DOCUMENT RESUME

ED 054 236	24	TM 000 875

AUTHOR	Rippey, Robert M.
TITLE	Scoreing and Analyzing Confidence Tests. Final Report.
INSTITUTION	Illinois Univ., Chicago.
SPONS AGENCY	Office of Education (DHEW), Washington, D.C. Bureau of Research.
BUREAU NO	BR-7-0578
PUB DATE	8 May 71
GRANT	OEG-5-9-230578-0070
NOTE	87p.
EDRS PRICE	EDRS Price MF-$0.65 HC-$3.29
DESCRIPTORS	Computer Programs, Correlation, Factor Analysis, *Guessing (Tests), High School Students, Item Analysis, *Scores, Scoring, *Scoring Formulas, Sex Differences, *Statistical Analysis, *Testing, Test Reliability, Tests, Test Validity, Weighted Scores
IDENTIFIERS	*Confidence Tests, Item Difficulty

ABSTRACT

Technical improvements, which may be made in the reliability and validity of tests through confidence scores, are discussed. However, studies indicate that subjects do not handle their confidence uniformly. (MS)

Final Report

Project No. 7-0578
Grant No. 5-9-230578-0070

Scoring and Analyzing Confidence Tests

Robert M. Rippey

University of Illinois at Chicago Circle

Chicago, Illinois

May 8, 1971

U.S. Department of
Health, Education, and Welfare

Office of Education
Bureau of Research

Final Report

Project No. 7-0578
Grant No. 5-9-230578-0070

Scoring and Analyzing Confidence Tests

Robert M. Rippey

University of Illinois at Chicago Circle

Chicago, Illinois

May 8, 1971

The research reported herein was performed pursuant to a grant with the Office of Education, U.S. Department of Health, Education, and Welfare. Contractors undertaking such projects under Government sponsorship are encouraged to express freely their professional judgement in the conduct of the project. Points of view or opinions stated do not, therefore, necessarily represent official Office of Education position or policy.

U.S. Department of
Health, Education, and Welfare

Office of Education
Bureau of Research

CONTENTS

	Page
Background	1
Development of a Scoring System	7
Studies of Different Scoring Systems	15
Design	15
Results and Discussion	16
Item difficulties	18
An Entropic Scoring Function	22
Properties of Tests Containing Intrinsic Items	25
Importance of Weighting	25
The Meaning of Confidence	26
Conclusions	38
Appendix A. Test Scoring Programs	
Appendix B. Sample Items from Tests Employing Euclidean Confidence Scoring	
Appendix C. Criterion Response Distribution Plotting Program	

	Page
Table 1. Five Scoring Functions	4
Table 2. 100-point Payoff Scheme	11
Table 3. Comparison of Reliabilities of Scores on Fourteen Administrations of Three Parts of STEP Writing Tests, Level 1.	17
Table 4. Mean Scores on Twelve Tests Using Four Different Scoring Functions	19
Table 5. Comparison of Reliabilities of Euclidean Scores, eighted and Unweighted	26
Table 6. Mean Scores on Nineteen Variables, Incentive Condition	28
Table 7. Mean Scores on Nineteen Variables, Relaxed Condition	28
Table 8. Mean Scores on PLN for Groups	31

Table 9. Correlation Coefficients for Nineteen Variables, Relaxed Condition 33

Table 10. Correlation Coefficients for Nineteen Variables, Incentive Condition 34

Table 11. Factor Matrix for Nineteen Variables, Incentive Condition 36

Table 12. Facotr Matrix for Nineteen Variables, Relaxed Condition 37

Figure 1. Response Triangle for Three-Option Item and Scoring Key 9

Figure 2. Graph of Scores for Functions 1 and 2 18

Figure 3. Graph of Scores for Functions 1 and 5 18

Figure 4. Graph of Scores Using Four Functions 20

SCORING AND ANALYZING

CONFIDENCE TESTS

Robert M. Rippey

Background

A conventional multiple choice item is a very special case of a much more general situation of decision making. Given a body of information, knowledge of a field, a limited number of options to choose among, a subject will in the conventional situation: consider the options, develop a set of preferences, and then make a single choice as instructed. Such a decision, however, is not particularly informative about the state of knowledge of an individual. The single, unqualified choice does not separate the confident subject from the timid one. Nor does it separate the lucky guesser from the qualified and certain expert. Furthermore, the search by test makers for questions having unique correct responses limits them to a small fraction of the possible questions which might be contrived in that broad area lying between warranted knowledge and aleatory opinion. Perhaps good guesses are as good as the same choice made with greater assurance. On the other hand, there may be some value in exploring more systematically some of the alternatives to dogmatic testing practices.

This additional information about confidence and distribution of belief may be important from several standpoints. It may lead to more accurate prediction of retention (Ahlgren, 1968). Furthermore, validity may be increased (Hambelton, Roberts, and Traub, 1970). Confidence information may be of importance in understanding the mechanisms involved in predicting performance in decision making involving complex sets of contingencies (Bruner, Goodnow, and Austin, 1956; Ward and Edwards, 1961; Kogan and Wallach, 1964).

Degree of belief in the options for an item may be represented by a set of weights. Although a set of weights contains more information than a single choice, the meaning of this additional information must be questioned, for these weights may mean different things to different subjects, and most certainly will mean different things depending upon the conditions of administration and the rewards or punishment expected as part of the testing procedure. It is perhaps because of uncertainty about the meaning of such weights that test makers have preferred almost exclusively items which had warranted unique answers. At least examples of tests of another sort are difficult to find.

By comparison with achievement tests, attitude and personality tests have eshewed unique right answers. The Strong Vocational Interest Blank, originally published in 1927, was an early example of an instrument involving response weighting (Strong, 1943). Others have followed (Swineford, 1941; Guttman, 1947). On the other hand, test makers have moved cautiously with uncertainty in the cognitive domain. Perhaps it is more acceptable to be uncertain how we feel than it is to be uncertain about what we know. Plato struggled briefly with the status of belief, dismissing the issue as follows (Plato, Republic):

> Indeed, I am content; I proceeded, to call as before the first division science, the second understanding, the third belief, and the fourth conjecture -- the two latter jointly constituting opinion, and the two former intelligence. Opinion deals with the changing, intelligence with the real; and as the real is to the changing, so is intelligence to opinion, so is science to belief, and understanding to conjecture. But the analogy between the objects of these mental acts, and the twofold division of the provinces of opinion and of intelligence, we had better omit, Glaucon, to prevent burdening ourselves with discussion far outnumbering all the former.

Although realizing that the area of uncertainty is always larger than the area of warranted knowledge, curriculum builders and test-makers have favored the certain side of the road, except in dire emergencies. It is interesting that much of the research in testing for uncertainty has come from the medical (Lewy & McGuire, 1966) and the military profession (Shuford, 1967).

In 1936, Soderquist suggested the application of confidence weights to true-false items and the application of penalties for misplaced confidence. Ebel (1965), in following this suggestion, obtained reliability increases of approximately .10.

A brief inroad into uncertainty was made by the Progressive Education Association in its Eight Year Study (Smith & Tyler, 1942). Items used in their investigation of outcomes of secondary schooling were not only classified as correct or incorrect, but also "Caution," "Insufficient data," "Beyond data," "Too certain," and "Too uncertain."

