

**JCNDE** JOINT COMMISSION  
ON NATIONAL  
DENTAL EXAMINATIONS



# 2020 Technical Report National Board Dental Examinations (NBDE Part I and II)



**Technical Report**  
**National Board Dental Examinations**

**2020**

## **Executive Summary**

### **Technical Report: National Board Dental Examinations (2020)**

The 2020 edition of the Technical Report for the National Board Dental Examinations (NBDE) is the main source of validity evidence available to state licensing boards and other users of dental examination results. Validity is the most important consideration for any examination program. For these dental examinations, 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 dentistry. The technical report contains both direct evidence and references to other documents and sources of information that contribute to this body of validity evidence. The background and historical information in this report allow users to understand the development of this program.

The NBDE 2020 Technical Report focuses on the National Board Dental Examination (NBDE) testing program and findings for the 2020 calendar year. The findings associated with the NBDE Part I exam program are summarized below for the last time, given the discontinuation of the NBDE Part I program on December 31, 2020.

The content of the Technical Report is presented to address professional standards regarding the validity of credentialing examinations (American Educational Research Association (AERA), American Psychological Association (APA), and the National Council on Measurement in Education (NCME), 2014). Successful completion of a credentialing examination indicates test-takers have achieved an acceptable level of performance in an area of knowledge. Some of the principal information presented in the Technical Report is summarized below.

- **Purpose:** The purpose of the NBDE is to assist state boards in determining the qualifications of individuals who seek licensure to practice dentistry. These qualifications include the ability to understand important information from the biomedical, dental, and clinical dental sciences and apply such information in a problem-solving context.
- **Content:** Content specifications are based on validity studies involving practice analyses conducted roughly every five years. Test construction teams are responsible for recommending minor modifications during the interim period. The American Dental Association's (ADA) Joint Commission on National Dental Examinations (JCNDE), with input from its Committee on Examination Development, 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 examinations using Joint Commission guidelines for writing high-quality multiple-choice items.
- **Standard Setting and Scoring:** Part I and Part II are criterion-referenced and not norm-referenced. This means examination results and pass/fail points are determined by specific criteria and not by the process sometimes known as "grading on a curve." A panel of expert educators and practitioners recommend the minimum passing score, which is ultimately determined by the JCNDE. The standards are maintained across examination forms through the use of equating procedures designed to control for subtle differences in the difficulty of items from one examination form to another. The equating process places exam results on a common metric regardless of which examination form was administered.
- **Administration:** The ADA maintains a high level of security on all examination materials. Strict precautions in place at the Joint Commission's offices

and testing centers help ensure test content remains secure. The Joint Commission offers Part I and Part II via computer at Prometric Professional Level Testing Centers throughout the United States, US territories, and Canada. Once eligible, candidates can schedule an examination for any business day.

In addition to the items above, this report provides information on examination history, administration, candidates' rights and responsibilities, and failure rates. A copy of the Technical Report is available for download on the ADA's website (<https://www.ada.org/en/jcnde/news-resources/technical-reports>).

## Table of Contents

1. Introduction .....	4
2. Purpose of the National Board Dental Examinations .....	5
3. Historical Perspective .....	6
4. The 2020 NBDE Program.....	7
5. Validity, Validation, and Validity Evidence .....	8
6. Professional Test Standards .....	8
7. Legal Issues.....	9
8. Validity Evidence in this Technical Report.....	9
9. Content Basis for the Examination .....	10
10. Item Development .....	13
11. Item Validation.....	17
13. Administration.....	23
14. Score Reliability.....	23
15. Standard Setting.....	24
16. Scaling, Equating, and Comparability of Test Forms .....	29
17. Scoring and Reporting Test Results.....	34
20. Validity Studies .....	38
21. Security .....	38
22. Guidelines for High-Stakes Testing.....	40
References .....	43
Appendix A.....	46
Appendix B.....	48
Appendix C.....	51
Appendix D.....	53

## 1. Introduction

High-stakes examination programs, such as those of the Joint Commission, must be concerned with validity. Validity refers to the degree to which logic and evidence support the use and interpretation of examination results in accordance with the purpose of the examination. With respect to the NBDE, the examination purpose involves providing boards of dentistry with information that helps them to understand the qualifications of individuals seeking licensure to practice dentistry, and specifically whether a candidate for licensure possesses the level of cognitive skills that is necessary to safely practice. The Joint Commission also has an obligation to inform state boards and communities of interest concerning its efforts to provide the highest quality examination programs possible. Established professional standards provide useful guidance to improve the quality of examinations. Testing programs must adhere to these standards and provide evidence their policies and procedures conform to them to help ensure confidence in the examination program.

This technical report focuses on the NBDE testing program and findings for the 2020 calendar year. This report also provides a comprehensive summary of validation efforts leading up to 2020 as well as background information, which allows the reader to understand the program's development to its present state.

Technical reports document the validity evidence that supports examination usage. The *Standards for Educational and Psychological Testing*, most recently published by AERA, APA, and NCME in 2014, provide professional standards for testing organizations. Chapter 7 of the *Standards* describes the importance of documented validity evidence in technical reports so examination users can evaluate the validity of examination results they interpret and use. Ten relevant standards from that chapter appear in Table 1.1. The Joint Commission endeavors to provide the highest quality examination programs possible.

**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 Examinations**

The first and most fundamental step in the development of any examination program is to establish a purpose. The purpose of the NBDE program is to measure whether a candidate possesses entry-level knowledge and cognitive skills adequate for the competent practice of dentistry. This knowledge includes the ability to recall important information from the biomedical, dental, and clinical dental sciences and apply such information in a problem-solving context.

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 ADA to provide and conduct written examinations to be used by state boards of dentistry for licensing dentists. These examinations were to provide a national standard for the biomedical and clinical dental sciences knowledge necessary for the competent practice of dentistry. The practical demonstrations of clinical skills were reserved for individual states to administer. The National Board's responsibilities included developing and administering National Board examinations and formulating the rules and regulations pertaining to them.

Current National Board Examinations bear little similarity to the first editions, which were administered in 1933 and 1934. Advances in examination methodology caused the most dramatic changes. The examination format was changed in the early 1950s from essay questions to multiple-choice questions. This led to the adoption of norm-referenced scoring procedures, and the National Board delegated examination construction to committees of dentists and dental hygienists who were subject-matter specialists. In the 1960s, the Council on National Board Examinations, which succeeded the National Board of Dental Examiners, was among the earliest testing agencies to employ computer scoring and use statistical techniques to identify candidates suspected of rule violations.

In the early 1980s, the JCNDE, which succeeded the Council on National Board Examinations, instituted the procedure of equating examinations by means of anchor items which would appear on more than one test form. This was done to implement a consistent standard for minimally acceptable performance across examination forms, and it ended the era of norm-referenced scoring. The pass rate on the examinations thereafter fluctuated only to the degree that candidates' abilities changed, or to the degree to which the standard itself was changed. In 1992, a comprehensive, case-based NBDE Part II replaced the NBDE Part II battery consisting of seven individual examinations. Also, at that time, a criterion-referenced method of setting the performance standard based on Rasch psychometric theory was instituted for Part II. In 2007, a comprehensive NBDE Part I examination replaced the traditional battery of four individual examinations. The comprehensive NBDE Part I currently consists of 400 items, including about 80 testlet-based items. Testlets involve a brief patient case in narrative form with a summary chart and series of associated multiple-choice items. Part I has been criterion-referenced since the early 1990s. In 2012, the Joint Commission moved to pass/fail reporting of results for candidates who passed the



examinations. Candidates who fail receive scores for remediation purposes. The Joint Commission regularly updates examination content to reflect advances in the biomedical and clinical dental sciences and keep the examinations current with the practice of dentistry.

State boards of dentistry and candidates came to accept the NBDE over time. The first candidates completed National Board examinations in 1934. For the five-year period from 1934 through 1938, an average of 70 candidates per year received National Board certificates. By 1938, 11 states accepted National Board results. State board participation remained low until the mid-1950s. By 1960, 33 state boards and the District of Columbia accepted National Board results, and by 1976, 48 states plus the District of Columbia, Puerto Rico, and the Virgin Islands accepted them. By 1990, all U.S. licensing jurisdictions accepted the National Board Examinations as fulfillment of the written examination requirement for licensure.

#### 4. The 2020 NBDE Program

NBDE 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 typical calendar year.

