American Board of Dental Examiners (ADEX)
Clinical Licensure Examinations in Dentistry

Technical Report Summary

July 11, 2017
Introduction

Clinical examination programs serve a critical role in the dental licensure process as independent evidence of minimum competency. This evidence is necessary for state boards of dentistry to be able to support decisions about candidates who seek admission to practice within a given jurisdiction. This decision may be based on multiple factors (e.g., graduation from an accredited training program, successful completion of clinical examinations of judgment and skill, background check) and does not rely on any one as the sole source of the licensure decision.

The primary function of the licensure process is to ensure the likelihood that a licensed individual can safely treat the public that they serve. An examination selected and used by a licensure board must validly identify those individuals who may be a threat to patient safety in their professional practice. All tests sample from the broader domain of knowledge, skills, and abilities (KSAs) that may be important to the profession. Within licensure, those KSAs are targeted to those that are necessary for entry-level, minimally competent practice. Licensing boards that use examination scores rely on a collection of evidence to support the program’s intended purpose. This validation process begins with an argument, a claim for validity and supporting evidence for using such test scores to make pass-fail licensure decisions. Multiple indicators supported by theory, policy, and practice are used to make the licensure decisions in dentistry. Candidates’ performance on ADEX’s clinical licensure examinations represents one component in the decision process.

Measurement experts recommend that testing programs undergo regular, independent evaluation of their programs against professional testing standards to monitor the quality (Buckendahl & Plake, 2006; Downing & Haladyna, 1996; Madaus, 1992). ADEX uses multiple approaches for evaluating its examination program both internally and externally. The organization receives feedback internally from an examination and quality assurance committees that is charged with ensuring that evidence based decisions are made with respect to content of the program, administration, psychometric characteristics, and ongoing program maintenance. ADEX has also called upon external measurement experts (e.g., Dr. Stephen Klein, Dr. Chad Buckendahl, Dr. Susan-Davis-Becker) to provide consultation on various psychometric concerns. This technical manual serves as documentation of the processes and publicly available results of ADEX’s development, data analysis, and maintenance activities for its clinical examinations in dentistry.

Because all readers may not be interested in the more detailed sections of this report, we provided the next section to brief respond to questions about how ADEX’s examinations align with the Guidance for Clinical Licensure Examinations in Dentistry (American Association of Dental Examiners [AADE], 2003).

[APA], & National Council on Measurement in Education [NCME], 2014). The Standards are the expectations promulgated by the broader testing community.

**Guideline 1: The purposes, interpretations and uses of the clinical examinations are clearly stated in order to make appropriate pass-fail decisions.**

The purpose of ADEX’s examinations is to support pass-fail licensure decisions regarding the clinical judgments and psychomotor skills for dentists. Diagnostic information that may be used for program evaluation or as evidence of outcomes for accreditation purposes are not within the intended scope of these examinations.

**Guideline 2: The knowledge, skills and abilities that are important in the clinical practice of dentistry or dental hygiene are identified.**

ADEX periodically conducts job/practice analysis surveys of practicing dentists (See Knapp & Knapp, 1995 for a description of practice analysis) to define the knowledge, skills, and abilities needed for entry-level licensure. These are typically conducted nationally every 5-7 years. The most recent study was conducted in 2011 with the next study beginning in 2017. Technical reports for these studies are proprietary with publicly available content outlines communicated to candidates.

**Guideline 3: Examination specifications are developed to provide a detailed description of the content of the examination and specify the scored tasks that are used to evaluate each discipline. The specifications should include scoring weights associated with each content area.**

ADEX’s clinical examinations in dentistry are developed from the results of its periodic practice analysis. The content specifications derived from the practice analysis are directly linked to the respective examination (e.g., diagnosis and treatment planning, restorative, fixed prosthodontics, endodontics). Because each discipline is within the scope of a licensed dentist, ADEX considers these as equally important to the licensure decision. Therefore, each of the examinations is considered as a separate decision that must be supported by evidence.

**Guideline 4: Scored tasks and scoring criteria are developed according to the examination specifications.**

Items and tasks represented on the examination are directly linked to the results of the practice analysis. Items on the clinical judgment examination are scored dichotomously (i.e., right or wrong). For tasks within the clinical skills examinations, the scoring criteria are defined for each task and a dichotomous decision is made for each subtask nested within a discipline-specific task.
Performance expectations are developed by subject matter experts on the exam committee using published, supporting literature.