Dressel and Schmid (1935) contrasted the results obtained from a multiple choice test when administered in several novel forms. Two modifications were:

A. The student was told to mark as many choices as he needed in order to make certain he had the right answer. This was scored as follows:

Number of choices Marked	1	2	3	4	5
Score if correct answer was marked	4	3	2	1	0
Score if correct answer was not marked	-1	-2	-3	-4	X

B. The student was told to mark a single answer but indicate degree of certainty, using a scale of 1 to 4 meant "at least certain." The item was scored as follows:

Certainty marked	1	2	3	4	Omit
Score if answer was correct	4	3	2	1	0
Score if answer was incorrect	-4	-3	-2	-1	

Neither of these systems actually assigned probabilities or degree of belief weights to the entire set of responses.

DeFinetti (1965) has discussed a number of the consequences of utilizing different scoring systems and argues strongly for the training of subjects in a more prudent strategy for dealing with uncertainty.

> Feedback, or learning from experience, does not only concern the reinforcement of such general ideas about the profitability of an undistorted forecast or response corresponding to personal evaluation of probability: the evaluation itself becomes improved by experience. It is particularly common for untrained people to reflect in their numerical evaluations of probabilities the effect of the usual way of thinking in rough terms of 'certain,' 'impossible,' 'unknown,' or 'completely indifferent,' giving values 1.00 amd 0.00 to the favored and rejected alternatives and 0.50 in the case of uncertainty between two, and so on. Experience forces them to realize how relatively often there are events which occur that can be too hastily classified as impossible, and they learn the advantage of giving these events an adequate small positive probability. It is chiefly because it provides the possibility of redressing such essential weaknesses in the machinery of human reasoning, and shows how workable measurement in the fields of belief can be developed, that I have felt obliged to emphasize so strongly the desirability of training in the use of the methods described in this paper.

Shuford and Massengill have argued that reliability and validity of tests can be increase using a class of scoring functions called reproducing scoring functions which maximize S's score if and only if his responses match his internal belief state. These functions were first studied by Toda (1963). The data available at this time does not conclusively support the viability of this model for responses of Ss. Therefore, another purpose of this study was to clarify the meaning of significance of the wieghts which subjects assign to confidence scored tests.

A number of scoring functions have been developed for scoring confidence test items. Five such functions are shown in Table 1.

TABLE I

Five Scoring Functions

1. Probability assigned to correct response

$$S = r_k$$

2. Logarithmic

If $r_k \geq .01$, $S = (2 + \log_{10}(r_k))/2$
If $r_k < .01$, $S = 0$

3. Spherical

$$S = r_k \Big/ \left(\sum_{i=1}^{n} (r_i)^2\right)^{\frac{1}{2}}$$

4. Euclidean

$$S = 1 - \frac{\left(\sum_{i=1}^{n} (r_i - k_i)^2\right)^{\frac{1}{2}}}{\sqrt{2}}$$

5. Inferred Choice

$S = 1$ if $r_k > r_i$ for all $i \neq k$;
$S = 0$ otherwise

r_i = Probability assigned to the i^{th} response

r_k = Probability assigned to the correct response

k_i = Criterion group mean probability assigned to the i^{th} response.

The probability assigned to the correct answer is the simplest and most intuitively obvious scoring function. Both the logarithmic and the spherical function possess the interesting property of allowing the student to maximize his score if and only if he does not guess (Shuford, Albert & Massengill, 1966). The Euclidean function will score items which do not unique correct responses. Thusitems can be constructed which call for answer which correspond to a distribution of preference representing the concensus of a group of experts. This function will also score items having unique correct answers.

Inferred choice is analogous to conventional multiple choice scoring. According to this rule, the subject receives 1 point if his maximum confidence is assigned to the correct option. Otherwise he receives nothing.

Although this method of scoring simulates the performance of subjects on conventional multiple choice tests, it does not duplicate it. Some subjects, confronted by absolute lack of preference guess. Others do not. The inferred choice function simulates the behavior of the subject who never guesses when he is absolutely uncertain, but who always makes a choice, even if his preference is slight. Since there are varieties of subject behavior on tests, this function will not always give results which are identical to the choices an individual subject might make. If a subject is instructed to answer every question, and if all subjects do this, scores obtained by the inferred choice function would always be less than or equal to scores obtained by subjects under the usual choice situation, since subjects would occasionally get an additional point due to guessing. If subjects were told not to guess, and the conventional penalty for guessing were applied, scores obtained by the inferred choice function would be less predictable, and could be either greater than or less than the scores obtained by the subject responding in the conventional manner.

The seriousness of this discrepancy would be proportional to the number of instances on a test where no dominant preference was shown for a single option. In an analysis of a random sample of answer sheets for the STEP test data used in this study, such a lack of preference was found in less than 15% of the responses. Since the items were unusually difficult for the subjects by design, it is likely that the amount of guessing would be less on other tests, more appropriate in difficulty for the subjects.

On a three option test, this would suggest that scores would be approximately 5% higher in terms of S behavior as compared with inferred choice scores.

It would, of course, be possible to simulate all manner of erratic S behavior in responding to multiple choice items. However, the inferred choice function does simulate the subject who does not guess. Any other simulations, involving random or systematic awarding of points in guess situations would lead to less reliable scores than the inferred choice function. In comparing both the reliability and the validity of functions against the standard of choice, it is probably best to use the inferred choice function as a standard since it does not contain any artificially induced error.

It would be possible, of course, to instruct S's to record both their probabilities and their choices. Thus, in the event of a split decision, the subject could flip his coin. Or the subject could also be instructed never to make his preference weights exactly equal. This should not be too unrealistic, since it is seldom, if ever, that the preference weights for a set of options would be entirely equal, no matter what the state of ignorance of S. Thus, although the inferred choice method does not simulate human behavior exactly, and this should be kept in mind, it is also unlikely that the inferred choice function produces less reliable or valid scores, or significantly lower scores, than would be given by scores obtained by conventional choice methods. Since the purpose of much research on confidence testing is to demonstrate the superiority of confidence methods over the conventional choice method, it seems that using the inferred choice function as a basis for comparison does not weaken the conclusions of such comparisons.

Although many arguments and some practice accept getting a fix on confidence, states of knowledge intermediate between certainty and chaos are not as readily accepted by some subject matter specialists. Therefore, when one examines achievement tests, it is unlikely that he will find many items dealing with incomplete information or uncertainty. The dearth of items requiring a distribution of belief over the available options may be due to a single technical consideration. Indeed, it can be argued that intrinsic items should not be written at all because an item which calls for a uniform distribution of confidence over all responses will not discriminate between the informed and the uninformed. Both groups would assign equal proabilities to all the options. Thus, the unweighted Euclidean function is inadequate by itself, for items not having single option responses are less efficient in detecting a state of no information than items having unique correct options. Nevertheless, this problem can be rectified by asking S for a distribution of belief and his confidence in his distribution, and subsequently incorporating both the distribution and the confidence measure into his score. The following function accomplishes this:

---

6. Weighted Euclidean function

$$S = C(1 - 2D / D_{max})$$

C = Confidence ($0 \leq C \leq 9$)
D = Distance from S's response to the criterion group response
D_{max} = Maximum distance attainable from the criterion groupresponse

---

If Ss use confidence weights varying from 0 to 9, scores will vary from -9 to +9. An examinee who expresses no confidence will be neither rewarded nor penalized for his distribution of preference. On the other hand, certainty about a single incorrect response may

may suffer a nine-point penalty. The results of a test containing a mixture of items may be scored in at least three ways: correctness, confidence, and appropriate use of caution. The last measure will be developed in a later part of this report.

