NCME Instructional Module on

Formula Scoring of Multiple-Choice Tests (Correction for Guessing)

Robert B. Frary
Virginia Polytechnic Institute and State University

Formula scoring is a procedure designed to reduce multiple-choice test score irregularities due to guessing. Typically, a formula score is obtained by subtracting a proportion of the number of wrong responses from the number correct. Examinees are instructed to omit items when their answers would be sheer guesses among all choices but otherwise to guess when unsure of an answer. Thus, formula scoring is not intended to discourage guessing when an examinee can rule out one or more of the options within a multiple-choice item. Examinees who, contrary to the instructions, do guess blindly among all choices are not penalized by formula scoring on the average; depending on luck, they may obtain better or worse scores than if they had refrained from this guessing. In contrast, examinees with partial information who refrain from answering tend to obtain lower formula scores than if they had guessed among the remaining choices. (Examinees with misinformation may be exceptions.) Formula scoring is viewed as inappropriate for most classroom testing but may be desirable for speeded tests and for difficult tests with low passing scores. Formula scores do not approximate scores from comparable fill-in-the-blank tests, nor can formula scoring preclude unrealistically high scores for examinees who are very lucky.

When multiple-choice tests began to be widely used, they were criticized because examinees could answer correctly by guessing. Many educators viewed any score gain from guessing as ill gotten. Also, multiple-choice scores were generally perceived to be "too high," because scores from comparable short-answer or fill-in-the-blank tests were lower.

Robert B. Frary is Professor and Director of Research and Measurement, Learning Resources Center, Virginia Polytechnic Institute and State University, Blacksburg, VA 24061. His specialization is educational and psychological measurement.

Series Information

ITEMS is a series of units designed to facilitate instruction in educational measurement. These units are published by the National Council on Measurement in Education. This module may be photocopied without permission if reproduced in its entirety and used for instructional purposes. Barbara S. Plake, University of Nebraska–Lincoln, has served as the editor for this module.

In response to these concerns, a scoring procedure was developed which lowered multiple-choice scores to levels more "in line" with scores from short-answer tests. Not surprisingly, this procedure was called the correction for guessing. This designation is a severe misnomer, because the resulting scores cannot reasonably be taken as approximations of those arising in the absence of guessing. Formula scoring is a more appropriate term and will be used in this instructional unit.

What Is Formula Scoring?

A simplistic view of guessing was employed in the original development of formula scoring. It was assumed that an examinee either knew the answer to a question or guessed at random among all of the choices. With 4-choice items, about 1 guess in 4 would be expected to succeed. This would result in guessing one item correctly for every 3 items guessed wrong. To reduce the test score by the number of points expected to be gained through guessing, 1 point would be subtracted for every 3 items an examinee guessed at but answered incorrectly.

This reasoning led to the formula

$$FS = R - W/(C - 1),$$

in which

FS = "corrected" or formula score
R = number of items answered right
W = number of items answered wrong
C = number of choices per item (same for all items).

When an examinee omits a response, the item is considered neither right nor wrong. To see how the formula works, consider an examinee who answers 25 right and 12 wrong on a 50-item, 5-choice test. In this case,

Number of items = 50
Number right = 25
Number wrong = 12
Number omitted = 13
Number of choices = 5

$$FS = 25 - 12/(5 - 1) = 22.$$

Note that, because the items have five choices, the formula assumes that this examinee guessed blindly on 15 items, getting 3 right and 12 wrong. Therefore, the examinee's score was reduced by 3 points. The reader is now asked to compute the formula score for an examinee who answers 18 right and 6 wrong on a 40-item, 4-choice test: ·

Number of items = ___
Number right = ___
Number wrong = ___ FS = ______.
Number omitted = ___
Number of choices = ___