Document (2020)
NBDE Part I Guide (2020) NBDE Part II 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 NBDE Examination Guides provide a description of the examination program in place for the pertinent calendar year. In this calendar year, two separate National Board Dental Examinations – Part I and Part II – provide publicly available information to dental licensing boards.

The Part I examination consists of four disciplines: (1) anatomic sciences, including gross anatomy, histology, and oral embryology, (2) biochemistry and physiology, (3) microbiology and pathology, and (4) dental anatomy and occlusion. Since 2007, Part I has consisted of one comprehensive examination covering the same disciplines, with items addressing the various disciplines intermingled throughout the examination.

Part II consists of one comprehensive examination covering the following disciplines: (1) operative dentistry, (2) pharmacology, (3) prosthodontics, (4) oral and maxillofacial surgery and pain control, (5) orthodontics and pediatric dentistry, (6) oral diagnosis, including oral pathology and dental radiology, (7) endodontics, (8) periodontics, and (9) patient management, which includes behavioral science, dental public health, and occupational safety.

The National Board Examinations have evolved since the publication of the first National Board Examination in response to changes in the sciences being tested.

### **Examination Dates**

The NBDE Parts I and II are administered year-round, every day that test centers are in operation.

### **Examination Centers**

Prometric administers NBDE Parts I and II at its Professional Level Testing Centers located throughout the United States, its territories, and Canada.

## **5. Validity, Validation, and Validity Evidence**

Validity is defined in the *Standards for Educational and Psychological Testing* as “the degree to which evidence and theory support the interpretations of test scores for purposed uses of tests” (AERA, APA, & NCME, 2014, p. 11). Validation involves the investigative process of creating a validity argument and collecting evidence relevant to this argument, the examination purpose, and the intended interpretation of results. When acquired validity evidence reveals weaknesses or deficiencies, the testing organization is expected to take steps to address the deficiencies to strengthen the validity of the test.

In the United States, all candidates for licensure as dentists must meet a number of criteria before they are licensed to practice in a state. Each state has the independent authority to issue the license. However, in dentistry, as in many other professions, national standards exist.

With the Part I examination, the intended interpretation of results concerns the biomedical and dental science knowledge dentists must possess; results indicate whether or not the candidate possesses an adequate level of such knowledge to safely practice. The intended interpretation of Part II examination results concerns the professional knowledge of clinical dental sciences—including professional responsibility and patient management abilities—dentists must possess to safely practice. Part II results are used to recommend passing or failing the candidate. This technical report presents validity evidence and additional references that support both the interpretation and use of exam 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—AERA, APA, and NCME—joined forces and resources to create the latest version of these standards (AERA, APA, NCME, 2014). These standards provide useful information to guide testing organizations in the validation of their test score interpretations and uses. Throughout this technical report, validity evidence is identified and connected to testing standards. Many sections of this technical report correspond to chapters in the *Standards for Educational and Psychological Testing* (AERA, APA, NCME, 2014).

In 2000, AERA issued a set of guidelines intended for use with high-stakes, high school graduation examination programs. Some of these guidelines apply to the NBDE. In section 22 of this technical report, these guidelines are reviewed against the validity evidence presented in this technical report.

## **7. Legal Issues**

All examination programs where results are used for high-stakes decisions run the risk of legal challenge based on validity. As a result, examination programs must be designed to withstand legal challenges.

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. In the process, it provides confidence 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**

This report is organized to address major categories of validity evidence. Each section contains narrative and validity documentation. In some instances, data are provided, as appropriate. In each major category, reference is made to one or more standards from the *Standards for Educational and Psychological Testing* (AERA, APA, & NCME, 2014). 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 references to other existing documents provide evidence that standards 1.0, 1.1, and 1.2 have been met, and they indicate the Joint Commission has acted responsibly in validating its examinations.

The rest of this report addresses the following important categories of validity evidence, presented with corresponding section numbers:

9. Content Basis for the Examination
10. Item Development
11. Item Validation
12. Test Design and Development
13. Administration
14. Score Reliability

15. Standard Setting
16. Scaling, Equating, and 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 methods for determining the content of a certification or licensure examination for any profession serve as a primary type of validity evidence. Table 9.1, which lists standards related to the content of such examinations, gives ample proof of the importance of the content basis for the NBDE Part I and NBDE Part II examinations. Key elements for validity evidence involve (1) the use of a practice analysis that identifies the knowledge and problem-solving skills necessary for safe practice of dentistry in the U.S., (2) examination specifications, and (3) the use of content experts for recommending minor modifications to the examination specifications in a series of validation processes.

### Examination Content

As noted previously, the dental examinations are organized into two parts, NBDE Part I and NBDE Part II. Each part is developed according to examination specifications. The examination specifications list topics included in each examination. The relevant examination specifications appear in Appendices A (pertaining to Part I) and B (for Part II).

**Part I.** NBDE Part I is a comprehensive computer-based examination usually taken after two years of dental school. The examination items focus on four disciplines in the biomedical and dental sciences: Anatomic Sciences; Biochemistry-Physiology; Microbiology-Pathology; and Dental Anatomy and Occlusion. Each discipline is examined in 100 multiple-choice items, intermingled throughout the examination. Approximately 20 percent of the 400 items are testlet-based.

**Part II.** NBDE Part II is a comprehensive, computer-based examination usually taken during the last year of dental school. It consists of a comprehensive, one-and-one-half day examination of 500 items in two components: 400 discipline-based, or case independent items, given on the first day, and 100 case-based items given on the second day. It covers the clinical dental sciences: Operative Dentistry, Pharmacology, Endodontics, Periodontics, Oral and Maxillofacial Surgery and Pain Control, Prosthodontics, Orthodontics, Pediatric Dentistry, and Oral Diagnosis, including Oral Pathology and Dental Radiology. It also covers Patient Management, including Behavioral Science, Dental Public Health and Occupational Safety. The 100 items based on patient cases might derive from any of the biomedical sciences and clinical dental sciences, including patient management.

**The Practice Analysis for Part I.** In 2001, validity evidence was acquired through a practice analysis conducted using the 63 *Competencies of the New Dentist*, developed by the American Dental Education Association (American Dental Education Association, 2001). The practice analysis findings suggested the Joint Commission should consider building a more clinically relevant NBDE Part I examination. Accordingly, the Joint Commission piloted a restructured 400-item examination composed of 80 percent discipline-based items and 20

percent testlets. The findings of the pilot were accepted, and the Joint Commission approved a resolution to implement a comprehensive Part I examination in 2007. An empirical study comparing the conjunctive and comprehensive examination formats determined that the change resulted in a more clinically relevant, interdisciplinary format that enhanced the test validity, a favorable outcome for the Joint Commission (Yang, Neumann, & Kramer, 2012).

**The Practice Analyses for Part II.** In 2017, the Joint Commission relied on the findings of a 2016 practice analysis survey to approve slight changes to the content specifications for the NBDE Part II examination. The specifications in place prior to that time were based on the results of a 2011 practice analysis (Tsai, Yang, Waldschmidt, & Chang, 2012). As part of the 2016 practice analysis, validation evidence was obtained by collecting ratings from a sample of active, full-time dentists who had been in practice for ten years or less, concerning the frequency and importance of 56 competencies judged relevant to patient care. The surveyed dentists 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 a total of 34,441 dentists. Of those, 2,542 (7.4%) provided valid responses. The mean frequency rating and mean importance rating were calculated for each competency. The mean frequency ratings ranged across competencies from 1.7 to 5.92. The mean importance ratings ranged from 3.22 to 4.83. 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 dentists' frequency and importance ratings, and the study's overall findings confirmed the validity of Part II. In June 2017, the Joint Commission's approved the practice analysis methodology and the revised content specifications. These revised content specifications were implemented in 2019.

**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.*

---

## **10. Item Development**

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

### **Who Writes Test Items?**

Based on the recommendations of the Committee on Examination Development, the JCNDE annually approves and reapproves test constructors into the NBDE Test Constructor Pool. An individual who has completed five years of service in the pool may be considered for re-approval as dictated by the needs of the examination program. Department of Testing Services (DTS) staff will place JCNDE approved test constructors onto specific Test Construction Teams (TCTs) based on the expertise of the individual and the needs of the TCT and examination program. A team is formed for each specific meeting, and disbands at the end of that meeting. These teams are flexible and may or may not consist of the same test constructors each year. Individuals are invited to attend a given meeting. Should they accept, they are considered part of the team for that calendar year. Teams may be rearranged as needed in the event that a given volunteer is not able to attend. If a volunteer

is invited but is unable to attend, an alternate volunteer may be identified and invited. Additionally, if a volunteer is invited to attend a meeting and does not respond in a timely manner, an alternative volunteer may be identified and invited to attend the meeting. This process helps ensure teams will always have a sufficient number of volunteers with the required expertise, so that meeting goals can be accomplished efficiently and effectively. Each test constructor receives the following materials: *Test Item Development Guide*, *Orientation Manual for Test Constructors*, *National Board Dental 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 all sections of the examination. When new test constructors attend their first meeting, returning test constructors informally discuss the process and serve as mentors to them during their initial service as test constructors.