DEVELOPMENT OF A SCORING SYSTEM

The study of confidence scored tests requires systems for scoring and computing the reliability and validity of such tests. A number of practical matters entered into the evolution of the programs which are displayed in Appendix A. There are four programs in all. Punched output from these programs can be used in conjunction with a multiple-regression and factor analysis program if desired. The programs are 1) a basic scoring program for eight different functions, 2) A key preparation program for preparation of a key from a set of responses from members of a criterion group, 3) A program for the weighted Euclidean function, and 4) A program for converting subject responses into risk taking indices. Descriptions of these programs follow.

1. <u>The basic scoring program</u>

Eight scoring function are available. Option One is a simple Euclidean distance function. Option Two is the Spherical function. Option Three is the logarithmic function. Option Four computes the inferred choice score. Option Five applies the Euclidean function to items having unique correct responses. Option Six assigns a score to an item which is equal to the probability the subject assigns to the correct answer. Option Seven, the Entropic function, will be discussed later. Option Eight is the Euclidean distance function with a correction for the maximum distance attainable from the criterion group response.

Limitations.

The program will score tests of up to seventy-two items. The number of response options may vary from two to six. Any number of subjects may be scored on each pass. Input formats are fixed.

Card Preparation.

Card 1 — Title card -- Cols 2-60

Card 2

Cols.	
1	Standard deviation of Criterion Group Scores 1 = No; 2 = Yes. (Applicable only to Option One)
4	Punched output of subject scores. 1 = Yes; otherwise blank.

Card 3

Cols.	
1	Punch a 1 in this column
2-3	Total number of items on test (Not more than 72)
4	Number of responses per item (2-6)
5-6	Number of cards per subject (Not more than 10)
7-8	Number of items per card

Card 4	Function Selection Card Punch a one in the column corresponding to the functions(s) selected.
Cols.	
1	Euclidean Function (Intrinsic Items)
2	Spherical Function
3	Logarithmic Function
4	Inferred Choice
5	Euclidean Unique
6	Probability assigned to the correct response
7	Entropic Function (Intrinsic Items)
8	Euclidean - Correctedfor Maximum Distance Attainable (Intrinsic Items)

Note -- Functions are of two types, intrinsic item and unique correct response. Functions of both types may not be requested on a single pass.

Key Card(s) — For intrinsic items -- use punched output from Key preparation Program

For items having unique correct responses -- Punch the number of the correct response for each item in the column corresponding to the number of the item. For example, if the correct response to item 7 is response option 3, a 3 would be punched in column 7.

Subject response cards

Cols.	
1-5	Identification number
6	Card sequence number
7---	Responses to each item. The weight assigned by S to each response.

Blank card

Jobs may be stacked. A card with a -1 punch in columns 1-2 should follow the last job.

Output

Output may be either printed, or printed and punched. It consists of individual items scores and total scores, item difficulties, and test reliability computed according to Hoyt's analysis of variance method. The punched output may be used as input to a standard statistical program for the purpose of obtaining inter-item correlations, correlations of items with total test and criterion scores, as well as factor analysis.

2. Key preparation from Criterion Responses

Card 1

Cols.	
1	Punch a 1 in this column

Cols.	
2-3	Number of items
4	Number of responses per item: may vary from 2 to 6
5-6	Number of cards per subject
7-8	Number of items per card
9-49	Title

Criterion group Responses

Blank card

Jobs may be stacked. A card with a -1 punched in columns 2-3 should follow the last job.

Output

Title card containing paramters for the scoring program (Card 3), and a set of cards containing the mean probabilities of the criterion group responses. These latter cards form the key for the scoring program. For items having unique correct responses, the key may be replaced by a single card.

3. Weighted Euclidean Scoring Function

This program score tests of up to thirty-two items. Items may have three options followed by a fourth measure of confidence. A sample answer sheet and instructions follow. A key may be prepared by calculating the mean responses assigned to the options by a criterion group, or by means of the separate program for key preparation. Two scoring options are available, the unweighted and the weighted option. The weighted option is preferred. The unweighted option assigns a score to each item according to the formula:

$$S = 1 - 2D / D_{max}$$

where d is the distance between the subjects answer (A) and the criterion group mean answer (B). D max is the maximum distance between any point on the probabilistic triangle and the criterion group point. See Figure 1.

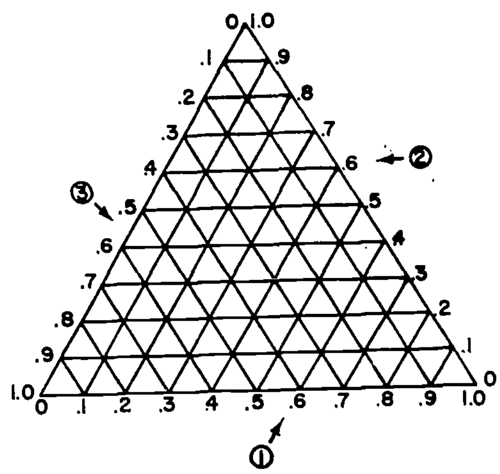
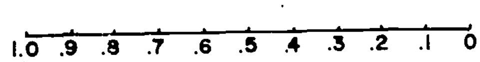


FIG. 1. Scoring key



In the weighted option, this score is multiplied by the degree of confidence, 0 through 9.

Card Preparation.

Parameter Card

Cols.	
1	Keying Option (See below)
2	Punched output option 1 = Yes; otherwise blank.
3	Weighting option. Unweighted = 1; otherwise blank.
4	1 if fewer than 17 items; 2 if 17 or more items.
5-6	Number of items. Less than or equal to 32.

Key

Option One:

Cols.	
1-6	Test number
7-70	Key responses (Mean criterion responses). Format 16(I3,1X)

Option Two:

Punched output from Key Preparation Program.

Title card Cols 2-60.

Responses

Cols.	
1-6	Identification Number
7-70	Responses, up to 16 items, in groups of 4 The fourth response represents confidence 0 = No confidence, 9 = Absolute certainty
71	Test attitude 0 = Hate the test; 9 = Love the test.

Blank card

Jobs may be stacked.

Last job should be followed by a card with -1 punched in columns 1-2.

A SIMPLIFIED LOGARITHMIC SCORING SYSTEM

For persons wishing to use a logarithmic scoring system without investing in computers or other feedback devices, the following system may be useful. It allows the use of the logarithmic function in a limited way. That is, the subject may use a fairly gross division of his confidence into ten categories. This approach is similar to DiFinetti's "Five Star" system (1965) which he used with young children. However, this system incorporates the logarithmic transformation,

$$S = (2 + \log_{10} r_k) / 2$$

Instructions and a sample sheet for the simplified logarithmic system follow. If it is desired to use a more precise assignment of probabilities, Table 2 may be used with modified instructions.

