

JCNDE JOINT COMMISSION
ON NATIONAL
DENTAL EXAMINATIONS



2020 Technical Report National Board Dental Hygiene Examination (NBDHE)



Technical Report
National Board Dental Hygiene Examination

2020

Executive Summary

Technical Report: The National Board Dental Hygiene Examination 2020

The 2020 edition of the Technical Report for the National Board Dental Hygiene Examination is the main source of validity evidence available to state licensing boards and other users of dental hygiene examination results. Validity is the most important consideration for any examination program. For the dental hygiene examination, validity refers to the degree to which logic and evidence support the use and interpretation of examination results for making pass/fail decisions affecting candidates for licensure to practice dental hygiene. The Technical Report contains both direct evidence and references to other documents and sources of information that contribute to this body of validity evidence. This report also provides background and historical information that allow users of the examination the opportunity to understand the factors that have contributed to the development of this program.

The NBDHE 2020 Technical Report focuses on the National Board Dental Hygiene Examination (NBDHE) testing program and findings for the 2020 calendar year.

The content of the Technical Report is presented in such a way as to address a series of standards regarding the validity of credentialing examinations (American Educational Research Association, American Psychological Association, and the National Council on Measurement in Education, 2014). Successful completion of a credentialing examination by individuals indicates that they have achieved an acceptable level of performance in an area of knowledge. Some of the principal information presented in this Technical Report is summarized below.

- **Purpose:** The purpose of the National Board Dental Hygiene Examination is to measure whether a candidate possesses an entry-level knowledge adequate for the competent practice of dental hygiene. This entry-level knowledge includes the ability to understand important information from the basic biomedical, dental, and dental hygiene sciences, and the ability to apply such information in a problem-solving context.
- **Content:** Traditionally, the examination's content specifications have been based on expert judgments that have been obtained in two ways. First, the dental hygiene and dental experts that serve on test construction teams have honed the specifications over the years by recommending changes. Second, educators and practitioners have routinely recommended modifications during conferences. Beginning in 2005, content specifications have been based on the findings of practice analyses conducted periodically. Test construction teams are responsible for recommending minor modifications during the interim period. The Joint Commission with its Committee on Dental Hygiene approves all changes to the content specifications.
- **Item and Examination Development:** Test construction teams are responsible for the development of items and forms/editions of the examination using Joint Commission guidelines for writing high-quality, multiple-choice items.
- **Standard Setting and Scoring:** The examination is criterion referenced and not norm-referenced. Specifically, examination results and the pass/fail standard are determined by specific criteria, not by the process sometimes known as "grading on a curve." Experts consisting of educators and practitioners establish the criteria. The standard is maintained across editions of the examination through the use of equating processes. Essentially, the equating process controls for subtle differences

- in the difficulty of items that appear on different examination forms, placing exam results on a common metric regardless of which exam form was administered.
- **Administration:** A high level of security is maintained on all examination materials. Strict precautions are in place at the Joint Commission's offices and testing centers to ensure that materials are not compromised before, during, or after test administration.

This report provides detailed information relating to the areas above along with information related to topics such as examination history, administration, the rights and responsibilities of candidates, and the failure rates of candidates. A copy of the Technical Report is available for download at ada.org/NBDHE.

Table of Contents

1. Introduction	4
2. Purpose of the National Board Dental Hygiene Examination	5
3. Historical Perspective	6
4. The 2020 NBDHE Program	6
5. Validity, Validation, and Validity Evidence	7
6. Professional Test Standards.....	8
7. Legal Issues.....	8
8. Validity Evidence in this Technical Report	9
9. Content Basis for the Examination	9
10. Item Development	13
11. Item Validation.....	15
12. Test Design and Development	15
13. Test Administration.....	19
14. Score Reliability.....	20
15. Standard Setting.....	21
16. Scaling/Equating/Comparability of Test Forms	26
17. Scoring and Reporting Test Results	31
18. Rights and Responsibilities of Test Takers	32
19. Threats to Validity.....	33
20. Validity Studies.....	34
21. Security	34
22. Guidelines for High-Stakes Testing	36
References.....	39
Appendix A: Dental Hygiene Examination General Specifications (2020).....	42

1. Introduction

High-stakes examination programs, such as those of the Joint Commission on National Dental Examinations (Joint Commission), should be concerned with validity, because validity relates to the degree to which logic and evidence support the interpretation and use of examination scores for making pass/fail decisions affecting candidates for licensure or certification. The Joint Commission also has an obligation to inform its constituency (i.e., state boards of dentistry) concerning its efforts to provide the highest quality examination programs possible.

In light of the recommendations from the Standards for Educational and Psychological Testing (American Educational Research Association [AERA], American Psychological Association [APA], and the National Council on Measurement in Education [NCME], 2014), the Joint Commission publishes annual technical reports concerning the National Board Dental and Dental Hygiene Examinations. This technical report focuses on the National Board Dental Hygiene Examination (NBDHE) testing program and findings for the 2020 calendar year. The objective of the Joint Commission's technical reports is to provide examination users with information to help them evaluate and understand the validity evidence collected in support of the examination results they interpret and use.

As suggested by the Standards, this technical report provides a comprehensive summary of information regarding the validation of the National Board Dental Hygiene Examination, including the development of the examination, the rationale underlying the examination, and evidence of validity and score reliability. Additionally, this report describes information concerning changes and enhancements to the examination program, as well as background and historical information that allows readers to fully understand the history and events that have contributed to the development of this examination program. This report shows that the Joint Commission endeavors to provide the highest quality National Board Dental Hygiene Examination program possible.

Ten standards from the Standards are relevant in the current context, concerning the supporting documentation that should be found in a technical report. These standards are presented in Table 1.1.

Table 1.1
Standards Pertaining to Supporting Documentation Found in a Technical Report

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 misuses of a test can be reasonably anticipated, cautions against such misuses should be specified.

7.2 The population for whom a test is intended and specifications for the test should be documented. If normative data are provided, the procedures used to gather the data should be explained; the norming population should be described in terms of relevant demographic variables; and the year(s) in which the data were collected should be reported.

7.3 When the information is available and appropriately shared, test documents should cite a representative set of the studies pertaining to general and specific uses of the test.

7.4 Test documentation should summarize test development procedures, including description 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.

7.5 Test documents should record the relevant characteristics of the individuals or groups of individuals who participated in data collection efforts associated with test development or validation; the nature of the data that were contributed; the nature of judgments made by subject matter experts; the instructions that were provided to participants in data collection efforts for their specific tasks; and the conditions under which the test data were collected in the validity study.

7.7 Test documents should specify user qualifications that are required to administer and score a test, as well as the user qualifications needed to interpret the test scores accurately.

7.8 Test documentation should include detailed instructions on how a test is to be administered and scored.

7.9 If test security is critical to the interpretation of test scores, the documentation should explain the steps necessary to protect test materials and to prevent inappropriate exchange of information during the test administration session.

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 language the test takers can understand and that assist them in understanding the test scores.

2. Purpose of the National Board Dental Hygiene Examination

The first and most fundamental step in the development of any examination program is to establish a purpose. The purpose of the NBDHE program is to measure whether a candidate possesses entry-level knowledge and cognitive skills adequate for the competent practice of dental hygiene. This entry-level knowledge includes the ability to understand important information from the biomedical, dental, and dental hygiene sciences, and the ability to apply such information in a problem-solving context. This examination purpose is in complete alignment with the Program duties of the Joint Commission.

The Joint Commission is the agency that oversees examination design, development, administration, scoring, and reporting. The Department of Testing Services of the American Dental Association provides operational and technical support with respect to the corresponding outlined activities. Prior to November 2019, the Joint Commission's Bylaws and Standing Rules represented focal governance documents for the Joint Commission, and provided descriptions of Joint Commission membership, as well as the committees that serve the Joint Commission and their roles. Beginning in November 2019, the JCNDE replaced the aforementioned two documents with the *Rules of the JCNDE* and the *Operational and Policy Manual* of the JCNDE, which serve as the Joint Commission's governance documents as the JCNDE moves forward.

Five standing committees serve the Joint Commission. Each committee is assigned a portion of the materials to be considered by the Joint Commission, and each committee is responsible for

making specific recommendations to the Joint Commission. The Committee on Administration deals with operations for both the dental and dental hygiene examinations. This includes security, examination rules and regulations, policies and procedures, and budget. The Committee on Dental Hygiene is responsible for National Board Dental Hygiene Examination content and examination specifications, test construction procedures, scoring procedures, dissemination of information about examination procedures and validity, and matters affecting finance. The Committee on Examination Development deals with the National Board Dental Examinations (Parts I and II and the Integrated National Board Dental Examination), their content and examination specifications, test construction procedures, scoring procedures, and reporting. It also concerns itself with the dissemination of information about the examination process and validity. The Committee on Research and Development focuses on research and development activities (e.g., psychometric investigations) related to both the dental and dental hygiene examination programs. The Committee on Communications and Stakeholder Engagement focuses on the communication needs of the JCNDE and corresponding communities of interest, as the JCNDE implements its examination programs.

3. Historical Perspective

The National Board of Dental Examiners was established in 1928 as a standing committee of the American Dental Association for the purpose of providing and conducting written examinations for use at the discretion of state boards of dentistry in licensing dentists. These examinations were to provide a national standard for the knowledge of basic and clinical sciences necessary for the competent practice of dentistry. The practical demonstrations of clinical skills were reserved for individual state examinations. The responsibilities of the National Board included not only developing and administering National Board examinations, but also formulating rules and regulations pertaining to those examinations.

When the National Board Dental Hygiene Examination program was introduced in 1962, there were four examinations consisting of 100 test items each. Each of the four examinations covered three subjects. In 1973, a single, comprehensive dental hygiene examination consisting of approximately 400 test items replaced the four examination battery. This comprehensive examination was organized around functions that a dental hygienist could be asked to perform. By 1990, all U.S. licensing jurisdictions accepted the National Board Dental Hygiene Examination as evidence of fulfillment of the written examination requirement for licensure.

In 1998, the examination was restructured to include 350 items, of which 150 were associated with patient case material. In 2012, the Joint Commission moved to pass/fail reporting of results for candidates who passed the examination. Candidates who fail receive their overall score and results feedback for remediation purposes.

4. The 2020 NBDHE Program

NBDHE administrations in 2020 took place in accordance with policies and procedures described in several governance and policy documents which are referenced throughout this report. These documents are scrutinized and reviewed on an annual basis (at minimum), with updates occurring to ensure policies and procedures remain appropriate and in accordance with industry best practices and the purpose of JCNDE examination programs. References to these documents within the text of this technical report should be interpreted accordingly, based on the list of documents found in the table below. The document policies cover administration periods within the indicated calendar year.