The *Test Item Development Guide* describes 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 committee. The *Guidelines* are provided to all test constructors and specialty boards, and anyone else requesting them through the Joint Commission office. A style manual 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 National Board examinations use multiple-choice formats. NBDE Part I uses both independent items and testlet-based items addressing biomedical and dental sciences. For NBDE Part II, the case-independent format assesses clinical dental sciences and patient management knowledge pertinent to licensing. The case-dependent format uses case materials consisting of patient dental and medical histories, dental charts, radiographs, and clinical photographs. These materials are used in Part II and serve as stimulus for case-associated questions. For Parts I and II, multiple-choice items each include a stem pairing a question or statement with a list of possible responses. National Board items have three to five possible responses.

## **The Process of Examination Revision**

National Board examinations are subject to a review and revision process to address unsatisfactory items. Test items are considered unsatisfactory if they are too easy, too difficult, or fail to discriminate between stronger and weaker candidates. In reviewing items, test construction teams (TCTs) look at two key factors: the proportion of individuals answering an item correctly (referred to as the *p-value*), and the point-biserial correlation (or  $r_{pb}$ ) between item and examinee performance. *P-values* provide information concerning item difficulty while  $r_{pb}$  point-biserial correlations provide information concerning item discrimination. The Joint Commission accepts a range of item difficulties, but items deemed too easy — where virtually all candidates answer them correctly — or too difficult — where virtually no candidates answer them correctly — are typically less useful from a measurement perspective. Item discrimination indicates the relation between the candidates who choose the correct answer and their performance across the total number of items. The Joint Commission considered the following ranges for item difficulty and discrimination.

**Table 10.2**  
**Discrimination and Difficulty Indices, Ranges and Interpretations**

	<b>Part I</b> <b>(<math>r_{pb}</math>)</b>	<b>Part II</b> <b>(<math>r_{pb}</math>)</b>	<b>Parts I and II</b> <b>(<math>p</math>)</b>
H	= .26 or greater	H = .26 or greater	E = .90 or greater
M	= .15 to .25	M = .08 to .25	M = .40 to .89
L	= less than .15	L = less than .08	D = .00 to .39
$r_{pb}$ :	H – High; M – Medium; L – Low		
$p$ :	E – Easy; M – Medium; D – Difficult		

For an item to be considered effective, Joint Commission standards dictate it must produce a difficulty index between 0.40 and 0.89, and a corresponding discrimination index of 0.15 or greater for NBDE Part I and 0.08 or greater for NBDE Part II. Items that do not meet these standards are scrutinized and become candidates for elimination or revision.

The following are steps for revising discipline-based NBDE Part I and NBDE Part II items.

1. The TCT reviews an analysis of item difficulty and discrimination. These two reports, which are generated after a set period, provide information on the results of the examination.
2. The TCT reviews the unsatisfactory items.
3. Joint Commission staff help to analyze the problematic items. The TCT decides whether to retain, revise, or remove the item. The revision process involves rewording the stem or changing the distractors.
4. Joint Commission staff note all changes. Revised items are returned to the item bank and subsequently field tested to see if they can be used in future examinations.

The following are steps for revising Part II case-dependent items.

1. Joint Commission staff determine the number of “good” and “poor” items in the cases and present a summary to the team.
2. The TCT determines whether a case is worth reviewing or revising based on the ratio of good to poor items. If the TCT decides to delete some case items, it retains the case materials — patient history, chart, radiographs, and photographs — for future use and writes new case items to replace the poor ones.
3. If the TCT determines a case can be improved with modifications, it reviews the patient history, dental chart, radiographs, and clinical photographs. Test constructors read all items aloud.
4. The TCT discusses each item. The facilitator helps to analyze the problematic items. The TCT decides whether to revise, replace, or completely remove the item. Revision involves rewording the stem or the distractors, or changing the distractors completely. Replacement involves writing an entirely new item. Removal means eliminating the item from the case.
5. Joint Commission staff note all changes. Revised cases are saved for future use.

Revising items replenishes item banks and familiarizes test constructors with the characteristics of acceptable examination items.

## 11. Item Validation

After an item is written, qualified personnel should validate the item through a review process. Downing and Haladyna (1997) recommend a series of reviews to improve the quality of items. Table 11.1 provides a short list of standards pertaining to item validation.

**Table 11.1**  
**Standards Pertaining to Item Validation**

---

*4.7 The procedures used to develop, review, and try out items and to select items from the item pool should be documented.*

*4.8 The test review process should include empirical analyses and/or the use of expert judges to review items and scoring criteria. When expert judges are used, their qualifications, relevant experiences, and demographic characteristics should be documented, along with the instructions and training in the item review process that the judges receive.*

*4.10 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.*

---

The *Standards* (AERA, APA, & NCME, 2014) indicate that items that count toward candidates' scores should exhibit sound psychometric characteristics. Specifically, item difficulty and discrimination should compare favorably with the Joint Commission's item-performance standards. Subsequent to their administration to a representative sample of candidates, item statistics are generated from the data and analyzed to confirm reasonable psychometric performance.

### **Evaluating and Revising Weak or Unacceptable Items**

The Joint Commission published a document in November 1995 to help test constructors review unsatisfactory items and revise or retire them. Retiring items results in openings in the item bank and an opportunity for test constructors to evaluate and improve test content.

## 12. Test Design and Development

The overall design of each examination is a crucial step in test development. Items chosen for each examination form must conform to the examination specifications in precise ways. Not only must content requirements be met, but also psychometric characteristics should be comparable across examination forms. Table 12.1 lists standards that pertain to examination 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 full participation of content expert teams and supervised by staff specialists from the Joint Commission's test development area. This process ensures the expertise of highly qualified, licensed dentists is fully used in item selection and examination design. Joint Commission staff provide technical support and guidance to ensure the desired technical qualities are achieved during the examination design phase.

The Joint Commission convenes numerous test construction teams. The details of test constructor eligibility, recruitment, and service are provided in this section. As noted earlier in this technical report, these teams also write and evaluate test items in the item development phase.

### **The Role of a Test Constructor**

The role of test constructors is fundamental to the validity and reliability of the National Board Dental Examinations. Test constructors evaluate examinations and—accompanied by appropriate justification—recommend changes to the Joint Commission through its Committee on Examination Development. Most recommended changes involve terminology or minor shifts in focus to the content specifications. Test constructors are responsible for constructing a clear, precise, and cohesive group of items for each examination and for providing content-related validity evidence.

Test constructors meet each year in discipline or case-based teams to engage in test development activities. The quality of the examinations relies on test constructors to use their subject-matter expertise, their familiarity with the curriculum in accredited dental schools, and their awareness of what is important in the practice of general dentistry in the construction of each new examination. Most of this work is done in team meetings.

## The Nature of Test Construction Teams

The Joint Commission relies on 18 test construction teams consisting of volunteers to develop the National Board Dental Examinations. Past experience in providing adequate content expertise determines team size. The 18 NBDE Part I and NBDE Part II TCTs are as follows, with the total number of test constructors appearing to the right.