Table 2

Amt. Bet	Amt. Paid	Amt. Bet	Amt. Paid	Amt. Bet	Amt. Paid	Amt. Bet	Amt. Paid
.01	0.00	.31	0.7457	.61	0.8927	.91	0.9795
.02	0.1505	.32	0.7526	.62	0.8962	.92	0.9819
.03	0.2386	.33	0.7593	.63	0.8997	.93	0.9843
.04	0.3010	.34	0.7658	.64	0.9031	.94	0.9866
.05	0.3495	.35	0.7721	.65	0.9065	.95	0.9889
.06	0.3891	.36	0.7782	.66	0.9098	.96	0.9912
.07	0.4226	.37	0.7841	.67	0.9131	.97	0.9934
.08	0.4516	.38	0.7899	.68	0.9163	.98	0.9956
.09	0.4771	.39	0.7956	.69	0.9194	.99	0.9988
.10	0.5000	.40	0.8010	.70	0.9226	1.00	1.0000
.11	0.5207	.41	0.8064	.71	0.9257		
.12	0.5396	.42	0.8116	.72	0.9287		
.13	0.5569	.43	0.8168	.73	0.9317		
.14	0.5731	.44	0.8218	.74	0.9346		
.15	0.5881	.45	0.8266	.75	0.9376		
.16	0.6021	.46	0.8314	.76	0.9404		
.17	0.6152	.47	0.8361	.77	0.9433		
.18	0.6276	.48	0.8406	.78	0.9461		
.19	0.6394	.49	0.8451	.79	0.9488		
.20	0.6505	.50	0.8495	.80	0.9516		
.21	0.6611	.51	0.8538	.81	0.9543		
.22	0.6712	.52	0.8580	.82	0.9579		
.23	0.6808	.53	0.8622	.83	0.9596		
.24	0.6901	.54	0.8662	.84	0.9622		
.25	0.6990	.55	0.8702	.85	0.9647		
.26	0.7075	.56	0.8741	.86	0.9673		
.27	0.7157	.57	0.8779	.87	0.9698		
.28	0.7236	.58	0.8817	.88	0.9723		
.29	0.7312	.59	0.8855	.89	0.9747		
.30	0.7386	.60	0.8891	.90	0.9771		

INSTRUCTIONS

In the following test, imagine that you are being paid for the amount which you learn. However, it will cost you something to attempt each item. On the other hand, you may bet on more than one of the choices.

Imagine that you have ten cents to place on the choices of each item. You may distribute each of these ten cents(points) in any way you like. However, you will only be paid on the basis of the number of points you put on the correct answer. The amount that you get paid can be as much as 1.00 but it might be as little as nothing. The amount you get paid for a given number of points is shown in the table below.

Amount bet	Amount paid
0	0.00
1	0.50
2	0.65
3	0.74
4	0.80
5	0.84
6	0.88
7	0.91
8	0.94
9	0.97
10	1.00

On your answer sheet, you will write down both how many points you bet, and also, what the table says you should be paid. Make certain you check to see that you do not use more than ten points for each item. Also, use the table accurately. Errors in using the table or in assigning the correct number of points will cause you to receive no points at all for that item. Below are some examples of answers to the question

2 + 2 = ?

1. 0
2. 2
3. 4
4. 6

Sample answers

	Student A		Student B		Student C		Student D		Student E	
1.	0	0	0	0	3	.74	10	0	5	.84
2.	0	0	0	0	2	.65	0	0	4	.80
3.	10	1.00	5	.84	2	.65	0	0	3	.78
4.	0	0	5	.84	3	.74	0	0	0	0

See if you can figure out what score each student would receive.

Student A would receive a score of 1.00 since he knew the correct answer and put it down.

Student B was unsure as to whether the answer was 3 or 4, so he bet five points on each and was awarded a score of 0.84

Student C couldn't tell which answer he liked best, but was honest about it, and divided his points as evenly as he could across all the answers. He received 0.65 for his honesty.

Student D didn't know the answer either, but he guessed at response number 1. Since this was wrong, he got nothing at all.

Student E also received a 0 score because he cheated. He used more than 10 points, and also did not use the table correctly.

On this kind of test you will make your highest score if you honestly report your confidence in each of the choices. The penalty for guessing is very great.

ANSWER SHEET

Name______________________ School______________________ Date______________________

Question	Points	Score	Question	Points	Score	Question	Points	Score
1.	a.______	______	8.	a.______	______	15.	a.______	______
	b.______	______		b.______	______		b.______	______
	c.______	______		c.______	______		c.______	______
	d.______	______		d.______	______		d.______	______
2.	a.______	______	9.	a.______	______	16.	a.______	______
	b.______	______		b.______	______		b.______	______
	c.______	______		c.______	______		c.______	______
	d.______	______		d.______	______		d.______	______
3.	a.______	______	10.	a.______	______	17.	a.______	______
	b.______	______		b.______	______		b.______	______
	c.______	______		c.______	______		c.______	______
	d.______	______		d.______	______		d.______	______
4.	a.______	______	11.	a.______	______	18.	a.______	______
	b.______	______		b.______	______		b.______	______
	c.______	______		c.______	______		c.______	______
	d.______	______		d.______	______		d.______	______
5.	a.______	______	12.	a.______	______	19.	a.______	______
	b.______	______		b.______	______		b.______	______
	c.______	______		c.______	______		c.______	______
	d.______	______		d.______	______		d.______	______
6.	a.______	______	13.	a.______	______	20.	a.______	______
	b.______	______		b.______	______		b.______	______
	c.______	______		c.______	______		c.______	______
	d.______	______		d.______	______		d.______	______
7.	a.______	______	14.	a.______	______			
	b.______	______		b.______	______			
	c.______	______		c.______	______			
	d.______	______		d.______	______			

STUDIES OF DIFFERENT SCORING SYSTEMS

Subsequent to the development of the scoring programs several questions were asked. These were 1) How do the several scoring functions compare with one another and 2) Are there differences in the ways in which subjects respond to confidence scored tests?

One of the fundamental assumptions underlying the logarithmic function, and its desirable matching property of maximizing the subject's score if and only if he responds with his actual degree of confidence is the need for the subject to have feedback at test time with response to the payoffs of his set of preference weights.

Shuford has suggested several testing aids such as a computer terminal or computational devices known as scorules. Unfortunately, scoring in these ways in likely to be costly, or time consuming. (Ebel, 1968) Furthermore, the necessity to provide the student with information about the scoring system is not only demanded by theory, but Rippey (1968b) and Romberg and Shepler (1968) both provide data which shows that the logarithmic function may at times produce less reliable scores than conventional testing when such information is not provided.

An alternative to the use of computers and scorules is the use of scoring functions which are simpler, and more transparent to the intuition of the subject. The simplest of these is a score which is equal to the proability assigned to the "correct" answer. Such a score contains more information per item than a conventional choice score. Furthermore, it offers the student some incentive not to guess, although not the optimal incentive promised by the logarithmic function. This leads to the question of the relative merits of various scoring functions with respect to reliability.

DESIGN

A total of 374 students, hereafter referred to as Sample A, from three Chicago suburban high schools were given three intact 30-item sections of the STEP Writing Test, Level 1[2], within schools. Tests were randomly assigned to students. No student took any form more than once.

Fourteen groups were tested. Each test was scored using the five scoring functions. Groups 7,8, and 9 had been previously tested twice and had some understanding of the properties of the logarithmic function. Groups 4, 5, and 6 had been tested once previously. All other groups were totally unfamiliar with confidence tests.

Prior to taking the test, the subjects were given the following statement:

> Each of the questions or incomplete statement in this test is followed by suggested answers. Assign a number from 0 to 9

[2] Permission for the use of this test was granted by Educational Testing Service.

to each suggested answer, depending on how strongly you feel the answer is correct. If you believe that only one suggested answer is correct, mark that answer with a 9 and mark the others with zeroes. If you like the suggested answers equally, assign the same number to each.

Next followed several examples of how S was to distribute the numbers under various patterns of degree of belief of the correctness of the several options. Finally was the statement:

> Your paper will be scored in such a way that you will get a higher score by estimating your degree of confidence and reporting it accurately. Guessing in any form will lower your score. If you are uninformed about the question and have no preference for the suggested answer, you will obtain your highest score by honestly distributing your confidence across all the options....

Tests were subsequently computer scored and reliability was estimating using Hoyt's analysis of variance procedure (Hoyt, 1941). This procedure is suitable for confidence tests whereas a number of other procedures such as K.R. 20 are not. This is because item scores range between 0 and 1. The Hoyt method underestimates reliability on short tests. Therefore the reliabilities are all conservative.

