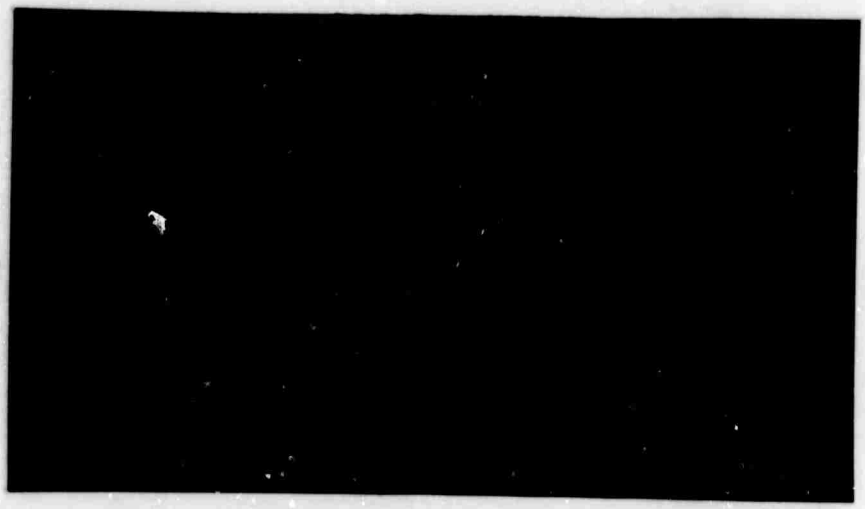
AFOSR 68-2157

AD 677205



1. This document has been approved for public release and sale; its distribution is unlimited.



THE SHUFORD-MASSENGILL CORPORATION

# WHAT PUPILS AND TEACHERS SHOULD KNOW ABOUT GUESSING

H. EDWARD MASSENGILL and EMIR H. SHUFORD, Jr.

AN INTERIM TECHNICAL REPORT OF THE WORK PERFORMED UNDER CONTRACT NUMBER AF 49 (638)-1744, ARPA ORDER NUMBER 833, BY THE SHUFORD – MASSENGILL CORPORATION, P.O. BOX 26, LEXINGTON, MASSACHUSETTS, 02173.

## FOREWORD

This is one in a continuing series of papers concerned with the theory and application of admissible confidence measurement techniques and one of a sub-series of papers concerned with the effects of guessing on the interpretation and use of objective test results in instructional settings. The research reported in this paper, prepared for the 1967 Meeting of The National Society for Programmed Instruction, was performed in support of the United States Air Force Office of Scientific Research contract number AF 49(638)-1744 sponsored by The Advanced Research Projects Agency of the Department of Defense (ARPA order number 833).

## *ABSTRACT*

*A recent advance in educational measurement has made it possible for the first time to obtain valid and reliable measurements of a pupil's degree of confidence in the answers to objective test questions. The meaning and qualitative significance of guessing on objective tests is examined by using decision-theoretic psychometrics. The origin of the "guessing problem" is traced to the conventional "number of rights" scoring system and three types of guessing are defined: blind guessing, partially blind guessing, and rational guessing. These three types of guessing are shown to affect both the reliability and validity of a pupil's test score as well as the ability of a teacher to classify pupils for remedial help. Four unsuccessful attempts to remedy the guessing problem are illustrated. This leads to an explanation of how admissible confidence measurement techniques completely eliminate guessing.*

## INTRODUCTION

In recent years it has become possible for the first time to obtain valid and reliable measurements of a person's degree of confidence (Shuford & Massengill, 1965). The basic principle of valid confidence measurement is that the measurement situation be so structured that it is in the person's best interest to honestly state his degrees of confidence. This principle fits very nicely into the framework of objective testing. The idea is to use a scoring system which makes it possible for a pupil to maximize his expected test score *if and only if* he honestly responds to each question with his degree of confidence in the correctness of each of the possible answers of the question (Shuford, Albert & Massengill, 1966).