The result should have been $FS = 16$. Now, suppose that a completely ignorant examinee guessed at random on all 40 items of the test in the last example. This examinee would be expected to answer one fourth, or 10, of the 40 items correctly (because there are 4 choices) and 30 wrong. Should this happen, the resulting formula score would be:

$$FS = 10 - 30/(4 - 1) = 0.$$

Using the above formula requires that all items have the same number of choices. Formula scoring can be applied, however, even if the number of choices varies. To accomplish this, each item on the test is assigned a score of 1 if the answer is correct, $-1/(C - 1)$ if it is wrong, and 0 if it is omitted. These adjusted scores for each item are then summed to obtain the formula score for the entire test. Adjusting each examinee's score on every item was impractical before the availability of computers. Consequently, in earlier times, formula-scored tests were effectively limited to having the same number of choices for all items. Some individuals have overgeneralized this constraint, incorrectly interpreting it to mean that all multiple-choice items on a given test should have the same number of items regardless of how the scores are to be computed.

What Should Examinees Be Told About Formula Scoring?

For a number of years after the introduction of formula scoring, it was common to instruct examinees simply not to guess if unsure of the answer. The late psychometrician Frederick B. Davis (1967) challenged this practice, noting that examinees who disregarded the instructions and did guess were likely to make higher formula scores than examinees of similar ability who followed the directions. This is the case when an examinee is able to eliminate one or more of the wrong choices and guesses among those remaining. For example, suppose that an examinee does not know the answer to any of 15 4-choice items but can eliminate 1 wrong choice on each. By randomly guessing among the remaining 3 choices on each item, this examinee would be expected to guess one third of the 15 items correctly. Formula scoring would adjust this examinee's score as follows:

Number of answers guessed right = $(1/3)15 = 5$
Number of answers guessed wrong = 10
Formula score on 15 items = $5 - 10/(4 - 1) = 1\ 2/3$.

Thus the examinee gains 1 2/3 points from guessing. Now the reader is asked to determine how many points an examinee would be expected to receive under formula scoring from guessing the answer to all 24 items of a test if each item has five choices and the examinee can eliminate two wrong choices per item:

Number of answers guessed right = ___
Number of answers guessed wrong = ___
Formula score on 24 items = ________.

This examinee should guess 8 right and 16 wrong and gain 4 points from guessing.

The profitable guessing just illustrated would obviously be unfair if other examinees followed formula-scoring instructions that called for refraining completely from guessing. Therefore, almost all formula-scoring instructions in use today, while advising avoidance of completely blind guessing, do encourage examinees to guess whenever they can eliminate a wrong choice. Davis recommended the following instructions:

> Your score will be the number of items you mark correctly minus a fraction of the number you mark incorrectly. You should answer questions even when you are not sure your answers are correct. This is especially true if you can eliminate one or more choices as incorrect or have a hunch or feeling about which choice is correct. However, it is better to omit an item than to guess wildly among all of the choices given. (1967, p. 43)

Regardless of the scoring method, examinees should be clearly informed of the answering strategy that will optimize their scores. For example, if a classroom or other test is to be scored number right, instructions such as the following should be provided:

> Your score on this test will equal the number of items you answer correctly. No points will be subtracted for wrong answers. Therefore, you should answer every question, even if your answer must be based on a guess.

It is important for teachers, guidance counselors, and others who deal with testing to make the examinees aware of the fact that these and similar instructions are designed to prevent score losses from failure to guess when it is to one's advantage to do so. Some examinees may need encouragement to guess under formula scoring when they can eliminate only one wrong choice, especially if they are not very certain of its incorrectness. Others can benefit from training in the use of their knowledge to eliminate at least some of the wrong answers to multiple-choice items.

One might wonder why examinees would need to be encouraged to guess when they have partial information, however uncertain. After all, formula scoring is *not* designed to penalize guessing but simply to adjust scores for the gains due to random guessing. One reason is that some examinees harbor the mistaken belief that they are likely to lower their scores by guessing at random. The prevalence of this incorrect belief has been confirmed through survey research. Furthermore, a number of other studies have shown that when examinees reconsider and answer items previously omitted, their responses are right more often than chance would predict. Therefore, it seems reasonable to conclude that at least some examinees may be expected to gain points under formula scoring if they routinely answer *every* item.

What Does Formula Scoring Accomplish?

