Educational Methodologies

Correcting for Guessing Increases Validity in Multiple-Choice Examinations in an Oral and Maxillofacial Pathology Course

Thomas J. Prihoda, Ph.D.; R. Neal Pinckard, Ph.D.; C. Alex McMahan, Ph.D.; Anne Cale Jones, D.D.S.

This Article

Abstract FREE
Full Text (PDF)
Alert me when this article is cited
Alert me if a correction is posted

Services

Similar articles in this journal
Similar articles in PubMed
Alert me to new issues of the journal
Download to citation manager
© Get Permissions

Citing Articles

Citing Articles via HighWire
Citing Articles via Google Scholar

Google Scholar

Articles by Prihoda, T. J.
Articles by Jones, A. C.
Search for Related Content

PubMed

PubMed Citation
Articles by Prihoda, T. J.
Articles by Jones, A. C.

▶ Abstract

- Top
- Abstract
- Methods
- Results
- Discussion
- Example
- Conclusion
- References

A standard correction for random guessing on multiple-choice examinations was examined retrospectively in an oral and maxillofacial pathology course for second-year dental students. The correction was a weighting formula for points awarded for correct answers, incorrect answers, and unanswered questions such that the expected value of the increase in test score due to guessing was zero. We compared uncorrected and corrected scores on examinations using a multiple-choice format with scores on examinations composed of short-answer questions. The short-answer format eliminated or at least greatly reduced the potential for guessing the correct answer. Agreement of corrected multiple-choice scores with short-answer scores (intraclass correlation coefficient 0.78) was significantly (p=0.015) higher than agreement of uncorrected multiple-choice scores with short-answer scores (intraclass correlation coefficient 0.71). The higher agreement indicated increased validity for the corrected multiple-choice examination.

The expectation for students to improve their grades by guessing on multiple-choice format examinations is well known. We have examined a method for correcting for random (no knowledge) guessing on multiple-choice question[1-3] by comparing uncorrected and corrected scores on examinations using a multiple-choice format with scores on examinations using short-answer questions in an oral and maxillofacial pathology course for dental students. We take as self-evident that the short-answer format eliminates or at least greatly reduces the potential for guessing the correct answer. The short-answer

questions were presented in a clinically relevant context to better simulate the situation students will face in providing patient care. Questions included a clinical history and projected Kodachrome® slides of clinical and microscopic pathology and appropriate radiographs.

This study represented a unique opportunity to compare scores from multiple-choice and short-answer examinations in a setting in which students were given the same number of questions in each of the two format types testing their knowledge over the same subject matter. The results of this study assessing four separate examinations during an oral and maxillofacial pathology course indicated that the corrected multiple-choice scores agreed significantly better with student performance on short-answer examinations than did the uncorrected multiple-choice scores.

▶ Methods

We investigated the standard correction for guessing (formula scoring) using scores on four examinations in the didactic oral and maxillofacial pathology course presented in the 2005 spring semester at the University of Texas Health Science Center at San Antonio. This course, given to all second-year dental students, was fifty-eight hours in length and consisted of fifty hours of lecture and four two-hour examinations. Each of the four examinations was divided into two one-hour examinations. The first hour of each examination consisted of twenty-five cases, each of which had a short clinical history and projected clinical, microscopic, and radiographic Kodachrome® slides. For each of the twenty-five cases, two short-answer questions were asked for a total of fifty questions. When all twenty-five cases had been presented, the students were given five minutes to look over their answers and make any changes or corrections. The short-answer examination was then collected and subsequently graded by the course director (ACJ). The short-answer questions were graded by looking for key words identified at the time of construction of the examination. If a student gave multiple answers, only the first answer was evaluated; no partial credit was awarded. The second hour of each examination consisted of fifty multiple-choice questions, each with one correct answer and four distractors. The multiple-choice questions were a mixture of clinical vignettes and ordinary didactic questions. Students were asked to choose the single correct answer for each question. At the end of the second hour, answer sheets were collected and graded electronically.

Since the multiple-choice examination and the short-answer examination each consisted of fifty questions, they were equally weighted during the calculation of each student's final grade. Each two-hour examination comprised 25 percent of the final grade. No comprehensive final examination was given. Students received final course grades based on averages calculated from the scores on the four one-hour multiple-choice examinations and the four one-hour short-answer examinations. These averages were used to assign course grades as A (90–100), B (80–89), C (70–79), or F (0–69).