The application of confidence measurement to educational testing calls for a reexamination of the much discussed "guessing problem" both in terms of the *meaning* and the *significance* of guessing. A look at textbooks dealing with educational testing (e.g., Ebel, 1965; Gulliksen, 1950; Cronbach, 1949; Guilford, 1936; Noll, 1957) reveals something less than a clear idea of what the term guessing includes as well as a general feeling that the presence of guessing in objective testing is not of great significance either to the pupil taking the test or to the teacher interpreting the results.

Our study of the guessing problem, based on decision theory and confidence measurement, has yielded an explicit definition of guessing. And working from this definition we have discovered that guessing is very significant both to the pupil, in terms of how it affects his test score, and to the teacher, in terms of how it affects her evaluation of the test results.

From a pupil's standpoint, total test score can often be materially *lowered* by failure to guess on a conventional objective test. From a teacher's standpoint, the ambiguity resulting from guessing (1) seriously degrades the value of selection, placement and counseling decisions, (2) significantly impairs test reliability and validity and (3) greatly limits the effectiveness of instruction.

In this paper we will consider both the meaning and the qualitative significance of guessing. We will not attempt to go into the quantitative aspects of this significance which have been reported at this meeting by Dr. Emir Shuford (Shuford & Massengill, 1967). (See also Shuford & Massengill, 1966b; 1966c.)

## THE MEANING OF "GUESSING"

First, let us examine the meaning of guessing. In order to do so we will define knowledge in a very specific way. We will equate a pupil's knowledge on a given test question with his degree of confidence that each of the possible answers to the question is the correct answer. In the following exposition we will use the multiple-choice format to illustrate our points and within this context we will discuss two-answer and three-answer

questions.

The three-answer question gives an additional insight into guessing which we do not get from the two-answer question. And, further, the results for the three-answer question can be generalized to questions with more than three answers.

## DEGREE OF CONFIDENCE REPRESENTATIONS

For a two-answer question, the pupil's total confidence is divided between the two possible answers. If we let the total confidence equal one and zero confidence equal zero, then the pupil's degree of confidence in the first answer can range between zero and one. And his degree of confidence in the second answer is, therefore, one *minus* his degree of confidence in the first answer. This means that we can use a straight line going from zero to one to represent all the possible degree of confidence patterns for a two-answer question.

Figure 1 shows four of the infinite number of patterns possible, with an arrow indicating the point on the line which represents that pattern. In the first pattern, the pupil is convinced that the second answer is correct, i.e., he has zero confidence in the first answer and complete confidence in the second. In the second pattern, the pupil has no idea which of the two answers is correct. In the third pattern, the pupil is moderately sure that the first answer is correct. And in the fourth pattern, he is completely certain that the first answer is correct.

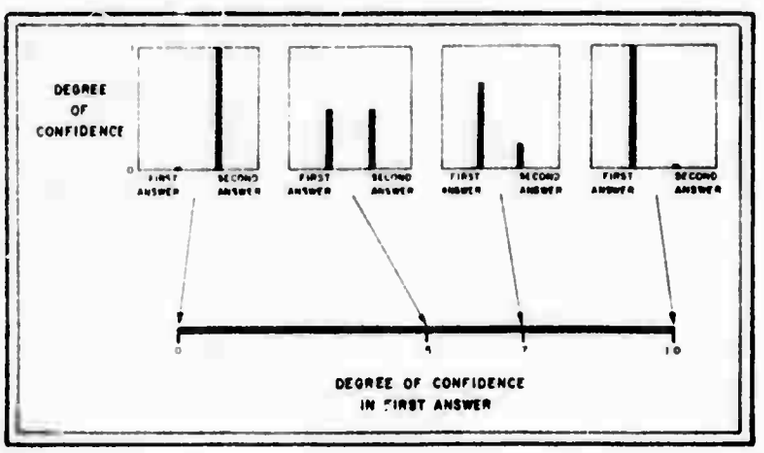



Figure 1. The use of points on a line to represent degree-of-confidence patterns for a two-answer question.

