence are noted in this technical report and remedies are advised. This technical report resembles a validation and, also, supports an evaluation of this testing program. For a positive evaluation, the argument has to be sound and compelling, the claims just, and the preponderance of evidence supporting each claim. Negative evidence should be inconsequential. Negative evidence leads to recommendations to study, assess, and eliminate or reduce the factors causing this negative evidence. Validity studies are often recommended (Haladyna, 2006). By studying negative evidence and seeking remedies, validity is increased.
Table 1 shows the constituent elements in validation.

Table 1: Validation of CRDTS's *NDHE*	
Argument	The *CRDTS' NDHE* is a clinical performance examination intended to measure dental clinical competence directly.
Claim About Validity	CRDTS claims that candidate scores from its *NDHE* represent dental clinical competence. The results of the test can be used with confidence by participating states, along with other criteria, to make licensing decisions for candidates.
Evidence Supporting the Argument	This technical report provides validity evidence of many types that are based on national test *standards*. CRDTS's documents cited in this report and found in the appendix offer validity evidence supporting this argument.
Evidence Weakening the Argument	CRDTS considers threats to validity and acts accordingly to diminish or eliminate each threat. By that, CRDTS strengthens the evidence supporting the argument and the claim for validity.
Lack of Evidence	If evidence is missing, CRDTS has the responsibility to gather such evidence in the future as it increases validity.

A Threat to Validity–Construct Representation

The target domain represents a large, ideal set of tasks that licensed dental hygienists typically perform in practice. The size of this target domain is a matter of professional judgment. Usually a survey of the profession provides the scope and content of the target domain. Administering the entire target domain to a candidate for licensure is impractical. Such a test would entail many days. CRDTS claims that its *NDHE* represents a fair and sufficient sampling

of tasks from the target domain. The domain of tasks was established via a survey of the profession, as is reported subsequently in this technical report. Such a survey is a necessary condition in developing a test like the *CRDTS' NDHE* (Raymond & Neustel, 2006, Raymond, 2016).

Construct representation designates the degree of match between the target domain and actual tasks on the *CRDTS' NDHE*. Because a survey of the profession assessed the target domain, CRDTS determined which tasks should be included in its *NDHE*. Thus, construct misrepresentation is not perceived as a threat to validity. This technical report provides evidence to support this claim.

Another Threat to Validity–Construct-irrelevant Variance (CIV)

CIV is a technical term for bias. It is systematic error. Such error falsely inflates or deflates a test score. CIV has many sources. For instance, a lenient examiner may overrate a candidate performance. An interruption in test administration may cause a candidate to lose time and fail to perform a task as intended, which results in a falsely reduced test score. Testing agencies have a responsibility to identify potential sources of CIV and eliminate or reduce each threat to validity. Throughout this technical report, potential sources of CIV are named, investigated, and reported. As the evidence shows, CIV is NOT a major threat to validity in this testing program.

Integrating Validity Evidence and the Judgment of Adequacy

> "*A sound validity argument integrates various strands of evidence into a coherent account of the degree to which existing evidence and theory support the intended interpretation of scores for specific uses*" (*Standards*, 2015, p. 21).

As the *Standards* (2014) state, validation is a never-ending process. This technical report provides a summary of validity at a point of time and offers a historical perspective when compared with previous and subsequent technical reports.

Validity Evidence Used in This Technical Report

To organize validity evidence, the following categories are presented: content, item quality, reliability, examination administration, selection, training, and retention of examiners and scoring, scaling and comparability, standard setting, score reporting, rights of test takers, security, and documentation. As noted previously, this body of evidence should be evaluated holistically.

Standards for Educational and Psychological Testing

The *Standards* (2014) update the previous *Standards* (1999). A large, representative committee of testing experts and other qualified volunteers participated in developing these *standards*. For this evaluation, the current *standards* are applied and cited in this technical report. All of the referenced *standards* influence the overall judgment of validity. The American Association of Dental Examiners (2003) 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.

Table 2 lists specific *standards* listed here and quoted throughout this report. In each section, a discussion and evidence are offered in support of these *standards*.

Table 2: s*tandards* Used in this Technical Report	
Chapter 1: Validity. This chapter identifies fundamental concepts and types of validity evidence that appear throughout this evaluation report.	1.0, 1.1, 1.2, 1.5, 1.7, 1.9, 1.11, 1.13
Chapter 2: Reliability. As a primary type of validity evidence, evidence is sought	2.0, 2.5, 2.7, 2.13, 2.14
Chapter 3: Fairness	3.0, 3.1, 3.2, 3.4,
Chapter 4: Test Design & Development	4.0, 4.1, 4.2, 4.3, 4.7, 4.8, 4.10, 4.12, 4.13, 4.16, 4.18, 4.20, 4.21
Chapter 5: Scores, Scales, Norms, Score Linking, and Cut Scores	5.0, 5.1, 5.5, 5.6
Chapter 6: Test Administration, Scoring, Reporting, and Interpretation	6.0, 6.1, 6.4, 6.5, 6.6, 6.8, 6.9, 6.10, 6.14, 6.15, 6.16
Chapter 7: Supporting Documentation	7.0, 7.1, 7.2, 7.4, 7.8, 7.10, 7.13
Chapter 8: The Rights and Responsibilities of Examination Takers	8.0, 8.1, 8.2, 8.6, 8.8, 8.9, 8.10, 8.11, 8.12
Chapter 11: Workplace Testing and Credentialing	11.1, 11.2, 11.3, 11.4, 11.13, 11.14, 11.16

As noted previously, the *Standards* promotes testing practices that can increase validity. The *Standards* are silent on policy issues. However, policy decisions can be informed by technical reports that consider the *Standards*.

The *Standards* have some important disclaimers:

1. Not all *standards* apply to a specific testing program. Thus, evidence need not be presented for every standard. In this technical report, *standards* were selected that bear on validity for a clinical performance test used as part of the criteria for licensing dental hygienists in states and other jurisdictions.

2. If there is a legal challenge to a test score interpretation or use, the use of *standards* provides a valuable basis for understanding and defending against a challenge. CRDTS can use

standards as a basis for its credibility if legal challenges were made on a test score interpretation or use.

Throughout this technical report, *standards* are quoted that apply to this testing program and relate to validity. Thus, readers are encouraged to consider that (1) *standards* are followed in test design, development, administration and scoring, and (2) the application of these *standards* with proper documentation in this report contributes to our evaluation of validity.