Each of the four examinations was equally spaced during the course and covered between eleven and thirteen hours of lecture material. When the individual short-answer and multiple-choice examinations were constructed, the questions were equally weighted to the topics that were presented prior to each of the four examinations. This was to ensure that a given topic was not stressed more often than another topic. The students were advised to add up the number of topics discussed in a given section and divide that number by fifty to arrive at an approximate number of questions per topic on both the multiple-choice and short-answer examinations.

The effect of correction for guessing was investigated after the completion of the course and official

grades awarded. Ninety students initially enrolled in the course during the 2004–05 academic year; two students who were failing the course after the completion of three examinations withdrew from the class before the fourth examination. The analyses presented in this report were based on the eighty-eight students who completed the course and had scores for all four examinations. This study was approved by the Institutional Review Board of the University of Texas Health Science Center at San Antonio.

In making the correction for guessing, we assumed that students were making a truly random choice. The correction that we applied was a modification to the ordinary grading method for multiple-choice examinations (number-correct or number-right scoring) where zero points are assigned for an incorrect answer and full credit is given for a correct answer.[4] Since each multiple-choice question has five possible answers, the standard correction for guessing consisted of awarding –1/4 for an incorrect answer, 0 for a question not answered, and +1 for a correct answer; these points were added and the sum divided by the number of questions. The probability of guessing a correct answer and being awarded +1 was 0.20, and the probability of guessing an incorrect answer and being awarded –1/4 was 0.80. Thus the expected value of the number of points gained due to guessing was $(0.20)(1)+(0.80)(-1/4)=0$. In general, for K possible answers per question, $-1/(K-1)$ is awarded for an incorrect answer, 0 for a question not answered, and +1 for a correct answer.[4] This correction for guessing is generally referred to as formula scoring[4] or the standard correction for guessing. Formula scoring is a special case[5] of choice weighting.

If all questions were answered, as was the case with our examinations, the correction for guessing was equivalent to applying a straight line adjustment such that a grade of 100 percent was unchanged and a grade of 20 percent was adjusted to zero. The equation of this straight line was Corrected Score (%)=1.25 [Uncorrected Score (%)–20].

An intraclass correlation coefficient[6] was used as a measure of agreement[7] between a multiple-choice score and a short-answer score. The Pearson correlation coefficient was not appropriate because it measures association, not agreement.[8] Perfect agreement occurs only if data points lie along the line of equality; perfect correlation occurs if data points lie along any straight line.

Principal component lines[9] were estimated from the variance-covariance matrix. The first principal component is the linear combination (straight line) of the variables that has the maximum variance among all (normalized) linear combinations; also, the first principal component is the line through the means, $(\bar{X}, \bar{Y})$, which minimizes the sum of the squared distances of the data points to the line.[10] We used principal components analysis because both the X and Y variables were random variables; in ordinary linear regression analysis, only the Y variable is considered to be a random variable, and the estimator of the line is biased if X also is a random variable.[11] Thus, the first principal component lines more accurately estimated the relation between these X and Y variables.

Variance components[12] were calculated for students, examinations, and error to study the sources of variability in scores. Reliability was the ratio of variance between students to total variance.

To compare means on each examination type for different typical grade classifications, we used analysis of variance for repeated measurements[13] with examination (four levels) as a repeated measures factor. The statistical model used an unstructured variance-covariance matrix.

A bootstrap procedure[14] with 1,000 samples was used to estimate confidence intervals and to compare agreement of uncorrected and corrected multiple-choice scores with short-answer scores, to estimate confidence intervals for the slope and intercept of the first principal component lines, to test that the slope

was 1.00 and the intercept 0.00, to test that the principal component lines for the uncorrected and corrected scores were the same, and to compare reliability coefficients of multiple-choice examinations and short-answer examinations. The bootstrap is a nonparametric procedure and thus does not depend on any particular probability distribution. The statistic of interest is calculated in bootstrap samples, of the same size as the original, that are generated by sampling with replacement from the original data. Thus, the bootstrap is a resampling procedure. If the resampling is repeated a large number of times, the empirical distribution of the statistic generated from many bootstrap samples approximates the actual distribution. The empirical distribution may be used to construct confidence intervals (95 percent confidence limits are the 2.5 and 97.5 percentiles of the empirical distribution) or perform hypothesis tests.