We need a two-dimensional figure to represent all of the possible degree-of-confidence patterns for a three-answer question. In this case the pupil's total confidence is divided among three possible answers. Figure 2 shows how an equilateral triangle can be used to represent the possible degree-of-confidence patterns for a question with three possible answers. Three patterns are illustrated.

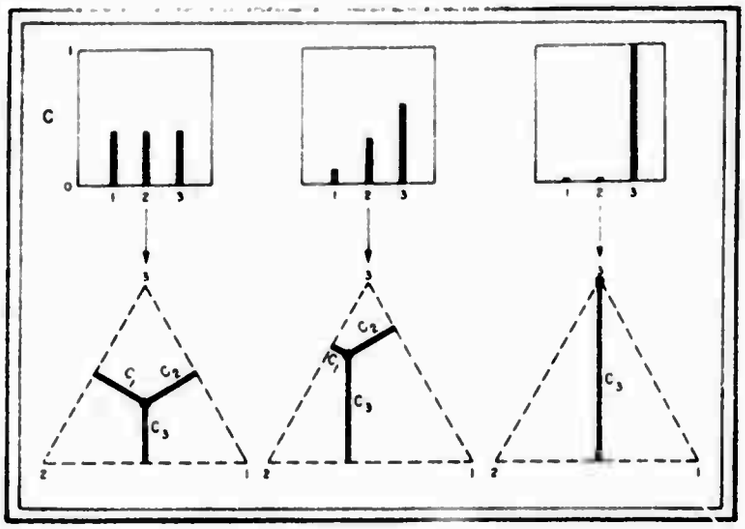



Figure 2. The use of points within an equilateral triangle to represent degree-of-confidence patterns for a three-answer question.

When the pupil has equal confidence in all three answers, the point in the middle of the triangle represents his pattern. When he has a high confidence in the third answer, a point in the upper portion of the triangle represents his pattern. When he has complete confidence in the third answer, the point at the top corner represents his pattern. Each point in the triangle corresponds to a particular confidence pattern for a three-answer question *and* there is a point for each possible pattern.

## THE ORIGIN OF GUESSING

Now we are ready to look at the origin of guessing. Until the development of valid confidence measurement procedures, we have had to be satisfied with the use of choice procedures in obtaining information about a pupil's knowledge on a test item. The most widely used choice procedure has been, of course, the conventional choice procedure in which a pupil receives one point if he picks the correct answer and zero points if he picks an incorrect answer or skips the question. It is just this scoring procedure which has led to the problem of guessing.

To see why this is so, we must look at the test situation from the point of view of the pupil. Most pupils want to maximize their test score given whatever knowledge they have at the time they take the test.

Given this desire, we can show that the pupil should respond to multiple-choice questions in the way illustrated in Figure 3. For a two-answer question he should choose the answer for which he has the largest degree of confidence. If he has equal confidence in the two answers, it is in his best interest to pick any one of the two. He should *never* skip a question.

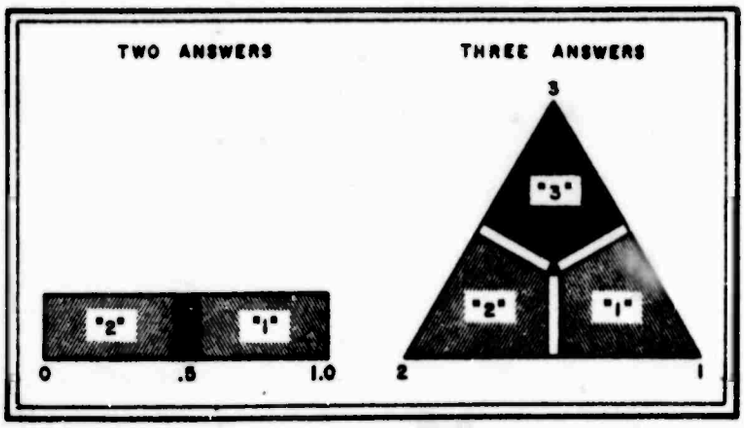



Figure 3. Optimal decision strategies for a pupil taking a conventionally-scored test.