Description of the *CRDTS' NDHE* Test

Dental hygiene schools typically provide classroom, laboratory, and clinical experiences. Topics covered in these programs include, anatomy, physiology, chemistry, microbiology, and pharmacology. Typical courses include nutrition, radiography, histology, periodontology, pathology, dental materials, clinical dental hygiene, and social and behavioral sciences. https://www.adha.org/resources-docs/72611_Dental_Hygiene_Education_Fact_Sheet.pdf, pp. 7-8

The best current source of information about this testing program comes from CRDTS' website: CRDTS.org. Detailed information about the examination can also be found in the 2016 *Dental Hygiene Candidate Manual* (CRDTS, 2016b). The appendix of this technical report contains many archived documents that attest to the development of the CRDTS' *NDHE* and the validation of interpretation and use of test scores.

CRDTS was established in 1972. As stated in its bylaws, state boards for dental hygiene licensing are its members. Its members meet annually in August. The CRDTS' *NDHE* is used to measure a candidate's clinical competence in dental hygiene in four distinct areas. Each area is represented by a subtest. Each candidate can achieve a total score as high as 100 points on the test. The cut score is set by legislation in participating states. It is usually 75 points. With the permission of candidates, scores are sent to appropriate member states and other participating states. These states use this information with other information to decide licensing each candidate.

Origin of Current Examination

In 1975, the ADHA initiated a national Clinical Evaluation Study to develop a criterion-based, model examination, funded by a federal grant, directed by a national Task Force of dentists and dental hygienists, and 22 field tests were conducted across the country and many CRDTS' examiners participated in the study, which concluded in 1978. CRDTS adopted the model exam as its dental hygiene examination in 1978, when they launched calibration exercises, and a statistical analysis program. In subsequent years, CRDTS refined the examination and collaborated with other testing agencies, but no organized effort to develop a national dental hygiene examination occurred until 2004 when ADEX was formed. Within ADEX, CRDTS participated in a developmental project that included dental hygienist representatives from all regional testing agencies that formed the basis of the current CRDTS examination, which CRDTS continues to administer and refine.

The ADEX was an umbrella organization formed to design national clinical dental and dental hygiene examinations. Evidence of the origin of the examination and its organization, structure, staff, and committees is presented in annual reports (ADEX, 2006, 2007, 2008; April 10, 2006; June 23, 2006; August 26, 2006; April 12, 2007; April 17, 2007; December 5-7, 2007; January 19, 2008a; January 19, 2008b; January 22, 2008; August 21, 2008).

As of June 30, 2009, CRDTS severed its association with ADEX but retained much of the examination design and structure, in which CRDTS had actively participated during its development over a four-year period. One report by ADEX (2008) provides an example of examination review and recommendations that bear on the current examination. Up to that point, documentation of validity was done by ADEX. After that point, the responsibility for subsequent

documentation and any modifications of the examination has been the responsibility of CRDTS.

CRDTS' *NDHE* Content

The CRDTS' *NDHE* test consists of four subtests as Table 3 shows.

Table 3: Subtests of the *CRDTS NDHE*			
Subtest	Scorable Items	Points/Item	Max. Points
1. Extra/Intra Oral Assessment	8	2	16
2. Periodontal Probing	12	1	12
3. Scaling/Subgingival Calculus Removal	12	5	60
4. Supragingival Deposit Removal	6	2	12
Total Examination Points/Maximum Score			100

As noted in the *NDHE Candidate Guide* (2016b), penalty points may be assessed by the Dental Hygiene Coordinator or Team Captain. Penalties are levied for improper treatment selection, improper treatment standards (e.g., professional demeanor, asepsis violation). Critical errors are those that could lead to patient injury or jeopardize overall patient treatment. Tissue trauma is a major category. A time penalty or unprofessional conduct may result in dismissal from the examination.

A compensatory scoring model is used. That is, the total score based on all four subtests is the criterion for a pass/fail decision. Performance on each sub-test is not considered in arriving at a total score. A conjunctive scoring model takes into account performance on each sub-test, where a pass-fail decision is made on the sub-test. The compensatory scoring model has the advantage of higher reliability than the conjunctive model.

VALIDITY EVIDENCE BEARING ON VALIDITY

The body of validity evidence in this technical report is organized in the following way. First validity *standards* are cited as the main concern. Then validity evidence is presented and discussed. In each instance, a claim is made in support of validity.

Validity

The *standards* cited in Table 4 deal directly with validity. Because some *standards* are quite lengthy, they were paraphrased and presented in italics. Also, some standards may seem repetitive. This is true because different sets of testing experts worked on different chapters yet shared the same concern for validity.

Table 4: *Standards* Generally Related to Validity	
1.0	Clear articulation of each intended test score interpretation for a specified use should be set forth, and appropriately 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 constructs-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.*

1. Competence of a dental hygienist is represented by a target domain of tasks.

2. A practice analysis (occupational analysis) is conducted regularly to ensure that the content of the four subtests has high fidelity with this domain of tasks (CRDTS, 2012).

3. All aspects of test development are refined and well described in this technical report and other documents cited in the appendix.

4. Threats to validity are regularly investigated, and attempts are made to reduce or eliminate these threats.

1. Content

The most fundamental type of validity evidence for a licensing test is content-related (Kane, 2006). As noted previously, a dental hygiene clinical examination should identify a target domain of tasks performed by a competent dental hygienist. Ideally, the tasks in the target domain are organized by important content topic descriptors. These tasks are prioritized according to relevance to the profession and how frequently the tasks are performed in regular professional practice. A good source of guidance for identifying such test content is through a survey of the profession, known as *practice analysis* (Raymond, 2015; Raymond & Neustel, 2006).

> *An investigation of a certain occupation or profession to obtain descriptive information about the activities and responsibilities of the occupation or profession and about the knowledge, skills, and abilities needed to engage successfully in the occupation or profession* (*Standards*, 2014, p. 222).

Table 5 presents *standards* bearing on content.

Table 5: *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.*

Chapter 11 of the *Standards* (2014) is devoted exclusively to *standards* affecting licensure tests, such as CRDTS's. Not only is CRDTS expected to define clinical competence in dental hygiene, but is also expected to show the validity of the constituent parts of dental hygiene competency as determined from the practice analysis. *Standards* 11.2, 11.3, 11.13 all address slightly different but complementary aspects of practice analysis as a basis for test specifications. The set of test specifications guides examination development. The test specifications are displayed in the *Dental Hygiene Candidate Manual*, CRDTS, 2016b).

Practice Analysis (Also known as Occupational Analysis or Job Analysis)

As stated previously, practice analysis was done to update the content of the *CRDTS' NDHE* (CRDTS, 2012). The results of the survey showed that the sample of respondents represented its population and the rating scales used to assess content were functional. The frequencies and ratings of criticality for patient oral health and systemic health of the patient were found within anticipated ranges. Adjustments were made in the point allocation to procedures and penalty points were reviewed for validity. From the practice analysis, the 2013 examination was slightly revised and continues to be administered in 2016. The use of the practice analysis is in keeping with the *Standards* (2014). Specifically, as stated in Table 5, the *standards* have been met. The Dental Hygiene Examination Review Committee reviewed the results of the occupational analysis and confirmed that the content comprising the four subtests was accurate (CRDTS, July 14, 2012).