▶ Results

Descriptive statistics for short-answer and uncorrected and corrected multiple-choice scores are given in Table 1⊡.

View this table:
[in this window]
[in a new window]

Table 1. Descriptive statistics for scores (%) obtained using short-answer and multiple-choice, uncorrected and corrected, questions for each of four examinations and averages (n=88)

As shown in Figure 1⊡, the average scores of four multiple-choice examinations, corrected for guessing, were clearly more in agreement with the short-answer scores than were the uncorrected multiple-choice scores. The agreement statistic (intraclass correlation coefficient) for the corrected scores was greater than the agreement statistic for the uncorrected scores for each of the four examinations and for the average (Table 2⊡). The agreement of the average of the four corrected multiple-choice examinations with the average of the four short-answer examinations (0.78, 95 percent confidence interval 0.69–0.85) was significantly (p=0.015, one-tailed test) greater than agreement of the average of the four uncorrected multiple-choice examinations with the average of the four short-answer examinations (0.71, 95 percent confidence interval 0.59–0.80). These results indicate increased validity due to applying the standard correction for guessing to multiple-choice examinations.
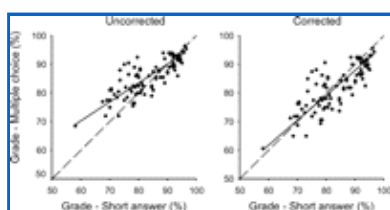

Figure 1. Scatterplots of average scores of four multiple-choice examinations, uncorrected for guessing (left panel) and corrected for guessing (right panel) with average of scores from four short-answer examinations

Note: Solid lines represent first principal components, and dashed lines represent equality.

View larger version (11K):
[in this window]
[in a new window]

View this table: **Table 2. Agreement statistics (intraclass correlation coefficients) of**

uncorrected and corrected scores from multiple-choice examinations with scores from short-answer examinations

The foregoing intraclass correlation coefficients describe the agreement among the scores for individual students. The agreement for the group also was better for the corrected scores as indicated by the lines of the first principal components (Figure 1⊡ and Table 3⊡). The line for the corrected multiple-choice scores was substantially closer to the line of equality in both slope (p=0.001) and intercept (p=0.001) than was the line for uncorrected scores. However, for both uncorrected and corrected multiple-choice scores, the slope was significantly different from one (p=0.002 for uncorrected and p=0.016 for corrected), and the intercept was significantly different from zero (p=0.002 for uncorrected and p=0.048 for corrected).

View this table:
Table 3. Slope and intercept for equations of straight lines for first principal components between uncorrected multiple-choice scores and short-answer scores and corrected multiple-choice scores and short-answer scores

We computed reliability separately for short-answer and multiple-choice examinations across the four examinations for the entire course. The reliability of the multiple-choice examinations (48.4 percent, 95 percent confidence interval 37.5–57.2) was not significantly (p=0.1225) different from the reliability of the short-answer examinations (43.1 percent, 95 percent confidence interval 32.5–53.6). Reliability was unaffected by the linear transformation used to correct for guessing if all questions were answered as was the case in our retrospective study; thus, reliability was the same for the uncorrected and corrected multiple-choice examinations.

The correction for guessing resulted in lower grades for students as indicated graphically in Figure 1⊡. This effect on the overall means is given in Table 1⊡. To further define the effects of correction, we classified students based on our classification of A (90–100), B (80–89), C (70–79), and F (0–69) using the average of the four short-answer examinations (corresponds to the horizontal axis of Figure 1⊡). The multiple-choice grades were lowered an average of 2.1, 3.8, 4.6, and 6.6 points for the A, B, C, and F categories respectively by the correction for guessing. The correction lowered scores more for those students with lower grades where presumably there was a greater degree of guessing.

