Comparisons of National Board Part II and NERB’s Written Examination for Outcomes and Redundancy


Abstract: The purpose of this study was to compare students’ in-school academic performance with their scores on the North East Regional Board (NERB) of Dental Examiners’ Dental Simulated Clinical Examination (DSCE; written) and the written examination given by the National Board of Dental Examiners, Part II (NBDE Part II). A second purpose was to evaluate whether requiring the DSCE in addition to the NBDE II is of benefit in assessing the qualifications of candidates for licensure. Results on the first attempt at passing the two examinations were compared for 836 graduates (1994-2002) of one dental school. Mean scores on the examinations and mean grade point average (GPA) of those who passed the two examinations did not differ. Mean GPAs of those who failed the two examinations also were not significantly different. Scores on the two exams and GPA significantly correlated with each other. Correlation coefficients were higher between GPA and scores on the NBDE II than between GPA and scores on the DSCE. The failure rates on first attempts were 6 percent on the DSCE and 7 percent on the NBDE II. The results of this study suggest that the DSCE does not materially assist in the licensure decision for candidates who passed the NBDE Part II, as the results on the two exams are very similar.

Dr. Ranney is former Dean and Professor, Department of Periodontics, Dental School, University of Maryland and Senior Policy Fellow, American Dental Education Association; Dr. Gunsolley is Professor and Chair, Department of Periodontics, and Ms. Miller is Director of Academic Support Services, both at the Dental School, University of Maryland. Direct correspondence and requests for reprints to Dr. Richard Ranney, Dental School, University of Maryland, 666 W. Baltimore Street, Baltimore, MD 21201; 410-706-7201 phone; 410/706-3028 fax; ranney@dental.umaryland.edu. This report is the work of the authors and is not intended to represent the position of the American Dental Education Association or the Dental School, University of Maryland.

Key words: licensure, national board dental examinations, NERB

Submitted for publication 9/10/03; accepted 11/12/03

Passing the National Board Dental Examinations (NBDE) developed by the Joint Commission on National Dental Examinations (JCNDE) of the American Dental Association (ADA) is a general requirement for licensure in all states in the United States. The NBDE consist of two parts. Part I addresses basic sciences and is taken by dental students after completing specified courses, generally at the end of the second year in a D.D.S. or D.M.D. program. Part II includes case-based questions and is taken midway in dental students’ fourth year. It is intended to evaluate students’ comprehension of important information from basic biomedical and dental sciences and their ability to apply such information in a problem-solving context. Part II is a comprehensive examination in the dental clinical sciences, with approximately 30 percent of the items pertinent to basic sciences and approximately 30 percent of the items pertinent to multiple clinical sciences. It is a written test as opposed to a clinical demonstration, but has been available electronically since 1999.

Because of the acceptance of NBDE Parts I and II by all states, nearly all clinical testing agencies for licensure have abandoned additional written examinations. One regional examining agency, the North East Regional Board of Dental Examiners, Inc. (NERB), has not however. Its examination includes a written test, termed the Dental Simulated Clinical Examination (DSCE), which is now available electronically. Students or graduates wishing to be licensed in a state that requires NERB’s examination therefore must pass both the NBDE Part II and the NERB DSCE.

As of January 2003, twenty-eight states accept results from more than one clinical testing agency. Some states (e.g., Maryland, Massachusetts) that recognize and accept passing scores from multiple testing agencies still require the DSCE for licensure. In Illinois in 1999 after that state decided to accept results of multiple testing agencies and in California before that, issues have arisen as to whether the NERB DSCE constitutes an unnecessary redundancy with the NBDE Part II. Content, stated purpose, test
reliability, scoring, and security of the two examinations were compared by two studies conducted for the California State Board of Dental Examiners and one contracted by the ADA and NERB, with somewhat contrasting results. Major findings from the California studies were that the NERB DSCE lacked sufficient reliability and was redundant in content with the NBDE Part II. However, the third study, conducted by Knapp & Associates, did not substantiate those results.