Structural Evidence

A major consideration in the design of any testing program is the theoretical and empirical structure of test data. Is dental hygiene clinical competence a single construct consisting of highly related tasks? Or is clinical competence a family of relatively independent tasks, each of which is important in dental hygiene practice?

Table 6 provides descriptive statistics for the four subtests of the *NDHE*. Note that the means and skewness in the distribution of test scores are very high. This high performance of candidates reflects the selection and training these candidates receive. The test is sensitive enough to detect high performance on the tasks in this test. Correlations among these four subtests range from -0.02 to 0.09. How should these correlations be interpreted? The four tasks are independent of each other. Each appears to measure an independent clinical skill of a dental hygienist.

Table 6: Descriptive Statistics for the Four Tests of the CRDTS' *NDHE*				
	Extra/Intraoral Assessment	Periodontal Probing	Scaling/Subgingival Calculus Removal	Supragingival Deposit Removal
Number of Candidates	2,201	2,201	2,201	2,201
Low Score	2	0	0	4
High Score	16	12	60	12
Mean	15.14	11.80	52.69	11.84
Standard Deviation.	1.45	0.72	9.73	0.70
Skewness	-2.29	-7.47	-2.36	-5.68

The mean of the total score was 91.46 and the standard deviation was 10.05.

Claim Supporting Validity

The practice analysis confirmed the choice of content of the *CRDTS' NDHE* (CRDTS, 2012). The study of structure shows the independence of the four subtests. CRDTS combines these four subtest scores minus penalty points to compute a total score. The evidence for the selected content of each of the four subtests is very strong.

2. Item Quality

Each task (test item) on each of the four subtests connects directly to the practice analysis. Professional judgment by highly qualified, licensed, experienced dental hygienists is crucial to supporting item development and validity. The set of test specifications is a public document. As noted previously, it can be found in the *Dental Hygiene Candidate Manual* (CRDTS, 2016b). The set of test specifications shows the number of items, types of items, criteria for scoring, distribution of points for each subtest, and other relevant information.

The *Standards* (2014) are very explicit about the role of item development in test development and validation. Table 7 lists relevant *standards* for item development.

Table 7: *Standards* 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.*

Historically and currently, the test items (tasks) for the current *NDHE* were the product of many years of development and refinement. CRDTS has recorded regular meetings of its Examination Review Committee that trace some of these refinements (CRDTS, 2006; August 26, 2006; 2007; August 24, 2007; 2008a; 2008b; August 22, 2008a; August 22, 2008b; August 22, 2008c; 2009; 2010, July 10-11, 2010; July 2011; July 9-10, 2011; July 14, 2012; July 13, 2013; July 12, 2014 July 11, 2015, July 9, 2016). The validity evidence needed to this examination includes the following:

1. Practice analysis identified the knowledge, skills, and abilities needed to practice safely and competently.
2. A set of test specifications was created that explicate this content.
3. Items are developed to match the test specifications.
4. Items undergo intensive review by subject-matter-experts on frequently held content subcommittee meetings, as documented in the above citations and in the appendix.
5. The scoring procedure is developed and is assigned a point value by the subject-matter-experts.
6. The item and the scoring protocol are field tested to assure its ability to discriminate between high- and low-performing candidates.
7. Most important, these items should have high fidelity with the criterion behavior intended–actual dental hygiene practice.

Fidelity

Tasks on any clinical performance test such as CRDTS should resemble those tasks performed by dental hygienists in practice. If the tasks possess fidelity with criterion behavior, part of the validity argument is that the content of the *CRDTS' NDHE* has high fidelity with the tasks performed by dental hygienists in practice. A review of these tasks and prior committee activities supports the fidelity argument. The tasks performed on the examination are identical or similar to tasks performed by dental hygienists on actual patients. The 2012 dental hygiene practice analysis provides evidence of the deliberate process to transform statements of tasks in practice to tasks on the dental hygiene examination. http://www.adha.org/resources-docs/2016-Revised-Standards-for-Clinical-Dental-Hygiene-Practice.pdf, pp. 6-10.

Weighting of the Four Subscales

This topic is very important because a weight is assigned to each subscale and this weighting process affect pass/fail decisions. If a candidate performs poorly on the most heavily weighted subtest, that alone, could result in a fail decision. In the development of the *NDHE*, CRDTS has carried out evaluations of different weighting systems and arrived at the present one (ADEX, April 5, 2005, CRDTS, April 12, 2005). Since the original examination was developed by ADEX, CRDTS has reviewed and revised the original weighting of test items. The weighting of any test item is a matter of professional judgment by subject-matter-experts. The decisions for the current weights for test items are the result of a deliberate process by the examination review committee during their frequent meetings. A useful reference is the *Dental Hygiene Candidate Manual* (2016b), which is available publicly on its website (CRDTS.org).

3. Reliability

Every test score has an unknown degree of random error and a true score. Random error can be positive or negative and large or small. There is no way to measure how much random error exists in a test score or to know the true score. All we can do is estimate the true score and estimate the degree of random error that exists. As the chart below shows, of four possible outcomes when making a pass/fail decision, two are correct decisions and two are incorrect decisions.

	Test score at or above 75	Test score below 75
True score at or above 75	Correct decision–pass	Incorrect decision–Fail (Type 1)
True score below 75	Incorrect decision–pass (Type 2)	Correct decision–Fail

We call these classification errors Type 1 and Type 2. Reliability affords us understanding of the risk of misclassifying candidates whose true scores are at or close to the cut score. For high- and low-scoring candidates, there is little risk of misclassification regarding passing or failing. It is those candidates whose scores are near the cut score that have the risk of being misclassified.

Because of random error, we cannot eliminate Type 1 or Type 2 errors. By making the test scores as reliable as possible, we can only minimize random error, and, by that, reduce the chances of committing Type 1 and Type 2 errors. Candidates due to poor performance place themselves in jeopardy of being a victim or beneficiary of a Type 1 or Type 2 error.

Several *standards* apply to reliability and are presented in Table 8:

Table 8: *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.

CRDTS has taken the following steps to ensure that reliability is high and the risk of misclassification is reduced.

1. CRDTS uses three examiners for each observation. This step ensures a high degree of internal consistency in ratings that is crucial in establishing reliability. Results of examiner consistency are reported in appropriate sections of this report for each of the four subtests.

2. CRDTS has many observations (test items) per test. Reliability benefits by having many observations for each of the four subtests.