Guessing results in a dilemma for psychometricians. Regardless of whether it is completely at random or occurs after the elimination of wrong choices, guessing tends to make score differences between examinees result partly from luck as opposed to true differences in ability. On the other hand, even under formula scoring, the examinees must do some guessing to maximize their scores. What formula scoring attempts to accomplish is some reduction in the amount of guessing that otherwise occurs, namely, guessing totally at random as a result of complete ignorance. If the score differences due to guessing of this sort are forestalled, the resulting scores should be more reliable and valid.

Empirical research has not consistently demonstrated advantages or disadvantages from the use of formula scoring. Generally, benefits and detriments have been found to be minimal. This variation in results is probably due to many factors. For example, if there is relatively little total ignorance among examinees with respect to the test items, one would not expect formula scores to be very superior—about as much guessing would go on under one type of scoring as the other. Another possibility is that examinees may largely disregard the admonition against guessing when totally ignorant. In this case also, the psychometric quality of formula scores will tend to be about the same as that of number-right scores.

Relationship of Formula Scores to Number-Right Scores

Suppose that a test was given under formula scoring instructions and then again under number-right instructions to the same examinees. Furthermore, imagine that, perhaps through some magic, they completely forgot about the first administration by the time of the second but were otherwise in the same state of preparation for the test as before. The concept of a *linear transformation* is useful for describing the relationship between the resulting formula scores and number-right scores.

A brief explanation of the term "linear transformation" is now provided. A linear transformation is accomplished by multiplying all examinees' scores by the same number or adding the same number of points to each score. These changes do not affect the relative standings of the examinees. Simple examples of linear transformations would be doubling every examinee's score or adding 20 to every score. A more complicated example would be to multiply everybody's score by 1.5 and then add 3. This linear transformation is illustrated in Figure 1 using hypothetical scores for five examinees. This modification of the scores is called a linear transformation because of the straight line in the resulting graph. All this linear transformation did was add 3 to everybody's score and make what used to be a 1-point score difference (between two examinees) into a score difference of 1.5 point. It should be clear that a linear transformation of scores does not give any examinee an advantage with respect to any other examinee. Therefore, the scores resulting from a linear transformation are equivalent to the original scores for comparing examinees with respect to each other. Moreover, knowing the transformation, one could return to the original scores if desired.

An interesting example of a linear transformation occurs when no examinee omits any item when taking a multiple-choice test. Then the resulting formula scores are a linear transformation of the scores obtained simply by counting the right answers. This fact is fairly well known, though of little

Student	Original Score X	Linearly Transformed Score 1.5X + 3
Joe	10	18
John	6	12
Sue	8	15
Alma	4	9
Mark	2	6

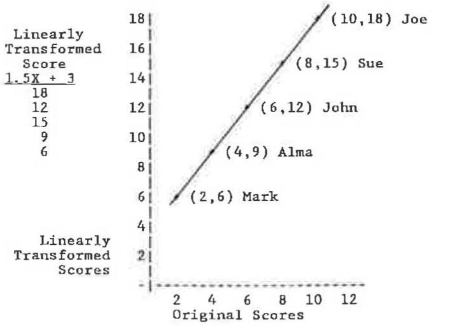

FIGURE 1. *Linear transformation of the scores of five examinees*

practical importance; there are almost always some omissions when formula scoring is applied.

There is a less well-known relationship between formula scores and number-right scores which does not depend on the absence of omissions. This relationship is true only for examinees who have perfectly average luck, but it is approximately correct for all examinees other than the very few with exceptionally good or bad luck. Specifically, *formula scores of examinees with perfectly average luck are a linear transformation of their number-right scores, provided that they make no omissions under number-right and guess whenever they can eliminate a wrong choice or more often under formula scoring.*[1] This means that, regardless of the extent to which examinees guess more than the formula-scoring instructions advise, the resulting scores provide the same information as number-right scores, except for minor differences due to luck at guessing. Thus, formula scoring offers no potential for providing basically different information about an examinee's ability than would be available from number-right scoring.