The average difference between uncorrected and corrected multiple-choice examinations and the short-answer examinations for each of the grade categories are given in Table 4⊡ (uncorrected multiple-choice $F(3,84)=23.6$, p<0.0001; corrected multiple-choice $F(3,84)=8.44$, p<0.0001). The uncorrected multiple-choice examination scores were significantly (p≤0.05) higher than the short-answer scores for the C and F categories; for the A and B categories, the uncorrected multiple-choice scores were not significantly different from the short-answer scores. For the F category, the corrected multiple-choice scores were not significantly different from the short-answer scores. The corrected multiple-choice examination scores were significantly higher than the short-answer examination scores for the C classifications. For the A and B categories, the corrected multiple-choice grade was significantly lower than the short-answer grade.

View this table:
Table 4. Average difference of uncorrected and corrected multiple-choice examinations and short-answer examinations by grade classification based on short-answer examinations

▶ **Discussion**

Figure 1⊞ and Table 2⊞ show that for individual students the agreement of the corrected multiple-choice scores with short-answer scores was significantly better than the agreement of uncorrected multiple-choice scores with short-answer scores. The principal component lines in Figure 1⊞ (that is, the single dimension that best summarizes the data from both examination formats) show that the corrected multiple-choice scores placed the group of students closer to the line of equality. These results indicate increased validity due to applying the standard correction for guessing to multiple-choice examination scores. While we cannot claim that the short-answer format better evaluates student knowledge based on these data only, we believe any question format that reduces the influence of guessing will be a better indicator of what students know or do not know on a given subject. In particular, the short-answer format examinations should provide a better measure of a student's ability to perform in clinical situations in which patients present without a set of possible choices for the diagnosis. Our use of validity refers to performance without guessing, that is, performance without "cuing." Diamond and Evans[1] report there are many studies with increased validity measures where formula scoring is used.

For medical students, Norman et al.[15] demonstrated significantly higher scores on examinations using multiple-choice questions compared to examinations using essay questions with slightly higher reliability for the multiple-choice examinations and similar measures of validity for the multiple-choice and essay examinations. In third- and fourth-year medical students, Veloski et al.[16] compared examinations using multiple-choice format questions (cued response) with examinations using uncued format questions. For the uncued questions, the students selected the answer from a numbered list of alphabetized choices so that these examinations could still be graded electronically by checking for the appropriate number. Their results indicated average scores from the cued multiple-choice examinations were 11 percent to 22 percent higher than average scores from uncued examinations. They concluded that the multiple-choice examination scores gave falsely inflated measures of abilities needed for clinical competency. Our results support this notion by showing that scores from multiple-choice format examinations when corrected for guessing better reflected the test scores on short-answer questions presented in a more clinically relevant manner. As shown in Figure 2⊞, instructors should realize that, when employing multiple-choice examinations without correcting for guessing, the standard for passing and for all grade levels is inflated, particularly at the lower end of the grading range. For example, if the minimum passing standard nominally is 70 percent on an uncorrected multiple-choice examination, then the correction for guessing shows the actual standard for passing in reality is only 62.5 percent.
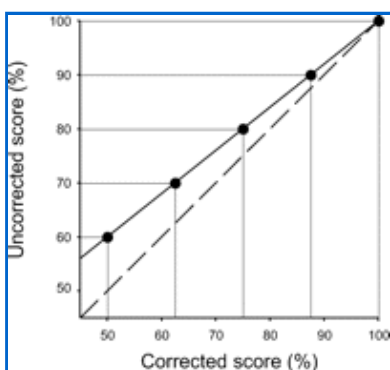

Figure 2. Relation of uncorrected score (Y-axis) and corrected score (X-axis) from a multiple-choice examination

Note: Circles highlight effect of correction at common grade cutoff points in the uncorrected score. Dashed line represents equality.

View larger version (12K):
[in this window]

The standard correction for guessing adjusts only for truly random guessing among the possible answers. If a student does not attempt an answer, zero points are awarded in the standard correction that we applied. It potentially would benefit a student to guess if they could eliminate one, two, or three of the distractors. Table 5 gives the expected gain per question if a student had partial knowledge and could eliminate one or more incorrect answers. Thus, if an instructor wishes to use this standard correction for guessing, it is imperative that all of the distractors be of uniformly high quality and should not obviously allow students to easily eliminate one or more irrelevant answers. Moreover, it also will be important to adequately shuffle the possible answers so that no pattern for the position (a, b, c, d, e) of the correct answer would be apparent from question to question.