To our knowledge no studies have made direct comparisons of academic performance with outcomes of the DSCE and NBDE Part II taken by the same individuals. A recent study of the results of NERB examinations in one school over a nine-year period found that the results of the DSCE related substantially better to measures of academic performance in school than did any of the clinical demonstration parts of the NERB examination. Our study was conducted to determine if the NBDE Part II obtained the same or different results compared with the DSCE in the same cohort of students. If clearly different, support for requiring the DSCE would be substantiated; if the results were not different, then arguments for redundancy of the two examinations would be supported.

**Methods**

Data consisted of results of NERB DSCE and NBDE II examinations that were given in the years 1994 through 2002 for the 836 D.D.S. graduates of a single dental school during that period, as well as those graduates’ respective total in-school academic performance as reflected by grade point average (GPA). All data were available from records maintained in the dean’s office. Only data from each graduate’s first effort at passing the DSCE and the NBDE II were considered. All but nine of the 836 graduates took the NBDE II. Data for the DSCE were available for 586 of the graduates. The others either did not take the examination or did not provide NERB with written permission to release their scores. Data were missing for both examinations for three of the 836 graduates. The first-time DSCE results for each graduate were the numeric scores and pass or fail result as reported to the school by NERB. Failure on the DSCE occurred if the score for the total examination was less than 75 or if the score on any one of its three subsections was below 62.5. The first-time NBDE II data were scores provided to the school by the JCNDE. Scores below 75 on the NBDE II were failures.

A significant difference between the two examinations is how high section scores affect low section scores. The DSCE uses partial compensatory scoring in which a lower score in one of the three sections of the examination may cause the candidate to fail, despite a high score in each of the other sections that would produce a fairly high overall test score if performance on all sections were used in overall score calculation. As an artifact, this scoring method allows for a failing candidate to attain a higher overall test score than a passing candidate. For these students, the score of only the failing section is reported. In contrast, with the NBDE II scoring system, pass/fail decisions and reports are based on the overall test score, allowing candidate strengths in one content area to compensate for relative weakness in another content area. Grade point average (GPA) for each student was calculated as the sum of grade points (1 to 4 scale, with 1 being low) for all work attempted in the D.D.S. program divided by the total credit hours for that work.

The relationship among the DSCE score, NBDE II score, and GPA were tested using Pearson product-moment correlations. Because the non-normal distribution of DSCE scores due to its scoring system could unfairly limit the correlation coefficients that could be obtained with that variable, Spearman’s rank correlation coefficient was also determined. To ensure that unequal numbers in the respective correlations of the two examination scores with GPA would not be statistically misleading, those correlations were performed with data from only those students who took both examinations (N = 580). Mean GPAs were compared between subject groups by two-tailed t-tests. Significant difference was concluded when p <0.05.

**Results**

The numbers of graduates passing or failing the two tests are listed in Table 1. Ninety-four percent of those for whom there were data for the DSCE had passing scores. The corresponding number for the NBDE II was 93 percent.

The mean score and standard deviation for the NBDE II were 81.6 and 4.94, respectively. The corresponding data for the DSCE were 82.8 and 7.93,
respectively. NBDE II scores were normally distributed (Figure 1); scores for the DSCE were not, but included several outliers at the low end (Figure 2). Scoring below 75 was failing for each exam. For the DSCE, there were six scores of 74, one 73, and then there was a gap in the distribution of scores with no scores between 73 and 56. The twenty-six other failing scores were distributed from 56 to 47. The gap and low-end outliers were artifacts of the truncation in DSCE’s scoring system that considered a severely failing score for one section to be the overall score.