### Part I - Basic Biomedical Sciences Teams

Anatomic Sciences-----	5
• 2 gross anatomists	
• 2 histologists (1 embryology expert and 1 neuroanatomy expert)	
• 1 full-time practitioner	
Biochemistry and Physiology -----	5
• 2 biochemists	
• 2 physiologists	
• 1 full-time practitioner	
Microbiology and Pathology -----	5
• 2 microbiologists (1 immunology expert)	
• 2 general pathologists	
• 1 full-time practitioner	
Dental Anatomy and Occlusion-----	4
• 3 dental anatomists	
• 1 full-time practitioner	

### Part I - Component B Teams

Testlet Development-----	9
• 4 full-time practitioners	
• 5 experts in each Part I discipline	
Consultant Review -----	2
• 1 dental sciences expert	
• 1 full-time practitioner	

### Part II - Component A (independent items) Teams

Endodontics-----	4
• 3 endodontists	
• 1 full-time practitioner	
Operative Dentistry -----	5
• 4 dentists (1 dental materials expert)	
• 1 full-time practitioner	

Oral and Maxillofacial Surgery and Pain Control -----	4
• 3 oral and maxillofacial surgeons (1 pain control expert)	
• 1 full-time practitioner	
Oral Diagnosis -----	6
• 2 oral pathologists	
• 2 oral and maxillofacial radiologists	
• 1 dentist with advanced education in oral diagnosis	
• 1 full-time practitioner	
Orthodontics and Pediatric Dentistry-----	6
• 3 orthodontists	
• 2 pediatric dentists	
• 1 full-time practitioner	
Patient Management -----	8
• 3 behavioral scientists (1 dentist)	
• 2 dental public health specialists	
• 1 dentist with advanced training in special needs	
• 2 full-time practitioners	
Periodontics-----	4
• 3 periodontists	
• 1 full-time practitioner	
Pharmacology-----	4
• 3 pharmacologists (1 dentist)	
• 1 full-time practitioner	
Prosthodontics -----	6
• 4 prosthodontists (2 fixed prosthodontic experts; 2 removable partial/ complete prosthodontics experts)	
• 1 dental materials expert	
• 1 full-time practitioner	

**Part II - Component B (case-dependent items) Teams**

Component B – Case Composition Team ----- 13  
This team, composed of dental discipline experts and practitioners, prepares the case-based items for Part II of the National Board Dental Examinations.

Case Selection Team ----- 4  
As an adjunct to Component B, this team does the preliminary work of screening new patient cases and identifying suitable cases for the examinations. In addition, it drafts and reviews the patient histories, dental charts and treatment plans associated with the cases.

Consultant Review Team ----- 2  
To ensure examination coherence and cohesion, this team reviews the discipline-based and case-based components of the Part II examination.

## **Criteria for Dental Test Constructors**

A document entitled *JCNDE Test Construction Teams and Selection Criteria* provides criteria for the selection of volunteers to serve on dental TCTs. To be considered for approval into the test constructor pool, a person must meet certain qualifications and must submit a completed personal data form.

The following are the criteria for test constructors in Anatomic Sciences, Biochemistry-Physiology, Microbiology-Pathology, Dental Materials, Pharmacology, and Patient Management, which includes Dental Public Health, Behavioral Science and Special Needs:

1. Dentist with a master's degree in that biomedical science or any professional with a doctoral degree in that biomedical science, and
2. Three years of experience within the last five years in teaching or in research in that biomedical science.

The following are the criteria for test construction teams in Dental Anatomy and Occlusion, Endodontics, Operative Dentistry, Oral and Maxillofacial Surgery and Pain Control, Oral Diagnosis, including Oral Pathology and Radiology, Orthodontics, Pediatric Dentistry, Periodontics, and Prosthodontics:

1. Dentist
2. In the case of special areas of dentistry, graduation from an accredited advanced education program in that specialty, and
3. Three years of experience within the last five years in teaching or research in that specialty.

To qualify for consideration in the NBDE test construction process as a full-time practitioner, a dentist must have experience practicing dentistry (inclusive of clinical teaching) 20 hours per week for at least 5 years.

## **The Selection of Test Constructors**

The Joint Commission annually advertises its need for test constructors. A letter explaining the online application materials, consisting of selection criteria, and a personal data form is emailed to dental schools, state boards of dentistry, constituent dental societies, and other institutions and individuals well in advance of the annual meeting of the Joint Commission. All applications are processed by staff and forwarded to the Joint Commission's Committee on Examination Development, which is responsible for recommending individuals for the NBDE test constructor pool.

The JCNDE annually approves and reapproves test constructors into the NBDE test constructor pool. Approval into the NBDE test constructor pool is for five 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 NBDE test constructor pool, staff select test constructors to attend TCT meetings for the upcoming year. These selections are based primarily on subject matter expertise, although geographic location is considered. Membership in the ADA is required for those in the clinical sciences.

## Test Constructor Responsibilities

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

1. Submit new 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 who have completed a year of service. The number of new items constructors are expected to submit varies according to the program's needs.
2. Attend each test construction meeting for the duration of the session.
3. Construct NBDE exams according to Joint Commission guidelines, specifications, and content outlines within the designated time frame.
4. Construct additional items for the item banks when necessary.
5. Assign ownership of all examination materials to the ADA and JCNDE, by agreeing to the terms of the copyright assignment.
6. Inform the Joint Commission of changes in the standard curricula, and suggest modifications in examination 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 the National Board examinations by declining any arrangement to assist with review courses or reviewing 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 ADA's policy on professional conduct. The policy includes prohibitions against sexual harassment and other forms of unlawful conduct.

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

## How National Board Dental Examinations Are Developed

Each team is charged with constructing a specific NBDE examination or portion of an examination. The Part I comprehensive National Board Dental Examination contains 400 items — about 320 discipline-based and 80 testlet-based items. The Part II comprehensive examination contains 500 items — 400 discipline-based items and 100 case-based items. Since Part I has been retired and Part II will be retired in 2022, these teams are no longer meeting. When meetings were taking place, the discipline-based teams met once per year, usually for three days. The Part I Testlet Development Team and the Part II Component B (case-based) Team each met three times per year. The Part II Case Selection Team and the Review Team each met twice per year, usually for two days.

Many test construction meetings involve a review of items that failed to meet JCNDE difficulty and discrimination standards. Items not meeting these standards are scrutinized and become candidates for elimination or revision. Next, test constructors review the *National Board Dental Examination Specifications* (see Appendices A and B) to ensure the discipline areas 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 *Test Item Development Guide* and the examination specifications. The final step is to draft new examinations using new items and existing items with acceptable statistical performance. Following the team meetings, test construction consultants and staff conduct final reviews.

### 13. Administration

Several important issues related to administration are addressed in this section and 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 JCNDE in its *Operational and Policy Manual of the JCNDE* describes eligibility requirements for candidates for the corresponding examinations who take it for the first time or who re-test. This publication also describes how candidates apply for the examinations.

The Joint Commission also describes the standardized procedures for administration, which are used by the third-party test centers that administer the examinations.

### 14. Score Reliability

Score reliability is an important indicator of examination quality. Test developers strive to ensure test scores provide a stable and precise measurement of a candidate's knowledge, skills, and abilities. Despite efforts to eliminate possible sources of measurement error, random factors can affect candidate performance and subsequent examination results. Reliability indices assess 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 ( $KR_{20}$ ) to report score reliability for NBDE Part I and NBDE Part II. This index provides internal consistency estimates for tests with items scored dichotomously (e.g., right or wrong). As shown in Tables C.1 and C.2 of Appendix C,  $KR_{20}$  values for the 2020 examinations range from 0.94 to 0.95 for Part I and from 0.88 to 0.92 for Part II.

Table 14.1 lists the reliability standards applicable to the Part I and Part II examinations. 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.*

---

## **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). Part I and Part II cut scores represent a collective judgment that candidates with scores below a particular skill level have an unacceptable likelihood of making serious errors in the practice of dentistry. The setting of cut scores may involve empirical study, but value judgments by content experts are inevitable. Judges involved in setting cut scores should be qualified, and documentation of their qualifications should be provided. The process for setting the cut score should be well described and documented. Table 15.1 provides standards that are relevant to setting the cut scores for Part I and Part II.

**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 nontest 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.*

---

In 2016, the Joint Commission began the process of transitioning its examination programs to new standards and standard setting procedures. With respect to the NBDE, the transition for Part I took place in 2016, and transition for Part II took place in 2017.

### **Standard-Setting procedures for NBDE Part I and NBDE Part II**

The current standards for the NBDE Part I and NBDE Part II were determined in October and November 2014 respectively using the Bookmark standard setting methods (Lewis, Mitzel, Mercado, & Schulz, 2012). Although held as two separate meetings involving separate individuals on two separate panels, the procedures used to set the cutoff standard were the same for both examinations. Specifically, the 2014 standard setting activities for both examinations took place at the ADA headquarters over a two-day period utilizing the following steps:

A standard setting committee was convened. The standard setting committee for the NBDE Part I was comprised of ten members: seven full-time practitioners and three dental educators each affiliated with an accredited dental school. The standard setting committee for the NBDE Part II was comprised of twelve members: eight full-time practitioners and four dental educators each affiliated with an accredited dental school.