3. CRDTS has special scoring rules for critical deficiencies. This scoring rule results in automatic failure if two or three examiners agree that a performance justifies a rating of zero–indicating a critical deficiency. This procedure is explained to candidates in their *2016 Dental Hygiene Candidate's Guide* (CRDTS, 2016b, p. 12). Critical errors are also discussed in the *2016 Dental Hygiene Examiner's Manual* (CRDTS, 2016a, p. 12).

Conventional reliability estimation depends on high internal consistency among item responses. That is to say, item responses need to be highly intercorrelated. Sometimes, a clinical performance test can consist of tasks that are not highly related. In this instance, a more appropriate technique for estimating reliability is stratified alpha (Haertel, 2006, pp. 76-78). Haertel asserts that conventional reliability methods greatly underestimate reliability. Whereas stratified alpha may be more accurate.

Another issue in assessing reliability is that if performance is heavily negatively skewed, there is little variation. Thus, conventional reliability estimates that depend on variability of test scores will be very low. The more important consideration is to estimate how much random error variance exists in these scores. The estimation of error variance results in the standard error of measurement (SEM). This estimate helps us understand the degree of risk involved in making pass/fail decisions for candidates whose scores are at or approach 75, the passing score.

Reliability estimates are reported in Table 9. As noted there, the first three subtests had differential coefficients. The third subtest, Scaling/subgingival deposit removal, had the greatest weight and, therefore, influenced the total score reliability to a large degree. The fourth subtest (Supragingival Deposit Removal) has very high performance and virtually no variation. Reliability using conventional methods cannot be used, because these methods depend on variability of test score. Random error appears to be extremely small. Therefore, the conclusion is that reliability is 1.00. This conclusion has no bearing or effect on the estimation of total score reliability.

Table 9: Reliability Estimate for Each Subtest	
Subtest	Alpha
1. Extra/intra Oral Assessment	0.57
2. Periodontal Probing	0.80
3. Scaling/subgingival Deposit Removal	0.87
4. Supragingival Deposit Removal	Not estimated
Reliability Estimate Using Stratified Alpha: 0.87	

Standard Error of Measurement

Reliability is not an end; it is a means to an end. The objective is to obtain an estimate of the degree of random error around the cut score. This helps states using test scores to assess the risk for misclassifying candidates whose true scores are close to the cut score of 75. Once reliability is properly estimated, the degree of random error is estimated and used to study the status of candidates whose observed scores falls at or near the cut score of 75. The standard error of measurement is 3.78. Constructing a zone of uncertainty around the cut score of 75, we observe the number of candidates whose scores fall between 71 and 79. These candidates are in jeopardy of falsely passing or failing this test due to random error. Of 2,186 with validated scores, 102 had scores in this range. Thus, by random error, any of these examinees may have been misclassified. However, these examinees are very low-scoring in comparison to the 1,911 examinees who scored 80 and higher. The overall mean of this group is 91.46.

The fact that so many candidates are observed in this zone of uncertainty leads to two arguments:

1. The test needs to be improved to make the zone of uncertainty smaller. However, the zone is quite small, so this would be very difficult to do.

2. Candidates who perform at or near 75 are justifiably at risk. These candidates need further training to improve their proficiency and retesting and should perform well above 75 at the next testing.

Claim Supporting Validity

The reliability estimate of 0.87 is very high considering the independence of the four subtests comprising the total score and the negative skewness and restrictedness of total test scores due to high performance of candidates on these tasks. The standard error of measurement is small. All *standards* regarding reliability are met. Validity evidence is very strong.

4. Examination Administration

Test administration is an important aspect of any testing program. McCallin (2006, 2015) provides a very detailed account of issues in examination administration and potential threats to validity. The *Standards* (2014) also provides guidance in several *standards*, shown in the table below.

Table 10: *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.

This standardized examination has been slightly revised each year upon recommendation from the Dental Hygiene Examination Review Committee. Regular meetings of this committee address many aspects of examination administration and document revisions in administration aimed at making the test better (CRDTS, 2006; August 26, 2006; 2007; August 24, 2007; 2008a; 2008b; August 22, 2008a; August 22, 2008b; August 22, 2008c; 2009; 2010, July 10-11, 2010; July 2011; July 9-10, 2011; July 14, 2012; July 13, 2013; July 12, 2014; July 11, 2015; July 9, 2016). A useful source of information about administration is the *2016 Hygiene Coordinator Notebook (CRDTS, 2016).* This notebook includes seven sections including (1) coordinator materials, (2) orientation materials, (3) calibration exercises, (4) assistant materials, (5) anesthesia materials, (6) miscellaneous forms, and (7) candidate and examination forms. The *2016 Dental Hygiene Examiner Materials* (CRDTS, 2016) is another source of information for dental hygiene examiners. The CRDTS *Dental Hygiene Candidate's Manual (2016b)* also provides specific information about administration that is suitable from the candidate's perspective.

Claim Supporting Validity

The examination administration has been developed and improved as documented in committee reports and in past technical manuals and evaluations. *Standards* for test administration are fully met.

5. Selection, Training, and Retention of Examiners and Scoring

Table 11 lists *standards* related to selection, training, and retention of examiners. Also, *standards* in this table addresses scoring.

Table 11: *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.

A committee was formed for establishing and maintaining an examiner preparation program. This committee defined the criteria for selection of examiners, reviews and monitors examiner reliability, assigns examiners to test sites, and selects chief examiners, coordinators, and team captains. Several documents attest to the well organized, efficient, and effective system for recruiting, training, evaluating, and retaining examiners. According to Klein (April 27, 2008), CRDTS has a well-established program for examiner training and calibration. Other external reviews have contributed to the continued development of this testing program (e.g., Klein, May 11, 2009; Littlefield, April 25, 2009).

The development of the scoring system is documented in reports (CRDTS, July 12, 2005). Additional documentation includes the following references (CRDTS, 2006; August 26, 2006; 2007; August 24, 2007; 2008a; 2008b; August 22, 2008a; August 22, 2008b; August 22, 2008c; 2009; 2010, July 10-11, 2010; July 2011; July 9-10, 2011; July 14, 2012; July 13, 2013; July 12, 2014; July 11, 2015; July 9, 2016).

Selection of Examiners

Many factors comprise criteria for examiner selection. Examiners must be in good standing with their state board, have an active practice, possess good health, make a commitment to participate in two or three examinations, accept CRDTS standards and evaluation criteria, accept the training regimen, receive a nomination to serve, and must observe an examination if a new examiner. All examiners must be subject-matter-experts.

Training and Evaluation of Examiners

Each examiner receives a copy of the most current *Dental Hygiene Examiner's Manual* (CRDTS, 2016a). All examiners receive an orientation and undergo a calibration exercise to ensure that their judgments are accurate and consistent. New examiners receive additional orientation.