Even with knowledge of this linear relationship, formula scores might mistakenly be thought to estimate the scores examinees would obtain if guessing were not possible. In other words, one might conclude that a formula score on a multiple-choice test is equivalent to the number-right score the examinee would obtain on a comparable fill-in-the-blank test. This conclusion would be in error. Previous illustrations have shown that when an examinee does not know the answers to some items but guesses after elimination of wrong choices, the expected formula score *over those items* will not be zero.

Interpreting Formula Scores

Interpretation of formula scores can be discussed conveniently according to whether they arise from a norm-referenced or a criterion-referenced test. Most classroom and standardized tests are norm referenced. This means that their scores cannot be interpreted as absolute measures of knowledge or achievement. Although they do not give information such as what percentage of some body of knowledge has been mastered by an examinee, scores from norm-referenced tests can be used to compare examinees with each other (e.g., for assignment of grades) or to relate an examinee's score to the score distribution for some reference group (e.g., a national sample of sixth grade students). In the case of scores from norm-referenced tests, there is absolutely no distinction between how formula or number-right scores should be interpreted. Although formula scores are generally lower than number-right scores from the same test, the information they provide about the relative standings of the examinees is equivalent.

A criterion-referenced test does provide scores that can be interpreted in an absolute sense. For example, consider a list of spelling words with 1,000 entries. A test might consist of 100 of these words chosen at random. If the words are dictated for the examinees to write and an examinee spells 85 correctly, it would be reasonble to estimate that this examinee is able to spell 85%, or 850, of the words on the list. If this criterion-referenced spelling test is transformed into a multiple-choice format, the scores will tend to be higher not only because of guessing success (even under formula scoring) but also because examinees can recognize answers when unable to produce them. However, there is no theoretical or a priori basis for predicting the extent of these phenomena. Therefore, determining the relationship of multiple-choice test scores to any "real-world" criterion requires experimental studies to determine the extent of recognition and guessing success, regardless of how the test is to be scored. This relationship will be different for formula scores and number-right scores, but either could be used to estimate an examinee's actual criterion performance.

Sometimes it is desired to interpret test scores on an absolute basis in the absence of any interest in relating them to a practical criterion, such as the proportion of words from a spelling list that an examinee can write correctly. For example, a multiple-choice test might consist of 100 items drawn at random from a large and comprehensive pool of items concerning American history. Then, a reasonable substitute criterion might be an examinee's score over the entire pool expressed as a percentage. Either number-right or formula scoring could be used for this purpose. A number-right score from the 100-item test expressed as a percentage would estimate the number-right percentage on all the items of the pool, and similarly for formula scoring. In this case, it was presumably of no interest to estimate the proportion of pool items an examinee could answer correctly in writing. Nevertheless, care must be taken not to think that a formula score, expressed as a percentage of the number of items, estimates this quantity.

Negative Aspects of Formula Scoring

The erroneous belief that formula scoring is designed to penalize guessing could certainly lead an examinee to adopt a test-taking strategy more conservative than that called for by the formula-scoring instructions. Undoubtedly, some examinees do harbor this belief. When they act on it, omitting items for which they can eliminate wrong choices, they tend to earn scores lower than they deserve. Other examinees may exhibit the same behavior because of personality factors, such as timidity or reticence. Various studies have yielded conflicting results as to the prevalence of this counterproductive omissive behavior. Regardless, it seems certain that inappropriate omissions harm some examinees. Therefore, any justification for the use of formula scoring must take into account the harm to these individuals.

Although the above discussion centered on examinees who omit to their detriment, an interesting reversal of this outcome is possible. Suppose that some examinees have misinformation, that is, for certain test items they believe that the answers are distractors. Then, according to the instructions, they should guess among the remaining choices, which will yield a wrong answer and a negative score component under formula scoring. In this case, the reticent examinee with misinformation (who refrains from guessing) is at an advantage. This outcome is no more desirable than the one discussed above. The advantage gained by the reticent examinees effectively places a burden on those who follow the instructions.