DOCUMENT (2020)
NBDHE Guide (2020)
Test Item Development Guide (2020)
Orientation Manual for Test Constructors (2020)
JCNDE Test Construction Teams and Selection Criteria (2020)
Rules of the JCNDE (2020)
Operational and Policy Manual of the JCNDE (2020)

The Joint Commission’s NBDHE Candidate Guide provides a description of the NBDHE Program. The dental hygiene examination is a comprehensive examination consisting of 350 multiple-choice items. The examination has two components: 200 discipline-based items and 150 case-based items (see examination specifications under Appendix A). Test items cover functions that a dental hygienist is expected to be able to perform. Only functions that may be delegated to a dental hygienist in a majority of states are included in the examination. Items are selected by test construction teams in accordance with the Dental Hygiene Examination Specifications. Test constructors are appointed to represent the following areas: (1) basic sciences, (2) oral radiology, (3) periodontics, (4) dental hygiene curriculum, (5) clinical dental hygiene, (6) community dental health, (7) oral medicine/oral diagnosis, and (8) special needs.

Formats Used

The dental hygiene examination is composed of multiple-choice items. Three multiple-choice formats are used. The case-independent format evaluates basic science knowledge pertinent to dental hygiene practice. The case-dependent format uses case materials consisting of a patient dental/medical history, a dental chart, radiographs, and photographs. These items serve as stimulus material for a series of case-associated questions. The National Board Dental Hygiene Examination also includes a testlet format. This format consists of one or two paragraphs describing a case study or problem from which at least four items are derived. In all cases, the key features of multiple-choice items are a stem pairing a question or statement with a list of possible responses. National Board items have at least three, but not more than five possible responses.

Released Items

The Joint Commission makes available for purchase sets of retired items and cases. These released items are no longer in use for licensure purposes. They are provided to familiarize candidates, educators, and others with the format and general nature of the examination. A disclaimer cautions users that some items may be outdated or no longer representative of content included in current examinations. All released items are the copyrighted property of the American Dental Association (ADA), and may not be reproduced in any format, whether paper or electronic, without the express written permission of the ADA.

5. Validity, Validation, and Validity Evidence

Validity is " the degree to which evidence and theory support the interpretations of test scores for purposed uses of tests" (AERA, APA, & NCME, 2014, p. 11). Every examination program has a purpose. To fulfill this purpose, an examination score has a desired interpretation and an

intended use. The sponsor of the examination program creates a logical argument and assembles validity evidence that is relevant to this argument. In some instances, acquired validity evidence reveals weaknesses or deficiencies. In these instances, the testing organization should take steps to address any deficiencies, with the goal of strengthening the validity argument. Essentially, validation is the investigative process of creating this argument and collecting evidence relevant to this argument and the examination purpose.

All candidates for licensure must meet many criteria before they are licensed to practice dental hygiene in a state within the US. Each state has the authority to issue the license, although in dental hygiene, as in many other professions, national standards exist.

The dental hygiene examination measures the professional knowledge that dental hygienists must possess, and their ability to apply this knowledge in settings involving patient care. Through their performance on the exam, candidates who achieve passing scores have demonstrated that they possess the required level of knowledge to safely practice. This technical report presents validity evidence and references to validity evidence that support both the interpretation and use of examination results.

6. Professional Test Standards

Large testing organizations responsible for developing, administering, and scoring examinations need criteria or standards upon which to judge their effectiveness. Three professional organizations have joined forces to create the latest version of such standards (AERA, APA, & NCME, 2014). These standards provide useful information to guide testing organizations in their validation efforts.

Throughout this technical report, validity evidence is identified and connected to testing standards. Many sections of this technical report correspond to specific chapters in the standards (AERA, APA, & NCME, 2014).

AERA (2000) has also issued a set of guidelines that are intended to be used with high-stakes, high school graduation examination programs. Some of these guidelines apply to the dental hygiene examination. In section 22 of this technical report, these guidelines are reviewed against the validity evidence presented in this technical report.

7. Legal Issues

Any examination program involving high-stakes decisions runs a risk of legal challenge based on validity. Thus, it is important that such an examination program be designed to withstand legal challenge.

This technical report represents an effective way to present the examination validity argument and validity evidence. This document organizes, describes, and presents a large array of validity evidence, and makes this information available to the public. This action speaks to the fact that the Joint Commission has acted responsibly in its duty to develop and administer an examination program capable of fulfilling its intended purpose.

8. Validity Evidence in this Technical Report

Validity and reliability are essential measurement concepts. Examinations are deemed valid when evidence exists to support the inferences for which the test has been designed. In the present case, the test provides accurate information concerning whether candidates have the requisite knowledge to safely practice dental hygiene. Examinations are considered reliable when they provide this accurate measurement on a consistent basis.

This report is organized to address major categories of validity evidence. In each major category, reference is made to one or more AERA, APA, and NCME (2014) standards. The first three standards are:

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.

This technical report and the corresponding references to other documents provide evidence that standards 1.0, 1.1 and 1.2 have been met. The information provided demonstrates that the Joint Commission has acted responsibly in validating the Dental Hygiene Examination.

Much of the rest of this report addresses important categories of validity evidence. These categories and their corresponding section number within this report are as follows:

9. Content Basis for the Examination
10. Item Development
11. Item Validation
12. Test Design and Development
13. Test Administration
14. Reliability of Test Scores
15. Standard Setting
16. Scaling/Equating/Comparability of Test Forms
17. Scoring and Reporting Test Scores
18. Rights and Responsibilities of Test Takers
19. Threats to Validity
20. Validity Studies
21. Security
22. Guidelines for High-Stakes Testing

9. Content Basis for the Examination

The content of a certification/licensure test of knowledge for any profession—and the procedures used to specify that content—represent a primary type of validity evidence. Table

9.1 lists standards related to the content of such examinations. Key elements constituting validity evidence include (1) a practice analysis that identifies the knowledge necessary for safe practice as a dental hygienist, (2) the development of examination specifications, and (3) the role of content experts in many procedures involved in identifying and codifying content.

Table 9.1
Standards That Apply to the Content Basis of the Examination

1.9 When a validation rests in part on the opinions or decisions of expert judges, observers, or raters, procedures for selecting such experts and for eliciting judgments or ratings should be fully described. The qualifications and experience of the judges should be presented. The description of procedures should include any training and instructions provided, should indicate whether participants reached their decisions independently, and should report the level of agreement reached. If participants interacted with one another or exchanged information, the procedures through which they may have influenced one another should be set forth.

1.11 When the rationale for test score interpretation for a given use rests in part on the appropriateness of test content, the procedures followed in specifying and generating test content should be described and justified with reference to the intended population to be tested and the construct the test is intended to measure or the domain it is intended to represent. If the definition of the content sampled incorporates criteria such as importance, frequency, or criticality, these criteria should also be clearly explained and justified.

1.12 If the rationale for score interpretation for a given use depends on premises about the psychological processes or cognitive operations of test takers, then theoretical or empirical evidence in support of those premises should be provided. When statements about the processes employed by observers or scorers are part of the argument for validity, similar information should be provided.

4.0 Tests and testing programs should be designed and developed in a way that supports the validity of interpretations of the test scores for their intended uses. Test developers and publishers should document steps taken during the design and development process to provide evidence of fairness, reliability, and validity for intended uses for individuals in the intended examinee population.

4.1 Test specification 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 In addition to describing intended uses of the test, the test specifications should define the content of the test, the proposed test length, the item formats, the desired psychometric properties of the test items and the rest, and the ordering of items and sections. These specifications should also specify the amount of time allowed for testing; directions for the test takers; procedures to be used for test administration, including permissible variations; any materials to be used; and scoring and reporting procedures. Specifications for computer-based tests should include a description of any hardware and software requirement.

4.6 When appropriate to documenting the validity of test score interpretations for intended uses, relevant experts external to the testing programs should review the test specifications to evaluate their appropriateness for intended uses of test scores and fairness for intended test

takers. The purpose of the review, the process by which the review is conducted, and the results of the review should be documented. The qualifications, relevant experiences, and demographic characteristics of expert judges should also be documented.

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.

4.12 Test developers should document the extent to which the content domain of a test represents the domain defined in the test specifications.

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 in support of the interpretation for the use of a test for employment decisions or credentialing, a close link between test content and the job or professional/occupational requirements should be documented.

11.13 The content domain to be covered by a credentialing test should be defined clearly and justified in terms of the importance of the content for credential-worthy performance in an occupation or profession. A rationale and evidence should be provided to support the claim that the knowledge or skills being assessed are required for credential-worthy performance in that occupation and are consistent with the purpose for which the credentialing program was instituted.

Practice Analyses

In 2017, the Joint Commission relied on findings from a 2016 practice analysis survey to approve minor changes to the content specifications for the NBDHE. The specifications in place prior to that time were based on the results of a 2009 practice analysis. As part of the 2016 practice analysis, validation evidence was obtained by collecting ratings from a sample of active, full-time dental hygienists who had been in practice for ten years or less, concerning the frequency and importance of 30 competencies judged relevant to patient care. The surveyed hygienists were asked to rate each competency with respect to its importance to patient care, and its frequency of use in patient care. The levels of the rating scale were defined as follows:

Importance to Patient care:

4. Extremely important
3. Very important
2. Important
1. Somewhat important
0. Not important

Frequency of Use in Patient Care:

5. More than 5 times per day
4. 3-5 times per day
3. 1-2 times per day

2. 1-4 times per week
1. Less than once per week
0. Never

The Joint Commission distributed the practice analysis survey to 43,743 dental hygienists. A total of 2,853 (6.5%) provided valid responses. The mean frequency rating and mean importance rating were calculated for each competency. The mean frequency ratings ranged across competencies from 1.91 to 5.93. The mean importance ratings ranged from 3.56 to 4.85. The multiplicative model (Kane, Kingsbury, Colton, & Estes, 1989) was used to provide an overall index of importance for each competency. The overall importance ratings were used to determine the number of items that should be devoted to each competency. The numbers of items devoted to the competencies were then distributed across individual content elements based on the judgments of experts. The revised content specifications reflected the surveyed hygienists' frequency and importance ratings, and the study's overall findings confirmed the validity of the NBDHE. In June 2017, the Joint Commission approved the practice analysis and revised content specifications. These revised content specifications were implemented in 2019.

Examination Specifications

The content specifications for the National Board Dental Hygiene Examination are based on practice analysis findings. In the interim period between practice analyses, refinements are made based on the judgments of subject matter experts. These judgments have been obtained in two ways. First, the dental hygiene and dental experts who serve on dental hygiene test construction teams hone the content specifications over interim years by recommending revisions. The experts who serve on these teams are educators and full-time practitioners. Second, educators and practitioners, including state board members, have routinely recommended modifications to the specifications during conferences or through correspondence with the Joint Commission. All recommendations, regardless of the source, are reviewed by the Joint Commission through its standing Committee on Dental Hygiene. The Joint Commission's Committee on Dental Hygiene includes four Joint Commissioners, three dental hygienists, and a student representative.

The 2020 National Board Dental Hygiene Examination specifications appear in Appendix A. In 2020, a shortened version of the NBDHE – referred to as the short-form-NBDHE – was made available in response to circumstances surrounding COVID-19. Content appearing in the shortened forms is proportionally representative of the NBDHE test specifications. Test specifications for these shortened forms also appear under Appendix A.

The Role of Content Experts

Content experts play a vital role that is rooted in the purpose of the National Board Dental Hygiene Examination, and the validity argument that supports its usage. The purpose of this examination is to assist boards of dentistry in determining the qualifications of dental hygienists seeking a license to practice. The examination presumes that candidates have successfully completed appropriate dental hygiene education in an accredited dental hygiene program or the equivalent. The test construction process serves as a complement to state or regional clinical (or practical) examinations in the licensure process.