1. The committee members received a thorough overview of the purpose and content of their respective NBDE 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.
2. 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.
3. 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.
4. Committee members reviewed a large set of examination items that had been placed into an Ordered Item Booklet (OIB) assembled as follows:
  - Each page of the OIB contained one item.
  - 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.
  - The items included in the OIB spanned a representative range of difficulty levels.
  - 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).
5. 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 (percentage failing) of using the cut score associated with the median bookmark placement.
  6. 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.
  7. At the conclusion of this process, panelists were asked to complete an evaluation questionnaire regarding their impressions of the process.
    - For the Part I standard setting, all panelists either agreed, or strongly agreed with the following statement: “Overall, I support the final group-recommended cut score as fairly representing the appropriate performance standard for the NBDE Part I.” On a five-point rating scale, ranging from 1=Strongly Disagree to 5=Strongly Agree, the mean rating for this question for the Part I standard setting committee was 4.6.
    - For the Part II standard setting, most 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 NBDE Part II”. On a five-point rating scale, ranging from 1=Strongly Disagree to 5=Strongly Agree, the mean rating for this question for the Part I standard setting committee was 4.92
  8. The final recommendation was presented to the Joint Commission for review and approval.

The recommended cut scores resulting from 2014 NDBE standard setting activities were reviewed and approved by the Joint Commission in 2015, and implemented in the fourth quarter of 2016 for the NBDE Part I and the first quarter of 2017 for the NBDE Part II.

### **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 the pass/fail point on the examination’s scale is reliable and valid (AERA, APA, NCME, 2014, p. 46-47). Testing programs typically adopt two methods to evaluate the reliability of the pass/fail point. The first method examines outcomes from standard setting activities (Cizek, Bunch, and Koons, 2004). The second method computes the probabilities of correct and consistent classifications of candidate performance on an examination

(Livingston and Wingersky, 1979; Hanson and Brennan, 1990; Livingston and Lewis, 1995).

With regard to 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 the second method, Hanson and Brennan (1990) procedures were used to analyze data and to provide results. The results are presented with two types of statistics: (1) classification accuracy—the probability of correct classification, false positive rate, and 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, equal in difficulty and covering the same content domain as the edition they actually took. These concepts are presented schematically in Tables 15.2 and 15.3.

**Table 15.2**  
**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 15.2, an accurate classification occurs when the theoretical decision made based on the average score obtained across all possible examination forms (i.e., the “true-score based decision”) agrees with the decision on the examination form actually taken (i.e., the “observed-score based decision”). False positive and false negative classifications refer to the mismatch between candidate “true-score based decisions” and “observed-score based decisions.” The false positive value is the proportion of candidates whose observed score would be misclassified as “pass” when they actually would have received “fail” based on their true score. The false negative value is the proportion of candidates whose observed score would be misclassified as “fail” when they actually achieved “pass” based on their true score.

**Table 15.3  
Classification Consistency**

		Decision based on a parallel form taken	
		Pass	Fail
Decision based on the actual examination form taken	Pass	Consistent Classification	Misclassification
	Fail	Misclassification	Consistent Classification

In Table 15.3, consistent classifications occur when two forms of an examination agree on the classification as either “pass” or “fail,” misclassifications occur when the decisions in the two forms differ.

In 2013, the Joint Commission conducted an investigation to understand classification accuracy and classification consistency levels for NBDE Part I and NBDE Part II (Yang and Waldschmidt, 2013). This study involved 1,000 candidates enrolled in accredited dental schools who took NBDE Part I and NBDE Part II for the first time in 2012. Results showing classification accuracy and consistency of the pass/fail point on the exams are presented in Table 15.4. This 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. Table 15.4 shows the reliability of the pass/fail points on the examination measurement scales were satisfactory across two study samples, with 97 percent classification accuracy and 96 percent classification consistency.

**Table 15.4  
Classification Accuracy and Consistency of the Pass/Fail Points  
on the NBDE Measurement Scales**

Examination	Part I	Part II
<b>Sample Size</b>	1,000	1,000
<b>Classification Accuracy</b>		
Correct Classification	.97	.97
False Positive	.01	.01
False Negative	.02	.02
<b>Classification Consistency</b>		
Consistent Classification	.96	.96
Misclassification	.04	.04

## 16. Scaling, Equating, and 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, and comparability.

**Table 16.1**  
**Standards Pertaining to Scaling, Equating, and 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.*

---

Different forms of the NBDE are available for administration. To ensure the scores of candidates completing different forms can be directly and meaningfully compared, some statistical adjustments are necessary. The Joint Commission ensures the comparability of scores using equating and score conversions methods. 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 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 standardized examination scores are equated, they are on a common measurement scale. Thus, the scores of candidates completing different forms can be meaningfully evaluated on the same scale using the same cut score of 75. In addition, because the mean scores obtained by different groups of candidates are expressed on the same metric, yearly trends in examination performance can be evaluated fairly, within standard setting cycles.

To equate two examination forms, certain requirements must 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 and identical content, and have been constructed for the same candidate population.

### **Equating Designs**

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

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 use of a single group could produce fatigue, practice, and order effects. This equating design is not feasible due to the length of the Part I and Part II examinations.

Random differences between equivalent groups may be controlled by the use of anchor items. Anchor items represent items 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 representative of the other items administered. Not only is this design feasible, it is widely used and accepted throughout large-scale examinations.

### **Statistical Methods for Adjusting Scores**

Once an equating design has been chosen, the next decision involves selecting 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 examination score level. Traditional methods, such as equipercentile equating, require entire examination score distributions. The use of cumulative distributions of examination scores introduces imprecision into the equating process. Rounding and interpolation errors may occur.



The IRT model currently used with the National Board examinations is called the Rasch model. This model is mathematically equivalent to the one-parameter logistic model, and is quite precise. Second, the Rasch model allows each candidate to complete a set of items different from those attempted by any other candidate, and still be scored on the same scale of measurement. This process can improve measurement accuracy for most candidates, but requires that IRT methods of equating be implemented. Third, Rasch equating allows for extensive cross-checking of item parameters. Because each equating event could introduce error into the estimation of item and person parameters, it is essential to review and evaluate 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 of Rasch equating 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. Under the Rasch model, the only characteristic of the item which is assumed to influence a response is its difficulty. The function used to determine the probability of a correct response of person  $v$  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  $\beta_v$  is the ability of person  $v$ , and  $\delta_i$  is the difficulty of item  $i$  (Wright and Stone, 1979). Both item difficulty and the ability of the person taking the test, or 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 item response model assumes all items in an examination measure the same construct, and that the logistic curve, defined by Equation 16.1 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 provides objective measurement. This means the estimate of a person's ability does not depend on the items attempted, and the estimate of an item's difficulty does not depend on the particular sample of individuals used in its calibration. When a set of items is administered to two samples, and calibrated separately for each, 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 examination 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 which, 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 examination 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  $\delta_{ix}$  is the difficulty of item  $i$  when calibrated with the items on Examination X;  $\delta_{iy}$  is its difficulty on the Examination Y scale; and  $K$  is the number of items in the anchor examination.

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 cross-check item difficulty parameters periodically to insure that the equating process remains accurate.

Person ability estimates,  $\beta_v$  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 allows for the comparison of group differences even though alternate forms may have been used for each administration.

National Board examinations are scored using the Rasch model and 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 examinee to be located on the same scale of measurement as that of the base year of the examination.

The following example illustrates how common-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 examination are shown in the next two columns. The linking constant is the difference between the mean item difficulties under the two calibrations. In the example, the linking constant is -0.36. Ideally, when the link is added to the new difficulty, 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 to determine how large a difference in difficulties one may expect 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 Diff. + Link</i>	<i>Squared Error</i>
	<i>Diff.</i>	<i>S.E.<sup>†</sup></i>	<i>Diff.</i>	<i>S.E.</i>		
1	-.88	.05	-.72	.04	-1.08	.0400
2	-.74	.05	-.42	.04	-.78	.0016
3	-.62	.05	-.28	.04	-.64	.0004
4	-.15	.04	.02	.04	-.34	.0361
5	.26	.04	.61	.04	.25	.0001
6	-.18	.04	<b>.05</b>	.04	-.31	.0169
7	-1.03	.05	-.08	.04	-.44	.3481**
<i>Sum</i>	-3.34		-.82		-3.34	.4432**
<i>Mean</i>	-.48		-.12		-.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 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 -0.27. Once satisfactory equivalence between the base year and current year anchor items has been established, the next step is to adjust the difficulties of 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 were administered in the base year. Because all the item parameters are grounded in the same scale of measurement used in the base year, estimates of person ability (determined from Equation 16.1 using the WINSTEPS Rasch scaling program) will be on that scale. Assuming examinations share common content specifications, this enables any person's score 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 scores may also be directly compared from examination to examination.

