

Use of Discriminant and Regression Analyses to Modify a Clinical Certification Board Examination

Jack D. Gerrow, D.D.S., M.S., M.Ed.; Marcia A. Boyd, D.D.S., M.A.; David A. Scott, Bsc, Msc, D.D.S.; André-Phillipe Boulais, B.A., M.Ed.

Abstract: The National Dental Examining Board of Canada (NDEB) conducts mandatory, high stakes, pass/fail, certification examinations for dental licensure. One of these examinations was a seven-part, simulated clinical examination in which candidates were required to perform procedures on typodonts. These requirements were two intracoronal and two extracoronal preparations, an amalgam restoration, a provisional crown, and a diagnostic wax-up. Feedback from candidates and examiners indicated that one or more of the requirements may not have been contributing effectively to the overall evaluation of candidates. The NDEB's Clinical Examination Committee therefore requested that an in-depth statistical analysis be performed to identify potential areas of concern and to provide a basis for modifying the examinations. The results of two examination sessions with a total of 168 candidates were subjected to both a discriminant and a logistic regression analysis. Every candidate had results for each of the seven requirements, and no candidate participated in both sessions of the examination. The discriminant analysis revealed that six of the seven requirements could be used to reliably assign examinees according to their true pass/fail classifications. Stepwise discriminant analysis resulted in a 98.81 percent classification success rate with a corresponding 2.50 percent false-positive classification error rate. The logistic regression analysis showed that five components correctly predicted 99.40 percent with a 1.25 percent false-positive rate. The Clinical Examination Committee concluded that one requirement (diagnostic wax-up) should be eliminated and that a second requirement (PFM preparation) be significantly modified and reevaluated. This study demonstrates the usefulness of statistical methods in the analysis and modification of a clinical certification board examination.

Dr. Gerrow is Associate Professor, Faculty of Dentistry, Dalhousie University, and Executive Director and Registrar, National Dental Examining Board of Canada; Dr. Boyd is Professor, Faculty of Dentistry, University of British Columbia, and Chief Written Examiner, National Dental Examining Board of Canada; Dr. Scott is Professor, Faculty of Dentistry, University of Alberta, and Chief Clinical Examiner, National Dental Examining Board of Canada; and Mr. Boulais is Manager, Evaluations Bureau, Medical Council of Canada. Direct correspondence and requests for reprints to Dr. Gerrow at the National Dental Examining Board of Canada, 100 Bronson Avenue, Suite 203, Ottawa, ON, K1R 6G8, Canada; 613-236-5912 phone; 613-236-8386 fax; jackg@ndeb.ca e-mail.

Key words: licensure, certification, evaluation

Submitted 4/7/98, returned for revisions 5/8/98, accepted 5/17/99.

The National Dental Examining Board of Canada (NDEB) was established by an Act of Parliament in 1952¹ to establish qualifying conditions for a national standard of competence for general dentists and to issue certificates to dentists who successfully meet this standard. To become certified to practice dentistry in Canada, graduates of undergraduate dental programs that are not accredited by either the Commission on Dental Accreditation of Canada or the American Dental Association's Commission on Dental Accreditation are required to successfully complete a written and a three-phase clinical examination.

One phase of this clinical examination requires that candidates perform identified procedures (requirements) on simulated patients (typodonts). These

requirements are established annually by the NDEB Clinical Examination Committee. Each of the requirements is weighted equally. At the time of this analysis the seven requirements of the examination were a class II amalgam cavity preparation, class III composite resin cavity preparation, an amalgam restoration, a porcelain fused to metal (PFM) crown preparation, a full gold crown (FGC) preparation, a provisional crown restoration, and a diagnostic wax reconstruction (wax-up). For the wax-up, candidates were required to reconstruct a molar and premolar into functional occlusion and contour on prepared casts mounted on a semi-adjustable articulator.

The criteria used by the examiners to evaluate the seven components are published and available to all candidates. Examiners are calibrated prior to each