RESULTS AND DISCUSSION

Reliabilities obtained from the multiple scorings of the 14 tests are shown in Table 3. The most obvious and immediate conclusion from these data is that the simplest of the functions, Function 1, has the highest reliability most of the time. In making the 56 possible comparisons of reliabilities of scores obtained using Function 1, with scores obtained using another function, scores obtained by using Function 1 are the most reliable 49 times. Furthermore, for the 14 independent situations, Function1 is the best 10 times. Assuming a null probability of 0.2 for Function 1 to be ranked first, the probability of method 1 being superior 10 or more times out of 14 is less than .001.

In order to examine the relative effectiveness of the five functions, an average raliability was computed over the 14 occasions. Since knowledge of the logarithmic scoring function had been given to subjects in Groups 7, 8, and 9 on prior occasions, the second set of reliabilities was computed for the complementary 11 groups. The averages are shown in Table 3 .

It is clear from the table that Function 1 leads to the most reliable score and Function 5, conventional choice scoring leads to the least reliable scores. Is there some model which would explain this ordering of functions? Assuming that any observed score is made of a true component and an error component, divide the error in two parts. The first part is a random variable whose range is proportional to some discrepancy between the score assigned to a

Table 3

Comparison of Reliabilities of Scores on Fourteen Administrations of Three Parts of STEP Writing Tests, Level 1.

Group	Administration Form	Administration Part	N	Function 1. $S=P_k$	2. Logarithmic	3. Spher-	4. Euclidean	5. Inferred Choice
1	1A	1	21	.811[a]	.706	.728	.670	.697
2	1A	2	19	.629	.676	.542	.338	.639
3	1B	1	22	.645	.197	.494	.464	.364
4	1A	1	22	.826	.487	.505	.647	.368
5	1A	2	19	.811	.060	.469	.740	.691
6	1B	1	20	.806	.594	.700	.672	.685
7	1A	1	19	.627	.756	.644	.581	.555
8	1A	2	23	.748	.489	.366	.539	.305
9	1B	1	20	.693	.563	.493	.483	.524
10	1A	1	38	.169	.439	.039	.184	-.063
11	1A	1	39	.641	.525	.489	.595	.521
12	1A	1	41	.755	.588	.621	.699	.604
13	1A	1	52	.687	.533	.561	.697	.535
14	1A	1	23	.712	.000	.160	.613	.202
Average[b] four 14 Groups				.69	.50	.49	.58	.47
Average[b] for Groups 1-6, 10-14				.68	.47	.49	.60	.48

[a]The largest reliability for each occasion is underscored.

[b]The customary _r_ to _z_ transformation was used in computing the average correlations.

subject for an item, and the subject's perception of the score he will receive for an item. The second part of the error then consists of the balance of uncontrolled variation which contributes to the unreliability on conventionally scored tests. Thus, most scores computed using esoteric scoring functions will have an error component which is due to the subject's lack of understanding of the scoring system. This error component is illustrated in Figures 1 and 2. If a subject is told or believes that the score he will receive is equal to the probability he assigns to that option which happens to be correct (this is the intuitively obvious conclusion), and his score is then computed by some other function, then a portion of the error component of his score will be proportional to this discrepancy. The discrepancy is shown in the two figures as the shaded area between the two curves.

This suggests that, other sources of error being equal, (the same subjects take the same tests on the same occasion), the rank order of reliabilities will be the same as the rank order of the discrepancies between Function 1 and the other functions. Such, in fact was the case.

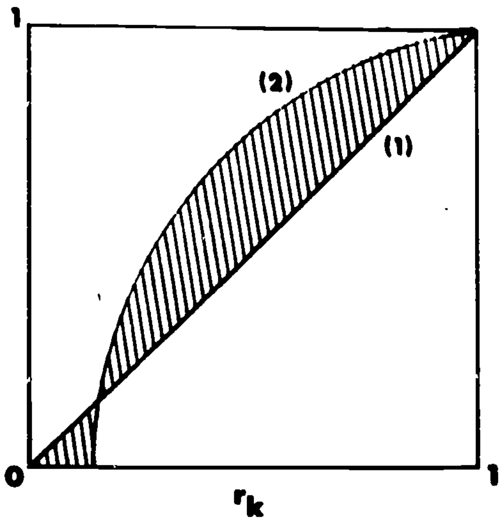


Figure 1 Graph of Scores for Functions 1 and 2.

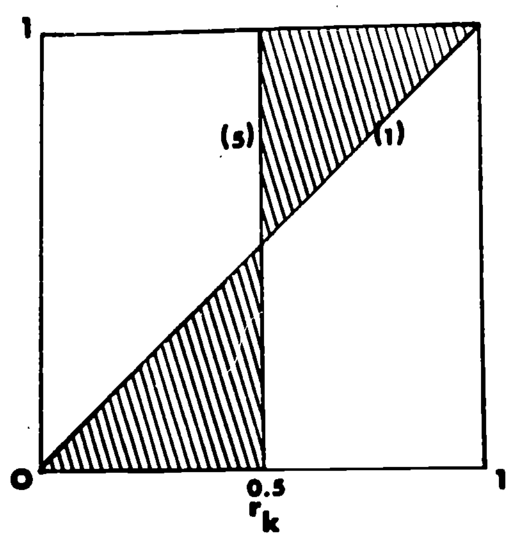


Figure 2 Graph of Scores for Functions 1 and 5.

The discrepancy between scores obtained from Function 1 and scores obtained from Functions 2, 3, 4, and 5 were obtained by integration of the various functions. The results of this integration were .06, .23, .25 and .28 for Functions 2 through 5 respectively. (The discrepancy is 0 for Function 1 as this function served as the standard.) Not only are the discrepancies and the reliabilities in exact reverse order, but also, the relationship between the discrepancies and the averaged reliabilities is essentially linear.

ITEM DIFFICULTIES

The simplest function, Function 1, produces a consistently high reliability. The Euclidean Function, which is similar to Function 1 in its scoring results, produces a comparitively high reliability.

However, the probability assigned to the correct answer has a serious disadvantage. Williams and Millman (1970) pointed out correctly that the simple probability function penalizes the student who honestly reflects his degree of belief and favors the student who shifts his confidence entirely to the option he prefers, even it his preference is slight. Therefore, with practice, a subject will learn to distort his responses in the direction of dogmatic choice. The objective of this portion of this study was to see if the Euclidean function might perhaps resolve this problem of confidence testing.

Two kinds of scores may be given Ss. The first is a score of correctness based on the distance of S's response from the response pattern agreed upon as correct. The second score, a score of appropriateness, is a count of the number of items to which S responded with certainty when certainty was called for and caution where it may be assumed his knowledge was incomplete.

Method

A shortened version of the STEP Writing Test Level 1A was administered to twelve groups of high school students (N = 303). Ss were told to honestly represent their confidence on their answer sheets. Incentives were not mentioned. The results were scored using four different scoring functions, the simple probability, Logarithmic, Euclidean, and Inferred Choice functions.