Each year analysis is done to report the accuracy and consistency of examiners (Ray, 2014, 2015, 2016). A very useful feature of these reports is the presentation of graphs showing degrees of leniency and severity in examiner judging. Such information can be very useful in refining training and improving examiner consistency or, if justified, removing examiners who are inconsistent. Such reports are very useful for quality control. These results are also used to evaluate examiners and to inform decision-making for future examiner assignments.

Scoring

The *Dental Hygiene Candidate Manual* (CRDTS, 2016b) provides the conditions for scoring including examiner ratings and penalty point assessments. All these decisions were reached by committee consensus and then approved by the Board.

Scoring is done on the site and ratings are recorded electronically. After every examination, there is verification and post examination review. All scores are rechecked. This effort seeks to uncover irregularities or errors in computing a candidate's score. All failing scores are subjected to manual verification by professional dental personnel.

Quality Control

All examiners are subjected to a multi-step process for standardization and calibration designed to produce accurate and consistent ratings of candidate performance. Exercises are designed and used during a two-day orientation of Hygiene Coordinators and Team Captains. Hygiene Coordinators contribute to the development of these exercises. Each year the exercises are reviewed, evaluated, and revised if necessary. Also, the *Dental Hygiene Candidate Manual* (2016b) is also revised as needed.

CRDTS maintains an Examiner Evaluation and Assignment Committee (EEAC) that meets annually to review examiner profile reports, with additional meetings as needed to assign examiner teams for every test site. The EEAC reviews every examiner's individual profile, makes decisions regarding their effectiveness, looks for emerging leadership qualities as Team Captains or Hygiene Coordinators. They also review each examiner's Peer Evaluations, which are part of the profile

reports. Every examiner is asked to evaluate their fellow team members at the close of each exam. These Peer Evaluations focus on the examiner's behavior, preparedness, adherence to protocol, and work ethic. The EEAC is empowered to change an examiner's assignment if they are not functioning well in a particular role, they may send letters to those examiners who are outliers in their profile reports, or terminate the examiner's assignments if their results or behavior is not appropriate. As stated previously, CRDTS has criteria for retaining examiners. Thus, examiners, who fail to rate accurately and consistently, are unlikely to be reappointed.

Examiner Consistency

CRDTS uses a technique for resolving differences in examiner consistency that is based on statistical wisdom. Instead of adding and averaging disparate examiner judgments, which may be biased and inconsistent, the median of three ratings is used. For small samples, the median is often recommended by statisticians instead of the mean or the mode as the best measure of central tendency. To be clear here is an example.

Examiner 1	Examiner 2	Examiner 3	Median
0	0	1	0

Although examiners might differ on a single task and by one-point, the majority judgment determines the score that the candidate receives. All tasks in this test are dichotomous, as no rating scales are used. Thus, examiner consistency can be measured and reported, but it is irrelevant to the actual score a candidate receives.

Claim Supporting Validity

The training and evaluation of examiners is a highly refined activity that has received considerable attention over many years. The *2016 Dental Examiner's Manual* (2016a) is an annual publication updated each year. It contains comprehensive information related to training and scoring. This document is supplemented with other materials used during training. Thus, training of examiners and their scoring is very effective. The *standards* supporting examiner quality have been met.

6. Scaling & Comparability

Chapter 5 of the *Standards* (2014) is devoted to scaling and comparability. Table 12 list *standards* related to this important topic.

Table 12: *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 validity of interpreting test scores is strongly dependent on having a test score scale that is constant from one examination administration to another. Considering that the cut score is also constant, it is important that the test be equally difficult and the content fixed on all occasions where the test is administered.

With multiple-choice test forms containing different items, equating these test forms is necessary so the scale is constant from one test form to another test form. The *NDHE* has only one form. The items are well known and disclosed to all candidates, who are allowed to practice. The administration is standardized at all testing centers. Therefore, it is claimed that the test score scale is the same for each testing center and for each group of candidates taking this test. Because the tasks are those that licensed dental hygienists must perform competently, there is transparency between the target domain and the test representing this target domain. Candidates have full knowledge of what tasks must be performed on this test and can prepare without any doubt as to what is to be tested.

The only variable is the set of examiners for any test. All examiners come from a common pool of examiners. All are highly qualified and extensively trained. Their ratings are calibrated before they rate performance. CRDTS has checks and balances for examiners, and a feedback system to examiners alerts them to instances of leniency or severity in rating and inconsistency (See Ray, 2014, 2015, 2016). Although the scoring system is complex, there is evidence of high examiner consistency and high reliability. These actions regulate and standardize the examination. The test score scale for each of the four subtests seems comparable from test site to test site and occasion to occasion.

Claim Supporting Validity

Scaling for comparability appears adequate given that this is a clinical performance test where the tasks are well known and frequently practiced by candidates. The use of three examiners helps achieve consistency and avoid bias. Examiners are well trained and calibrated. All tasks are standardized. Although scoring is very complex, it too is standardized. The test score scales for each part are the same. *Standards* addressing scaling and comparability are claimed to be met.

7. Standard Setting

Table 13 below lists four relevant *standards*. This section provides evidence relating to these four *standards*. Note that some *standards* are repetitious because they come from different sources and are located in different chapters in the *Standards* (2014).

Table 13: *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.

Most state dental practice acts and/or rules and regulations specify a passing score for clinical licensure examinations. Typically, such laws set a passing score of 70 or 75. For the first 20 years of its existence, CRDTS designed its examinations so that few, if any, scores fell between 69 and 75. From time to time as CRDTS has collaborated with other testing agencies in a test development project, defining a uniform cut score has been an issue. In 1993, CRDTS was working with the Northeast Regional Board in a test development project known as CORE. In order to establish a uniform cut score that would be acceptable in any state, CRDTS reweighted their four-point rating scale to achieve that purpose, as recorded in the November, 1993 minutes of the CRDTS' Steering Committee (CRDTS, October 16, 1993; November, 19-20, 1993). Similarly, in the fall of 2003, CRDTS changed its dental hygiene passing score from 70 to 75, as recorded in the Examination Review Committee minutes of September 2003 and the Steering Committee minutes of October, 2003 (CRDTS, September 7, 2003; October 3, 2003). This action established uniformity with the majority of states and also with the cut score utilized by the Joint Commission on National Dental Examinations.

Claim Supporting Validity

As CRDTS' has engaged in a long-term study for establishing a cut score that involves responsible test agencies and states with the help of subject-matter experts, the way cut scores were established is in compliance with these *standards*.

8. Score Reporting

Table 14 below shows *standards* addressing score reporting.

Table 14: *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.