For a three-answer question, he should choose that answer for which he has maximum confidence. But there are situations in which he may have maximum *and* equal confidence in all of the possible answers (the pattern represented by the point in the middle of the triangle). Here he should pick any one of the three answers. There are also cases where he may have maximum *and* equal confidence in two of the possible answers. Here he should choose any one of the two.

For a multiple-choice test given with the conventional scoring system, which uses the number of right answers as the pupil's test score, the pupil should always pick an answer regardless of how little he knows, even if he is completely uncertain.

This means that the teacher never knows, by looking at a pupil's answer for a given question, whether he was completely sure, moderately sure, uncertain between two or more of the possible answers, etc.

## THREE TYPES OF GUESSING

The above considerations allow us to explicitly define guessing. Actually we can define three types of guessing. (Figure 4 illustrates the portions of the confidence surface related to the three types of guessing.) The first is *blind* guessing (Ebel, 1965, p. 223ff). This occurs when the pupil is completely uninformed about the question and chooses an answer at random. For a two-answer question his confidence in each possible answer is 1/2. For a three-answer question his confidence in each answer is 1/3.

Second, there is *partially* blind guessing. This situation is defined only for questions with three or more possible answers and occurs when the pupil has maximum *and* equal confidence in some but not all of the answers and picks one at random. The third type of guessing we call *rational* guessing (Ebel, 1965, p. 230). Rational guessing occurs in those situations in which a pupil has maximum, but not complete, confidence in one of the answers and responds with that answer.

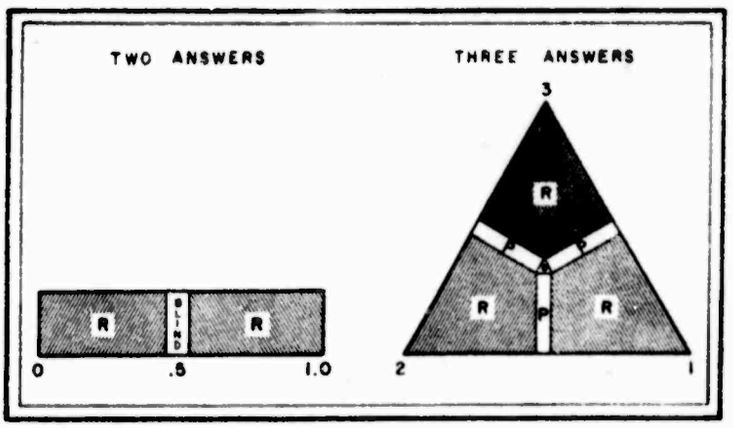



Figure 4. Confidence patterns associated with each of the three types of guessing.

For the two-answer question only two points, the end points, are not within one of the definitions of guessing. For the three-answer question only three points, the three corners, are not within one of the definitions.

## EFFECTS OF GUESSING

A study of the types of guessing defined above reveals three qualitative effects of guessing on the interpretation of multiple-choice test results. First, in terms of item score, the pupil whose answer results from blind guessing or partially blind guessing has a *chance* of obtaining the same score as the pupil who is absolutely certain of the answer. Thus, blind and partially blind guessing introduce *unreliability* into the score. Second, a pupil who has maximum, but less than complete, confidence in the correct answer will *always* obtain the same score as the pupil who is completely certain of the correct answer. Thus, rational guessing lowers the *validity* of the item score.

Finally, guessing makes it impossible for the teacher to classify the pupils into more than two categories for remedial help. The teacher can only classify the pupil as being correct or incorrect. "Correct" can include being completely uninformed, partially informed, moderately informed and completely informed. "Incorrect" can include being completely uninformed, partially misinformed and completely misinformed.