**Guideline 5: Policies and procedures are defined and published to standardized examination administration. This administrative protocol addresses legal issues, fair testing practices and patient protection.**

ADEX’s examiner administration manuals are published by the agencies that administer its examinations and are consistent with professional expectations for fair testing practices. Examination administration for the clinical judgment examination occurs at Prometric testing centers. For exam administration of the clinical skills examinations, staff members are available at each exam site to ensure that procedures are followed. Individuals serving in these administrative roles are trained prior to the beginning of each examination cycle to assist in monitoring the procedural requirements of the examination.

**Guideline 6: The examining agency provides candidates with clear and comprehensive information about the examination program, including application requirements, examination content, performance expectations, reporting of results, and an appeals process.**

ADEX’s candidate manuals are also produced by the agencies that administer its examinations and are modeled after the examiner manual. The candidate manual is intended to provide transparent information about the examination process. Specifically, the manual provides candidates with the relevant information they need for the examination. Eligibility, registration, content, scoring criteria, expected performance, score reports, and formal appeals procedures are also documented in the candidate manual.

**Guideline 7: The rationale and justification for the use of conjunctive or compensatory scoring are documented.**

ADEX’s rationale for a conjunctive decision rule across disciplines in dentistry is based on both judgmental and empirical evidence. The dental profession recognizes that there are distinct disciplines within its profession. This acknowledgement is observed in the organization of departments within dental schools in addition to specialty certifications within the profession. These distinctions suggest that there are separate, unique skills, each of which are important for an entry-level dentist who will see patients in practice having a range of potential needs.

Empirically, the decision to use a conjunctive decision rule that requires demonstrations of minimum competency within each discipline is supported by observed correlations between different sections of the examination. For example, candidates that perform well in restorative do not necessarily perform well in endodontics. The lack of a strong statistical relationship suggests that these
disciplines may different constructs or dimensions of the profession. Although the frequency of performance of procedures within a discipline will differ, it does not diminish a state board’s need to ensure that when a practitioner is called upon to perform those skills, they are at least minimally competent to do so.

Combining these examinations into a single score creates problems both theoretically and psychometrically. Theoretically, if the disciplines are distinct, then by combining scores, it creates a confound where a potential user would not be able to interpret the representation of the construct. Psychometrically, this creates a problem of multidimensionality where the measurement and interpretation of scores are not derived from a single, well-defined construct.

**Guideline 8: Passing score studies are designed, performed and documented.**

ADEX used modifications of the Angoff (1971) method for establishing the passing scores for its examinations. The Angoff methodology is a commonly used test-based approach to recommending passing scores that considers the ease or difficulty of the components of the tasks in the decision process. In implementing this study, panels of subject matter experts make judgments about items or tasks referencing the definition of minimum competency to guide their decisions.

**Guideline 9: Policies for examiner selection and retention are defined and published.**

ADEX’s examiner policies and procedures are published in the examiner administration manual. Data from examiners’ performance and examiner policies are reviewed dynamically during administration to provide formative feedback and annually to make summative determinations of performance. The evaluation committee uses data from the previous examination cycle to evaluate both policies for examiner selection, assignment, and retention; and individual examiner performance.

**Guideline 10: An examiner training program is established and implemented. The program introduces examiners to appropriate applications of the agency’s scoring criteria and assesses their ability to apply the criteria. The methodology of examiner standardization and its results are documented.**

ADEX requires examiners to complete training and calibration prior to serving in operational scoring situations. The training approach is consistent with common practice in training raters for scoring performance tasks. Specifically, examiners receive information about the scoring criteria and procedures. Examiners are then required to pass a qualification test to independently demonstrate their understanding and application of the scoring criteria prior to operational scoring.
Guideline 11: Post-examination analyses are routinely conducted. Reliability and other factors affecting validity are investigated.

ADEX conducts psychometric analysis to evaluate the validity of decisions made on test scores. First, decision consistency analyses on the clinical skills exams are conducted to determine the level of agreement across independent raters. Because rater error represents the greatest potential source of variance in the candidates’ scores on a performance task, these estimates of reliability provide evidence to support ADEX’s confidence in the pass-fail decisions. These data also inform the quantitative analyses of individual examiner performance. Second, internal consistency analyses, decision consistency, and item analyses are conducted for the DSE examination, a computerized objectively structured clinical exam (OSCE) of clinical judgments.