Figure 1. Cavity Preparation for Direct (Amalgam, Composite Resin) Restoration

External Outline Form	Internal Form	Finish
<p>P</p> <ul style="list-style-type: none"> based on the location and extent of the decay present, adequately extended for convenience of preparation and restoration, for removal of decalcification, caries and contiguous fissures. Proximal margins clear adjacent teeth. 	<p>P</p> <ul style="list-style-type: none"> extended to provide adequate bulk for strength adequate retention pupal and cervical floors parallel to occlusal plane when appropriate resins—rounded internal line angles 	<p>P</p> <ul style="list-style-type: none"> complete caries removal smooth cavosurface margins with all unsupported enamel removed cavity free of debris minor amount of softened (affected) dentin remaining on pulpal or axial walls
<p>Y</p> <ul style="list-style-type: none"> supporting tissues unnecessarily traumatized minor damage to adjacent tooth significant over-extension under-extended proximal walls flared creating acute cavosurface angle 	<p>Y</p> <ul style="list-style-type: none"> unnecessary removal of internal tooth structure inadequate retention and/or resistance form inadequate depth of preparation incomplete removal of an existing restoration (if present) water coolant not used when indicated 	<p>Y</p> <ul style="list-style-type: none"> minor amount of infected dentin on pulpal or axial walls excessive roughness of cavity wall and/or cavosurface margins unsupported enamel that needs minor correction presence of debris or calculus
<p>Z</p> <ul style="list-style-type: none"> unnecessary over-extension, e.g., beyond line angles mutilation of hard and/or soft tissue damage to adjacent tooth requiring restoration (in Clinical III candidate to be dismissed and provisional restoration to be placed by an examiner) 	<p>Z</p> <ul style="list-style-type: none"> grossly excessive removal of tooth structure avoidable pulpal exposure cavity devoid of retention and resistance form (in Clinical III candidate to be dismissed and provisional restoration to be placed by an examiner) 	<p>Z</p> <ul style="list-style-type: none"> caries remaining at DEJ enamel grossly undermined or unsupported gross caries remaining (in Clinical III candidate to be dismissed and provisional restoration to be placed by an examiner)

examination session, and evaluation is carried out by teams of two examiners who function independently. Examiners use the published criteria (Figure 1) to assign a grade for each requirement on a three-point scale² (Figure 2). The Chief Examiner reviews all evaluations and adjudicates any differences. A pass/fail result for each candidate is determined by comparing the grade for the seven requirements against a predetermined grade derivation grid (Figure 3).

Feedback from candidates and examiners indicated that one requirement, the wax-up, may not have been contributing effectively to the overall evaluation of candidates. Therefore, to identify areas of concern and to provide a basis for modifying the examination, the NDEB's Clinical Examination Committee undertook an investigation to determine which requirements or combination of requirements could best predict pass/fail outcomes and do so with the least amount of error. Given the challenge of choosing a subset of requirements that could increase the

efficiency of the examination, as well as retaining the reliability of the pass/fail decisions, a choice must be made as to which type of error has the most negative impact given the fact that success can lead to certification. For this reason it was decided to minimize false-positive errors. The purpose of this study, therefore, was to determine which set of requirements would yield the least number of false-positive errors while most reliably establishing pass/fail classifications.

Method

The sample consisted of performance outcomes for 168 candidates from two examination sessions. Every candidate had results for each of the seven requirements, and no candidate participated in both examination sessions. Of the 168 candidates in these two sessions, eighty-eight had passing results, and eighty had failing results. Two basic statistical ap-

Figure 2. NDEB Grading System

The 3-point ("P", "Y", "Z") grading system used for clinical examinations:

- P The candidate has performed the clinical procedure to an acceptable standard. Improvements or modifications could be made but are not necessary.
- Y The candidate needs to make minor corrections or additions to make the clinical procedure acceptable.
- Z The candidate needs to make major corrections or additions to make the clinical procedure acceptable, or the candidate has made a non-correctable mistake or demonstrated a significant lack of knowledge or skill.

proaches were used to determine the contribution of each requirement to the reliability of pass/fail decisions. Stepwise discriminant analysis³ using a forward inclusion of variables was performed to identify which combinations of test components could best be used to classify candidates into their respective true pass/fail classification. Default F-values for entry and removal were 3.84 and 2.71 respectively. Interaction effects were not explored because of problems associated with the interpretability of any results. The logistic regression analysis⁴ was applied to establish which combinations of variables could best predict pass/fail outcomes. The pass/fail outcome was the criterion variable, with the seven evaluations being the predictor variables. The statistical analyses were performed on a Toshiba Terera using SP55 Version 7.5 for Windows and on a Compaq Workstation using SP55 Version 8.0 for Windows NT 4.0.

Results

The results of the discriminant analyses for the seven requirements are shown in Table 1. This analysis reveals that the provisional restoration was the single most reliable predictor of classification of pass/fail status since it correctly classifies candidates in 76 percent of the cases and also minimizes false-positive errors. However, the rates of correct classifications for this requirement in isolation are not sufficient given the nature of the decisions which are to be made. Since the decisions impact on the safety of the public, it is imperative that a high degree of decision reliability be retained. To attain decision accuracy of at least 95 percent for overall correct classifications and a false-positive errors rate of 3 percent

Figure 3. Grade Derivation Grid for Clinical II Examination

- 3 Y + 4 P = P
- 1 Z + 2 Y + 4 P = P
- 4 or more Y = F
- 2 or more Z = F
- 1 Z + 2 or more Y = F

or less, a subset of components that maximize correct classification and minimize false-positive was established using a stepwise discriminant analysis. This procedure allows a progressive increase in decision accuracy (incremental reliability) as components are added and their effect monitored.