The examination assesses candidates' knowledge and problem solving skills in areas such as the scientific basis for dental hygiene practice, the provision of clinical dental hygiene services, and community health/research principles. The examinations require candidates to demonstrate

their knowledge and abilities by answering test items that sample the content domain. Through this demonstration, the measurement process discriminates between those who have the required knowledge level to safely begin practice and those who do not.

10. Item Development

The most essential building block of any examination is the item. The development and validation of test items is one of the most important steps in examination development. The Joint Commission greatly values item development and validation, and continues to invest considerable resources in both activities. Relevant standards are provided in Table 10.1. Section 11 addresses item analysis and evaluation.

Who Writes Test Items?

The JCNDE annually approves and reapproves test constructors into the NBDHE test constructor pool. Approval into the NBDHE test constructor pool is for three years, after which the test constructor would need to reapply and receive JCNDE re-approval to remain in the test constructor pool. On an annual basis, and based on those individuals approved within the NBDHE test constructor pool, staff select test constructors to attend TCT meetings for the upcoming year. Details concerning the necessary qualifications for test constructors, the structure of the teams, duties, and other relevant information appear in Section 12. This information can also be found in Joint Commission governance documents.

Once staff select test constructors to attend TCT meetings for the upcoming year, letters are sent to new test constructors to acknowledge their appointments and familiarize them with the examination development process. Each test constructor receives the following materials: Test Item Development Guide, Orientation Manual for Test Constructors, Dental Hygiene Examination Specifications, and DTS Contributor Agreement Form.

Test Constructors review the examination specifications and ensure they are reflected in examination development. They are also responsible for constructing a clear, precise, and cohesive group of items for each examination. Consultants review final drafts of the examination to ensure the consistency and coherence of both the case-independent and case dependent sections of the examination.

When new test constructors come to their first meeting, returning test constructors serve as mentors and provide guidance and instruction. The Test Item Development Guide describes the different item formats and general guidelines for writing items. The Orientation Manual for Dental and Dental Hygiene Test Constructors describes the responsibilities of the test constructors, the general item-development process, selection criteria, and the composition of each team. The Guidelines are provided to all test constructors and specialty boards, and anyone else requesting them through the Joint Commission's office. A manual of style serves as a guide in the production of the final examinations.

Table 10.1
Standards Relevant to Item Development and Validation

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.

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.

4.9 When item or test form tryouts or field tests are conducted, the procedures used to select the sample(s) of test takers as well as the resulting characteristics of sample(s) should be documented. The sample(s) should be as representative as possible of the population(s) for which the test is intended.

4.10 When a test developer evaluates the psychometric properties of items, the model used for that purpose (e.g., classical test theory, item response theory, or another model) should be documented. The sample used for estimating item properties should be described and should be of adequate size and diversity for the procedure. The process by which items are screened and the data used for screening, such as item difficulty, item discrimination, or item differential functioning (DIF) for major examinee groups, should also be documented. When model-based methods (e.g., IRT) are used to estimate item parameters in test development, the item response model, estimation procedures, and evidence of model fit should be documented.

4.11 Test developers should conduct cross-validation studies when items or tests are selected primarily on the basis of empirical relationships rather than on the basis of content or theoretical considerations. The extent to which the different studies show consistent results 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.

Item Formats

Standard 4.2 refers to identifying item formats in the examination specifications. The dental hygiene examination uses three multiple-choice formats. The case-independent format evaluates basic science and clinical knowledge that is pertinent to patient care. The case-dependent format uses case materials consisting of a patient dental/medical history, a dental chart, radiographs, and photographs. The dental hygiene examination also includes testlets. Testlets consist of one or two paragraphs describing a case study or problem, from which three to six items are derived. In all cases, the key features of multiple-choice items are a stem pairing a question or statement with a list of possible responses. National Board items have between three and five possible responses.

The Process of Examination Revision

The National Board Dental Hygiene Examination is subject to a review and revision process to address unsatisfactory items. Test items are unsatisfactory if they are too easy, too difficult, or fail to discriminate between stronger and weaker candidates. In reviewing test items, Test Construction Teams (TCTs) look at two key factors: the proportion of individuals answering an item correctly (i.e., the p-value) and the point-biserial correlation (rpb) between item and examinee performance. P-values provide information concerning item difficulty while point-

biserial correlations provide information concerning item discrimination. The Joint Commission accepts a broad range of item difficulties, but items that are too easy (virtually all candidates answer correctly) or too difficult (virtually no candidates answer correctly) are typically less useful from a measurement perspective. The following item difficulty and discrimination ranges were considered by the Joint Commission.

Table 10.2
Discrimination and Difficulty Indices Ranges and Interpretations
Dental Hygiene Examination

<i>Discrimination (r_{pb})</i>	<i>Difficulty (p-value)</i>
H = .26 or higher	E = .90 or higher
M = .08 to .25	M = .40 to .89
L = under .08	D = .00 to .39
<i>(r_{pb}): H — High; M — Medium; L — Low</i>	
<i>(p-value): E — Easy; M — Medium; D — Difficult</i>	

For an item to be considered effective, Joint Commission standards dictate that it must produce a difficulty index between .40 and .89, and a corresponding discrimination index of .08 or higher for the dental hygiene examination. Test items that do not meet these standards are scrutinized and become candidates for elimination or revision.

11. Item Validation

After an item is written, Downing and Haladyna (1997) recommend a series of reviews to improve the quality of the item. Evidence should be presented that qualified personnel have conducted these reviews. Standards 4.7, 4.8, 4.9, and 4.10 (previously presented in Table 10.1) show those standards pertaining to item validation.

The Standards (AERA, APA, & NCME, 2014) suggest that those test items that count toward candidates' scores should exhibit sound psychometric characteristics. Specifically, item difficulty and discrimination should compare favorably with the item-performance standards set by the Joint Commission. Item statistics are obtained through analysis of data from administrations involving a representative sample of candidates.

Evaluating and Revising Weak or Unacceptable Items

The Joint Commission (November 1995) has published a document that is intended to help test constructors review unsatisfactorily performing items and then revise or retire such items. This activity replenishes the item bank and helps test constructors sharpen their ability to evaluate and improve items.

12. Test Design and Development

The design of each examination is a very important step in test development. The set of items chosen for each examination must conform to the examination specifications. Not only must

content requirements be met, but also the psychometric characteristics should be comparable. Table 12.1 lists standards that pertain to test design and development.

Table 12.1
Standards Relevant to Test Design and Development

4.0 Tests and testing programs should be designed and developed in a way that supports the validity of interpretations of the test scores for their intended uses. Test developers and publishers should document steps taken during the design and development process to provide evidence of fairness, reliability, and validity for intended uses for individuals in the intended examinee population.

4.7 The procedures used to develop, review, and try out items and to select items from the item pool 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

Examinations are designed with the direct participation of test construction teams, and their development is supervised by a staff of specialists from the Joint Commission's test development area. This process ensures that the expertise of highly qualified dental and dental hygiene educators, licensed dentists, and dental hygienists is fully used in the selection of items and the examination design. Staff specialists provide technical support and guidance to ensure that the desired technical qualities of the examination are achieved during this design phase.

The Joint Commission convenes several Test Construction Teams (TCTs). The details of eligibility, recruitment, and service are provided in this section. As noted earlier in this technical report, these teams also write and evaluate examination items as part of the item development phase of the overall examination development process.

The Role of a Test Constructor

The role of test constructors is fundamental to both the validity and reliability of the dental hygiene examination. Test constructors are responsible for revising the examination specifications based on current knowledge and practices within the field of dental hygiene and implementing these specifications in their item selection for each examination. This is one method toward ensuring test validity. In addition, test constructors are responsible for constructing a clear, precise, and cohesive group of test items for each examination. This directly influences score reliability. Therefore, it is essential to the quality of the examination that test constructors use their subject-matter expertise, their familiarity with the curriculum in accredited programs, and their awareness of what is important to the practice of dental hygiene, in the construction of each new examination. Most of this work is accomplished through collaborative and meticulous work in meetings. Test constructors meet in discipline or case-based teams each year to engage in their examination development activities.

The selection of test constructors from a pool of volunteers is based on a set of qualifications. Evidence that these qualifications have been met include credentials demonstrating subject matter expertise and peer recognition of expertise. Maintaining geographical diversity is also a factor in placing test constructors on specific teams. A description of the selection criteria for dental hygiene test constructors appears below.

Criteria for Dental Hygiene Test Constructors

A document entitled JCNDE Test Construction Teams and Selection Criteria provides criteria concerning the qualifications of test constructors. The Dental Hygiene Examination is constructed by teams of consultants with subject-matter expertise in the following eight areas.

Basic Sciences: The basic sciences include anatomy, histology, physiology, biochemistry and nutrition, microbiology and immunology, pathology, pharmacology, and oral biology.

- Doctoral degree in a biomedical science, or a dentist or dental hygienist with an advanced degree in a biomedical or dental science
- At least three years' experience within the last five years teaching a biomedical or dental science to dental hygiene students

Radiology:

- Dentist or dental hygienist with a baccalaureate degree from an accredited program
- Formal education in dental radiology beyond what was provided in a dental or dental hygiene program
- At least three years' experience within the last five years teaching radiology

Periodontics:

- Graduate of an accredited dental or dental hygiene program with advanced formal education or training in periodontics
- At least three years' experience within the last five years teaching or practicing periodontics

Oral Medicine/Oral Diagnosis:

- Dentist with advanced clinical training
- At least three years' experience within the last five years teaching oral medicine/oral diagnosis

Special Needs Professional:

- Dentist or dental hygienist with advanced clinical training
- At least three years' experience within the last five years in teaching a clinical science

Dental Hygiene Curriculum:

- Dental hygienist who has graduated from an accredited program
- Advanced degree, preferably in dental hygiene
- At least three years' experience within the last five years teaching dental hygiene students
- Experience in curriculum design as a dental hygiene program director, member of a dental hygiene curriculum committee, or accreditation consultant for dental hygiene

Clinical Dental Hygiene:

- Dental hygienist who has graduated from an accredited program
- Baccalaureate degree in dental hygiene, education, or a biomedical science
- At least three-years' experience, preferably within the last five years, teaching and practicing clinical dental hygiene; full-time or part-time in private practice or faculty practice.

Community Dental Health:

- Dentist or dental hygienist who has graduated from an accredited program
- Advanced degree in public health or related field

- At least three years' experience within the last five years in a public health position or teaching community and public health courses to dental or dental hygiene students

Dental Hygiene Test Construction Teams

Six test construction teams work together to develop the content of the dental hygiene examination. The number of test constructors on a team ranges from two to eight consultants to the Joint Commission.

Component A Teams

Component A consists of three discipline-based teams addressing: 1) basic biomedical sciences, 2) clinical dental hygiene, and 3) community dental health. The Dental Hygiene I Team consists of three basic science experts, and one dental hygiene curriculum expert. The Dental Hygiene II Team consists of three periodontics experts, one dental hygiene curriculum expert, two clinical dental hygiene experts, and one oral and maxillofacial radiologist or dental hygienist with formal education in radiology. The Dental Hygiene III Team consists of one dental hygiene curriculum expert, one clinical dental hygiene expert, and two community dental health experts.