Guideline 12: A program is developed and implemented for ongoing examination of examiner ratings. The examining agency provides examiners with feedback on their individual rating performance. Policies and procedures are defined for remediation or discontinuance of examiners based on analysis of their performance.

ADEX provides onsite feedback to examiners regarding their performance. The committee that evaluates examiner performance has established policies for determining continuance, remediation, or removal for individual examiners.

Guideline 13: A technical report is prepared to summarize the year’s results and activities and to recommend improvements for future examinations. Results of analyses and studies are reported to a board that is empowered to set policy and authorize revisions to the examination. Results are published as appropriate. Qualified measurement specialists, external to the examining agency, should review the examination program periodically.

ADEX produces annual technical reports and periodic validation or evaluation studies that respond to ongoing questions for the program. ADEX has also used external measurement experts such as Dr. Stephen Klein, Dr. Chad Buckendahl, and Dr. Susan Davis-Becker to observe, advise, and comment on their procedures and practices since its inception.

Guideline 14: What procedures are in place to ensure that the same examination is given from site to site? Are the same protocols followed at every exam administration?

The ADEX candidate and examiner administration manuals document the requirements at each site. Training for the administrative and leadership staff, examiners, and candidates also help to ensure consistency across sites. The logistics of some exam sites necessitate some differences in location or flow of
the examination, but these do not alter the exam conditions for the candidate or the constructs being measured.

**Guideline 15: What are the examiner calibration procedures? What exercises are required? How often are examiners calibrated?**

ADEX conducts training for onsite leadership staff prior to operational examinations. The training procedures emphasize conceptual understanding of the rating criteria, discussion of the criteria, application of the criteria in practice, and administrative procedures. Following the training, examiners must pass a qualification test prior to operational scoring.

The next section of the manual describes the specific activities that ADEX undergoes in their development, validation, and maintenance of its clinical testing program.

**ADEX Test Development**

**I. Validity**

The purpose of ADEX’s examinations is to support licensure decisions about the clinical competence of entry-level candidates in dentistry. These licensure decisions are reported as pass/fail and do not differentiate across different degrees of these decisions. It is important for users to understand that a licensure testing program is not directly intended to provide information for programmatic or curricular improvement for accreditation purposes; these additional uses are beyond the scope of the information gathered in the exam.

Validation studies have been conducted to gather supporting evidence for the ADEX testing program and how it relates to its intended purpose. Licensure testing programs necessarily focus validation efforts on the content representation of their examinations. The content for ADEX’s dental examinations is based on a content specification methodology often called an occupational, job, or practice analysis (Knapp & Knapp, 1995). ADEX conducts occupational analysis surveys of licensed, practicing dentists to define the knowledge, skills, and abilities (KSAs) that may be needed for entry-level practice. Through this periodic study, the KSAs that are important in the clinical practice of dentistry are identified and prioritized. The occupational analysis surveys are conducted nationally every 5 to 7 years. Results of the occupational analysis provide empirical evidence that is used to support a test developer’s decision about the content that will be on their examination. The professional judgment of the subject matter experts that serve on ADEX’s examination committee is also relevant because they help to provide interpretations of how to measure the skills identified in the job analysis.

The validation framework for subsequent studies includes many concepts, principles, and procedures of test development. These include but are not limited to the *Standards for Educational and Psychological Testing* (AERA, APA, & NCME, 2014) and the *Guidance for Clinical Licensure Examinations in Dentistry* (AADE, 2003). The referenced guidelines
contribute to the organization’s overall judgment of validity. ADEX evaluates the quality of its program against these guidelines.

ADEX has engaged in a systematic decision process to support their policy decisions. In all testing programs there is an intersection of policy and measurement science (i.e., psychometrics) that often requires priorities to be set based on those considerations. ADEX’s decisions are based on professional judgment supported by empirical evidence from multiple sources.

II. Occupational analysis

An occupational analysis survey was most recently conducted by ADEX in 2011. The results of the survey were used to identify the categories of most frequent use and most importance as judged by the surveyed dental professionals. Because the primary goal of the survey was to collect information that would generalize nationally, practitioners from across the country were sampled. The sample included respondents with varying degrees of experience. Periodic updates of the occupational analysis are conducted to reflect current knowledge, skills, and abilities of a profession in the test specification of the examination program. The full technical report for the occupational analysis is proprietary. ADEX is currently beginning the process of conducting its next occupational analysis in fall 2017.