Thus guessing affects the reliability and validity of the pupil's score *and* the teachers ability to classify pupils for remedial help.

## ATTEMPTS TO REMEDY THE "GUESSING PROBLEM"

There have been many attempts to remedy the guessing problem (Shuford & Massengill, 1966a). The basic approach is to divide the confidence surface into more areas. Figure 5 shows four of the better known attempts to remedy the guessing problem:

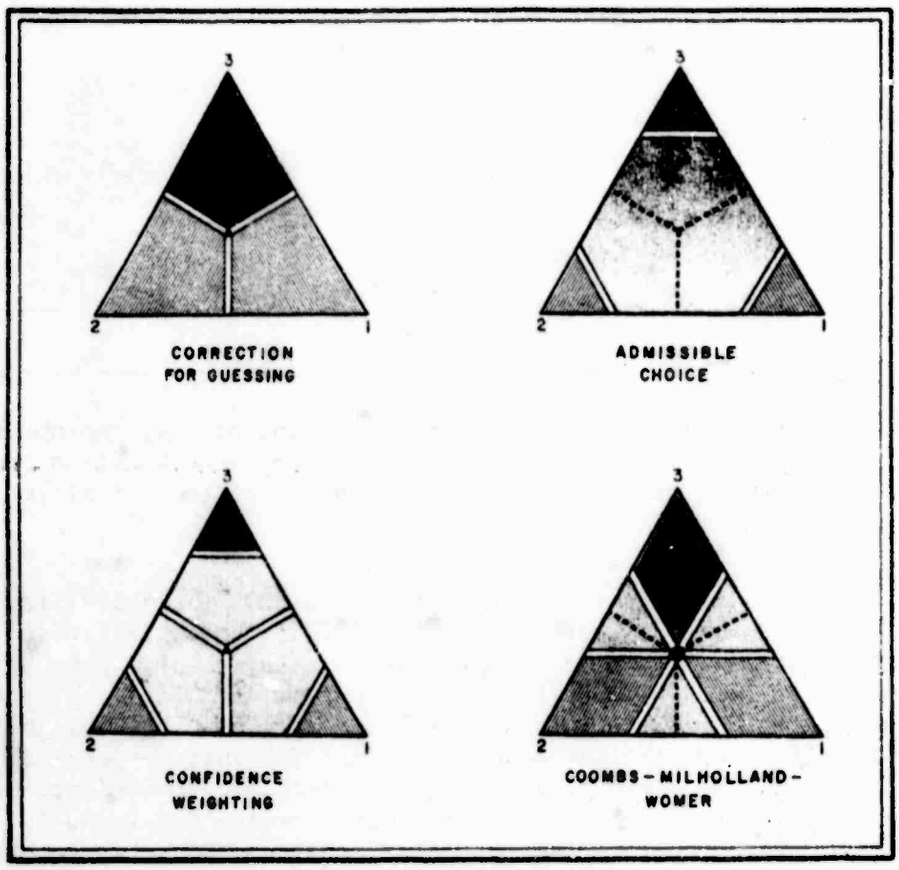



Figure 5. Four attempts to remedy the effects of guessing.

1. The correction-for-guessing system, i.e., the number of rights *minus* the number of wrongs divided by n-1 (Gulliksen, 1950).

2. The admissible choice system (Massengill & Shuford, 1965; Shuford & Massengill, 1966a).

3. The confidence-weighting system (Ebel, 1965).

4. The Coombs-Milholland-Womer procedure (Coombs, Milholland, & Womer, 1955).

Though these procedures represent a step in the right direction i.e., that of more finely dividing the confidence surface, none of them eliminates the effects of guessing which we have described above.

There is, however, a special way of dividing the confidence surface so that guessing is completely eliminated. We will illustrate with a two alternative question.