Formula scoring can increase the time required to take a test. This effect may not be the same for all examinees. For some, it may be slight, but others may agonize over many items, trying to decide whether they have an adequate basis for guessing. If testing time is adequate for all examinees, there is probably no ill effect. If some examinees have omissions after their last responses, however, there is reason for concern. If they were slowed by formula scoring more than other examinees and could have at least eliminated some wrong choices on unreached items, they have been penalized. Less deliberative examinees with the same levels of knowledge will finish more items and make higher scores. As was the case with respect to omitting after elimination of wrong choices, justification for use of formula scoring must take into account the harm to individuals it prevents from reaching all test items regardless of the number of examinees involved.

Failure of examinees to finish because of formula scoring may also cause problems with respect to test validity. If important test content near the end of the test is not answered, the scores will not reflect knowledge in this area.

A more subtle negative aspect of formula scoring is related to misconceptions prevalent in the general public and among educators. For example, it is widely believed that formula scoring eliminates score gains due to lucky guessing. Of course, it does nothing of the kind. An examinee with an exceptional run of good luck will do as well (relatively) on a formula-scored test as on one scored number right. Only to the extent that the instructions curtail guessing does formula scoring blunt the effect of an examinee's good luck. (Then, among the lucky, the audacious gain more than the compliant.) Such a misconception can engender false confidence in testing, which can lead to overinterpretation and misuse of scores. The belief that formula scores represent what examinees would have made in the absence of guessing can have a similar effect.

When Should Formula Scoring Be Used?

There are circumstances in which the use of formula scoring should be considered notwithstanding its undesirable characteristics. These are all cases in which large proportions of examinees are unable to eliminate even a single wrong choice on many items. Two basic examples are highly speeded and very difficult tests.

Highly speeded tests. These are tests for which speed of response is a major aspect of the ability being measured. On a test of this kind, most examinees should make very high scores if given unlimited time. If scored number right, examinees would be foolish not to answer unreached items at random at the end of the time limit. This action would add substantial score components varying with the luck of the examinees.

Difficult tests with low score requirements. For some tests, there may be substantial proportions of examinees who have no basis at all for answering many items but who nevertheless might be expected to gain acceptable or adequate scores. In this case, reducing the massive amount of random guessing they would have to do under number-right scoring may be desirable. Tests of this type are often qualifying tests for employment or admission to educational programs. Note that invoking formula scoring under these circumstances will not prevent a lucky guesser who disregards the instructions from making an unrealistically high score.

Formula scoring is generally inappropriate for classroom testing. If instruction has been reasonably effective and the test represents a reasonable sample of what examinees should know, it is unlikely that very many examinees will be totally ignorant on very many items. Then formula scoring will have little effect, because there will be very few omissions. For the same reason, professional licensing examinations should usually be scored number right. An exception might be a licensing examination with a very low passing score, such as some in use for teacher certification.

The decision on whether to use formula scoring also depends on the following factors:

1. Ability of examinees to follow instructions. Some groups of examinees may be largely unable to understand the relatively complex instructions concerning when to guess and when not to do so.

2. Resources used by formula scoring. Invoking formula scoring requires the provision of more complex examinee instructions, additional steps in scoring, and efforts to inform score users of the nature and effect of formula scoring.

3. The extent of improvement of the psychometric characteristics of the test scores (reliability and validity). Even in the two examples given above, where formula scoring might be desirable, its use is not justified if meaningful improvement of the psychometric characteristics of the scores does not result.

4. The extent of harm to examinees. Before formula scoring is used repeatedly or extensively in any testing situation, research should be done to determine the extent of inappropriate omissive behavior, either failure to guess after classifying choices as wrong or failure to reach items at the end of a test intended to be unspeeded.

Note

[1]The actual linear transformation is presented here. Provided that examinees make no omissions under number-right instructions and guess at least as often as the formula-scoring instructions advise, their expected (average luck) formula scores and number-right scores are related as follows:

$$E(FS) = (C/(C - 1))E(NR) - N/(C - 1),$$

in which $E(FS)$ and $E(NR)$ are the expected formula and number-right scores, C is the number of choices, and N is the number of test items.