Table 2 summarizes the results obtained through the stepwise discriminant analysis procedure. Requirements 7, 1, 2, 6, 3, and 4, in that order, were included in the final discriminant function, while Requirement 5, the evaluation of the wax-up, failed to add any significant increment in pass/fail decision precision. The six requirements resulted in a correct classification of 98.81 percent with only 2.50 percent false positives.

It is customary in discriminant analysis to test the resulting discriminant function for validation. While the above results seemed to yield respectable outcomes in terms of percentage of total cases correctly classified, as well as a relatively low false-positive rate, these outcomes were based on the total sample of candidates. One approach to test the validity of the discriminant function is to repeat the process using a sub-sample of candidates and obtain a discriminant function that could be applied to calculate predicted pass/fail status of candidates not used in calculations of the discriminant function. The true discriminatory power of the function is then put to the test because it is applied to a separate sample.

A random sample of 148 candidates was drawn from the original 168. This sub-sample was then used

Table 1. Results of discriminant analysis

Requirement	%	
	Correctly Classified	False-Positive
1. Amalgam	73.18	19.05
2. Composite Resin Preparation	74.94	12.50
3. PFM Crown	73.86	17.86
4. FGC Crown	72.39	15.48
5. Wax-up	67.61	23.81
6. Amalgam Restoration	72.05	19.05
7. Provisional Crown Restoration	76.36	9.52

Table 2. Stepwise discriminant analysis using requirements 7, 1, 2, 6, 3, and 4

Observed	Predicted		Total
	Pass	Fail	
Pass	88	0	88
Fail	2	78	80
TOTAL	90	78	168

Percent of total cases correctly classified: 98.81%

Percent of false-positives: 2.50%

to calculate the discriminant function to be subsequently applied to the remaining twenty candidates. Table 3 shows the pass/fail outcomes at the end of the discriminant analysis. Requirements 7, 1, 2, 4, 6, and 3, in that order, entered the discriminant function. Percent of correct classification (98.65) and percent of false-positives (2.78) are very similar to the analysis using the total 168 candidates.

The following standardized discriminant function coefficients were obtained for each of the test measures and applied in the discriminant function (DF), which is also shown below.

$$\begin{aligned}
 \text{Requirement 1 } (m_1) & \quad 0.6997690 (w_1) \\
 \text{Requirement 2 } (m_2) & \quad 0.8485214 (w_2) \\
 \text{Requirement 3 } (m_3) & \quad 0.6752133 (w_3) \\
 \text{Requirement 4 } (m_4) & \quad 0.6918394 (w_4) \\
 \text{Requirement 6 } (m_6) & \quad 0.6839596 (w_6) \\
 \text{Requirement 7 } (m_7) & \quad 0.8414111 (w_7) \\
 \text{DF} = & -5.920 + w_1m_1 + w_2m_2 + w_3m_3 + w_4m_4 + \\
 & w_6m_6 + w_7m_7
 \end{aligned}$$

where m_1 through m_7 represent values on the various predictor variables Requirement 1 through 7 (excluding Requirement 5) and w_1 through w_7 are the weights associated with each of the respective variables. The value -5.920 is a constant.

Table 3. Pass/fail outcomes for validation subset using requirements 7, 1, 2, 4, 6, and 3

Observed	Predicted		Total
	Pass	Fail	
Pass	76	0	76
Fail	2	70	72
TOTAL	78	70	148

Percent of total cases correctly classified: 98.65%

Percent of false-positives: 2.78%

The application of this formula will predict the pass/fail status of candidates not included in the subsample. If a negative score is obtained, a fail status is assigned and a positive score will yield a pass status. The status of all twenty candidates was established this way. Table 4 presents the results of the validating procedure. Perfect prediction was achieved, therefore yielding no false-positive errors of classification with this method of cross-validation.

Table 4. Validation prediction using requirements 7, 1, 2, 4, 6, and 3

Observed	Predicted		Total
	Pass	Fail	
Pass	12	0	12
Fail	0	8	8
TOTAL	12	8	20

Percent of total cases correctly classified: 100%

Percent of false-positives: 0%

In addition, a leave-one-out or "bootstrapping" method of cross-validating the discriminant function was performed. The results of this cross-validation are shown in Table 5. The leave-one-out method is a more robust validation method and resulted in 97.02 percent of the cases being correctly classified with a 6.25 percent false-positive rate.