Figure 6 shows the essence of the procedure for a two-answer question. No-

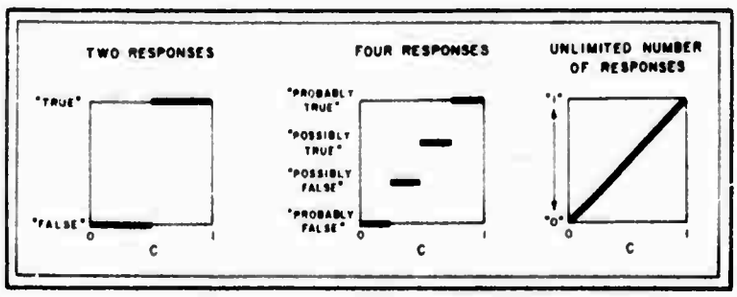



**Figure 6.** The progression from two to an unlimited number of actions resulting in the elimination of guessing.

tice that as the number of choices is increased from two to four, the pupil's degree of confidence can be more precisely pinpointed by knowing his answer. Thus for four choices, the scoring system makes it in the best interest of the pupil to say that an answer is *probably true* if his degree of confidence is greater than .75; that it is *possibly true* if his degree of confidence is between .5 and .75. Thus if he answers with *possibly true*, the teacher knows that he was at least 75% certain that the correct answer was "true".

But we needn't stop with four choices. We can continue to add choices until there is a choice corresponding to each possible confidence pattern. In this case, the pupil's answer *is* his confidence pattern, i.e., his degree of confidence in the first answer and his degree of confidence in the second answer. This means that a teacher can precisely pinpoint a pupil's degree of knowledge for a question by knowing his answer to the question.

And we needn't stop with a two-answer question. Figure 7 shows the result for a three-answer question. The triangle is no longer divided into large segments but is now a surface of points, each point corresponding to a particular confidence pattern. And, of course, the same thing can be done with a question containing any number of possible answers.

It is evident that when the pupil responds with his confidence pattern all three of the effects of guessing discussed above are eliminated. But in arriving at this response system we have had to use a very special scoring procedure, viz., one which makes it in the best interest of the pupil to respond with his confidence pattern no matter what his degree of knowledge. Such a scoring system is called an admissible confidence measurement procedure.

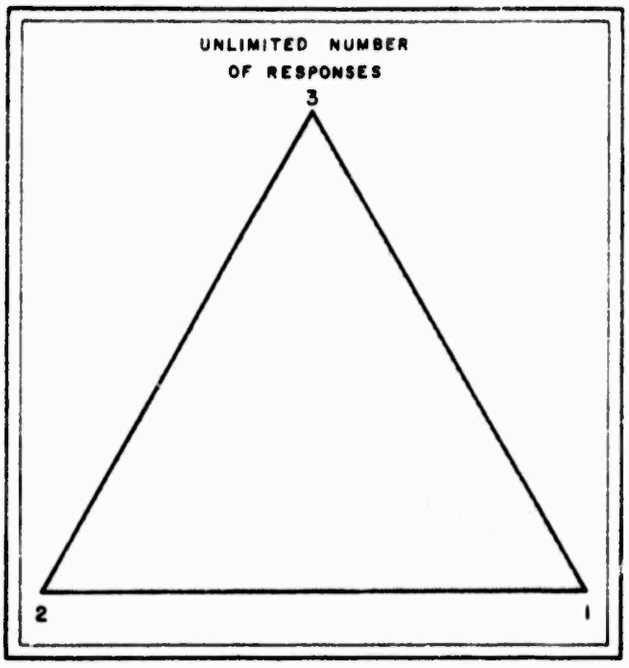



Figure 7. The three-answer response surface for admissible confidence testing.

The use of an admissible confidence measurement procedure means that the pupil is *never* put in a guessing situation of any type: blind, partially blind, or rational. It means that a pupil's item score is completely reliable and valid in terms of his degree-of-confidence pattern. And, most important, it means that the teacher can develop her own classification schemes. In other words, she can divide the confidence surface in any manner that she desires. No predetermined scheme is forced on her.