Mean GPAs of graduates who passed and those who failed each of the examinations are given in Table 2. The mean GPAs of those who passed the two examinations respectively were not significantly different. Mean GPA of those who failed the DSCE appeared higher than that of those who failed the NBDE II, but the difference was not statistically significant. For both tests, mean GPA differed significantly between those who passed and those who failed (DSCE: $t_{32} = 6.767; p <0.01$; NBDE II: $t_{60} = 15.426; p <0.01$). Nine percent of the DSCE failures were in the top half of their class. Five percent of the NBDE II failures were in the top half of their class.

Scores on the DSCE, the NBDE II, and the GPA were highly correlated with each other. The correlation coefficients appear in Table 3. Using the Pearson product-moment, the highest correlation coefficient was between NBDE II scores and GPA. Because the imposed artifact of discontinuity in DSCE scores due to scoring method makes parametric correlations to those scores a bit unfair since they are not normally distributed, the nonparametric Spearman correlations are also shown. By that method the correlation between scores on the two examinations appears very close to that between the NBDE II and GPA, but the correlation between the DSCE and GPA is lower than either of those. However, for all comparisons, the correlations were significant ($p <0.01$).

### Discussion

All states in the United States require candidates to pass the NBDE II to qualify for licensure. This study found a significant correlation between scores on the NBDE II and NERB’s DSCE and between scores on each test and GPA. The NBDE II scores tended to correlate better with GPA than did DSCE scores, but they were similar and both were well correlated with academic performance in school. Appropriate analysis of reported DSCE scores was limited because its truncated scoring system suppressed scores of some sections for some candidates (those failing because of a single section score below 62.5). However, the correlations also pertained if those scores were eliminated from the analysis. No matter how the analysis was performed, using Pearson product-moment correlation or the nonparametric Spearman rank correlation, and with or without the outliers produced by the scoring system of

### Table 1. Numbers of graduates passing or failing the NBDE Part II and the NERB DSCE, 1994-2002

<table>
<thead>
<tr>
<th></th>
<th>Failed DSCE</th>
<th>Passed DSCE</th>
<th>Totals</th>
</tr>
</thead>
<tbody>
<tr>
<td>Failed NBDE II</td>
<td>11</td>
<td>25</td>
<td>36</td>
</tr>
<tr>
<td>Passed NBDE II</td>
<td>22</td>
<td>523</td>
<td>545</td>
</tr>
<tr>
<td>Totals</td>
<td>33</td>
<td>548</td>
<td>581</td>
</tr>
</tbody>
</table>

*Data do not include those for eight graduates who did not take the NBDE II and 250 for whom there were no DSCE data (three graduates had missing data for both tests).*
the DSCE, the correlations were significant and they were higher between the NBDE II and GPA than between the DSCE and GPA. Thus, there clearly was no superiority for NERB’s examination over the National Board examination in correlations with academic performance at one dental school. Of course, one cannot be certain whether our results would generalize to other schools until confirmed or refuted by analysis elsewhere.

We are not aware of any other peer-reviewed, published study that directly compared the DSCE with the NBDE II. However, the report by Knapp & Associates to the ADA and NERB5 included an analysis of studies previously contracted by the California State Board of Dental Examiners,6,7 as well as their own review of the two tests. The California studies had recommended to the dental board that the DSCE not be used since its reliability was not within usually acceptable range for a high stakes test and its content was redundant with the NBDE II. On the other hand, NERB has cited the study by Knapp & Associates to say that the California findings were incorrect with respect to both content and reliability. Specifically, NERB’s ad hoc long-range planning committee stated that since Knapp & Associates reported different content between the two examinations, they were not redundant and the reliability “was sufficient.” The same interpretations were provided to the ADA’s House of Delegates.8

As our current results provide no reason to require both examinations in terms of outcomes, resolution of the contrasting claims about content and examination reliability is important to the evaluation of whether requiring the DSCE in addition to the NBDE II is of benefit in assessing the qualifications of candidates for licensure. To that end, we assessed the listings of DSCE and NBDE II content and examined the methods and conclusions regarding reliability of the DSCE, directly in a copy of the report by Knapp & Associates to the ADA, obtained by request.