Component B Teams

Component B consists of two case-based teams and a review team. The Component B Team, which develops the 150 case-based items for each form of the dental hygiene examination, has one basic science expert, one radiologist, one periodontics expert, one dental hygiene curriculum expert, one clinical dental hygiene expert, one community dental health expert, one dentist with advanced education in oral medicine/oral diagnosis, and one dentist or dental hygienist with advanced clinical training in special needs.

The Case Selection Team consists of four test constructors who review new patient cases and identify suitable cases for examinations. The team also drafts and reviews the patient histories, dental charts, and treatment plans that will be the basis for writing test items by the Component B Team.

The Consultant Review Team consists of four test constructors who review final drafts of the examinations to ensure consistency and coherence of both the Component A and Component B sections of each examination form.

Test Constructor Responsibilities

The following is a list of the responsibilities of every test constructor:

1. Submit new test items for the National Board item banks, according to Joint Commission guidelines, specifications, and content outlines by the designated time. This requirement applies to test constructors after they have completed their first year of service. The number of new items expected may vary according to the needs of the program.
2. Attend each test construction meeting for the duration of the session.
3. Construct National Board Dental Hygiene Examinations according to Joint Commission guidelines, specifications, and content outlines, within the designated time frame.
4. Construct additional items for the item banks as necessary.

5. Assign ownership of all test materials to the American Dental Association and the Joint Commission on National Dental Examinations by agreeing to the terms of the Copyright Assignment.
6. Inform the Joint Commission of changes in the standard curricula, and suggest modifications in test specifications and content outlines.
7. Consider special issues and make recommendations at the request of the Joint Commission.
8. Safeguard the security and confidentiality of National Board Examinations by declining any arrangement to assist with review courses or review books pertaining to the examinations while serving as a test constructor, and for at least one year following the final term of appointment.
9. Comply with the American Dental Association's policy on professional conduct. The policy includes prohibitions against sexual harassment, as well as other forms of unlawful conduct.

An Orientation Manual for Dental Test Constructors provides basic information to new test constructors.

How Dental Hygiene Examinations Are Developed

Many test construction meetings involve a review of individual items and their statistics. These statistics include the difficulty of the item, the proportion of candidates choosing each option, and the item discrimination index (i.e., the point-biserial correlation between the scored item response and the total score). Items that produce statistics below the standards set by the Joint Commission are reviewed. Next, test constructors review the Dental Hygiene Examination Specifications to ensure that discipline areas as represented on the examination continue to reflect the current perspective and practice with respect to the subject matter. Then test constructors finalize the draft examinations by reviewing all items, according to the examination specifications.

13. Test Administration

Although administration may seem like one of the more mundane aspects of an examination program, several important issues are addressed in this section and are linked to testing industry standards. Table 13.1 provides a short list of relevant standards.

Table 13.1
Standards Pertaining to Administration

4.15 The directions for test administration should be presented with sufficient clarity so that it is possible for others to replicate the administration conditions under which the data on reliability, validity, and (where appropriate) norms were obtained. Allowable variations in administration procedures should be clearly described. The process for reviewing requests for additional testing variations should also be documented.

4.16 The instructions 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 item format or major area in the test's classification or domain should be provided to

the test takers prior to the administration of the test, or should be included in the testing material as part of the standard administration instructions.

The Joint Commission on National Dental Examinations in its *Operational and Policy Manual for the JCNDE* describes the eligibility requirements for first time and repeating candidates for the dental hygiene examination. This publication also describes how candidates apply for the examination.

A *Guide* is published annually. The *Guide* provides detailed information regarding the format and content of the examination, candidate eligibility requirements, examination regulations, and the scoring of the examination. The *Guide* also provides examples of item formats and sample items. The *Guide* is available online at the Joint Commission's website (www.ada.org/en/jcnde/examinations/nb-guides). For those individuals interested in obtaining additional and more detailed information regarding the format and content of the dental hygiene examination, a released set of items is also available through the Joint Commission's office.

14. Score Reliability

Score reliability is an important indicator of examination quality. When test developers design an examination, they want to ensure that test scores provide stable and precise measurement of candidates' knowledge, skills, and/or abilities. Despite efforts to eliminate possible sources of measurement error, random factors can affect candidate performance and subsequent examination results. Reliability gauges the degree to which random error affects scores. Low score reliability indicates the strong presence of random sources of measurement error, whereas high score reliability indicates the absence of such sources of error.

The Joint Commission uses Kuder Richardson Formula 20 (i.e., KR20) to report score reliability for the NBDHE. This index provides internal consistency estimates for tests with items that are scored dichotomously (e.g., right or wrong). As shown in Table 14.2, KR20 values range from 0.86 to 0.93 for the 2020 examinations.

Table 14.1 lists the reliability standards that are applicable to the National Board Dental Hygiene Examination. The Standards highlight the importance of reporting the reliability of test-based decisions for high stakes licensing examinations. A strategy that is commonly used to increase reliability is to lengthen examinations. Having uniformly high quality items also contributes to reliability.

Table 14.1
Standards that Apply to Reliability

2.3 For each total score, subscore, or combination of scores that is to be interpreted, estimates of relevant indices of reliability/precision should be reported.

2.14 When possible and appropriate, conditional standard errors of measurement should be reported at several score levels unless there is evidence that the standard error is constant across score levels. Where cut scores are specified for selection or classification, the standard errors of measurement should be reported in the vicinity of each cut score.

11.14 Estimates of the consistency of test-based credentialing decisions should be provided in addition to other sources of reliability evidence.

Table 14.2 provides statistics for all versions of the full-length and short-form NBDHE in 2020.

Table 14.2
Statistics for the Dental Hygiene Examination

	2020*
Total Number of Candidates in the Reference Group	6,797
Standard Score Mean	82.79
Standard Score Standard Deviation	6.35
Mean Raw Score Percentage	72.27
Reliability KR ₂₀ (Range)	.86 - .93

*The 2020 results are based on administrations of the full-length and short-form NBDHE.

Reference Group: The reference group is comprised of all students enrolled in accredited schools who took the examination for the first time. The performance of this group serves as an important reference point for all candidates taking the examination.

Scale Score Mean: The mean score is the average scale score by candidates in the reference group.

Standard Deviation: The standard deviation provides a measure of the spread in scores.

Mean Score (%): The mean score (%) is the average percentage of test items answered correctly by candidates in the reference group.

Reliability (KR20): KR20 is a measure of internal consistency reliability for measures that are scored dichotomously (e.g., right/wrong). Perfect score reliability produces a reliability coefficient of 1.00; however, no set of scores is perfectly reliable in psychological and educational testing. The higher the coefficient, the more reliable are the scores.

15. Standard Setting

A critical step in the development of any pass/fail examination is the setting of the cut score that separates passing and failing candidates (AERA, APA, NCME, 2014, p. 100-101). The NBDHE cut score represents a collective judgment that those who fall below a particular skill level have an unacceptable likelihood of making serious errors in their dental hygiene work. The setting of a cut score may involve empirical study, but value judgments by content experts are ultimately necessary. Judges involved in setting a cut score should be qualified, and documentation of qualifications should be provided. The process for setting the cut score should also be well described and documented. Table 15.1 provides standards that are relevant to setting the cut score for the dental hygiene examination.

Table 15.1
Standards Pertaining to Standard Setting

5.21 When proposed score interpretation involves one or more cut scores, the rationale and procedures used for establishing cut scores should be documented clearly.

5.22 When cut scores defining pass-fail or proficiency levels are based on direct judgments about the adequacy of items or test performances, the judgmental process should be designed so that the participants providing the judgments can bring their knowledge and experience to bear in a reasonable way.

5.23 When feasible and appropriate, cut scores defining categories with distinct substantive interpretations should be informed by sound empirical data concerning the relation of test performance to the relevant criteria.

11.4 When multiple test scores or test scores and non-test information are integrated for the purpose of making a decision, the role played by each should be clearly explicated, and the inference made from each source of information should be supported by validity evidence.

11.16 The level of performance required for passing a credentialing test should depend on the 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.

The Joint Commission periodically conducts standard-setting activities to ensure the appropriateness of the minimum passing score for the National Board Dental Hygiene Examination (NBDHE). These activities have not only set the passing score for future examinations, but have also historically provided support for passing scores that are currently in use. In 2015, the Joint Commission began transitioning to new standards for the NBDHE. The standard setting activities took place in May of 2015 with the updated standard applied to the examination beginning in January 2017.

Standard-Setting procedures for NBDHE

A modified bookmark procedure (Lewis, Mitzel, Mercado, & Schultz, 2012) was used to establish a new passing standard for the NBDHE. The two-day meeting took place on May 11-12, 2015 at the ADA's Chicago offices. The procedures used during the meeting involved the following steps:

1. A standard setting committee was convened. The standard setting committee for the NBDHE was comprised of twelve members: five practitioners, four dentists, and three educators.
2. The committee members received a thorough overview of the purpose and content of the NBDHE exam. This included a description of the test blueprint, test construction methods, scaling, scoring, and of reporting methods. Committee members were also provided with historical information about candidate performance. Finally, committee members completed an abbreviated version of their exam which was representative of a full version with respect to content, difficulty level, and item formats.
3. The committee members engaged in a complete and thorough discussion of the characteristics and behaviors of the "just qualified" (i.e., minimally competent) candidate, and of the importance of individual content elements on the exam.

4. Following the discussion phase, committee members were trained in the Bookmark standard setting method, and were given an opportunity to practice the method using provided practice materials.
5. Committee members reviewed a large set of examination items that had been placed into an Ordered Item Booklet (OIB) assembled as follows:
 - a. Each page of the OIB contained one item.
 - b. Items within the OIB were presented in ascending order of difficulty such that the item on the first page was the least difficult and the item on the last page was the most difficult.
 - c. The items included in the OIB spanned a representative range of difficulty levels.
 - d. After reviewing the OIB, each committee member was asked to independently “bookmark” the page number in the OIB of the last item for which a minimally competent candidate would have at least a two thirds (67%) chance of answering correctly. A cut score for the examination was derived from the median of the committee members’ bookmark placements using the method described by Lewis et al. (2012).
6. After making their judgments, committee members engaged in group discussion regarding their ratings and the rationales for their judgments. During this phase committee members were provided with information about the bookmark placements of the other committee members, and the anticipated impact (resulting percentage failing) of using the cut score associated with the median for the group bookmark placement.
7. Steps 4 and 5 described above were repeated three times. After each replication of the process, committee members were provided an opportunity to ask questions, express concerns, and engage in group discussion. The final recommended cut score for each of the examinations was based on the cut score derived in the third round of the process.
8. At the conclusion of this process, panelists were asked to complete an evaluation questionnaire regarding their impressions of the process. Presented with a five-point rating scale (ranging from 1 - Strongly Disagree to 5 - Strongly Agree) all panelists strongly agreed with the following statement “Overall, I support the final group-recommended cut score as fairly representing the appropriate performance standard for the NBDHE”.
9. The final recommendation was presented to the Joint Commission for review and approval.