This allows a teacher to use classification schemes appropriate to her particular instructional strategies. One such pattern, given by none of the other procedures described in this paper, is shown in Figure 8. Here the pupil is classified on each item as being well informed, i.e., having a high degree of confidence in the correct answer; moderately informed; partially informed; completely uninformed; or misinformed.

## SUMMARY

Now we can summarize our advice to pupils and teachers. To the pupil: you should *always* guess when taking a test using the conventional "number of rights" scoring system or the "correction-for-guessing" scoring system. The only type of scoring system for which we could advise you never to guess is an admissible confidence scoring system.

And for the teacher: first, great care should be exercised in the use of

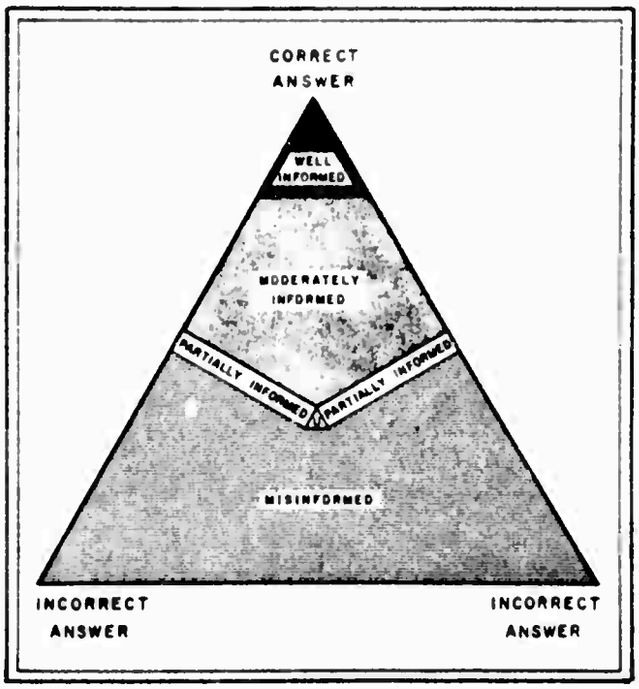



Figure 8. A classification scheme derived from admissible confidence testing responses.

choice test results, both for inferences about ability and for diagnostic decisions. Second, the new confidence procedures, the only procedures which completely eliminate guessing, are now available with answer, scoring and classification aids which make their use both natural and simple for pupils and teachers. Thus, it is now possible by using these confidence procedures (1) to improve the reliability and validity of your test scores and (2) to obtain the flexibility of being able to use classification schemes appropriate to your particular instructional strategies.

## REFERENCES


Coombs, C. H., Milholland, J. E. & Womer, F. B. (1955) The assessment of partial knowledge. *J. ed. psychol. Meas.*, 16.

Cronbach, L. J. (1949) *Essentials of psychological testing.* New York: Harper & Brothers.

Ebel, R. L. (1965) *Measuring educational achievement.* Englewood Cliffs: Prentice-Hall Inc.

Guilford, J. P. (1936) *Psychometric methods.* New York: McGraw-Hill.

Gulliksen, H. (1950) *Theory of mental tests.* New York: John Wiley & Sons.

Massengill, H. E. & Shuford, E. H., Jr. (1965) A note on admissible choice procedures. Report No. 2, The Shuford-Massengill Corporation, Box 26, Lexington, Mass., Dec.

Noll, V. H. (1957) *Introduction to educational measurement.* Boston: Houghton Mifflin Company.

Shuford, E. H., Jr. Albert, A. & Massengill, H. E. (1966) Admissible probability measurement procedures. *Psychometrika*, 31.

Shuford, E. H., Jr. & Massengill, H. E. (1965) On communication and control in the educational process. ESD-TR-65-568, Decision Sciences Laboratory, L. G. Hanscom Field, Bedford, Mass.