Score Reports to Candidates

Two types of candidate score reports are sent. The report of a failing score is provided to the candidate with a justification/critique in the lower portion of the report. The report of a passing score simply provides the total score with a recommendation to a state to pass the candidate if other criteria for licensure have been met. Both score reports provide the total score.

Score Reports to Jurisdictions

For every examination site, a complete score report is prepared that presents scores for all candidates and other information.

Score Reports to Dental Hygiene Schools

A comprehensive score report is sent annually to each dental hygiene school (CRDTS, January 2016). This report provides an overview of the examination including its purposes. For every school the mean score on the examination is given along with the percent of candidates who passed and a quintile ranking. All subscores are reported in the same fashion.

Such scores can be validly used to identify strengths and weaknesses in each dental hygiene school's program and curriculum. However, it seems invalid to make comparisons among dental hygiene schools as to performance unless all candidates entering each program are of equal ability.

Claim for Validity

Score reports are responsibly and effectively designed. *Standards* cited above are claimed to have been met.

9. Rights of Test Takers

Chapter 8 of the *Standards* (2014) is devoted to the rights of test takers. Table 15 below lists *standards* relevant to the rights of test takers.

Table 15: *Standards* Related to the Rights of Test Takers	
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.
8.2	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.
8.5	*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.*
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 Hygiene Candidate Manual* (CRDTS, 2016b) contains many topics of importance to candidates. CRDTS website (http://www.crdts.org/) is also a source of information for candidates. The dental hygiene section offers information about application and eligibility, the calendar for administration, examination content, scoring, forms and manuals, online application, and orientation. Relevant forms are provided online. Under frequently asked question, the appeals process is described, and a summary description of this process is also provided in the *Dental Hygiene Candidate Manual*. CRDTS has a review petition/appeals process for failing candidates who want to inquire about the accuracy of scoring. CRDTS will not re-score the examination, but will consider any appropriate evidence that points to alternative results. As described previously, failing scores are verified. A candidate who fails any examination receives a report itemizing deficient performances. Applicants may question a failing score using the formal procedures that CRDTS has established and described in the *Dental Hygiene Candidate Manual.*

Claim Supporting Validity

The claim is that these *standards* are met. The *Dental Hygiene Candidate Manual* (2016b) is the best source of information supporting this claim.

10. Security

CRDTS has taken many steps to ensure security in examination development, administration, scoring, and reporting. The following *standards* apply to security.

Table 16: *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.

CRDTS Central Office is located in Topeka, Kansas. There are six full-time employees using telecommunication to interact with other staff and CRDTS officers and CRDTS web servers via password-protected access to manage and process important or confidential documents.

The office is located on the lower floor of a two-story building with a rear-entry access for pickups and deliveries. The office is locked the majority of the time; a glass front door allows for visibility at the front desk to see who is seeking entry. At least one full-time employee is in the office on all weekdays.

Examination materials, such as *Progress Forms and Flow Sheets*, that are part of the candidate's permanent record are pre-printed with each candidate's individual sequential ID number and a 10-digit computer ID number that is a secure coded version of their social security number. In addition, the electronic equipment for scoring the exam is pre-loaded with each candidate's ID numbers, and the examiner ID numbers and names for all examiners assigned to the test site. This is done to ensure that all exam results are correctly identified.

CRDTS maintains a large supply of metal trunks with built-in combination locks for shipment of material to and from exams. CRDTS maintains an insured shipping contract with labels and pickup/ delivery dates prearranged for shipment to and from test sites. There are two laptops, two routers and extra ESD's sent to every site so that there is a backup system in case of a breakdown in electronic equipment. Although most test sites have their own wireless networks, all evaluation results are fed from the ESD's into the laptop computer through a preprogramed, encrypted, secure, dedicated wireless system. Once the hardware is set up by CRDTS staff at the test site, the secure, encrypted wireless network becomes active and the results are constantly monitored by an IT proctor throughout the exam to make sure that all evaluations have been uploaded and all candidates have complete results. At the end of the exam, the test site file is downloaded to a jump drive, then uploaded onto CRDTS' secure scoring website preparatory for final scoring and release of results. Therefore, the CRDTS Electronic Scoring system provides multiple options for the backup of results: exam laptop, jump drive, individual ESD's, secure website; if necessary, exam results can be reconstructed from any of these devices at any point during or after the evaluation process.

Multiple dedicated terminal servers are maintained at the offices of Premiere One in Topeka, Kansas. All these servers are protected with individual Barracuda Web Filters that provide content filtering, virus filtering, spyware download filtering, spyware communication blocking, spyware detection, and spyware removal from Windows desktop computers. The documents and functionality of each server is continuously backed up for easy recreation and access to important files and documents.

Test scores are processed on the secure website by staff members who gain varying levels of password-protected access. Once the scores are verified and processed, they are released to the candidate files within the appropriate exam site. Candidates may then see their scores via password-protected access to their file. Additionally, the appropriate educational institution is provided with a list of candidate scores for their students of record via the CRDTS Archive and Document system. Reports and documents are provided via a secure encrypted link e-mailed to the respective schools in a PDF format. These documents and reports are also archived according to a predetermined filing system within the Archive and Document system. This system is maintained and managed via a separate terminal server using the security and encryption protocols outlined above. It is the means by which CRDTS can manage, share and archive all documents and items necessary in a secure environment with continuous back-up that ensures security and continuity.

It must be recognized that security issues for clinical performance examinations differ greatly from security concerns with the potential exposure of written test items and a subsequent threat to validity. With clinical performance examinations, the "answers" to the test are published in advance in the performance criteria that appear in the candidates' manuals. Candidates have access in school to manikin modules that are identical to CRDTS' test modules, so they can practice test procedures as often as they wish before the exam. To ensure that candidates do not bring prepared teeth into the exam and substitute them for the testing modules that CRDTS provides, CRDTS orders its test modules with an exclusive color of acrylic gingiva that is available only to us. For patient-based procedures, all treatment selections, with current radiographs, are checked and recorded in advance so that teeth cannot be prepared in advance by the candidate or another experienced practitioner.

Claim Supporting Validity

CRDTS has a well-developed system ensuring security in all phases of examination planning, development, administration, scoring, and reporting. All *standards* concerning security have been met.

11. Documentation

Chapter 7 of the *Standard*s (2014) states:

> *"The objective of the documentation is to provide test users with the information needed to help them assess the nature and quality of the test, the resulting scores, and the interpretations based on the test scores"* (p. 123).

The table below provides *standards* related to documentation. Most of the *standards* in this table duplicate other *standards* throughout this report. As shown in the appendix, CRDTS has a large collection of documents attesting to meetings, publications, manuals, studies, and reports bearing on test development and validation. Throughout this technical report, references are made to these supporting documents that address and add to validity.