The recommended cut scores resulting from 2015 NBDHE standard setting activities were reviewed and approved by the Joint Commission in 2016 and implemented in early 2017.

Table 15.2 provides a summary of failure rates and numbers of examinations administered to candidates from accredited and non-accredited dental hygiene programs during the ten-year period beginning with 2010. As shown, graduates of accredited programs tend to perform substantially better than graduates from non-accredited programs. Candidates who took the examination for the first time tended to perform better than candidates who were repeating the examination. Please note that beginning in 2010, only the computer-based examination was administered. A standard setting activity occurred in September of 2010, and was implemented for the 2011 examinations. The latest standard setting activity occurred in May of 2015 and the revised standard was implemented for the 2017 examinations.

From 2011 to 2012, the total number of examinations administered decreased from 7,705 to 7,646 (a decrease of 0.8%). The number of examinations increased slightly to 8,156 in 2014; slightly decreased to 7,945 in 2015; then steadily increased again to 8,664 in 2019; and slightly decreased to 8,139 in 2020.

Table 15.2
Summary of Failure Rates for the Dental Hygiene Examination

Year	Accredited				Non-Accredited				Total	
	First-time		Repeating		First-time		Repeating		First-time and Repeating	
	Number	% Failing	Number	% Failing	Number	% Failing	Number	% Failing	Number	% Failing
2011*	6,968	5.2	492	46.5	194	23.7	51	60.8	7,705	8.7
2012	6,882	4.2	486	47.1	236	26.7	42	50.0	7,646	7.9
2013	7,014	4.8	487	45.6	231	17.8	54	55.6	7,786	8.1
2014	7,357	4.8	527	47.4	204	23.0	68	63.2	8,156	8.5
2015	7,227	4.4	499	46.3	179	22.9	40	55.0	7,945	7.7
2016	7,397	5.1	506	41.7	214	27.6	45	35.6	8,162	8.1
2017*	7,262	6.2	677	49.8	253	33.2	81	46.9	8,273	11.0
2018	7,360	5.8	654	46.2	328	34.8	88	44.3	8,430	10.4
2019	7,316	7.9	852	49.1	377	35.5	119	51.3	8,664	13.8
2020	6,938	9.7	764	51.3	302	44.0	135	60.7	8,139	15.7

* A new standard was introduced this year, based on updated standard setting activities.

Classification Accuracy and Classification Consistency

When scores on an examination are used as a basis for making pass/fail decisions, it is critical to ensure that the pass/fail point on the examination's scale is reliable and valid (AERA, APA, NCME, 2014, p. 46-47). Two methods are typically adopted by testing programs to evaluate the reliability of the pass/fail point on an examination's scale. The first method is to examine outcomes from standard setting activities (Cizek, Bunch, and Koons, 2004). The second method is to compute the probabilities of correct and consistent classifications of candidate performance on an examination (Livingston and Wingsky, 1979; Hanson and Brennan, 1990; Livingston and Lewis, 1995).

With regard to evaluating the reliability of the pass/fail point on the National Board Dental Hygiene Examination measurement scale through the use of the first method, the following statistics support the conclusion that the passing point is reliable: (1) the error of measurement is lowest at the pass/fail point on the measurement scale, (2) the spread of scores covers the entire scale, (3) failure rates are reasonably consistent with the judgments of standard-setting committee members, and (4) trends in failure rates are reasonably stable across years.

With regard to evaluating the pass/fail point on the examination' measurement scale through the use of the second method, procedures developed by Hanson and Brennan (1990) were used to analyze data and provide results. The results are presented in terms of the following: (1) classification accuracy– the probability of correct classifications, the false positive rate, and the false negative rate, and (2) classification consistency- the probabilities of consistent classification and misclassification. The accuracy of decisions is the extent to which decisions would agree with those that would theoretically be made if candidates could be tested with all possible editions of the examination. The consistency of decisions is the extent to which decisions would agree with the decisions that would have been made if candidates had taken parallel editions of the examination that were equal in difficulty and covered the same content domain as the edition they actually took. These concepts are presented schematically in Tables 15.3 and 15.4.

**Table 15.3
Classification Accuracy**

		Decision made on examination form actually taken (Observed Score)	
		Pass	Fail
True status based on average score obtained from all possible examination forms (True score)	Pass	Correct Classification	False Negative
	Fail	False Positive	Correct Classification

In Table 15.3, an accurate classification is regarded as occurring when the theoretical decision made on the basis of the average score obtained across all possible examination forms (i.e., the true score based decision) agrees with the decision made on the basis of the examination form actually taken (i.e., the observed score based decision). False positives and false negatives occur when there is a mismatch between candidate true score-based decisions and observed score based decisions. The false positive value is the proportion of candidates misclassified as achieving “Pass” with respect to their observed score when they actually would not have achieved “Pass” with respect to their true score. The false negative value is the proportion of candidates misclassified as “Fail” with respect to their observed score when they actually achieved “Pass” with respect to their true score.

**Table 15.4
Classification Consistency**

		Decision would be made on a parallel form taken	
		Pass	Fail
Decision made on the actual examination form taken	Pass	Consistent Classification	Misclassification
	Fail	Misclassification	Consistent Classification

In Table 15.4, consistent classifications occur when two forms of an examination agree on the classification of a candidate as either “Pass” or “Fail”, whereas misclassifications occur when the decisions made by the two forms of the examination differ.

Evaluation: Data used in the classification analyses were based on the item responses from 5,567 candidates who were enrolled in accredited dental hygiene programs and who took the NBDHE for the first time. Results with respect to the classification accuracy and consistency of the current pass/fail point on the National Board Dental Hygiene Examination are presented in Table 15.5. The table also includes false positive and false negative rates. The sum of the correct classifications, false positives, and false negatives is equal to 1. This is also true for values associated with consistent classifications and misclassifications. As shown, the probability of correct classifications is higher than that of consistent classifications for the National Board Dental Hygiene Examination. This finding is consistent with the findings from Young and Yoon (1998) and Yang and Waldschmidt (2013), in which both studies found that the probabilities of classification consistency were less than those of classification accuracy after

comparing results across several examination programs. Table 15.5 also shows that the reliability of the pass/fail points on the examination measurement scales were satisfactory on the study sample, with 96% for classification accuracy and 93% for classification consistency.

Results from outcome assessment of the standard-setting studies and analyses of decision accuracy and consistency of the pass/fail point affirmed that the pass/fail point on the National Board Dental Hygiene Examination measurement scale is highly reliable and valid. To further affirm that candidate performance on the National Board Dental Hygiene Examinations is accurately represented, the Joint Commission confirms the pass/fail point on the examinations periodically through standard-setting activities. In addition, the Joint Commission conducts monthly quality control procedures to monitor examination failure rates. Examination responses from candidates are audited for accuracy before score reports are distributed. The Joint Commission also conducts audits for candidates who express questions or concerns regarding their scores.

Table 15.5
Classification Accuracy and Consistency of the Pass/Fail Point on the National Board Dental Hygiene Examination's Measurement Scale

Examination	Dental Hygiene	
Sample Size	5,567	
Classification Accuracy	Correct Classification	0.96
	False Positive	0.00
	False Negative	0.04
Classification Consistency	Consistent Classification	0.93
	Misclassification	0.07

16. Scaling/Equating/Comparability of Test Forms

The Standards (AERA, APA, & NCME, 2014) devote chapter five to discussions on the comparability of test forms. When different forms of the same examination are used, the psychometric equivalence of these forms is of vital importance. Table 16.1 lists the relevant standards that apply to scaling/equating/comparability.

Table 16.1
Standards Pertaining to Scaling/Equating/Comparability

5.12 A clear rationale and supporting evidence should be provided for any claim that scale scores earned on alternate forms of a test may be used interchangeably.

5.13 When claims of form-to-form score equivalence are based on equating procedures, detailed technical information should be provided on the method by which equating functions were established and on the accuracy of the equating functions.

5.14 In equating studies that rely on the statistical equivalence of examinee groups receiving different forms, methods of establishing such equivalence should be described in detail.

5.15 *In equating studies that employ an anchor test design, the characteristics of the anchor test and its similarity to the forms being equated should be presented, including both content specifications and empirically determined relationships among test scores. If anchor items are used in the equating study, the representativeness and psychometric characteristics of the anchor items should be presented.*

5.19 *When tests are created by taking a subset of the items in an existing test or by rearranging items, evidence should be provided that there are no distortions of scale scores, cut scores, or norms for the different versions or for score linkages between them.*

5.20 *If test specifications are changed from one version of a test to a subsequent version, such changes should be identified, and an indication should be given that converted scores for the two versions may not be strictly equivalent, even when statistical procedures have been used to link scores from the different versions. When substantial changes in test specifications occur, scores should be reported on a new scale, or a clear statement should be provided to alert users that the scores are not directly comparable with those on earlier versions of the test.*

The dental hygiene examination is a criterion-referenced examination and not a norm referenced one. In other words, scores and candidates' pass/fail status are determined by specific criteria, not by the process sometimes known as "grading on a curve." Using an objective measurement methodology in concert with the judgment of experts in biomedical sciences, clinical dental hygiene sciences, and community dental health, a minimum passing score is set on a single, base year edition/form of the examination. However, different forms of the dental hygiene examination are available for administration. In order to ensure that the scores of candidates completing different examination forms can be directly and meaningfully compared, some statistical adjustments are necessary. Raw scores can permit meaningful comparisons of examinees who have completed the same examination form; however, comparing raw scores obtained under different examination forms can be inappropriate unless certain statistical assumptions are met. Because raw score distributions can vary across forms, raw scores must be transformed in order to permit meaningful comparison of candidates across forms. The process of statistically adjusting scores to enable comparisons across forms is known as *test equating*.

Once scores are equated, they are on a common measurement scale. Thus, the scores of candidates completing different forms can be evaluated on the same scale using the same cut score of 75. In addition, because the mean scores obtained by different groups of candidates may be expressed on the same metric, yearly trends in test performance can be evaluated fairly, within standard setting cycles.

To equate two examination forms, certain requirements need to be met (Lord, 1980). First, both examinations must assess the same content. Second, the equation used to adjust scores should remain the same regardless of the groups used. And third, the correspondence between the scores must be symmetric—that is, it should make no difference whether examination X is adjusted to the scale of examination Y, or vice-versa. The equating procedures presented here fall within the context of *horizontal* score transformations. That is, the alternative forms of the examination are of similar difficulty, have identical content specifications, and have been constructed for the same population of candidates.

Equating Designs

Many different data collection designs have been used for equating (Petersen, Kolen, and Hoover, 1989). All require that either the same group (or equivalent groups) of candidates complete both forms of the test, or that a group of common items, called *anchor* items, appear on both forms of the test.

In the simplest of these designs, the same group of candidates completes both examinations. Because only one group is used, possible between-group differences in ability cannot influence the equating, as might occur when multi-group designs are used. However, the single group would be subject to fatigue, practice, and order effects. This equating design is not feasible due to the length of the dental hygiene examination.