References

Davis, F. B. (1967). A note on the correction for chance success. *Journal of Experimental Education, 3,* 43–47.

Annotated Bibliography

Abu-Sayf, F. K. (1979, June). The scoring of multiple-choice tests: A closer look. *Educational Technology,* 5–15.

This article presents an excellent review of competing test scoring methods and some insightful comments about formula scoring.

Albanese, M. A. (1986). The correction for guessing: A further analysis of Angoff and Schrader. *Journal of Educational Measurement, 23,* 225–235.

This article provides extensive further analysis of the research reported by Angoff and Schrader (1984). Methodological concerns discussed include treatment contamination and the effect of using volunteer subjects. In addition, analyses are presented which suggest that, contrary to the conclusions of Angoff and Schrader, estimated success rates on items omitted by subjects under formula scoring were above chance for one phase of their study.

Angoff, W. H., & Schrader, W. B. (1984). A study of hypotheses basic to the use of rights and formula scores. *Journal of Educational Measurement, 21,* 1–17.

Two equivalent-groups designs were used with large samples of examinees to evaluate whether omissiveness under formula scoring tended to cause scores lower than in the absence of this behavior. The authors concluded that there was no reason to believe that omissions replaced by responses in their study would be correct more often than chance would predict.

Cross, L. H., & Frary, R. B. (1977). An empirical test of Lord's theoretical results regarding formula scoring of multiple-choice tests. *Journal of Educational Measurement, 14,* 313–321.

A single-group design was employed to test whether items omitted under formula scoring would be answered with above-chance-level success under number-right instructions. This was found to be the case, and the gains from items omitted under formula-scoring instructions were found to be related to personality factors.

Diamond, J., & Evans, W. (1973). The correction for guessing. *Review of Educational Research, 43,* 181–191.

This comprehensive review recounts a wide range of seemingly contradictory results from the use of formula scoring. Limitations on use of the results are detailed as are various methodological deficiencies. Also reviewed are opinions and commentary concerning formula scoring.

Ebel, R. L., & Frisbie, D. A. (1986). *Essentials of educational measurement* (4th ed.). Englewood Cliffs, NJ: Prentice-Hall.

This text is an example of several current ones which provide comprehensive and correct discussions of formula scoring. Included is the alternative formula-scoring procedure under which points are added for omissions rather than subtracted for wrong answers. Criteria are given for deciding whether to use formula scoring which would rule out this use in most cases.

Frary, R. B. (1982). A simulation study of reliability and validity of multiple-choice test scores under six response-scoring modes. *Journal of Educational Statistics, 7,* 333–351.

Formula scoring was one of the six response-scoring modes investigated in this simulation study involving unspeeded tests of moderate difficulty. No discernable reliability or validity advantage was observed for formula scoring.

Lord, F. M. (1975). Formula scoring and number-right scoring. *Journal of Educational Measurement, 12,* 7–12.

This frequently cited article points out that expected formula scores are a linear function of expected number-right scores provided that examinees make no number-right omissions and guess at least as often under formula scoring as the instructions indicate. Lord goes on to prove that if omissions under formula scoring would be replaced by random guesses under number-right scoring, then the formula scores would be superior estimators of examinee ability.

Rowley, G. L., & Traub, R. E. (1977). Formula scoring, number-right scoring, and test-taking strategy. *Journal of Educational Measurement, 14,* 15–22.

Data are presented which suggest that when examinees believe themselves to be totally ignorant on test items, they nevertheless can answer such items at better than a chance rate of success. Interesting commentary is provided on the motivation for preferring number-right or formula scoring. This may be the first published article commenting to the effect that if there are even a few examinees whose judgment is faulty in this manner, they are placed at an unfair disadvantage and that the validity of their scores is thus lessened because of measuring something different from what was intended.

Self-Test

Select the single *best* answer for each of the following questions. Your score will be the number of right answers. Answer all questions regardless of knowledge.