Conclusions

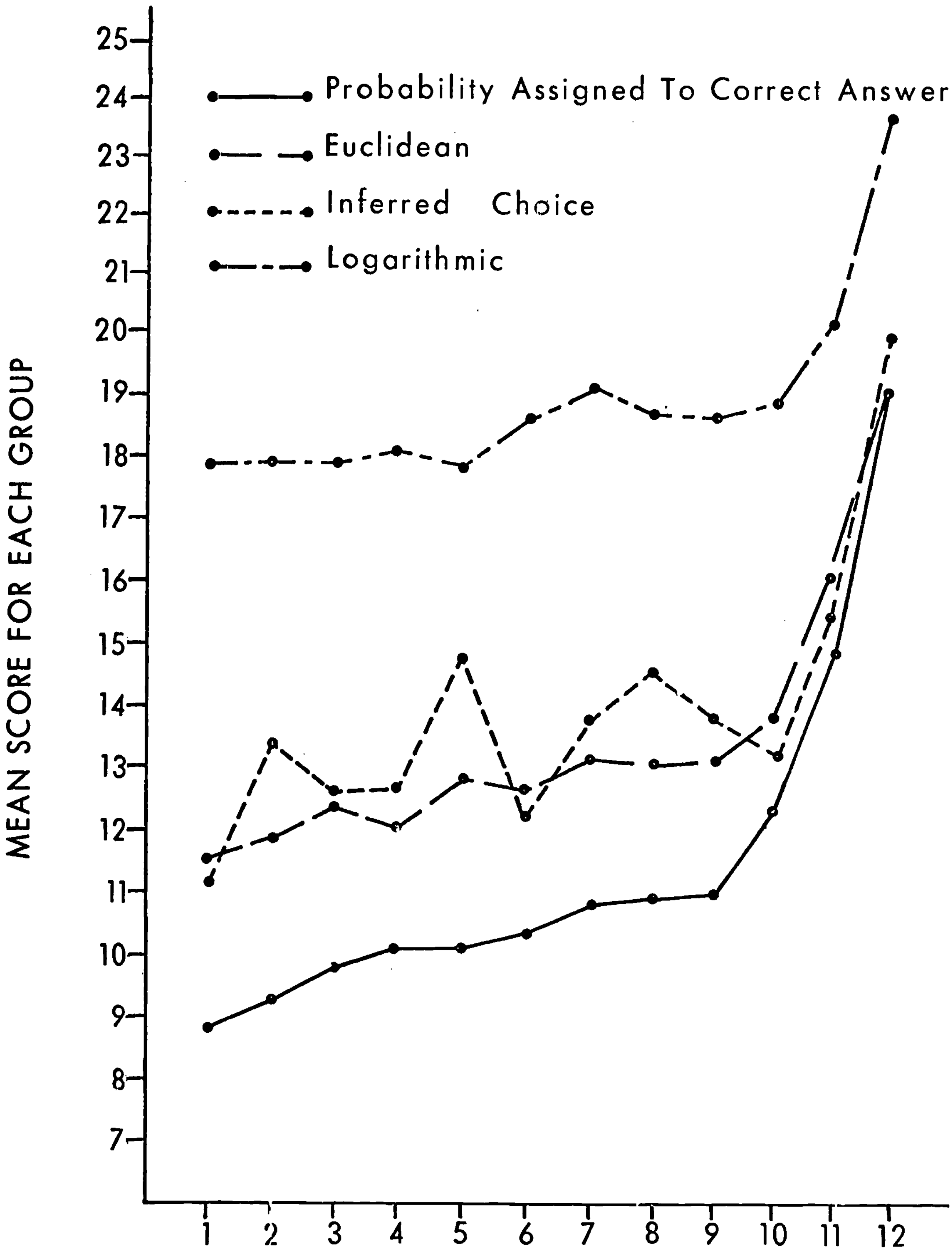
As predicted, the Logarithmic function gace the highest mean scores for all twelve groups and the probability assigned to the correct answer gave the lowest scores consistently to all twelve groups. On the other hand, the mean group scores obtained using the Euclidean function were slightly higher in four instances and slightly lower in eight instances, than the score derived by assigning full confidence to the preferred option (Inferred Choice score). These scores are displayed in Table 4.

Table 4

Mean Scores on Twelve Tests Using Four Different Scoring Functions

Rank Order	N	Probability Assigned To Correct Answer	Log Function	Inferred Choice	Euclidean
1	22	9.77	18.91	12.23	12.54
2	19	10.37	18.98	14.46	12.92
3	20	10.82	18.99	13.60	13.45
4	20	11.07	19.28	13.70	13.11
5	21	11.09	18.83	15.75	13.96
6	19	11.39	19.55	13.31	13.73
7	23	11.79	20.02	14.82	14.17
8	19	11.95	19.76	15.53	14.01
9	22	12.00	19.54	14.85	14.11
10	41	13.32	19.44	14.25	14.89
11	39	15.72	21.13	16.55	16.68
12	38	19.95	24.57	20.99	19.95

Graph Of Scores Using Four Scoring Functions



RANK ORDER OF GROUP MEAN SCORES USING PROBABILITY ASSIGNS TO CORRECT ANSWER AS THE SCORING FUNCTION

Thus the Euclidean function did not consistently penalize the honest student, although it did not reward him, as did the Logarthmic function. The high reliability of the Euclidean function therefore suggests that it might be a preferred method of scoring confidence weighted tests. It has the further advantage of being applicable to items not having unique correct responses. Therefore, it can be used with a wider variety of test items than any of the other functions. Furthermore, a test consisting of a mixture of items having unique correct responses saves the test constructors from the ethical problem of asking the subject to be honest about his belief when following such instructions may lower Ss score.

It is true that all of the scoring functions will rank the subjects in approximately, although not identically the same order. However, if a subject ignores the instructions of the proctor and refuses to follow the instructions to assign probabilities and instead gives his full weight of choice to his preferred option, even though that option is only slightly preferred, he will get a higher score. The point of the Euclidean and the Logarithmic functions is that they do not favor the subject who ignores the instructions to assign his degree of belief to his options.

If, thus far, it seems that I am recommending the Euclidean function on all occasions, let me say that I am not. The situation as I see it is as follows.

The Logarithmic function is recommended on the basis of its reproducing property of maximizing scores if and only if S assigns his degree of belief to the options accurately. The theory is based upon the maximization of S's expected utility for his response spectrum. This assumes, however, that the utility of a response is proportional to the score S receives. If in the case where S is being paid dollars for the points he earns--fine. However, there are many testing situations where utility is not proportional to the test points. In research, for example, if tests are not counted toward student grades, or if they receive no extrinsic or intrinsic rewards for their performance, utility of response is independent of score. In such situations, it is likely that the use of the probability assigned to the correct answer as a scoring function might be preferred because of its high reliability with naive Ss.

Nevertheless, where the stakes are high, where S's career or income, or the fate of his regiment depend upon an accurate assessment of his state of knowledge, the logarithmic function is to be preferred in spite of the cost and trouble it entails.

On the other hand, the Euclidean function does not deter S from reporting his probabilities as does the function which scores proportionally to the probability assigned to the correct answer. Its high reliability with naive Ss, and the other advantages mentioned do not make it a panacea to the problems of confidence testing, but the advantages do make it a reasonable alternative in some situations.

Not all knowledge is warranted knowledge. Training and education must prepare subjects for situations involving partial knowledge and incomplete information. However, technical problems have thus far impeded the development of tests which reflect such situations. The Euclidean function can achieve high reliability on confidence weighted items without expensive or complex feedback devices, without penalizing the honest students, and without compromising the integrity of the test constructor.

An Entropic Scoring Function

The Euclidean function seems to lead to high reliabilities, does not penalize Ss for responding with their degrees of belief. It has an intuitive appeal. Nevertheless, it is not tied in with the mainstream of information theory, nor does it provide any particular incentive for S to struggle with his probabilities.