Table 17: *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.*

Claim Supporting Validity

Throughout this technical report, these documents are cited in reference to *standards*. By that, it is argued that validity is served and improved. The annual technical report alone stands as a single authoritative source of validity evidence matched to *standards*.

VALIDITY EVIDENCE BEARING ON SPECIFIC SUBTESTS

This final section of the technical report focuses on item quality and examiner consistency for each of the four subtests.

1. For item quality, mean item difficulty and discrimination are reviewed. As tables will show, the means and discriminations are very consistent, as they should be.

2. For examiner consistency, it is more complicated. All task performances are evaluated by three examiners. Although most observations have perfect agreement, a single disagreeable rating will not harm the candidate's score as the median is used instead of the mean. The reports of rater consistency appear in tables reflecting a high degree of agreement among examiners.

1. Extra/intra Oral Assessment

The extra/intra oral assessment subtest is a set of eight tasks weighted 16 points. Each task is scored 2 or 0. As noted previously, the examinee score is the median value of three examiners. Average performance as reported in Table 6 of this technical report is 95%. This result shows very high performance of the candidates. Table 18 provides some descriptive statistics for the eight items.

Table 18: Descriptive Statistics for the Extra/intra Oral Assessment

	1	2	3	4	5	6	7	8
Mean	1.85	1.97	1.93	1.87	1.91	1.89	1.88	1.81
S. D.	0.52	0.25	0.38	0.49	0.43	0.45	0.46	0.57
Disc.	0.40	0.27	0.37	0.42	0.42	0.42	0.46	0.51

As shown in this table, performance on each of the eight items is remarkably consistent, ranging from 1.86 to 1.93 (93% to 96%). The items all contribute reasonably well to the total extra/intra oral assessment score.

Rater consistency is 86.25%. As noted previously, this is very high. Disagreement among raters is handled by taking the median instead of the mean, so no candidate score is jeopardized by rater inconsistency or bias.

2. Periodontal Probing

This subtest consists of 12 items scored 0-1. Thus, a candidate can earn up to 12 points on this subject. Table 19 reports the set of descriptive statistics for this subtest.

Table 19: Descriptive Statistics for Periodontal Probing												
	1	2	3	4	5	6	7	8	9	10	11	12
Mean	0.97	1.00	0.98	0.98	0.99	0.99	1.00	1.00	1.00	1.00	1.00	1.00
S. D.	0.18	0.06	0.18	0.14	0.07	0.11	0.16	0.06	0.12	0.18	0.08	0.12
Disc.	0.52	0.40	0.50	0.53	0.43	0.46	0.54	0.44	0.52	0.51	0.45	0.45

Performance on these 12 tasks (test items) is nearly perfect. The means for 7 to 12 were 1.00, but this is a rounded value as some zero scores exist. The items are remarkably consistent in the performance levels. The discrimination indexes are relatively high. These items contribute consistently to the total periodontal probing score.

3. Scaling/subgingival Calculus Removal

This subtest consists of 12 tasks (test items). Each item earns a maximum of five points. This subscale is worth 60 points on the *NDHE.*

Table 20: Descriptive Statistics for Scaling/Subgingival Calculus Removal												
	1	2	3	4	5	6	7	8	9	10	11	12
Mean	4.30	4.48	4.40	4.39	4.34	4.38	4.31	4.34	4.27	4.35	4.29	4.38
S. D.	1.74	1.53	1.63	1.62	1.68	1.63	1.71	1.67	1.74	1.66	1.74	1.63
Disc.	0.48	0.43	0.45	0.52	0.56	0.54	0.54	0.54	0.55	0.54	0.53	0.55

The item means varied between 4.31 and 4.51, and the standard deviations are similar. The discrimination indexes were consistently high across all 12 items. Overall, the items appear to perform as expected.

Examiner consistency for this subtest was 83.10%. Although this subtest had the lowest degree of perfect examiner consistency, this level of consistency is high. Further, candidates' scores are not jeopardized due to the strategy of using the median of three ratings instead of the mean.

4. Supragingival Deposit Removal

This subtest consists of 12 tasks. Each task has a maximum score of 1 point. Thus, this subscale contributes a total of 12 points to the total score. Table 20 presents another set of descriptive item statistics.

Table 20: Descriptive Statistics for Supragingival Deposit Removal

	1	2	3	4	5	6	7	8	9	10	11	12
Mean	0.98	0.98	0.99	0.98	0.98	0.98	0.98	0.98	1.00	0.99	0.96	0.90
S. D.	0.22	0.17	0.12	0.12	0.12	0.13	0.12	0.14	0.00	0.16	0.28	0.44
Discrimination	--	--	--	--	--	0.18	--	0.17	0.39	0.55	0.68	0.81

Performance on these items is nearly perfect. The peculiar set of discrimination indexes is due to the fact that the first seven items have nearly perfect performance. The last four items had some negative one values that increased the item's discrimination ability.

SUMMARY OF VALIDITY EVIDENCE

CRDTS has designed and improved the *NDHE* over many years. As this technical report attests, *standards* have been met. Moreover, the argument for validity presented in this technical report and the evidence assembled supports the validity of interpreting a test score as a measure of clinical dental competency.

To summarize this evidence:

1. The *standards* cited in this technical report address validity directly and are well linked to the development of *CRDTS' NDHE*, its development, design, administration, scoring, and reporting.

2. Although the content consists of four independent skill areas, CRDTS has determined that the sum of performance constitutes sufficient evidence for a pass/fail decision.

3. Item development includes the creation of tasks and scoring protocols. As noted in documentation in this technical report, these tasks and scoring protocols are reviewed annually and occasionally improved.

4. Examiner consistency is high and this fact contributes to reliability. The resulting standard error of measurement helps develop a zone of uncertainty around the cut score of 75. The candidates having scores in this zone are low-performing. The risk of misclassification is unavoidable in test development and scoring due to random error. Candidates scoring in this zone of uncertainty need to improve their clinical skills if they wish to eliminate this risk.

5. Examination administration is standardized. Documents previously cited in this report show that administrative procedures are reviewed annually for polishing and fine-tuning.

6. Examiners are carefully selected, trained extensively, validated, monitored, and retrained if scoring is not consistently high.

7. Scoring is very systematic with high degree of quality control.

8. Scores are reported responsibly.

9. Security procedures are carried out and reviewed annually for improvement.

10. All validity evidence is well documented in this report or other documents cited in the appendix.

References

American Association of Dental Examiners (AADE). *Guidance for Clinical Licensure Examinations in Dentistry.* A document approved by the AADE General Assembly on October 23, 2003, revised September 2004 and March 2005. Chicago: Author

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

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.

Haertel, E. H. (2006). Reliability. In R. L. Brennan (Ed.) *Educational Measurement*, 4th edition, pp. 65-110. Westport, CN: Praeger.