### **Developing Score Conversions for Test Forms from the Item Bank**

The NBDE Part I and Part II examinations currently being administered are linear forms developed directly from the Joint Commission's item banks. For these forms, score conversions are developed using statistical methods based on the Rasch measurement model.

With regard to the estimation of candidate ability, items for the examinations are drawn from item banks according to the content requirements of the individual examination. Each new form of the examination is composed of a unique combination of items. An examination form assembled based on items selected in this way requires modifying the way converted scores are estimated (i.e., as compared to relying on an intact form). The Rasch model provides a way to establish person abilities, even when the items are not drawn from a previously used print edition.

Because the Rasch model evaluates person ability and item difficulty using the same units of measurement (i.e. logits), person ability may be estimated using the following two steps. (See *Best Test Design* by Wright and Stone, 1979, p.27 for details). First, the difficulties of the selected items are averaged. Part I and Part II item banks contain both Rasch parameters and traditional item statistics such as the percent of candidates responding correctly. The variance of the Rasch difficulties of these items is also computed.

Second, compute a Rasch ability for each possible raw score using the following formula:

$$B_v = H + (1 + w^2 / 2.89)^{1/2} \ln [r_v / (L - r_v)], \text{ in which} \quad [ 16.3 ]$$

- $B_v$  is the ability estimate in logits for candidate v
- H is the average difficulty of the items
- $w^2$  is the variance of the item difficulties
- ln is the natural logarithm of the term in brackets
- $r_v$  is the number of correct answers for candidate v
- L is the number of items in the examination

The standard error of  $B_v$  is:

$$SE(B_v) = (1 + w^2 / 2.89)^{1/2} [L / r_v(L - r_v)]^{1/2} \quad [ 16.4 ]$$

The above formula can be applied to all raw scores from 1 to (L-1). For zero and perfect scores, approximations can be applied: a "raw score" of 0.5 is substituted for zero scores (where all items were answered incorrectly), and (L-0.5) is substituted for perfect scores (where all items were answered correctly).

Once an estimate in logits has been calculated for every possible raw score, conversion tables are used to translate raw score scales to the converted score scales in use for all editions of Part I and Part II.

This approach has been successfully used with a variety of examination programs, including admission and licensure examinations. For separate forms of the examinations, raw-score to scale-score conversions have been developed. The performance of candidates across these forms was consistent with performance on previous intact forms. Comparable performance across forms serves as a source of evidence to support the use of this approach for scale score development.

## **17. Scoring and Reporting Test Results**

Standards pertaining to scoring and reporting of examination 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 refers to responsible transmission of scores. Standards 6.14 and 6.15 refer to record keeping.

**Table 17.1**  
**Standards Pertaining to Scoring and Reporting of Test Results**

---

*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 Examinations**

Procedures for scoring examinations are presented in the *Operational and Policy Manual of the JCNDE*, referenced previously. Quality control procedures are in place to facilitate accurate scoring. Each candidate's raw and scale scores 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 Reports**

Candidate score reporting is more fully discussed in the *Operational and Policy Manual of the JCNDE*. Factors that affect a candidate's score include the number of correct answers selected by the candidate and the score scale conversion for the examination form.

The score scale and minimum passing score are determined by a standard-setting

committee using a criterion-referenced method. The minimum passing score identified by the committee is assigned a scale score of 75. Scale scores range from 49 to 99. Under some circumstances, a zero is reported. A score below 75 is considered a failing score and does not earn National Board credit. Part I and Part II examination responses are audited for accuracy before score reports are distributed.

As noted previously, in 2012 the Joint Commission moved to pass/fail reporting of results for candidates who passed the examinations. Candidates who fail NBDE Part I receive a score report that contains the following information: the candidate's comprehensive scale score, the number of items on the examination, the number of items the candidate answered correctly, and the national means for the four disciplines. Candidates who fail NBDE Part II receive a score report that contains their comprehensive scale score in addition to performance information for each discipline. The discipline subscore information reported to failing NBDE Part II candidates is represented graphically and placed on a common measurement scale so that performance on different disciplines can be meaningfully and visually compared. This allows failing candidates to assess their relative performance in the different disciplines and identify disciplines where they are most in need of remediation.

It also should be noted score reporting for examinations occurring prior to 2012 remains unchanged. Scale scores will continue to be reported for these administrations.

## **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 National Board Dental Examination candidates. Because so much is at stake in taking these examinations, the Joint Commission should ensure candidates for licensure receive 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 examination information be made available to all candidates. Generally, a candidate Guide or webpage is the most suitable way to accomplish this. Standard 8.9 refers to cheating, and standard 8.12 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.*

---

## **Guides for National Board Dental Examinations**

To help satisfy the standards appearing in Table 18.1, the Joint Commission publishes annual examination guides for the National Board Dental Examinations. These documents provide detailed information related to the Joint Commission's examination policies, the format and content of the examination, eligibility requirements, examination regulations, the appeal process, examination scoring, and examples of item formats. Each year the guides are updated and amended as necessary. The guides are available through the Joint Commission's website: <http://www.ada.org/en/jcnde/examinations/nb-guides>

### **19. Threats to Validity**

According to Messick (1989), two major threats to validity are construct-irrelevant variance (CIV) and construct under representation (CUR). This next 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 examination scores. Haladyna (2002) identifies many sources of CIV, including nonequivalent examination forms, cheating on an examination, improper examination preparation, errors in scoring examination results, and faulty items.

The Joint Commission periodically releases editions of National Board examinations or collections of items to familiarize candidates with National Board item formats. However, the Joint Commission recommends candidates use textbooks and lecture notes as their primary sources of study material. Released dental examinations are available in most dental school libraries and the ADA's library, which is located at the ADA headquarters. In addition, copies of released examinations can be purchased from the American Student Dental Association. The Joint Commission discourages superficial learning as a basis for examination preparation.

The Joint Commission does not discriminate based on race, color, religion, gender, age, sex, national origin, disability, sexual orientation, or marital status. If performance on examination items inappropriately reflects these factors as opposed to the focal construct of interest, CIV is present.

#### **Construct Under Representation**

Another threat to validity is construct underrepresentation. This occurs when an examination does not adequately represent the domain of knowledge intended. This bias leads to inadequate construct coverage, and can cast doubt on the meaning of an examination score and its legitimacy in making a pass/fail decision. The procedures used to define the domain of knowledge to be tested and determine the examination specifications go quite far in assuring the public and the dental community the NBDE Part I and NBDE Part II examinations do not underrepresent biomedical science and professional knowledge

deemed essential for entry-level dentists.

## 20. Validity Studies

Studies are undertaken to investigate significant threats to validity and provide new sources of validity evidence, which can strengthen the argument for using examination results to inform licensure decisions. Validity studies of various types and scope are described below.

Practice analyses can be effective for updating examination specifications and ensuring examination content is current (Kramer & Neumann, 2003; Tsai, Yang, Waldschmidt, & Chang, 2012), while standard-setting studies are conducted to confirm the standard (i.e., passing score) that separates passing and failing candidates.

Other studies, which examine the content and content structure of NBDE exams, are used to confirm the content-related validity of the examinations. Kramer and DeMarais (1992) have confirmed that the NBDE are unidimensional. This unidimensionality is essential because the measurement model (e.g., Rasch) used for constructing and scoring the National Board Dental Examinations assumes that these examinations are unidimensional.

## 21. Security

Breakdowns in examination security can threaten validity. Table 21.1 provides a list of security standards. The Joint Commission has policies and procedures in place to provide for 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 test developers, test administrators, and examination users. Examination security is



maintained through test development and test administration procedures in a variety of ways. Policies of the Department of Testing Services address issues related to examination security and are reviewed regularly by the Joint Commission and its staff.

**Security Audit:** In 2008, Caveon Test Security, an independent organization, conducted a security audit of the Department of Testing Services, which is the department within the ADA 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 certain materials as secure. These include the following:

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 for assignment 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: Test 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 review 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 conflicts of interest. Volunteers are prohibited from releasing information about 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 Test 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:** Examinations are administered by Prometric at its nationwide, professional level testing centers. The National Board Dental Examination Guides describe procedures for identification of candidates, including requirements for positive identification through biometrics. Candidates' conduct 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 (AERA) is the largest organization in the world devoted to the scientific study of education. In 2000, it issued a brief publication of guidelines for designing and using examinations in high-stakes educational settings. Some guidelines are also appropriate for the Joint Commission's NBDE programs. This section presents selected guidelines and provides a brief discussion of each guideline for the NBDE Part I and NBDE Part II examinations.