Although the feedback devices required to bring about an accurate application of the Logarithmic function may be time consuming and moderately expensive, perhaps Ebel's point is unduly negative. For the average classroom teacher, perhaps accurate assessment of subjective probabilities is unimportant. On the other hand, to understand subjective probabilities and to interpret their meanings in terms of human behavior, it may be premature to rule out the development of their assessment until we know much more about them. Therefore, it may be desirable to proceed with the extension of such assessment without regard to simplicity.

There are other ways of matching probabilities than comparisons of Euclidean Distances. For example, the product of the vector of subject responses and criterion responses, $\bar{r} \cdot \bar{k}'$ ranges between 1 and 0 depending on the closeness of the two vectors. On the other hand, the product

$$\bar{k} \cdot (\overline{\log r})' = \sum_{i=1}^{n} k_i \log r_i$$

becomes $\sum_{i=1}^{n} r_i \log r_i$ when $k_i = r_i$. (1)

(1) is the formula for the entropy of a logical spectrum (Watanabe, 1969a). Gibb's theorm states that, given two sets of real numbers, $\{k_i\}$, $\{r_i\}$, where $i = 1,n$ and $k_i \geq 0$, $r_i \geq 0$, and

$$\sum_{i=1}^{n} k_i = \sum_{i=1}^{n} r_i$$

then

$$\sum_{i=1}^{n} k_i \log r_i \leq \sum_{i=1}^{n} r_i \log r_i$$

and equality holds only where $k_i = r_i$.

This formula may be applied to a multiple choice item as follows: Given a set of criterion response probabilities to a logical spectrum A, $\{k_i\}$, and a set of subject responses to the same logical spectrum, $\{r_i\}$,

where
$$\sum_{i=1}^{n} k_i = \sum_{i=1}^{n} r_i = 1$$

a logical spectrum can be established by providing the option "some other answer" in an intrinsic type item, or by informing the subject that there is one and only one correct answer for a conventional item. This latter information makes it possible for the subject to conclude that not only is there one and only one correct answer which is given to him as an option, but also that none of the other options, or for that matter any other option not stated is open to him as a choice.

If the subject's score is then made equal to

$$\sum_{i=1}^{n} k_i \log r_i .$$

it immediately follows from Gibb's theorem (Watanabe, 1969) that this score is maximized only under the condition that the subject's probability coincides with the criterion group probabilities. Furthermore, if an item has a unique correct response, the vector $\overline{k}_i$ will contain a single 1 and all other entries will be 0. In this case, then, the scoring function becomes equivalent to the logarithmic scoring function. Thus the entropic scoring function is applicable both to intrinsic items and items having unique correct responses.

However, the question still remains as how to give S enough information about the system that an accurate assessment of his subjective probabilities may be made. It must be remembered that Toda's original work (1963) postulated a computer giving the S information about the payoff of his responses, and that this payoff, depending on the scoring function, might not be proportional to his response probability. The S would then be allowed to modify his probabilities to achieve a maximum expected utility (Ramsey, 1931).

In order to receive appropriate feedback, the S taking an n option item is given n payoff values, one for each of the vertices of the response n-hedron. In the case of an intrinsic item, the subject could be given the same kind of information for any number of points on or within the response n-hedron, including the vertices. On the other hand, the sampling could be left under the control of the subject who could request a computation of his payoff based on his own response and a variety of assumed criterion responses which he might choose to experiment with.

This may seem like a complex process. On the other hand, it is probably realistic since accurate probability assessment in the face of incomplete information is in most cases difficult, and in many cases not even rational. On the other hand, it seems that the number of points to be tested for feedback from a three-option intrinsic item would not be unreasonably greater than the number of points to be tested on a six-point conventional item. On the six option conventional item, each item would need feedback concerning six vertices. On a three option intrinsic item, one could probably get by nicely with six points--the three vertices, the centroid, and two other interior points.

One characteristic of the entropic scoring function is that S would get a very low score (minus infinity) for neglecting a vertex which had any possibility of being correct. For example, if the criterion response vector were (1, 0, 0) and one chose as his probabilities

$$\begin{bmatrix} 0 \\ \frac{1}{2} \\ \frac{1}{2} \end{bmatrix}$$

his score would be

$$\overline{k} \cdot \overline{\log r}' = (1, 0, 0) \begin{bmatrix} \text{Log } 0 \\ \text{Log } \frac{1}{2} \\ \text{Log } \frac{1}{2} \end{bmatrix} = -\infty$$

Although this may seem like an extreme penalty, it is perhaps realistic, and certainly is in keeping with the methods of science. Since, as Platt (1963) suggests, "Any conclusion that is not an exclusion is insecure and must be rechecked." Shuford, Albert, and Massengill (1966) were sufficiently uncomfortable with such a severe penalty that they limited the minimum score that could be obtained with an item. This is partly a matter of taste, and partly a question of one's objective in testing. It is not uncommon in medical school examinations to have an item or so with a response which results in failure of the entire test. Perhaps if the captain of the Titanic, and the commanding officer at Pearl Harbor had been tested in this way while in college, they would have been less likely to completely rule out the possibility of icebergs that far South or Japanese planes that far East.

All entropic function scores are negative or zero. If this is bothersome, a positive constant could be added, and a lower limit could keep the scores within a range of ± 1. In the computer program, the following scoring function was used as Option 7.

$$S = 10. + \sum_{i=1}^{n} k_i \log r_i$$

with a minimum of 0 applied to S.

Properties of Tests Containing Intrinsic Items

The data reported earlier utilized questions which had unique correct responses. Items having such answers have been used for a long time. Intrinsic items which do not have unique correct answers have not been used to any appreciable extent. In fact, such items are not easy to find. Several tests of this sort were produced in the course of this investigation. One written by the author, one written by Sallie Churchill, and one written by Leona Peterson may be found in Appendix . It is not always easy to find a sample of subjects for such items. Therefore, in some of the data, intrinsic items were simulated using very difficult items from the STEP Writing Test, Level 1, by instructing Ss that the items might or might not have unique correct responses. In the data reported, samples will be designated by numbers. Sample 1 was a group of 125 freshmen from one of the Chicago Junior Colleges. These subjects took the 15 simulated intrinsic items. Samples 3 and 5 also took the 15 simulated intrinsic items. Ss were sophmores and juniors from a suburban high school. Sample 2 consisted of 263 students in Graduate Social Work programs in five universities. These Ss took the Test of Troublesome Management Situations which is found in the appendix. This test has two separate keys, with one set of responses representing a diagnostic-treatment approach to social work, and another set of responses representing a mediating approach to social work. The third option represents a nondescript response, not based on any particular philosophical or theoretical position.

In all four samples, Ss were asked to respond to each item with a distribution of preference. In addition, they were asked to make an independent assessment of their confidence in this distribution. The items were scored using both the weighted and the unweighted Euclidean functions.

Importance of Weighting

Whether or not confidence weighting is useful is a question which is not yet clear. Tests containing intrinsic items should probably be scored with confidence weighting until evidence shows that the confidence weighting is unnecessary. This study developes some empirical evidence that weighting may not be necessary.

In an analysis of the responses from several of the data samples, weighting had practically no effect on reliability. This lack of difference is shown in Table 5.

Thus, although the reliability is consistently higher with weighting, the difference is very slight.

The reliability of the writing tests may seem low. However, the Spearman Brown correction for test length increases the reliability of the least reliable set of scores (.347) to 0.68, which is not far from the reported reliability of 0.74 for STEP Level 1 (Educational Testing Service, 1957). The maximum corrected reliability for the sample 4, weighted, is 0.80.