Random differences between equivalent groups may be controlled by the use of anchor items. Anchor items are administered to both groups in the design and may or may not be counted in computing total scores. Performance on the anchor items can be used to make statistical adjustments to each of the examination forms so that an estimate can be made of how the combined group of candidates would score on both forms of the examination. Because the anchor items serve as the link among the alternate forms, the format and content of the anchor items should be similar to those of the other items administered. Not only is this design feasible, it is widely used and accepted throughout large-scale testing.

Statistical Methods for Adjusting Scores

Once an equating design has been chosen, the next decision is to select an appropriate statistical method for producing equivalent scores on the parallel forms. The three most commonly used techniques are linear equating, equipercentile equating, and item response theory (IRT). Equivalence of scores is defined differently in each method, and each makes different assumptions about the data and the distributions of scores.

The IRT method has many advantages that warrant its use. First, IRT approaches to equating are rooted at the item level rather than the total score level. Traditional methods, such as equipercentile equating, require entire total score distributions to be used. The use of cumulative distributions of scores introduces imprecision into the equating process. Rounding and interpolation errors may occur. The IRT model currently used to score and equate the National Board examinations is the one-parameter or Rasch model. This model is very precise. Second, the Rasch model allows each candidate to complete a set of items different from that attempted by any other candidate, and still be scored on the same measurement scale. This process, now used in adaptive testing, has the potential to improve measurement accuracy for most candidates, but it requires that IRT methods of equating be implemented. Third, the Rasch model allows for extensive cross-checking of item parameters. Because each equating event may introduce error into the estimation of item and person parameters, it is essential to double check item parameters by linking them through various paths back to the scale of the base year. This precaution prevents item difficulties from drifting too far away from the correct scale, but is cumbersome to do with any method other than Rasch equating. The versatility and precision associated with using the Rasch model to equate enables the item bank to be managed more easily and updated more accurately.

IRT postulates that the response of an individual to an item is a function of that person's ability and certain characteristics, or parameters, of the item. From the perspective of the Rasch model, the only characteristic of the item which can influence a response is its difficulty. The function used to determine the probability that person v will respond correctly to item i is shown below (Wright & Stone, 1979):

$$P\{x_{vi}=1 | \beta_v, \delta_i\} = \frac{\exp(\beta_v - \delta_i)}{1 + \exp(\beta_v - \delta_i)} \quad [16.1]$$

where β_v is the ability of person v , and δ_i is the difficulty of item i (Wright and Stone, 1979). Both item difficulty and person ability are expressed in the same unit of measurement, called the logit. A logit may be defined as the natural log odds of a correct response to an item chosen to represent the center (or "zero" point) of the measurement scale.

The Rasch model assumes that all the items in a test measure the same construct, and that the logistic curve defined by Equation 1 is a satisfactory representation of the data. Items that do not fit the model can be detected statistically and discarded. An important reason for using the Rasch model is that it is regarded as providing objective measurement. This means that the estimate of a person's ability does not depend on the items attempted, and that the estimate of an item's difficulty does not depend on the particular sample of individuals used in the item's calibration. Thus, when a set of items is administered to two samples, and calibrated separately for each group, the two resulting sets of Rasch item difficulties will be linearly related. Therefore, a set of common items (i.e., anchor items) present in each of two different test forms administered to two different samples, may serve a linking function. Determining the linear relationship between the linking items on the different forms yields a constant that, if added to the difficulties of the anchor items as calibrated in examination Y, will transform them to the scale of examination X. The same constant, added to the difficulties of the remaining items of examination Y, also places these remaining items on the examination X scale of measurement because the same linear relationship applies to all the items, even those present on only one of the test forms.

The necessary constant used to transform the item difficulty parameters of examination Y onto the scale of examination X is given by Wright and Stone (1979):

$$G_{XY} = \frac{\sum^K (\delta_{ix} - \delta_{iy})}{K} \quad [16.2]$$

where δ_{ix} is the difficulty of item i when calibrated with the items on Examination X; δ_{iy} is its difficulty on the Examination Y scale; and K is the number of anchor items.

After two examinations have been linked in this manner, the same procedure may be repeated to link one of the examinations with yet another examination using a (possibly) new set of linking items. In this way, many alternate versions of an examination may be equated, enabling examination performance to be evaluated and meaningfully compared over periods of several years. Large inventories of items (item banks) may also be built up systematically over time using the chaining process. A certain degree of error, however, accompanies each linking step, so it is advisable to crosscheck item difficulty parameters periodically to insure that the equating process remains accurate.

Person ability estimates, β_v also expressed on the logit scale, may be transformed by the same constant used to place items on a common scale. Equating the ability scales makes possible the comparison of group differences even though alternate examination forms may have been used for each administration.

The dental hygiene examination is scaled according to the Rasch model using the unconditional maximum likelihood estimation procedure (Wright & Panchapakesan, 1969) employed in the

WINSTEPS computer program (Linacre, 2002). Output includes person and item parameters scored in logits and indices of how well the responses of each person and item fit the model. Included among the items is a set of linking or anchor items. As discussed above, links enable each item and each candidate to be located on the same scale of measurement as that of the base year of the examination.

The following simple example illustrates how common (anchor) item equating is carried out. Table 16.2 presents item statistics for seven anchor items appearing on two separate administration forms. The first column shows item difficulties scaled on the base year logit scale. Standard errors show how accurately item difficulty has been estimated. The corresponding statistics for the new test form are shown in the next two columns. The linking constant is simply the difference between the mean item difficulties under the two calibrations. In this example, the linking constant is -0.36. Ideally, when the linking constant is added to the item difficulty of the new form, the sum should equal the corresponding base year difficulty for each item. However, error due to sampling and imperfect measurement usually results in a discrepancy between these two values. If the difference is too large for a given item, it should not be included in the equating process. Wright and Stone (1979) provide a statistical chi-square test that allows one to determine how large a difference in difficulties may be expected by chance.

Table 16.2
Difficulties of Anchor Items Calibrated on Two Test Administrations

<i>Item</i>	<i>Base Year</i>		<i>New Testing</i>		<i>New Testing</i>	<i>Squared</i>
	<i>Diff.</i>	<i>S.E.</i>	<i>Diff.</i>	<i>S.E.</i>	<i>Diff. + Link</i>	<i>Error</i>
1	-0.88	.05	-0.72	.04	-1.08	.0400
2	-0.74	.05	-0.42	.04	-0.78	.0016
3	-0.62	.05	-0.28	.04	-0.64	.0004
4	-0.15	.04	0.02	.04	-0.34	.0361
5	0.26	.04	0.61	.04	0.25	.0001
6	-0.18	.04	0.05	.04	-0.31	.0169
7	-1.03	.05	-0.08	.04	-0.44	.3481**
Sum	-3.34		-0.82		-3.34	.4432**
Mean	-0.48		-0.12		-0.48	

** $p < .01$

In the example, Item 7 produced a difference in difficulties greater than would be expected by chance alone. As a result, the overall fit of the equating was not acceptable.

When an unsuitable item is detected, the equating process must begin again with the offending item removed. This requires that the mean item difficulties be recalculated for the remaining items, a new linking constant determined, and the discrepancies between the old and new calibrations recalculated. In this case, the new linking constant was equal to -0.27. Once a satisfactory equivalence between the base year and current year anchor items has been established, the next step is to adjust the difficulties of all the remaining items in the new examination by adding the linking constant to them. This adjustment places all the items on the original base year scale, even though none of the non-anchor items was administered in the base year. Since all item parameters are now grounded in the same scale of measurement originally used in the base year, estimates of person ability (determined from Equation 1 using the WINSTEPS Rasch scaling program) will be on that scale. Assuming examinations share common content specifications, this enables any person's logit to be meaningfully compared to

that of any other person, regardless of the year in which they completed the examination, and regardless of which particular items were included on that examination. Mean logits may also be directly compared from examination to examination. Scale scores may also be directly compared, as long as the scores fall within the same standard setting cycle (recall that the scale score measurement scale is adjusted with each standard setting event, such that the passing score point is assigned a value of 75).

17. Scoring and Reporting Test Results

Standards pertaining to scoring and reporting of results appear in Table 17.1 below. Quality control in scoring is an important, yet often invisible, feature of any examination program. Standards 6.8 and 6.9 refer to scoring and potential scoring errors. Standard 6.10 refers generally to making responsible interpretation of scores available to recipients of these scores. Standard 6.16 makes certain that the transmission of scores is done responsibly. Standard 6.14 and 6.15 refers to record keeping.

Table 17.1
Standards Pertaining to Scoring and Reporting of Test Scores

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. When scoring of complex responses is done by computer, the accuracy of the algorithm and processes should be documented.

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 source of scoring errors should be documented and corrected.

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 scores represent, the precision/reliability of the scores, and how scores are intended to be used.

6.14 Organizations that maintain individually identifiable test score information should develop a clear set of policy guidelines on the duration of retention of an individual's records and on the availability and use over time of such data for research or other purposes. The policy should be documented and available to the test taker. These users should maintain appropriate data security, which should include administrative, technical, and physical protections.

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.

Scoring of the National Board Dental Hygiene Examination

Procedures for scoring the National Board Dental Hygiene Examination are presented in the Joint Commission's Examination Regulations. Quality control procedures are in place to facilitate accurate scoring. The raw and scale scores for each candidate are determined by comparing the candidate's responses to the examination's answer key, computing a raw score, and converting the raw score to a scale score. Each week the roster of candidates scheduled to complete board examinations is compared with the candidates appearing in result files, to ensure no result files are missing.

Candidate Scores and Score Reports

Two factors that affect a candidate's score are (1) the number of correct answers selected by the candidate and (2) the conversion scale for the examination form. For the dental hygiene examination, there is no penalty for selecting an incorrect response. The scale score of 75 is used by the Joint Commission to reflect the minimum passing score on the examination. A score below 75 is considered a failing score. Scale scores range between 49 and 99.

As noted previously, beginning with administrations occurring in 2012, Joint Commission policies dictate that examinees who pass examinations will simply receive notification of their passing status; examinees who fail examinations will receive score information for remediation purposes. Score reporting for examinations occurring prior to 2012 remains unchanged (i.e., scale scores will continue to be reported for these administrations). In 2019, the Joint Commission approved an updated score report template for candidates who fail the NBDHE. The information for each discipline in the new score report is presented as a graphic instead of a table of numeric values. The discipline subscores are represented graphically and placed on a common measurement scale so that scores from different disciplines can be meaningfully visually compared. This allows failing candidates to assess their relative performance in the different disciplines and identifies disciplines where they are most in need of remediation. It also allows candidates to compare subscores across examination attempts. The new failing candidate report was implemented in the fourth quarter of 2019.

18. Rights and Responsibilities of Test Takers

Chapter 8 of the Standards (AERA, APA, NCME, 2014) addresses the issue of fairness and the interests of candidates who complete the dental hygiene examination. Because so much is at stake in taking these examinations, the Joint Commission strives to ensure that candidates for licensure receive the utmost in fair treatment in the preparation, administration, and scoring of examinations. Table 18.1 below provides four relevant standards. Standards 8.1 and 8.2 require that examination information be made available to all candidates. The National Board Dental Hygiene Examination Guide, published by the Joint Commission annually, is the most suitable way of accomplishing this. Standard 8.9 refers to cheating, and standard 8.13 refers to challenges and other conflicts in examination scoring.