View this table: **Table 5. Chances of guessing correctly and incorrectly and expected gain per question based on partial knowledge, that is, the ability to eliminate incorrect answers**
[in this window]
[in a new window]

Diamond and Evans[1] reported on the need for specific instructions to be given to students about guessing to allow examinations with correction for guessing to retain reliability. Students must be informed that a correction for guessing will be applied and must be shown the effect of guessing without knowledge or even with partial knowledge (the ability to eliminate one or more incorrect answers) as well as the potential benefits of partial knowledge. Later in this report, we present an example illustrating that consideration of only the expected gain is inadequate to make a decision regarding guessing. Such an example should be given to students as part of the discussion of correcting for guessing.

The reliability for the short-answer examination and the multiple-choice examination in our study was similar. Lord[4] argues that formula scoring will always improve reliability provided the student leaves at least one question unanswered. Intuitively, this can be understood as removing some random guessing component from the score and, thus, focusing on the student's actual knowledge. This advantage of formula scoring has been empirically supported and discussed in several recent studies.[17–22] Thus the use of formula scoring (corrected multiple-choice examinations) not only results in increased validity but also saves faculty time that would have to be spent grading short-answer examinations. Increasing the number of questions that are included in the multiple-choice examinations potentially would result in greater reliability. This addition would not increase faculty time spent in grading but would require additional time in test preparation.

In striking contrast to those students with lower grades in this course, we observed that those students with high grades performed better on the short-answer examinations than on the multiple-choice examinations. This may reflect deficiencies or confusion in the multiple-choice examinations that are detected by the better-scoring students with substantial knowledge of the subject matter. This interpretation is consistent with results from factor analysis[23] where an additional small dimension of knowledge was supported with uncued questions in testing students.

Choppin[3] points out that correction for guessing addresses three concerns: 1) guessing introduces a random factor into test scores that adversely lowers reliability and validity, 2) expected correct guesses inflate estimation of students' abilities, and 3) the inflation from guessing can be an unfair advantage for students who guess frequently when compared to students with equal ability who do not guess. Applying

the correction for guessing reduces the advantage for students who guess frequently. Our study clearly shows the inflated grades on multiple-choice examinations. Thus, the multiple-choice examination scores were brought into better agreement with the short-answer scores by the standard correction for guessing, indicating increased validity of the corrected multiple-choice tests. This increased validity supports the side of the controversy in recent literature[17]–[22] that favors the use of formula scoring. We have not performed the exercise of applying a correction for guessing to a multiple-choice examination after thoroughly informing students of the procedure and comparing these grades with a short-answer examination. Nonetheless, the results presented here would predict a positive result for such an undertaking—that is, increased validity.

▶ **Example**

The following example illustrates the effects of the standard correction for guessing and will provide considerations that must be addressed by students who contemplate guessing. Suppose that on a fifty-question multiple-choice examination with five possible answers per question, a student had the following result: five questions not answered, eight incorrect answers, and thirty-seven correct answers. We assume that the questions answered incorrectly represent misunderstanding of the material; that is, the student thought he or she knew the correct answer but, in fact, did not. On the questions not answered, the student admitted a complete lack of knowledge. The corrected score is computed using the formula previously described in this article as follows:

$$\text{Corrected Score} = \frac{0 \bullet 5 + (-\tfrac{1}{4}) \bullet 8 + (1) \bullet 37}{50} = \frac{0 - 2 + 37}{50} = \frac{35}{50} = 0.70 = 70.0\%$$

If the student had instead tried to guess the correct answer to the five questions left blank, we would *expect* the student to answer one correctly, yielding twelve incorrect answers and thirty-eight correct answers with an uncorrected score of 38/50=76%. Applying the correction algorithm to the hypothetical result with twelve wrong answers and thirty-eight correct answers yields:

$$\text{Corrected Score} = \frac{0 \bullet 0 + (-\tfrac{1}{4}) \bullet 12 + (4) \bullet 38}{50} = \frac{0 - 3 + 38}{50} = \frac{35}{50} = 0.70 = 70.0\%$$