Knapp & Associates’ conclusions included the following: “Both tests are similar in content but assess different aspects of clinical knowledge. Therefore, the examinations as they were administered up to 1998 were not redundant, had different purposes and were intended to be given at different points in the credentialing process.” They also stated, “Content categories for both examinations are similar. There are moderate differences in the relative weights assigned to some of the content categories. There are large differences between the examinations in the percentage of questions that are practice-related.”5

However, those conclusions are questionable from the data in the report. Its Tables 4 and 5 showed that neither test ignored content areas that are in the other. The proportional differences in emphasis were that the DSCE had a higher percentage of questions in oral pathology, periodontics, and prosthodontics, whereas the NBDE II had relatively more questions in restorative/operative, orthodontics/pediatrics, and pharmacology. A conclusion from those proportional differences that one of the tests is more “practice-related” is impossible to make without validation. In fact, Knapp & Associates also concluded under the heading of validity as related to purpose, “However, the documentation submitted by either program was not sufficient to determine how the test content outlines of the programs, the weighting of the content, and question types were linked to the actual tasks performed and the knowledge needed by dentists who are entering practice.”5

The proportional content differences between the two examinations, in fact, seemed to lie mainly in the NBDE II being broader than the DSCE. When only the case-based part of NBDE II was considered, little if any content difference existed. The fact that neither test ignored content that was present in the other refutes any potential claim for uniquely relevant content. As to the timing in the credentialing

Table 2. Mean GPA ± standard deviation for graduates passing or failing the DSCE and NBDE Part II

<table>
<thead>
<tr>
<th>Test</th>
<th>Pass</th>
<th>Fail</th>
</tr>
</thead>
<tbody>
<tr>
<td>DCSE</td>
<td>3.11 ± 0.41</td>
<td>2.72 ± 0.33</td>
</tr>
<tr>
<td>NBDE II</td>
<td>3.10 ± 0.40</td>
<td>2.62 ± 0.24</td>
</tr>
</tbody>
</table>

Table 3. Pairwise correlations (Rho) among scores on the DSCE, scores on the NBDE II, and GPA

<table>
<thead>
<tr>
<th>Comparison</th>
<th>Pearson</th>
<th>Spearman</th>
</tr>
</thead>
<tbody>
<tr>
<td>NBDE II by All DSCE Scores</td>
<td>0.59</td>
<td>0.73</td>
</tr>
<tr>
<td>NBDE II by GPA*</td>
<td>0.75</td>
<td>0.75</td>
</tr>
<tr>
<td>All DSCE Scores by GPA*</td>
<td>0.48</td>
<td>0.60</td>
</tr>
<tr>
<td>Restricted** DSCE Scores by GPA</td>
<td>0.60</td>
<td>0.59</td>
</tr>
</tbody>
</table>

*Used data only from those students who took both examinations (N = 580)
**Restricted to normally distributed scores (>72)
process, both examinations claim the purpose of aiding in determination of qualifications for licensure. Knapp & Associates concluded, however, that neither had used methods to analyze content that would allow determination about validity as related to purpose. Therefore, arguments on validity do not appear to favor one test over the other, nor to prove that one supplements or corrects deficiencies in the other.

Knapp & Associates acknowledged in their report that they are not content experts. It might be desirable for a task force of experts to examine the two tests critically to determine what, if any, uniqueness exists and to better assess content validity than has been done by either organization to date. We understand that the JCNDE is conducting such a validation process at the current time, but the results are not available to us as of this writing. Even should NERB do likewise and find that its test does indeed have unique content, given the similarities in outcomes found in this study and the fact that tests for licensure in other health care disciplines in the United States are national rather than regional or local, it would make more sense for dental students, the profession overall, and the public that any truly unique content in the DSCE be absorbed within a broader NBDE II than to require two tests that are largely the same in content and results.