Table 18.1
Standards Addressing Rights and Responsibilities of Test Takers

8.1 Information about test content and purposes that is available to any test taker prior to testing should be available to all test takers. Shared information should be free of charge and in accessible formats.

8.2 Test takers should be provided in advance with as much information about the test, the testing process, the intended test use, test scoring criteria, testing policy, availability of accommodations, and confidentiality protection as is consistent with obtaining valid responses and making appropriate interpretations of test scores.

8.9 Test takers should be made aware that having someone else take the test for them, disclosing confidential test material, or engaging in any other form of cheating is unacceptable and that such behavior may result in sanctions.

8.12 In educational and credentialing testing programs, a test taker is entitled to fair treatment and a reasonable resolution process, appropriate to the particular circumstances, regarding charges associated with testing irregularities, or challenges issued by the test taker regarding accuracies of the scoring or scoring key. Test takers are entitled to be informed of any available means of recourse.

NBDHE Guide

The single most effective means for satisfying the standards appearing in Table 18.1 is the publication of a program guide. The Joint Commission publishes a National Board Dental Hygiene Examination Guide on an annual basis. The Guide is available online at ada.org/NBDHE. The Guide contains detailed information related to the format and content of the examination, eligibility requirements, examination regulations, the appeal process, and the scoring of the examination. There are also sections in the Guide that provide examples of item formats and sample items.

19. Threats to Validity

According to Messick (1989), two major threats to validity are construct-irrelevant variance (CIV) and construct under representation (CUR). This part of the technical report discusses validity evidence bearing on these two major threats.

Construct-Irrelevant Variance (CIV)

This threat to validity involves systematic error in scores that is contributed by extraneous sources. Haladyna (2002) identifies many such sources of CIV, including non-equivalent examination forms, cheating, improper examination preparation, errors in scoring results, and faulty items. Previous sections of this report address some of these issues.

The Joint Commission periodically releases editions of National Board examinations or collections of items in order to familiarize candidates with National Board item formats. However, the Joint Commission recommends that candidates use textbooks and lecture notes as their primary sources of study material. Released examinations are available at most dental hygiene programs and the library of the American Dental Association. The Joint Commission discourages superficial learning as a basis for examination preparation.

The Joint Commission on National Dental Examinations does not discriminate based on race, color, religion, gender, age, sex, national origin, disability, sexual orientation, or marital status. CIV can become present when examination items reflect such discrimination.

Construct Under-Representation

Another threat to validity is construct under-representation. When an examination does not completely represent the domain of knowledge intended in the initial creation of the content, this bias leads to inadequate construct coverage. This can cast doubt on the meaning of examination scores and the legitimacy of their use in making pass/fail decisions.

The procedures used to define the domain of knowledge to be tested and to determine the examination specifications go quite far in assuring the public and the dental community that the Dental Hygiene examination does not under-represent areas of professional knowledge that are deemed essential for entry-level dental hygienists.

20. Validity Studies

Studies are undertaken to investigate significant issues that threaten validity. Studies are also undertaken to provide new sources of validity evidence that strengthen arguments in favor of using examination results to inform licensure decisions.

Validity studies can vary significantly in type and scope. The practice analysis is one type of validity study, conducted periodically to update the examination specifications and ensure that examination content is current. The most recent practice analysis for the dental hygiene examination was conducted in 2016. The findings of the practice analysis along with the corresponding revised examination specifications were reviewed and approved by the Joint Commission at its meeting in 2017. The revised examination specifications were implemented in 2019.

Standard-setting studies are essential when a new set of examination specifications is used that has potential implications for the cut score. Kramer and DeMarais (1992) provide a good example of this kind of study. The most recent standard setting activities were conducted in May of 2015. The purpose of the activities was to establish an appropriate standard or cut score based on dental hygiene cognitive skills requirements for safe practice.

Reliability studies are considered a primary type of evidence that is relevant to validity. Such studies are conducted on a routine basis.

Studies of the content and content structure of the examination are also important. For example, since the Rasch model used by the NBDHE assumes the unidimensionality of item responses, evidence should be obtained that is consistent with this requirement.

21. Security

Breakdowns in examination security can threaten an examination's validity. Table 21.1 provides a list of standards that pertain to security. The Joint Commission has policies and procedures in place to provide for examination security.

Table 21.1
Standards Pertaining to Security

6.7 Test users have the responsibility of protecting the security of test materials at all times.

8.6 Test data maintained or transmitted in data files, including all personally identifiable information (not just results), should be adequately protected from improper access, use, or disclosure, including by reasonable physical, technical, and administrative protections as appropriate to the particular data set and its risks, and in compliance with applicable legal requirements. Use of facsimile transmission, computer networks, data banks, or other electronic data-processing or transmittal systems should be restricted to situations in which confidentiality can be reasonably assured. Users should develop and/or follow policies, consistent with any legal requirements, for whether and how test takers may review and correct personal information.

10.18 Professionals and others who have access to test materials and test results should maintain the confidentiality of the test results and testing materials consistent with scientific, professional, legal, and ethical requirements. Tests (including obsolete versions) should not be made available to the public or resold to unqualified test users.

General Principles: Effective examination security procedures are critical to the success of any examination program. Responsibilities for examination security are clearly defined for examination developers, administrators, and users. Examination security is maintained throughout the development and administration processes in a variety of ways. Policies of the Department of Testing Services address issues related to examination security and are reviewed periodically by the Joint Commission and its staff.

Security Audit: In 2008, Caveon Test Security, an independent, external organization, conducted a security audit of the Department of Testing Services, which is the department within the American Dental Association that conducts examination programs for the Joint Commission. The audit was conducted to identify potential security risks, propose specific measures to ameliorate or diminish any potential risks, and provide recommendations to support security planning. The findings of the audit supported the department's overall security measures.

Identification of Secure Materials: The Joint Commission has identified the following materials as secure:

1. individual items and case materials (e.g. radiographs, clinical photographs, and dental charts in development, in camera-ready copy, and in electronic files for transmission to administration sites);
2. scoring materials (e.g., item analyses, answer keys, and statistical analyses);
3. computer scoring software;
4. standard setting materials and meeting notes;
5. item banks; and
6. candidate personal information.

Departmental Procedures:

- Policies and legal issues: All items and examinations are copyrighted to establish ownership and restrict their use or dissemination through unauthorized means. Policies and procedures for handling secure materials require continuous secure custody of materials and a chain of evidence attesting to the status and location of secure materials.
- Personnel: The team that maintains the security of examination materials includes Joint Commission staff, vendors, and volunteers.

- Personnel who handle examination materials must be screened at the time of hire or selection to disqualify individuals who could represent an unacceptable risk.
- All staff members are trained in procedures for handling secure materials and are required to comply with policies on confidentiality and conflict of interest.
- Staff: The examination development staff maintain security on examination materials during the development process.
- Vendors: All vendors are responsible for maintaining security of examination materials. Joint Commission staff reviews vendors' operations to ensure compliance with security policy. All service agreements with vendors require adherence to the Joint Commission's security procedures.
- Volunteers: Volunteers who assist in the development of items and editions of the examination must complete agreements regarding confidentiality, copyright assignment, and conflict of interest. Volunteers are prohibited from releasing information regarding examination content.
- Facilities and storage of examination materials: access to the offices of the Joint Commission is restricted. Security of materials stored or transmitted in electronic format includes technology for password protection, encryption, firewalls, etc.
- No factor is more critical to effective examination administration and security than an adequate test administration facility.

Security of Examination Materials in Electronic Format: Departmental and vendor computers are protected with firewalls, login identifications, passwords and other forms of security. Access to electronic files is limited to authorized individuals.

Testing Procedures: The examination is administered by Pearson VUE at its nationwide, professional level testing centers, unless additional test facilities are authorized by the JCNDE. The NBDHE Guide describes procedures for identification of candidates, including requirements for multiple forms of positive identification. Conduct of candidates is closely monitored during the testing appointment. Examination regulations and testing center policies are designed to deter cheating and breaches of security.

Policies and Procedures for Dealing with Breaches in Security: The Joint Commission provides specific procedures for observing and reporting breaches in security and communicates them to test administrators. The Joint Commission promptly investigates reports of security breaches and ensures examination items are removed from use when it determines security has been breached. When the source of a security breach is identified, the Joint Commission takes legal action or imposes appropriate sanctions.

22. Guidelines for High-Stakes Testing

The American Educational Research Association is the largest organization in the world devoted to the scientific study of education. In 2000, it issued a brief publication that listed guidelines for designing and using high-stakes examinations. These guidelines are intended for educational examinations given in high-stakes settings, such as for high-school graduation. However, some of the guidelines are also appropriate for the dental hygiene examination. This section presents a selected set of these guidelines and provides a brief discussion of each guideline for the dental hygiene examination.

Protection against High-Stakes Decisions Based on a Single Examination

Can a single examination prevent a candidate from practicing as a dental hygienist after other criteria for licensure have been met? The dental hygiene examination program provides repeated opportunities for candidates to prepare for and pass the dental hygiene examination. The decision to license a dental hygienist is based on meeting many criteria. Since public welfare and safety are at issue, state boards bear a heavy responsibility for using examination information alongside other information for making licensing decisions.

Adequate Resources and Opportunity to Learn

The Joint Commission has no responsibility for the educational preparation of dental hygienists. This task falls to schools. Failure to provide adequate opportunities to learn professional knowledge can lead to a candidate failing the dental hygiene examination. The Joint Commission publishes on its website a list of reference texts and resources for the examination.

Validation for Each Separate Intended Use

For each use of examination results, validity evidence is collected. The Joint Commission adheres to this guideline, as this technical report has demonstrated.

Full Disclosure of Likely Negative Consequences of High-Stakes Testing Programs

Where credible scientific evidence suggests that a given type of examination program is likely to have negative side effects, examinations developers and users should make a serious effort to explain these possible effects to policy makers.

The above guideline does not appear relevant to National Board examination programs.

Alignment between the Examination and the Curriculum

It is the responsibility of dental hygiene educational programs to align student learning with the knowledge, skills, and abilities that national practice analyses have determined represent the core knowledge required of practicing dental hygienists. NBDHE content is aligned with the core knowledge of practicing dental hygienists that serves as the source for curriculum development.

Validity of Passing Scores and Achievement Levels

The Joint Commission has determined its passing scores using methodology that is consistent with the Standards for Educational and Psychological Testing (2014).

Opportunities for Meaningful Remediation for Examinees Who Fail High-Stakes Examinations

The Joint Commission bears no responsibility for remediation, but schools may choose to provide remediation if a candidate fails. The Joint Commission provides a list of reference materials that may be useful to candidates but does not endorse any specific review courses.

Appropriate Attention to Candidates with Disabilities

In testing individuals with disabilities based on the Americans with Disabilities Act, steps should be taken to ensure that examination results accurately reflect standing on the intended construct rather than any disabilities and their associated characteristics that are extraneous to the intent

of the measurement. The Joint Commission complies with federal regulations bearing on examination administration involving candidates with disabilities. Joint Commission reports do not identify candidates who may have received testing accommodations for an examination.