Shuford, E. H., Jr. & Massengill, H. E. (1966a) Decision-theoretic psychometrics: a logical analysis of guessing. Report No. 4, The Shuford-Massengill Corporation, Box 26, Lexington, Mass., Nov.

Shuford, E. H., Jr. & Massengill, H. E. (1966b) Decision-theoretic psychometrics: the effect of guessing on the quality of personnel and counseling decisions. Report No. 5, The Shuford-Massengill Corporation, Box 26, Lexington, Mass., Nov.

Shuford, E. H., Jr. & Massengill, H. E. (1966c) Decision-theoretic psychometrics: the worth of individualizing instruction. Report No. 6, The Shuford-Massengill Corporation, Box 26, Lexington, Mass., Nov.

Shuford, E. H., Jr. & Massengill, H. E. (1967) The relative effectiveness of five instructional strategies. Report No. 8, The Shuford-Massengill Corporation, Box 26, Lexington, Mass., April. (Paper read at the 1967 Meeting of The National Society For Programmed Instruction.)

UNCLASSIFIED

Security Classification

## DOCUMENT CONTROL DATA - R & D

(Security classification of title, body of abstract and indexing annotation must be entered when the overall report is classified)

1. ORIGINATING ACTIVITY (Corporate author)

The Shuford-Massengill Corporation
Box 26, Lexington, Mass. 02173

2a. REPORT SECURITY CLASSIFICATION

UNCLASSIFIED

2b. GROUP

REPORT TITLE

WHAT PUPILS AND TEACHERS SHOULD KNOW ABOUT GUESSING

4. DESCRIPTIVE NOTES (Type of report and inclusive dates)

Technical Report, May 1967

5. AUTHOR(S) (First name, middle initial, last name)

H. Edward Massengill & Emir H. Shuford, Jr.

| 6. REPORT DATE | 7a. TOTAL NO. OF PAGES | 7b. NO. OF REFS |
|---|---|---|
| May 1967 | 10 | 13 |

| 8a. CONTRACT OR GRANT NO | 9a. ORIGINATOR'S REPORT NUMBER(S) |
|---|---|
| AF 49(638)-1744 | SMC R-7 |
| b. PROJECT NO<br>920F - 9719 | |
| c. [redacted] 6.154501R | 9b. OTHER REPORT NO(S) (Any other numbers that may be assigned this report) |
| d. 681313 | AFOSR 68-2157 |

10. DISTRIBUTION STATEMENT

[redacted]

1. This document has been approved for public release and sale; its distribution is unlimited.

| 11. SUPPLEMENTARY NOTES | 12. SPONSORING MILITARY ACTIVITY |
|---|---|
| TECH, OTHER | Air Force Office of Scientific Research<br>Office of Aerospace Research (SRLB)<br>United States Air Force |

ABSTRACT

A recent advance in educational measurement has made it possible for the first time to obtain valid and reliable measurements of a pupil's degree of confidence in the answers to objective test questions. The meaning and qualitative significance of guessing on objective tests is examined by using decision-theoretic psychometrics. The origin of the "guessing problem" is traced to the conventional "number of rights" scoring system and three types of guessing are defined: blind guessing, partially blind guessing, and rational guessing. These three types of guessing are shown to affect both the reliability and validity of a pupil's test score as well as the ability of a teacher to classify pupils for remedial help. Four unsuccessful attempts to remedy the guessing problem are illustrated. This leads to an explanation of how admissible confidence measurement techniques completely eliminate guessing.

DD FORM 1 NOV 65 1473

UNCLASSIFIED

Security Classification

| KEY WORDS | LINK A | | LINK B | | LINK C | |
|---|---|---|---|---|---|---|
| | ROLE | WT | ROLE | WT | ROLE | WT |
| guessing | | | | | | |
| test scoring | | | | | | |
| degree of confidence | | | | | | |
| instruction | | | | | | |