The issue of test reliability is easier than validity to evaluate from currently available information. Knapp & Associates reported that their findings on reliability did not confirm the California studies’ conclusions of inadequate reliability for the DCSE. However, they succeeded in obtaining reliability coefficients for the DSCE of similar magnitude to those for the NBDE II only by extrapolating the DSCE to be longer than it is or by considering only the case-based part of the NBDE II. The estimates realized by “correcting” for number of questions assumed that the quality of the added questions would be the same as for the existing ones. That is a questionable assumption, but even if it were to be right, the “correction” for the limited number of questions in the DSCE was a flawed use of the Spearman-Brown prophecy formula, which is used appropriately to determine how much to lengthen a test to achieve the desired level of reliability, not to declare a sufficient level of reliability on a short version. Because the DSCE does not include those additional test questions, its calculated reliability coefficients are best judged within the conditions of the test as it exists rather than being based on more acceptable conditions that do not exist. Thus, the original conclusions for California are probably correct: the internal reliability of the DSCE is not within the usually acceptable range for a high stakes test.

It is also pertinent that NERB uses a “partially compensatory” scoring system; thus, Knapp & Associates correctly concluded that the reliability estimates for the DSCE’s subsections remain important. Given that the three subsections have to be shorter than the test as a whole, their respective internal reliabilities are likely substantially lower than for the test as a whole. This raises serious questions about the propriety of DSCE’s scoring system, in which failure can result from a test with low reliability. At the very least, in the context of a decision as to whether to require both the DSCE and the NBDE Part II, it is clear that the DSCE has no superiority in reliability to the NBDE Part II.

Both the DSCE and the NBDE II received criticism from Knapp & Associates for the method by which cut scores were established. For the NBDE Part II, there is a standardization of scores from year to year, whereas for the DSCE a new reporting scale is established each year. It was recommended that both examinations would have a more defensible cut score if they adopted a process whereby each question is judged against criteria and individual judgments of experts are calculated to give a passing score. Therefore, on the basis of scoring, again neither test was clearly superior to the other. Further, it is inappropriate in a high stakes licensure examination to have a definitive cut score on one section of an examination that indicates failure, when in fact there are questions about whether the cut score was established in an appropriate manner and when reliability is suspect as discussed above.

The National Dental Examining Board of Canada recently assessed the concurrent validity of its written examination and OSCE by correlating scores with students’ performance in their dental programs. That study concluded that the findings supported the concurrent validity of the tests, based on statistically significant correlations ranging from 0.43 to 0.46 (p <0.01). The correlations between academic performance and scores on the NBDE II and the DSCE found in the present study were as high or higher than those found by the Canadians for their tests.

It presently costs a candidate $650 to take the DSCE as a single section of NERB’s clinical exami-
nation, plus the costs to the candidate and school for obtaining necessary documentation to establish eligibility if the candidate did not take NERB’s examination previously. If the best that one could do by requiring that of dentists who had already passed the NBDE II is described by the data in this report, the cost-effectiveness and overall desirability of such a requirement are highly suspect.

Conclusion

On the basis of no established superiority in validity or cut score, questionable if any relevant content differences, the highly significant positive correlations between the two tests and between each test and GPA, the apparently lower reliability of the DSCE, and the cost to individuals and institutions for no clearly demonstrable benefit, it does not seem reasonable to require the DSCE in addition to the NBDE Part II as part of testing competency for dental licensure. As other doctoral-level health care professions in the United States use the national testing model, the burden of proof for use of a regional examination in lieu of or in addition to the NBDE II should be on the regional testing agency. From available evidence, NERB has not provided that for the DSCE.

Acknowledgments

The authors acknowledge the benefit of consultant advice from Leon J. Gross, Ph.D., for psychometric interpretation of data in this paper. The work was supported in part by a Senior Policy Fellowship provided to Dr. Ranney by the American Dental Education Association.

REFERENCES