Haladyna, T. M. (2006). *Validity studies.* In S. M. Downing & T. M. Haladyna (Eds.) *Handbook of test development*, pp. 131-154. 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.

Kane, M. T. (2006). Content-related validity evidence. In S. M. Downing & T. M. Haladyna (Eds.) *Handbook of test development*, pp. 131-154. Mahwah, NJ: Lawrence Erlbaum Associates.

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,* pp. 567-584. (2nd ed.).

Raymond, M. & 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.

Raymond, M. (2016). Job analysis, practice analysis, and the content of credentialing examinations In S. Lane, M. Raymond, & T. M. Haladyna (Eds.) *Handbook of test development*, pp. 144-164. (2nd ed.).

Appendix: Archive of Cited Documents Providing Validity Evidence

American Board of Dental Examiners (ADEX). (2006). *2006 Annual Report.* Author.

American Board of Dental Examiners (ADEX). (2007). *2007 Annual Report*. Author.

American Board of Dental Examiners (ADEX) (April 10, 2006) *ADEX Examination Committee Report.* Author.

American Board of Dental Examiners (ADEX). June 23, 2006). *ADEX Examination Committee*.

American Board of Dental Examiners (ADEX). August 26, 2006). *ADEX Examination Committee*. Author.

American Board of Dental Examiners (ADEX). (April 12, 2007). *Memorandum from Quality Assurance Committee*. Author.

American Board of Dental Examiners (ADEX). (April 17, 2007). *ADEX Examination Committee*. Author.

American Board of Dental Examiners (ADEX). December 5, 6, 7, 2007). *ADEX Chief Examiner's Report*. Author.

American Board of Dental Examiners (ADEX). (January 19, 2008a). *Minutes of ADEX Meeting*. Author.

American Board of Dental Examiners (ADEX). January 19, 2008b). *Examination Committee Meeting Minutes*. Author.

American Board of Dental Examiners (ADEX). (January, 22, 2008). *Minutes of ADEX Board of Directors*. Author.

American Board of Dental Examiners (ADEX). (August 21, 2008). *ADEX Quality Assurance Committee.* Author.

American Board of Dental Examiners (ADEX). (2008). *ADEX Dental Hygiene Committee Report to CRDTS Dental Hygiene ERC*. Chicago: Author.

CRDTS (October 16, 1993). *Examination Review Committee Minutes*. Topeka, KS: Author.

CRDTS (November 19-20, 1993). *Executive Committee Meeting Minutes*. Topeka, KS: Author.

CRDTS (September 7, 2003). *Examination Review Committee Minutes*. Topeka, KS: Author.

CRDTS (October 3, 2003). *Executive Committee Meeting Minutes*. Topeka, KS: Author.

CRDTS (2006). *Analysis of the 2005 Dental Hygiene Examination*. Topeka, KS: Author.

CRDTS (August 26, 2006). *CRDTS Dental Hygiene Examination Review Committee Meeting*. Topeka, KS: Author.

CRDTS (2007). *Analysis of the 2006 Dental Hygiene Examination*. Topeka, KS: Author.

CRDTS (August 24, 2007). *CRDTS Dental Hygiene Examination Review Committee Meeting*. Topeka, KS: Author.

CRDTS (2008a). *Analysis of the 2007 Dental Hygiene Examination*. Topeka, KS: Author.

CRDTS (2008b). *Background and current issues with the ADEX Dental Hygiene Examination.* Topeka, KS: Author.

CRDTS (August 22, 2008a). *CRDTS Dental Hygiene Examination Review Committee Meeting*. Topeka, KS: Author.

CRDTS (August 22, 2008b). *Recommendations for the CRDTS Hygiene Examination Review Committee.* Topeka, KS: Author.

CRDTS (August 22, 2008c). *Attachment A & Attachment B.* Topeka, KS: Author.

CRDTS (2009). *Analysis of the 2008 Dental Hygiene Examination*. Topeka, KS: Author.

CRDTS (July 11-12, 2009). *CRDTS Dental Hygiene Examination Review Committee Meeting*. Topeka, KS: Author.

CRDTS (2010). *Analysis of the 2009 Dental Hygiene Examination*. Topeka, KS: Author.
CRDTS (July 10-11, 2010). *CRDTS Dental Hygiene Examination Review Committee Meeting*. Topeka, KS: Author.
CRDTS (2011). *Analysis of the 2010 Dental Hygiene Examination*. Topeka, KS: Author.
CRDTS (July 9-10, 2011). *Dental Hygiene Examination Review Committee Meeting*. Topeka, KS: Author.
CRDTS (2012). *Dental Hygiene Examination Occupational Analysis*. Topeka, KS: Author.
CRDTS (July 14, 2012). *Dental Hygiene Examination Review Committee Meeting*. Topeka, KS: Author.
CRDTS (July 13, 2013). *Dental Hygiene Examination Review Committee Meeting*. Topeka, KS: Author.
CRDTS (2014a). *Dental Hygiene Examiner Materials*. Topeka, KS: Author.
CRDTS (2014c). *Dental Coordinator Notebook*. Topeka, KS: Author.
CRDTS (July 11, 2015) *Dental Hygiene Examination Committee Meeting*. Topeka, KS.
CRDTS (January 2016). *2016 Annual report to program directors on dental hygiene examination results*. Topeka, KS: Author.
CRDTS (2016a). *Dental Hygiene Examiner's Manual*. Topeka, KS: Author.
CRDTS (2016b). *Dental Hygiene Candidate Manual*. Topeka, KS: Author.
CRDTS (July 11, 2015) *Dental Hygiene Examination Committee Meeting*. Topeka, KS.
CRDTS (July 19, 2016). *Dental Hygiene Examination Review Committee Meeting*. Topeka, KS: Author.
Klein, S. P. (April 27, 2008). *Technical report on the ADEX/CRDTS Dental Hygiene Licensing Examination: Class of 2007*.
Klein, S. P. (May 11, 2009). *Technical report on the ADEX/CRDTS Dental Hygiene Licensing Examination: Class of 2008*. Santa Monica, CA: Author.
Littlefield, J. (April 25, 2009). Comments on ADEX's White Paper Dated March 18, 2009. San Antonio, TX: Author.
Ray, L. (2014). *Examiner Profile Statistics. Dental Hygiene Clinical Examination*. Topeka, KS: *Examination*. CRDTS, Topeka, KS.
Ray, L. (August 2015). *Examiner Profile Statistics. Dental Hygiene Clinical Examination*. Topeka, KS: Author.
Ray, L. (August 2016). *Examiner Profile Statistics. Dental Hygiene Clinical Examination*. Topeka, KS: Author.