1. The purpose of formula scoring is to
   a. reduce score differences between examinees due to luck.
   b. prevent lucky guessers from getting unrealistically high scores.
   c. produce scores approximating those from a comparable non-multiple-choice test.
   d. penalize those who guess contrary to instructions.
2. Instructions for a formula-scored test should
   a. advise the examinees not to guess at all.
   b. advise the examinees to guess only when they have narrowed the answer to two choices.
   c. advise the examinees to answer every question unless totally ignorant.
   d. avoid reference to how the test will be scored.
3. On a formula-scored test, an examinee who needs a score much higher than his or her ability would predict should
   a. follow the formula-scoring instructions as closely as possible, because this is the only way to maximize one's score.
   b. guess at every opportunity regardless of knowledge.
   c. guess less than the instructions advise to avoid losing score points.
4. Suppose a group of examinees took a test many, many times both under formula scoring and number-right instructions. Each examinee received two final scores: the average of all the number-right scores and the average of all the formula scores. Which of the following best indicates the relationship between these two final averages?
   a. Any examinee's average formula score could be estimated very accurately from his or her average number-right score.
   b. Any examinee's average number-right score could be estimated very accurately from his or her average formula score but not vice versa.
   c. There is no reasonable basis for estimating an average number-right score from an average formula score because one contains a guessing component that the other score does not.
5. If an examinee gets 30 right and omits 6 on a 40-item test with 5-choice items, the formula score [$FS = R - W/(C - 1)$] is
   a. 28.5 b. 28.8 c. 29.0 d. 29.2 e. none of these.
6. Suppose that, on a formula-scored test, all of the examinees answer all of the items.
   a. The test could be scored number right and provide exactly the same information about the examinees as if it were formula scored.
   b. The formula scores will nevertheless be corrected for guessing and hence more indicative of the examinees' knowledge.
7. Guessing much more often than the formula-scoring instructions suggest is likely to result in a score
   a. somewhat lower than in the absence of this behavior.
   b. somewhat higher than in the absence of this behavior.
   c. rather close to the score that would have occurred in the absence of this behavior.
8. An examinee can eliminate two wrong choices on each of 12 5-choice items. The examinee then guesses at the answer to all 12 items. How many guesses would you expect to be correct?
   a. none b. two c. three d. four e. six
9. Formula scoring can be used only for tests or sections of tests where all the items have the same number of choices.
   a. True b. False
10. An examinee refrains from guessing on a 5-choice test except when able to narrow the answer to two choices. How well is this examinee likely to do under formula scoring compared to others who guess MORE OFTEN than the instructions advise?
    a. Better than the others
    b. Worse than the others
    c. About the same as the others
11. Formula scoring is most likely to be desirable when a substantial proportion of examinees
    a. can eliminate one or two wrong choices on a lot of items but do not actually know the answers to these items.
    b. is completely ignorant with respect to large proportions of the test items.
    c. can pass the test only if they are lucky at guessing.
12. Use of formula scoring can be unfair to some examinees.
    a. True b. False
13. An examinee who never guesses will obtain a formula score equal to his or her number-right score.
    a. True b. False
14. For a true-false test, the formula score is determined by subtracting the number of wrong answers from the number right.
    a. True b. False
15. Interpretation of scores from a norm-referenced test depends on whether the test was administered under formula-scoring directions or number-right directions.
    a. True b. False

Answers to the Self-Test

1. a	4. a	7. c	10. b	13. a
2. c	5. c	8. d	11. b	14. a
3. b	6. a	9. b	12. a	15. b

Teaching Aids Are Available

A set of teaching aids, designed by Robert B. Frary to complement his ITEMS module, "Formula Scoring of Multiple-Choice Tests (Correction for Guessing)," is available at cost from NCME. These teaching aids consist of a test of "general knowledge" designed to illustrate the appropriateness of guessing on a formula-scored test. As long as they are available, they can be obtained by sending $2.00 to: **Teaching Aids, ITEMS Module #4, NCME, 1230 17th St., NW, Washington, DC 20036.**

TEACHING AIDS: Formula Scoring of Multiple-Choice Tests (Correction for Guessing)

Robert B. Frary

Virginia Polytechnic Institute and State University

Recommended Procedures

I. Preliminary Activity

The TEST OF GENERAL KNOWLEDGE is designed orient participants and stimulate discussion. It should be administered without comment before any actual discussion of formula scoring and may be duplicated for this purpose. Complaints about difficulty or absurdity of the items should be passed off lightly, perhaps with the suggestion that a discussion of the test will follow.