III. Exam Development

Use of subject matter experts (SMEs)

ADEX uses SMEs in multiple phases of the examination development process – occupational analysis, task development and review, scoring criteria development and review, operational examiners, and standard setting panelists. These SMEs include educators and practitioners.

Exam Specifications and Decision Rules

Examination specifications are developed to describe the content of the examination and specify the scored tasks that are used to evaluate each domain (e.g., diagnosis and treatment planning, prosthodontics, endodontics). ADEX’s dental examination is weighted based on a decision to consider each domain as equally important to the licensure decision (see also, Fortune & Cromack, 1995).

The rationale for using a conjunctive decision rule is based on both judgmental and empirical evidence. The dental profession, from a judgmental perspective recognizes that there are distinct domains within its profession. The philosophy is further supported through different departments within dental schools as well as different advanced certifications and associations within the profession itself. These distinctions suggest that there are separate, unique skill sets that comprise the profession, each of which are equally important for a minimally competent, entry level dentist.
Empirically, the decision to use a conjunctive decision rule that requires demonstrations of minimum competency within each discipline is based on low, observed correlations and pass-fail decisions between different sections of the examination. Meaning, candidates who fail to pass the operative examination are not necessarily the same ones who fail the prosthodontics examination. The historical lack of a meaningful statistical relationship suggests that these domains represent different constructs or dimension of the profession. Although ADEX recognizes that the population characteristics for licensed dentists may differ in terms of the number of procedures they perform, it does not diminish a state board’s need to have a reasonable level of confidence that when they are called upon to perform those skills, they are at least minimally capable of doing so.

ADEX’s examiner and candidate manuals represent the organization’s published policies and procedures regarding standardization of the examination administration. This administrative protocol addresses fair testing practices and patient protection. The administration (Examiners) manual published by ADEX is consistent with fair testing practices and staff members are available at each exam site to ensure that procedures are followed.

**Task and item development**

Task and item development and review relates to task or item writing, scoring criteria development, pilot testing, and analyses of these pilot data to improve the quality of the tests. These activities involve the use of subject matter experts drafting and editing tasks, items, and scoring criteria that are then reviewed.

Tasks represented on the examination are directly linked to the results of the job analysis. The scoring criteria for minimum competency are defined for each task and dichotomous decisions (pass or fail) are made for each sub-task subsumed within a given procedure. These criteria are documented in both the examiners’ and candidates’ manuals. Because decisions are made for each examination, it is important to have sufficient measurement opportunities for candidates to demonstrate their abilities. Subkoviak (1984) suggests that a minimum of 6 to 8 score points are necessary to make dichotomous decisions (i.e. pass-fail) with an acceptable level of decision consistency reliability. Recognizing this need, each of the task-based procedures of the clinical skills examinations contain sufficient observable behaviors to support breadth of coverage of the respective domain.

Post-examination analyses are conducted at multiple levels: dynamically during the examination for formative purposes and annually to inform summative decisions. Analyses conducted during examinations are used to provide formative feedback to examiners on their performance (e.g., inter-rater agreement, rater performance relative to other examiners assigned to the site, decision consistency). Annual data analyses provide both formative and summative about the level of inter-rater agreement/decision consistency, individual rater performance, and psychometric analyses of the DSE (i.e., computerized OSCE). These analyses provide information about the confidence the organization has in the pass-fail decisions for candidates.

ADEX has a history of maintaining a deliberate process for exam development and implementation. This process includes discussing the feasibility of a concept, providing adequate
notice to stakeholder groups, collecting field (pilot) test information, and evaluating that information prior to implementation. This strategy allows ADEX to collect the necessary validity evidence to support or refute a given concept. It also reinforces the organization’s commitment to evaluating proposals that seek to improve or change their testing program. This policy assists the organization respond purposefully when confronted with the clinical licensure testing community.