Since there is no difference in the outcome regardless of whether a student guessed or left a question unanswered, why should a student not make random guesses? While we may expect students to answer one question correctly, there is a chance that they will guess the correct answer less frequently than expected. Similarly, they might guess better than expected. These probabilities of these different outcomes are given in Table 6▪. There is about a one-third chance that the student will lower his or her grade and less than a one-third chance (0.262) that this student will improve his or her grade by guessing. Leaving questions unanswered does not expose the student to the risk of lowering the grade by achieving less than the expected success due to guessing. This is perhaps a critical decision for those students at the cutoff point for passing (70 percent) in our oral and maxillofacial pathology course.

View this table:
[in this window]
[in a new window]
Table 6. Probability of answering various numbers of five questions correctly by guessing and resulting corrected score on a fifty-question test with thirty-seven questions answered correctly, eight questions answered

incorrectly, and the remaining five questions answered by guessing

Many instructors would consider the unanswered questions as incorrect and award a score of 37/50 or 74 percent. If students know that no correction for guessing will be applied, they would be foolish not to answer all questions. Computing a raw score ignores the different information that may be contained in the unanswered questions compared to the incorrectly answered questions.

Suppose the student could eliminate one, two, or three possible answers for each of the five questions left blank. The probabilities of a correct guess for each question are 1/4, 1/3, or 1/2, respectively. The probabilities of various numbers of correct answers under these circumstances are given in Table 7◼. Although students would lessen their chances of a lower (and failing grade) and improve their chances of a higher grade due to guessing, it still would seem prudent for students to avoid the risk of the lower grade even if they can eliminate two incorrect choices. Only if they could eliminate three incorrect answers would the decision to guess be a wise one. Decisions made by students having greater knowledge and thus higher grades than our example student might well make different decisions. That is, they might be more willing to gamble to achieve a higher grade.

View this table: [in this window] [in a new window]	**Table 7. Probability of answering various numbers of five questions correctly by guessing if one, two, or three possible answers can be eliminated from each of five questions guessing and resulting corrected score on a fifty-question test with thirty-seven questions answered correctly, eight questions answered incorrectly, and the remaining five questions answered by guessing**

Suppose this student had a better knowledge of what he or she didn't know and did not give an answer to five of the eight incorrect answers. That is, the student had the same number of correct responses (thirty-seven) with ten unanswered and three incorrect. The corrected score is:

$$\text{Corrected Score} = \frac{0 \bullet 10 + (-\frac{1}{4}) \bullet 3 + 37}{50} = \frac{0 - \frac{3}{4} + 37}{50} = \frac{36.25}{50} = 0.725 = 72.5\%$$

In this case, students would be rewarded for recognizing when they cannot do more than make a random guess.

▶ Conclusion

By comparing uncorrected and corrected for guessing scores on multiple-choice examinations with scores on short-answer examinations, we demonstrated that dental students have been guessing at a level close to that anticipated due to random guessing. In this retrospective analysis, applying the standard correction for guessing increased the validity of the multiple-choice examination in that the corrected scores agreed better with the scores on short-answer examinations presented in a more clinically relevant context.

▲ **Top**
▲ **Abstract**
▲ **Methods**
▲ **Results**
▲ **Discussion**
▲ **Example**
▪ Conclusion
▼ **References**

This study suggests that instructors using multiple-choice examinations should either correct for guessing or take into account the effect of guessing in setting the standard for minimal passing and, in fact, for all grade levels.

▶ **Footnotes**

Dr. Prihoda is Associate Professor, Dr. Pinckard is Professor, Dr. McMahan is Professor, and Dr. Jones is Professor—all in the Department of Pathology, University of Texas Health Science Center at San Antonio. Direct correspondence and requests for reprints to Dr. Anne Cale Jones, Department of Pathology, University of Texas Health Science Center at San Antonio, 7703 Floyd Curl Drive, San Antonio, TX 78229-3900;   210-567-4122   phone;   210-567-2303   fax; jonesac@uthscsa.edu .