### **Protection Against High-Stakes Decisions Based on a Single Test**

Can a single examination prevent a candidate from practicing dentistry after other criteria for licensure are met? The National Board Dental Examinations provide repeated opportunities for candidates to prepare for and pass these examinations. In addition, licensure decisions are based on many other criteria. Since public welfare and safety are at issue, state boards bear a heavy responsibility in using this examination information appropriately with other information for making licensing decisions.

### **Adequate Resources and Opportunity to Learn**

The Joint Commission assumes no responsibility for dentists' educational preparation. This task falls to dental schools and the students themselves. Failure to acquire basic scientific and professional knowledge can lead to a candidate failing the Part I or Part II examinations. 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 follows this guideline, as discussed in this technical report.

## **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, examination 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 Test and the Curriculum**

It is the responsibility of dental schools to align student learning with the knowledge, skills, and abilities that national practice analyses have determined represent the core knowledge required of practicing dentists. NBDE content is aligned with the core knowledge of practicing dentists, which serves as the source for curriculum development.

## **Validity of Passing Scores and Achievement Levels**

The Joint Commission determines its passing scores using methodology consistent with the *Standards for Educational and Psychological Testing* (AERA, APA, NCME, 2014).

## **Opportunities for Meaningful Remediation for Candidates Who Fail High-Stakes Tests**

The Joint Commission bears no responsibility for remediation, but dental schools may choose to provide remediation if a candidate fails. The Joint Commission provides a list of reference materials for candidates but does not endorse any specific review courses. The Joint Commission has updated its score reporting to assist failing candidates in remediation efforts.

## **Appropriate Attention to Language Differences Among Examinees**

The NBDE is written in English. To the degree that test takers do not possess the prerequisite English reading skills, their examination results may reflect construct-irrelevant variance due to language deficiency.

## **Appropriate Attention to Candidates with Disabilities**

In examining 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 have received testing accommodations for an examination.

## **Sufficient Reliability for Each Intended Use**

Reliability refers to the consistency or precision of examination scores. Scores reported for individuals or schools must be shown to be sufficiently free from error to support each

intended interpretation. Accuracy should be examined for results as they are actually used. For example, information about the reliability of raw scores may not adequately describe percentiles. This technical report provides solid evidence regarding the adequacy of reliability estimates.

## References

- American Dental Education Association (2001). The competencies of the new dentist. *Journal of Dental Education*, 65(7) 659-661.
- 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. 89-108). 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, American Dental Association. (November 1995). *Guidelines for Committee Revision of Weak or Unacceptable Items for National Board Dental Examinations*. Chicago, IL: Author.
- Joint Commission on National Dental Examinations, American Dental Association. (2020). *National Board Dental Examination Part I Guide*. Chicago: Author.
- Joint Commission on National Dental Examinations, American Dental Association. (2020). *National Board Dental Examination Part II Guide*. Chicago: Author.
- Joint Commission on National Dental Examinations, American Dental Association. (2020). *Orientation Manual for Dental Test Constructors*. Chicago: Author.
- Joint Commission on National Dental Examinations, American Dental Association. (2020). *Test Item Development Guide: National Board Dental Examinations*. Chicago: Author.
- Joint Commission on National Dental Examinations, American Dental Association. (2020). *JCNDE Test Construction Teams and Selection Criteria*. Chicago: Author.
- Joint Commission on National Dental Examinations, American Dental Association. (2020). *Procedures for test administration*. Chicago: Author.
- Joint Commission on National Dental Examinations, American Dental Association. (2020). *Rules of the JCNDE*. Chicago: Author.

- Joint Commission on National Dental Examinations, American Dental Association. (2020). *Operational and Policy Manual of the JCNDE*. Chicago: Author.
- 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. (1992). Reliability and validity of the pilot national dental examination. *Journal of Dental Education*, 56(4), 236-241.
- Kramer, G.A., & Neumann, L.M. (2003). Confirming the validity of Part II of the National Board Dental Examinations: A practice analysis. *Journal of Dental Education*, 67(12), 1286-1298.
- 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.
- Stone, G.E. (2004). Objective standard setting (or truth in advertising). In E. V. Smith, JR. & R. M. Smith (Eds.). *Introduction to Rasch measurement: Theory, models and applications* (pp.445-459). Maple Grove, MN: JAM Press.
- Tsai, T-H., Yang, C-L., Waldschmidt, D.M., & Chang, S-R. (2012, February). *2012 National Board Dental Examination Part II Practice Analysis (TR-02-12)*. American Dental Association, Chicago, IL.
- 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). 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.

Yang, C.-L., Neumann, L.M., & Kramer, G.A. (2012). Assessing Context Effects on Test Validity of the National Board Dental Examination Part I. *Journal of Dental Education*, 76 (4), 395-406.

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

### NBDE Part I Examination Specifications: 2020

The National Board Dental Examinations are administered in two parts. The comprehensive NBDE Part I consists of 400 test items. For each discipline, approximately 80% of the items are intermingled, discipline-based and approximately 20% are interdisciplinary testlet-based items. A testlet consists of the patient scenario and a set of items from the various disciplines associated with the scenario. Test items for the comprehensive Part I are drawn from the following disciplines:

1. Anatomic Sciences
2. Biochemistry and Physiology
3. Microbiology and Pathology
4. Dental Anatomy and Occlusion

One item from each of the disciplines listed above will be designated for the testlets under the topic, "Professional Ethics and Patient Management." These items will require a basic understanding of professional ethical principles in patient management.

#### ANATOMIC SCIENCES [100]

- 1.0. Gross Anatomy\*
- 2.0. Histology
- 3.0. Oral Histology
- 4.0. Developmental Biology
- 5.0. Professional Ethics/Patient Management

---

\* The following topics will be considered under each category of gross anatomy: bone; muscles; fascia, nerves (peripheral and autonomic); arteries, veins, and lymphatics; spaces and cavities; joints and ligaments; and endocrines and exocrines

---

#### BIOCHEMISTRY-PHYSIOLOGY [100]

- 1.0. Biological Compounds
- 2.0. Metabolism
- 3.0. Molecular and Cellular Biology
- 4.0. Connective Tissues
- 5.0. Membranes
- 6.0. Nervous System



## **BIOCHEMISTRY and PHYSIOLOGY [100] (continued)**

- 7.0. Muscle
- 8.0. Circulation
- 9.0. Respiration
- 10.0. Renal
- 11.0. Oral Physiology
- 12.0. Digestion
- 13.0. Endocrines
- 14.0. Professional Ethics and Patient Management

## **MICROBIOLOGY and PATHOLOGY [100]**

- 1.0. General Microbiology
- 2.0. Reactions of Tissue to Injury
- 3.0. Immunology and Immunopathology
- 4.0. Microbiology, Immunology, and Pathology of Specific Infectious Diseases
- 5.0. Systemic Pathology
- 6.0. Growth Disturbances
- 7.0. Professional Ethics and Patient Management

## **DENTAL ANATOMY AND OCCLUSION [100]**

- 1.0. Tooth Morphology
- 2.0. Pulp Cavity Morphology
- 3.0. Calcification and Eruption
- 4.0. Principles of Occlusion and Function
- 5.0. Clinical Considerations—Tooth Morphology and Anomalies
- 6.0. Professional Ethics and Patient Management

## Appendix B

### NBDE Part II Examination Specifications: 2020

#### EXAMINATION SPECIFICATIONS

The NBDE Part II is a comprehensive examination consisting of 500 items. For each discipline, approximately 80% of the items are stand-alone, while approximately 20% are interdisciplinary and case-based. A case consists of a patient scenario, patient history, and a set of discipline based items relevant to the scenario. NBDE items are developed by test construction teams composed of subject-matter experts in accordance with examination specifications approved by the JCNDE.

The Universal/National System for tooth notation that has been adopted by the American Dental Association is used on all National Board Examinations. This system is a sequential tooth numbering system, designating the permanent dentition (numbers 1-32), and the primary dentition (letters A-T).