**Passing Scores**

ADEX used modifications of the modified Angoff (1971) method to establish passing scores/performance on both the DSE and the clinical skills examinations. The Angoff method involves panelists with content knowledge and familiarity with the target candidate (i.e., the minimally competent candidate) making judgments on each item or scoring element in the examination. These judgments are about the likely performance of the target candidate on each item (e.g., with dichotomous decisions, proportional estimates). The cut scores have remained consistent for the examinations through periodic review of the policy definition of minimum competency as applied to the respective examination. For the DSE, an additional step occurs to ensure the consistency of the passing score. Specifically, ADEX conducts equating when new forms of the exam are introduced. Equating is a statistical process through which the empirical characteristics of an anchor form of the exam is compared with the new form to adjust for any change in the difficulty of the form.

**IV. Reliability**

One of the primary sources of validity evidence is derived from estimates of measurement error, also characterized as reliability. Within classical test theory an observed score (i.e. one that a candidate may demonstrate on an exam) is a combination of the candidate’s true score and some unknown error component. The error component may be comprised of both systematic and random error. One of the goals in testing is to control systematic error – also known as bias – through standardization procedures, so that each candidate experiences the same exam under comparable conditions. By controlling these errors, we are left with random error which we estimate using different methods. The nature of the measurement often dictates the type of reliability estimate we use because different exam types have different potential sources of error in the scores. ADEX recognizes that its exam components have different potential sources of error. For example, the use of multiple raters on the clinical skills examinations shows an understanding that rater variation represents the greatest potential source of error on these item types. Thus, a two-part decision rule is used when determining candidates’ scores on these items. First, for an error to be “true”, at least two of the three examiners need to independently document the specific error. Second, at the point of the pass-fail decision, the sum of “true” errors needs to be greater than the minimum allowable for a candidate to fail that section of the exam.

According to the *Standards* (AERA et al., 2014), because the inter-rater agreement on the pass-fail decision is the most critical to the confidence ADEX has in its raters’ scores, two analyses are conducted for each of the task-based sections of the exam to evaluate this component of reliability. First, ADEX determines the candidate’s pass-fail decision using the
decision rules described above for each of the three, independent raters. Within each group of raters across candidates, an evaluation of whether an examiner was individually in agreement or disagreement with the ultimate decision for a candidate on a given procedure. If the rater’s individual decision agreed with the pass-fail decision on the task/procedure, this is deemed an agreement. If the rater’s individual decision was not in agreement with the pass-fail decision, ADEX considers this a disagreement. ADEX has established a minimum requirement of 85% decision consistency for each procedure with a goal of 90%.

A second analysis focuses on individual rater performance. This series of analyses aggregates examiners’ performance within a procedure across sites to determine if there are any systematic variations in how an examiner is applying the scoring criteria. These analyses consider the number of performances that an examiner has rated and determines the proportion of the time that the examiner agreed with the ultimate pass-fail decision. Because there are fewer total performances when conducting these analyses by examiner, a minimum number of performances was established for decision-making purposes. ADEX examiners are required to perform at an 85% decision consistency level or higher. Examiners who do not meet this minimum requirement are referred to the evaluation committee for remediation and potential removal if their performance fails to meet the requirement after remediation.

Because the items on the DSE exam are objectively scored, the variation among examiners would not provide information on the greatest potential source of error in the candidates’ scores. Thus, a different analysis is employed. For selected response items, a statistical calculation, coefficient alpha, is often used to determine the internal consistency of the items (i.e., the ratio of the intercorrelations of items relative to total variation) used to measure the construct. This method is a generalized approach to split-half reliability that allows both dichotomous and polytomous items to be considered in the calculation. These analyses are conducted for each form of the examination to determine the acceptability of the form performance. Follow-up item analyses that evaluate the item difficulty and discrimination (from a classical test theory perspective) are then used by the examination committee to make decisions about item revision, suspension, or removal.

Summary

In summary, ADEX has been attentive to the expectations outlined in the Standards for Educational and Psychological Testing (AERA, APA, & NCME, 2014) and the AADE Guidelines (2003). The most critical components of a licensure exam that includes performance tasks rest in the content of the exam, the specification of the content and scoring criteria, the standardization of the examination for all candidates, the calibration of qualified examiners to apply the scoring criteria as intended, the operational application of the scoring rules, defensible decisions about the pass-fail status of candidates, and empirical evidence to support the quality of the examiners. Tests by themselves are not valid, reliable, or fair. Validity is a property of scores and how we use those scores to make intended inferences or decisions. It is only by gathering the validity evidence of each of these important components that we can support the important decisions made in a licensure program.
References