▶ **REFERENCES**

1. Diamond J, Evans W. The correction for guessing. Rev Educ Res 1973;43:181–91.
2. Crocker L, Crocker L, Algina J. Introduction to classical and modern test theory. New York: Holt, Rinehart, and Winston, 1986.
3. Choppin BH. Correction for guessing. In: Keeves JP, ed. Educational research, methodology, and measurement: an international handbook. Oxford: Pergamon Press, 1988: 384–6.
4. Lord FM. Formula scoring and number-right scoring. J Educ Measurement 1975;12:7–12.
5. Davis FB, Fifer G. The effect on test reliability and validity of scoring aptitude and achievement tests with weights for every choice. Educ Psychol Measurement 1959; 19:159–70.
6. Shrout PE. Measurement reliability and agreement in psychiatry. Stat Methods Med Res 1998;7:301–17.[Abstract/Free Full Text]
7. Fleiss JL. Statistical methods for rates and proportions. New York: John Wiley & Sons, 1973.
8. Bland JM, Altman DG. Statistical methods for assessing agreement between two methods of clinical measurement. Lancet 1986;327:307–10.
9. Morrison DF. Multivariate statistical methods. Belmont, CA: Brooks/Cole Thomson Learning, 2005.
10. Seber GAF. Multivariate observations. New York: Wiley, 1984.
11. Draper NR, Smith H. Applied regression analysis. New York: John Wiley & Sons, 1998.
12. Shrout PE, Fleiss JL. Intraclass correlations: uses in assessing rater reliability. Psychol Bull 1979;86:420–7.
13. Winer BJ. Statistical principles in experimental design. New York: McGraw-Hill, 1971.
14. Efron B, Tibshirani RJ. An introduction to the bootstrap. New York: Chapman & Hall, 1993.
15. Norman GR, Smith EKM, Powles ACP, Rooney PJ, Henry NL, Dodd PE. Factors underlying performance on written tests of knowledge. Med Educ 1987;27:297–304.
16. Veloski JJ, Rabinowitz HK, Robeson MR. A solution to the cueing effects of multiple-choice questions: the Un-Q format. Med Educ 1993;27:371–5.[Medline]
17. Muijtjens AMM, van Mameren H, Hoogenboom RJI, Evers JLH, van der Vleuten CPM. The effect of a "don't know" option on test scores: number-right and formula scoring compared. Med Educ 1999;33:267–75.[Medline]
18. Burton RF. Misinformation, partial knowledge, and guessing in true/false tests. Med Educ 2002;36:805–11.[Medline]
19. Downing SM. Guessing on selected-response examinations. Med Educ 2003;37:670–1.[Medline]
20. Burton RF. Guessing in selected-response tests. Med Educ 2004;38:112.[Medline]
21. Downing SM. On guessing corrections. Med Educ 2004; 38:113.
22. McHarg J, Bradley P, Chamberlain S, Ricketts C, Searle J, McLachlan JC. Assessment of progress tests. Med Educ 2005;39:221–7.[Medline]
23. Thissen D, Wainer H, Wang X. Are tests comprising both multiple-choice and free-response items necessarily less unidimensional than multiple-choice tests? an analysis of two tests. J Educ Measurement 1994;31(2):113–23.

This article has been cited by other articles:

Journal of Dental Education   ▶ HOME

D. W. Chambers, T. J. Prihoda, R. N. Pinckard, C. A. McMahan, and A. C.
Jones
Correcting for Guessing on Multiple-Choice Exams
J Dent Educ., February 1, 2007; 71(2): 193 - 196.
[Full Text] [PDF]

This Article

▶ **Abstract** FREE
▶ **Full Text (PDF)**
▶ **Alert me when this article is cited**
▶ **Alert me if a correction is posted**

Services

▶ **Similar articles in this journal**
▶ **Similar articles in PubMed**
▶ **Alert me to new issues of the journal**
▶ **Download to citation manager**
▶ © Get Permissions

Citing Articles

▶ **Citing Articles via HighWire**
▶ **Citing Articles via Google Scholar**

Google Scholar

▶ **Articles by Prihoda, T. J.**
▶ **Articles by Jones, A. C.**
▶ **Search for Related Content**

PubMed

▶ **PubMed Citation**
▶ **Articles by Prihoda, T. J.**
▶ **Articles by Jones, A. C.**