A second set of analyses was conducted to establish a prediction equation similar to the discriminant analysis method. Logistic regression analysis is a statistical procedure by which prediction of the criterion variable outcome is sought through the inclusion of one or several predictor variables in a regression equation. For this study, prediction of passing or failing the Clinical II Examination (the criterion variable) is sought by including a smaller subset of predictor variables. Logistic regression was used for this study since the dependent variable (pass/fail status) is a dichotomous variable. A stepwise procedure was also used as described for the discriminant analysis method.

While logistic regression shares some of its theoretical basis with discriminant analysis, the statistical procedures of regression are somewhat different. The application of logistic regression yielded slightly different results. The final step of the logistic regression procedure yielded the following order of entry into the logistic equation: 7, 2, 4, 1, and 6. The final step excluded Requirements 3 (porcelain fused to metal crown preparation) and 5 (wax-up). The final prediction outcomes are shown in Table 6.

Table 5. Results of leave-one-out cross-validation

		P/F	Predicted Group Membership		Total
			0	1	
Original	Count	0	78.00	2.00	80.00
		1	0.00	88.00	88.00
	%	0	97.50	2.5	100.00
		1	0.00	100.00	100.00
Cross-validated	Count	0	75.00	5.00	80.00
		1	0.00	88.00	88.00
	%	0	93.75	6.25	100.00
		1	0.00	100.00	100.00

According to these logistic regression results, reliable decisions can be achieved using Requirements 1, 2, 4, 6, and 7 with 99.40 percent correctly classified and a rate of 1.25 percent false-positives.

Following a review of the results of this study, the Clinical Examination Committee eliminated the wax-up, changed both the tooth selection and the written directions for the PFM crown preparation, and added a new requirement. The results of the revised examination will be analyzed in a similar manner.

Discussion

The results obtained from the logistic regression analysis are similar to those obtained from the discriminant analysis procedure in that the same test components were retained in the final equations and were introduced in a similar order. The difference is that, in the logistic regression analysis, Requirement 3 and Requirement 5 were not included in the final equation, whereas in the discriminant analysis only Requirement 5 was excluded. Based on the sample of candidates provided, it would seem that reliable decisions could be accomplished even when excluding results obtained from Requirement 5 and perhaps even Requirement 3.

Given the nature of the statistical analyses used in this type of study, it could be recommended that similar analyses be conducted in the future when a larger sample size would be available. It would be most useful to perform these analyses when the sample size reached at least twice the number of candidates; however, this is not possible as the examination has been modified. It would have enabled the implementation of a validating exercise similar to the

Table 6. Prediction from logistic regression analysis using requirements 7, 2, 4, 1, and 6

Observed	Predicted		Total
	Pass	Fail	
Pass	88	0	88
Fail	1	79	80
TOTAL	89	79	168

Percent of total cases correctly classified: 99.40%

Percent of false-positives: 1.25%

one outlined in the discriminant analysis section but using twice the number of subjects and allowing the predicting group to reach a ratio of twenty to one for the number of candidates versus the number of predictor variables. This higher number would yield a high degree of confidence in the results. Another worthwhile procedure would be to repeat the above studies using a different criterion variable, one that is independent of the test measures being used to establish pass/fail status. This change would avoid the problem of criterion variable contamination. While it would probably be difficult to obtain, an independent judgment of candidate competence would be more valid.

Much of this discussion has focused on the reliability of the decisions that can be achieved based on a subset of test measures. However, the notion of validity as it relates to clinical practice must not be overlooked. While it is beyond the scope of this study, it is worth noting that the decision to include or exclude tests in the final decision-making process should not be based on reliability studies alone. The issue of validity should also play an important role. The nature of the total examination and how it is being used should be considered. If it is felt that what is being measured by the test content of Requirement 5 (or for any test component being excluded in a reliability study) is very important in terms of clinical practice, then perhaps areas for improvement could be identified to render this measurement tool more reliable rather than eliminating it.

REFERENCES

1. Revised Statutes of Canada, Ch. 69, 1952.
2. Gerrow JD, Boyd MA, Doyle G, and Scott D. Clinical evaluation in prosthodontics: practical methods to improve validity and reliability at the undergraduate level. *J Prosthet Dent* 1996;75:676.
3. Stevens J. Applied multivariate analysis for the social sciences. Hillsdale, NJ: Lawrence Erlbaum, 1992:273-302.
4. Darlington RB. Regression and linear models. Toronto: McGraw-Hill, 1990:441-61.