Sufficient Reliability for Each Intended Use

Reliability refers to the stability or precision of examination scores. It must be shown that scores reported for individuals or for schools are sufficiently reliable to support each intended interpretation. Reliability should be examined for the results actually used. This technical report provides solid evidence regarding the adequacy of reliability estimates.

References

- American Dental Association. (1998). *Accreditation Standards for Dental Hygiene Programs*. Chicago, IL: Author.
- American Dental Education Association. (2000). *Competencies for Entry into the Profession of Dental Hygiene* [Online]. Available: <http://www.adea.org/News/competencies.htm>
- American Educational Research Association. (2000). *Guidelines for High Stakes Testing*. Washington, DC: Author.
- American Educational Research Association, American Psychological Association. National Council on Measurement in Education. (2014). *Standards for Educational and Psychological Testing*. Washington, DC: Author.
- Cizek, G.J., Bunch, M.B., & Koons, H. (2004). Setting performance standards: contemporary methods. *Educational measurement: issues and practice*, 25(4), 31-50.
- Downing, S. M., & Haladyna, T. M. (1997). Test item development: Validity evidence from quality assurance procedures. *Applied Measurement in Education*, 10, 61-82.
- Haladyna, T. M. (2002). Supporting documentation: Assuring more valid test score interpretation and uses. In J. Tindal & T. M. Haladyna (Eds). *Large-Scale Assessment Programs for All Students: Validity, technical adequacy, and implementation* (pp. 89108). Mahwah, NJ: Lawrence Erlbaum Associates.
- Hanson, B. A., & Brennan, R. L. (1990). An investigation of classification consistency indexes estimated under alternative strong true score models. *Journal of Educational Measurement*, 27, 345–359.
- Joint Commission on National Dental Examinations. (November 1995). *Guidelines for Committee Revision of Weak or Unacceptable Items for National Board Dental Examinations*. Chicago, IL: American Dental Association.
- Joint Commission on National Dental Examinations. (2020). *National Board Dental Hygiene Examination Guide*. Chicago, IL: American Dental Association.
- Joint Commission on National Dental Examinations. (2020). *Orientation Manual for Dental Test Constructors*. Chicago, IL: American Dental Association.
- Joint Commission on National Dental Examinations. (2020). *Test Item Development Guide: National Board Dental Examinations*. Chicago, IL: American Dental Association.
- Joint Commission on National Dental Examinations. (2020). *JCNDE Test Construction Teams and Selection Criteria*. Chicago, IL: American Dental Association.
- Joint Commission on National Dental Examinations. (2020). *Procedures for test administration*. Chicago, IL: American Dental Association.

- Joint Commission on National Dental Examinations. (2020). *Rules of the JCNDE*. Chicago, IL: American Dent Association.
- Joint Commission on National Dental Examinations. (2020). *Operational and Policy Manual of the JCNDE*. Chicago, IL: American Dental Association.
- Kane, M.T., Kingsbury C., Colton, D., Estes, C. (1989). Combining data on criticality and frequency in developing test plans for licensure and certification examination. *Journal of Educational Measurement* 26, 17-27.
- Kramer, G. A., & DeMarais, D. R. (1997). Construct validity of the National Board Dental Hygiene Examination. *Journal of Dental Education*, 61, 709-716.
- Lewis, D. M., Mitzel, H. C., Mercado, R., Schulz, E. M. (2012) The Bookmark standard setting procedure. In G. J. Cizek (Ed.) *Setting performance standards: Foundations, methods, and innovations* (pp. 225-254). New York: Taylor and Francis.
- Linacre, J. M. (2002). *WINSTEPS: Rasch-model computer programs* [computer software]. Chicago, IL. MESA Press.
- Livingston, S. A., & Lewis, C. (1995). Estimating the consistency and accuracy of classifications based on test scores. *Journal of Educational Measurement*, 32, 179–197.
- Livingston, S. A., & Wingersky, M. S. (1979). Assessing the reliability of tests used to make pass/fail decisions. *Journal of Educational Measurement*, 16, 247-260.
- Lord, F. M. (1980). *Applications of item response theory to practical testing problems*. Hillsdale, NJ: Lawrence Erlbaum Associates.
- Messick, S. (1989). Validity. In R. L. Linn (Ed.), *Educational measurement* (3rd ed., pp. 13-104). New York: American Council on Education and Macmillan.
- Petersen, N.S., Kolen, M. J., & Hoover, H. D. (1989). Scaling, norming, and equating. In R. L. Linn (Ed.), *Educational Measurement* (3rd ed.: pp 221-262). New York: American Council on Education-Macmillan Publishing Company.
- Tsai, T-H. (2006, March). Reliability of classification decision at proficient cut on New York State English as a Second Language Achievement Test. *Presentation delivered at the meeting of New York State Education Department, Albany, NY.*
- Wright, B. D., & Panchapakesan, N. (1969). A procedure for sample-free item analysis. *Educational and Psychological Measurement*, 29, 23-48.
- Wright, B. E. & Stone, M. (1979). *Best Test Design*. Chicago: MESA Press.
- Yang, C.-L., & Waldschmidt, D.M. (2013, June). Comparison of classification accuracy involving pass/fail decisions using three dichotomous item response theory models. (*INBDE Research Report No. 8*). Joint Commission on National Dental Examinations, American Dental Association. Chicago, IL.

Young, M.J., & Yoon, B. (1998, April). Estimating the consistency and accuracy of classifications in a standard-referenced assessment. (*CSE Technical Report 475*). Center for the Study of Evaluation, Standards, and Student Testing, Los Angeles, CA: University of California at Los Angeles.

Appendix A: Dental Hygiene Examination General Specifications (2020)

The National Board Dental Hygiene Examination is a comprehensive examination comprised of 350 multiple choice test items. Although performance on the examination is reported by means of one official score, the examination has two components. The discipline-based Component A includes 200 items addressing three major areas:

- I. Scientific Basis for Dental Hygiene Practice
- II. Provision of Clinical Dental Hygiene Services
- III. Community Health / Research Principles.

Component B includes 150 case-based items that reference 12 to 16 dental hygiene patient cases. These cases present information dealing with adult and child patients by means of patient histories, dental charts, radiographs, and, at times, intra- and extra-oral photographs. Each examination includes at least one case regarding patients of the following types: Geriatric, Adult-Periodontal, Pediatric, Special Needs, and Medically Compromised. A compromised patient is one whose health status may require modification of standard treatment or special consideration.

The case-based items address knowledge and skills required in:

- 1.0. Assessing patient characteristics
- 2.0. Obtaining and interpreting radiographs
- 3.0. Planning and managing dental hygiene care
- 4.0. Performing periodontal procedures
- 5.0. Using preventive agents
- 6.0. Providing supportive treatment services
- 7.0. Professional responsibility

The distribution of items in Component A is defined by the following outline. The number in brackets after each topic indicates the number of questions assigned to this topic.

DETAILED TEST SPECIFICATIONS FOR THE FULL-LENGTH NBDHE: 2020

Case Independent Items [200]

SCIENTIFIC BASIS FOR DENTAL HYGIENE PRACTICE [61]

1.0. Anatomic Sciences [14]

- 1.1. Anatomy
 - 1.1.1. Head and neck anatomy
 - 1.1.2. Dental anatomy
 - 1.1.2.1. General anatomy
 - 1.1.2.2. Root anatomy
- 1.2. Histology and Embryology

2.0. Physiology [4]

3.0. Biochemistry and Nutrition [6]

4.0. Microbiology and Immunology [9]

5.0. Pathology [13]

5.1. General

5.2. Oral

6.0. Pharmacology [15]

PROVISION OF CLINICAL DENTAL HYGIENE SERVICES [115]

1.0. Assessing Patient Characteristics [20]

1.1. Medical and dental history

1.2. Head and neck examination

1.3. Periodontal evaluation

1.4 Oral evaluation

1.5. Occlusal evaluation

1.6. General

2.0. Obtaining and Interpreting Radiographs [13]

2.1. Principles of radiophysics and radiobiology

2.2. Principles of radiologic health

2.3. Technique

2.4. Recognition of normalities and abnormalities

2.5. General

3.0. Planning / Managing Dental Hygiene Care [34]

3.1. Infection control (application)

3.2. Recognition of emergency situations and provision of appropriate care

3.3 Individualized patient education

3.3.1. Planning of individualized instruction

3.3.2. Provision of instruction for prevention and Management of oral disease

3.3.2.1. Dental caries

3.3.2.2. Periodontal diseases

3.3.2.3. Oral conditions

3.4. Anxiety and pain control

3.5. Recognition and management of comprised patients

3.6. Dental hygiene treatment strategies

3.6.1. Diagnosis

3.6.2. Treatment plan

3.6.3. Case presentation

3.7. General

4.0. Performing Periodontal Procedures [20]

4.1. Etiology and pathogenesis of periodontal diseases

4.2. Prescribed therapy

4.2.1. Periodontal debridement

4.2.2. Surgical support services

4.2.3. Chemotherapeutic agents

4.2.4. General

4.3. Reassessment and maintenance (e.g. implant care)

5.0. Using Preventive Agents [9]

5.1. Fluorides - systemic and topical

5.1.1. Mechanisms of action

5.1.2. Toxicology

5.1.3. Methods of administration

5.1.3.1. Water fluoridation

5.1.3.2. Self-administered

5.2. Pit and fissure sealants

5.2.1. Mechanisms of action

5.2.2. Self-administered

5.3. Other preventative agents

6.0. Providing Supportive Treatment Services [7]

6.1. Properties and manipulation of materials

6.2. Polishing natural and restored teeth

6.3. Making of impressions and preparation of study casts

6.4. Other supportive services (e.g. tooth desensitization)

6.5. General

7.0. Professional Responsibility [12]

7.1. Ethical principles, including informed consent

7.2. Regulatory compliance

7.3. Patient and professional communication

7.4. General

COMMUNITY HEALTH/RESEARCH PRINCIPLES [24]

1.0. Promoting Health and Preventing Disease within Groups [6]

2.0. Participating in Community Programs [9]

2.1. Assessing populations and defining objectives

2.2. Designing, implementing, and evaluating programs

3.0. Analyzing Scientific Literature, Understanding Statistical Concepts, and Applying Research Results [9]

TEST SPECIFICATIONS FOR SHORT-FORM-NBDHE: 2020

In 2020, the JCNDE released the short-form-NBDHE in response to circumstances surrounding COVID-19. The short-form-NBDHE is a shortened version of the full-length NBDHE comprised of 155 total questions. Content appearing in the short-form-NBDHE is proportionally representative of the current NBDHE test specifications. The specifications for the short-form-NBDHE are provided below.

Discipline-Based Component (85 questions)

The discipline-based component contains questions addressing three major areas:

- Scientific Basis for Dental Hygiene Practice (29%)
- Provision of Clinical Dental Hygiene Services (59%)
- Community Health/Research Principles (12%)

Case-Based Component (70 questions)

The case-based component includes questions involving dental hygiene patient cases.

**Joint Commission on National Dental Examinations
Department of Testing Services
211 East Chicago Avenue
Chicago, Illinois 60611-2637**

© 2021. Joint Commission on National Dental Examinations, American Dental Association.
All rights reserved.