In the June 2018 issue of the Journal of Periodontology the American Academy of Periodontology (AAP) published updates to its Classification of Periodontal and Peri-Implant Diseases and Conditions. Candidates completing examinations published through the Department of Testing Services—including those overseen by the Joint Commission on National Dental Examinations or by the American Dental Association—continue to be based on the prior (i.e., 1999) AAP disease classification. Candidates will be notified when examinations are updated to reflect the new terminology.

#### Discipline-Based Component (400 items)

The exam items that comprise the discipline-based component are derived from the following disciplines:

##### Endodontics\* (36 items)

- Case Clinical Diagnosis, Case Selection, Treatment Planning, and Patient Management
- Basic Endodontic Treatment Procedures
- Procedural Complications
- Traumatic Injuries
- Adjunctive Endodontic Therapy
- Post-Treatment Evaluation

\* The American Association of Endodontists' Glossary of Endodontic Terms is used in reference to endodontic pathoses. The endodontics diagnostic terminology adopted by AAE as described in the December 2009 Journal of Endodontics [Volume 35, Number 12, p.1634], is incorporated in the NBDE Part II.

##### Operative Dentistry (44 items)

- Dental Caries
- Examination, Diagnosis, & Treatment Planning
- General Operative Procedures
- Preparation of Cavities
- Restoration of Prepared Cavities

##### Oral And Maxillofacial Surgery/Pain Control (52 items)

- Surgery
- Anxiety and Pain Control

- Medical Assessment and Emergency Care
- Treatment Plan
- Diagnosis

**Oral Diagnosis (42 items)**

- Oral Pathology
- Oral Radiology

**Orthodontics/Pediatric Dentistry (44 items)**

- Individual Tooth Pathology
- Supporting Tissue Pathology
- Dentofacial Variations
- Behavior
- Systemic Pathology

**Patient Management (53 items)**

- Communication and Interpersonal Skills
- Anxiety and Pain Control
- Health Behavior Change
- Disabled and Medically Compromised
- Epidemiology
- Prevention of Oral Diseases
- Evaluation of Dental Literature
- Infection Control
- Materials and Equipment Safety
- Professional Responsibility/Liability
- Practice Management

**Periodontics (46 items)**

- Diagnosis
- Etiology
- Pathogenesis
- Treatment Planning
- Prognosis
- Therapy
- Prevention and Maintenance

**Pharmacology (35 items)**

- General Principles
- Central Nervous System
- Autonomic
- Cardiovascular
- Local Anesthetics
- Chemotherapy
- Endocrines/Immunosuppressants
- Analgesics
- Antihistamines and Autocoids

**Prosthodontics (48 items)**

- General Considerations
- Complete and Removable Partial Denture Prosthodontics
- Fixed Partial Prosthodontics

### **Case-Based Component (100 items)**

The case-based component of the NBDE Part II presents items dealing with actual patients. The patient cases are developed to include the following approximate distribution: adult patients – 70%; child patients – 30%. A minimum of 15% of case-based exam questions will address the medical management of compromised adults and children. A compromised patient is defined as a person whose health status requires modification of standard treatment. Each case presentation in the examination consists of:

1. synopsis of a patient's health and social histories,
2. patient dental charting,
3. diagnostic radiographs, and
4. clinical photographs of the patient (when necessary).

Each case contains from 10 to 15 questions about various aspects of the patient's dental care. These questions-- totaling 100 across all cases-- might derive from any of the biomedical sciences and clinical disciplines, including Patient Management. The proportion stemming from any particular discipline depends upon the nature of the case itself. For example, the case of an elderly adult might be based upon Maxillofacial Surgery/Pain Control, Prosthodontics, and Operative Dentistry, whereas a child's case might derive from Orthodontics, Pediatric Dentistry, and Patient Management. In responding to case-based items, the candidate must:

1. interpret the findings and information provided.
2. identify problems and make diagnoses.
3. select materials, technique, and armamentarium.
4. apply treatment.
5. evaluate progress and complications.
6. establish procedures for prevention and maintenance.

## Appendix C

### Examination Summary Statistics

The tables that follow provide information related to the quality of the NBDE Part I and Part II examinations. The terms used in the tables are described below.

**Reference Group:** The reference group is comprised of all students enrolled in schools with approved accreditation status who took the examination for the first time. This group does not include graduates. This reference group's performance establishes standards for all candidates who will take 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 number or percentage of test items answered correctly by candidates in the reference group.

**Reliability ( $KR_{20}$ ):**  $KR_{20}$  is a measure of internal consistency reliability for items scored dichotomously. Perfect score reliability would produce a reliability coefficient of 1.0, but no set of scores is perfectly reliable. The higher the coefficient, the more reliable the scores.

**Table C.1  
NBDE Part I Statistics\***

Total Number of Candidates in the Reference Group	2,586
Scale Score Mean	81.80
Scale Score Standard Deviation	5.91
Mean Score (%)	71.55
Reliability KR20 (Range)	.94 - .95

\* NBDE Part I has been a comprehensive examination since 2007 and is completely computerized. The statistics reported in this table reflect the aggregated results for all Part I forms administered in 2020.

**Table C.2  
NBDE Part II Statistics\***

Total Number of Candidates in the Reference Group	6,183
Scale Score Mean	81.1
Scale Score Standard Deviation	4.79
Mean Score (%)	76.0
Reliability KR <sub>20</sub> (Range)	.88 - .92

\* NBDE Part II administrations have been completely computerized since 2006. The statistics reported in this table reflect the aggregated results for all Part II forms administered in 2020.

## Appendix D

### Trends in Number of Test-Takers and Failure Rates

Tables D.1 and D.2 on the following pages present the numbers and failure rates for first-time and repeating candidates taking the NBDE Part I and NBDE Part II from accredited and non-accredited dental schools during the 10-year period beginning with 2011. The numbers include both current students and graduates.

**Table D.1**  
**Numbers and Failure Rates for First-time and Repeating Candidates**  
**NBDE Part I**

		<b>Accredited</b>				<b>Non-Accredited</b>				<b>Total</b>	
<b>Year</b>	<b>First-time</b>		<b>Repeating</b>		<b>First-time</b>		<b>Repeating</b>		<b>First-time and Repeating</b>		
	<b>Number</b>	<b>% Failing</b>	<b>Number</b>	<b>% Failing</b>	<b>Number</b>	<b>% Failing</b>	<b>Number</b>	<b>% Failing</b>	<b>Number</b>	<b>% Failing</b>	
2011	5,068	4.5	396	33.6	1,713	32.2	921	62.2	8,098	18.3	
2012	5,497	6.1	344	39.2	1,721	38.3	842	68.1	8,404	20.3	
2013	5,574	6.3	504	30.6	1,912	40.0	944	63.1	8,934	20.0	
2014	6,041	3.7	337	26.3	2,211	31.9	988	56.4	9,617	16.5	
2015	6,092	3.4	308	28.6	2,329	33.4	939	57.6	9,668	16.7	
2016*	6,260	5.2	340	33.5	2,351	33.0	1,022	59.1	9,973	18.2	
2017	5,995	10.6	669	33.5	2,289	37.2	1,044	67.2	9,997	24.1	
2018	6,180	12.1	819	39.7	2,226	44.3	1,036	70.1	10,261	27.1	
2019	5,432	10.6	972	35.3	2,372	48.6	1,409	66.7	10,185	29.6	
2020	2,616	10.6	650	35.1	2,167	47.1	1,474	61.5	6,907	35.2	

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



**Table D.2**  
**Numbers and Failure Rates for First-time and Repeating Candidates**  
**NBDE Part II**

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	5,312	5.1	395	28.9	1,050	29.6	471	48.4	7,228	12.8
2012	4,803	5.6	363	29.2	1,216	31.3	410	49.5	6,792	14.1
2013	5,338	6.5	465	22.5	1,190	40.0	513	52.8	7,506	15.3
2014	5,704	7.4	543	21.4	1,557	37.3	593	45.2	8,397	16.5
2015	5,834	7.5	604	22.7	1,630	42.0	783	48.8	8,851	18.5
2016	6,034	8.7	682	24.1	1,861	34.2	913	45.0	9,490	18.3
2017*	6,138	8.3	712	23.9	1,698	34.4	879	45.3	9,427	17.6
2018	5,769	7.9	670	23.4	1,759	23.7	766	39.4	8,964	14.8
2019	5,985	9.7	653	20.1	1,562	23.3	605	47.4	8,805	15.5
2020	6,227	7.4	673	21.8	1,206	26.4	511	41.3	8,617	13.2

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

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