II. Discussion of Test

Participants should receive a copy of the answer key and score their tests using the formula FS = R - W/3. The discussion can be initiated by asking whether any participants had omissions. This may be followed by participant identification of wrong choices intended to be obviously wrong. A list of these may be provided. The discussion should try to determine why there were any omissions, given the obviously wrong answers and the test instructions to guess in this instance. (If there were no participant omissions, the discussion can explore why more naive examinees might have omitted items.) The purpose is to highlight misconceptions common about formula scoring and provide good motivation for the discussion to follow.

III. Instruction

The instructional module is designed to be compatible with a conventional lecture approach to instruction. However, the instructor should feel free to improvise. If possible, it it would be desirable to administer the TEST OF GENERAL KNOWLEDGE at one session and provide copies of the instructional module for study prior to the next session.

IV. Evaluation

After instruction, the multiple-choice self-test for the module should be administered with number-right scoring directions (answer all items regardless of knowledge). Participants should be provided an answer key and permitted to keep their copies of the test. A discussion of the answers is highly desirable. If the instructor wishes to use the responses for evaluation, separate answer sheets should be provided.

TEST OF GENERAL KNOWLEDGE

INSTRUCTIONS: READ CAREFULLY

Your score will be the number of questions you answer correctly MINUS A FRACTION OF THOSE YOU ANSWER INCORRECTLY. If you do not know the answer but can eliminate one or more wrong choices or have a hunch, you should guess. However, you should not guess at random among all the choices of a question.

1. What is the capital of Burkina Faso (formerly Upper Volta)?
   1) Ouagadougou
   2) Nouakchott
   3) Douala
   4) Cape Town

2. Which of the following is an Indo-European language?
   1) Polish
   2) Finnish
   3) Chinese
   4) Hungarian

3. Which one of the following has NO error of grammar or usage?
   1) Give the report to whoever is on the committee.
   2) Everybody should get their hats and coats.
   3) She don't have any money.
   4) This checkout is for customers with 12 items or less.

4. The water pollution process due to phosphates in detergents is called
   1) enzymatic action.
   2) eutrophication.
   3) acid rain.
   4) biodegrading.

5. Which of the following states has the largest land area?
   1) Washington
   2) North Dakota
   3) Delaware
   4) Oklahoma

6. Which of the following is the oldest Land Grant university?
   1) University of Georgia
   2) University of Hawaii
   3) Pennsylvania State University
   4) University of Massachusetts

7. Between which pair of cities is the airplane flight distance the longest?
   1) Chicago to Detroit
   2) Albuquerque to Atlanta
   3) Los Angeles to Tulsa
   4) Phoenix to New Orleans

8. The calculus was developed independently by
   1) Newton and Gauss
   2) Liebniz and Gauss
   3) Newton and Liebniz
   4) Newton and Einstein

9. Suriname used to be called
   1) Portugese Guinea
   2) Cambodia
   3) French Guiana
   4) Dutch Guiana

10. Which of the following was U.S. President for the shortest amount of time?
   1) Chester A. Arthur
   2) Gerald Ford
   3) Millard Fillmore
   4) Lyndon Johnson

KEY TO TEST OF GENERAL KNOWLEDGE

1. 1)	3. 1)	5. 2)	7. 4)	9. 4)
2. 1)	4. 2)	6. 1)	8. 3)	10. 2)

WRONG CHOICES INTENDED TO BE OBVIOUS ON TEST OF GENERAL KNOWLEDGE

1. 4)	3. 3)	5. 3)	7. 1)	9. 2)
2. 3)	4. 3)	6. 2)	8. 4)	10. 4)