TABLE 5

COMPARISON OF RELIABILITIES OF EUCLIDEAN SCORES, WEIGHTED AND UNWEIGHTED

	Sample	Test	N	Reliability Weighted	Reliability Unweighted
1	Junior College Freshmen	15 Item Writing Test	125	.396	.347
2	Graduate Social Work Students	24 Item Problem Management Test	297	.755	.743
	High School Sophmores and Juniors				
3	Sample A	15 Item Writing Test	263	.401	.396
4	Sample B	15 Item Writing Test	135	.493	.478

The Meaning of Confidence

In order to explore the nature of confidence ratings which Ss assign to test items, a variety of data were used by Grozelier for her master's paper (1970). Two hundred sixty-three Sophomore and Junior students from a high school in a suburb of Chicago were randomly assigned to two groups and administered the achievement test in writing described previously. One group was told that the test they were taking would count toward their grade in English. The other group was told that the test was being administered for research purposes and would not be counted on their grades. The teachers were given the grades of the subjects in the incentive group, and they had agreed to utilize them in grading, although the amount of weight to be given to the results was not specified. They were instructed in the system of scoring to be used as follows:

> Each of the questions in this test is followed by suggested answers. Assign a number from 0 to 9 to each suggested answer, depending on how strongly you feel that the answer is correct. If you believe that only one suggested answer is correct, mark that answer with a 9 and mark the other(s) with zeros. If you like the suggested answers equally, assign the same number to each. The sum of the three responses should add up to 9. . .

Then additional information was given by the test supervisors on the way it was going to be scored:

> If your answer is closer to the right answer, you will get a positive score. If it is closer to the wrong answer you will get a negative score. The scores vary from -1 to +1. They are multiplied by your certainty (answer D).

The test itself was preceded by a six item practice test at the end of which subjects were given the right answer for each question and could ask any question about the instructions. They were told that for the practice test there was one single right answer, but for the test itself, there might or might not be more than one single right answer to each item.

Grozelier's data and additional variables were used in the analysis which follows. In addition to the achievement test, Ss were asked to fill out a personal data sheet, and were given a test of 5 personality variables.* From these instruments the following variables were measured:

1. Sex: Male = 1, Female = 2
2. Year in School: 1 = Sophmore, 2 = Junior
3. Score: Mean weighted Euclidean score on the 15 item writing test
4. Attitude: 0 = maximum dislike for test, 9 = maximum liking
5. Confidence: 0 = minimum confidence in responses, 9 = maximum
6. Autonomy: Scale score from Personality Research Inventory
7. Harm Avoidance: Personality Research Inventory
8. Impulsivity: Scale from Peraonality Research Inventory
9. Order: Scale from Personality Research Inventory
10. Succorance: Scale from Personality Inventory
11. Social Class: (on a three-point scale) Low = 1, Middle = 2, Upper = 3
12. Appropriateness of Confidence (WPLN)
13. Propensity to gamble (PLN)
14. Appropriateness of Confidence on an item of medium difficulty
15. Gambling propensity on an item of medium difficulty
16. Appropriateness of Confidence on an easy item, #7
17. Gambling propensity on an easy item
18. Appropriateness on a difficult item, #13
19. Gambling propensity on a difficult item.

Some explanation is necessary on the computation of variables 12 through 19.

The propensity to gamble, PLN, for an item was equal to the sum of the squares of the differences between numerical response for each of the responses and three, divided by six. That is,

$$PLN = \left(\sum_{j=1}^{3} (r_j - 3)^2\right)/6 \text{ for the } i^{th} \text{ item,}$$

where $0 \leq r_j \leq 9$

and $\sum_{j=1}^{3} r_j = 9$, j = option number

Since subject responses ranged from 0 to 9 for the three options, S who had no preference for the options, and who expressed this lack of preference by responding (3,3,3) to the three options would receive a PLN

---

*Scales Au, Ha, Im, Or, Su, from Douglas Jackson, Personality Research Form, Form AA, Research Psychologists Press Inc. 1965.

TABLE 6
INCENTIVE GROUP

	VARIABLE	MEAN	ERROR	N	ST.DEV.	ERROR
1	Sex	1.545	0.048	110	0.500	0.
2	Year	1.664	0.045	110	0.475	0.015
3	Score	4.344	0.306	110	3.209	0.197
4	Attitude	3.336	0.230	110	2.409	0.123
5	Confidence	7.912	0.085	110	0.891	0.089
6	Autonomy	9.229	0.345	109	3.597	0.193
7	Harm Avoidance	7.321	0.299	109	3.118	0.179
8	Impulsivity	10.376	0.291	109	3.036	0.181
9	Order	10.000	0.370	109	3.866	0.237
10	Succorance	9.899	0.381	109	3.979	0.198
11	Social Class	1.764	0.062	110	0.649	0.047
12	Approp. Confidence	1.527	0.071	110	0.739	0.053
13	Propensity to Gamble	7.130	0.092	110	0.963	0.069
14	Approp. on Med. Diff.	2.447	0.222	110	2.328	0.119
15	Gamble on Med. Diff.	6.571	0,319	110	3.349	0.117
16	Approp. on Easy Item	0.292	0.090	110	0.942	0.232
17	Gamble on Easy Item	8.890	0.070	110	0.730	0.283
18	Approp. on Hard Item	2.174	0.218	110	2.291	0.164
19	Gamble on Hard Item	6.346	0.298	110	3.126	0.150

TABLE 6
RELAXED GROUP

	VARIABLE	MEAN	ERROR	N	ST.DEV.	ERROR
1	Sex	1.477	0.048	111	0.502	0.
2	Year	1.631	0.046	111	0.485	0.012
3	Score	3.680	0.344	111	3.619	0.229
4	Attitude	3,027	0.244	111	2.574	0.122
5	Confidence	7.324	0.149	111	1.570	0.232
6	Autonomy	9.410	0.355	105	3.642	0.185
7	Harm Avoidance	7.571	0.355	105	3.642	0.198
8	Impulsivity	11.448	0.298	105	3.051	0.186
9	Order	8.667	0.415	105	4.251	0.263
10	Succorance	9.371	0.349	105	3.574	0.253
11	Social Class	1.982	0.067	110	0.704	0.063
12	Approp. Confidence	1.638	0.119	111	1.256	0.175
13	Propensity to Gamble	6.586	0.116	111	1.218	0.095
14	Approp. on Med. Diff.	2.432	0.239	111	2.521	0.176
15	Gamble on Med. Diff.	5.414	0.328	111	3.454	0.120
16	Approp. on Easy Item	0.383	0.142	111	1.501	0.354
17	Gamble on Easy Item	8.482	0.186	111	1.960	0.358
18	Approp. on Hard Item	2.220	0.218	111	2.296	0.216
19	Gamble on Hard Item	5.241	0.329	111	3.469	0.115

equal to zero. On the other hand, S showing a complete preference for a single option (propensity to gamble) would receive PLN = (36 + 9 + 9)/6 = 9. Thus PLN is an index of the subject's tendency to select a single option. PLN for a test would then consist of the average value of PLN over all the items.

Appropriateness of confidence compares S's PLN with his expressed confidence in the item. For the i^{th} item, appropriateness of confidence (WPLN) is the absolute value of the difference between S's PLN for that item and his confidence measure, D_i:

$$WPLN_i = \quad PLN_i - D_i$$

Theoretically, a person with no knowledge should declare $D_i = 0$ and distribute his responses (3,3,3). This would make PLN = 0 and D = 0. Thus a score of 0 on WPLN indicates congruence between PLN and D_i. An S who is certain of his response would mark one option with a nine and the other options with zeroes. This would make PLN = 9. If he was that certain, he should also mark D = 9, again giving WPLN = 0. Larger values of WPLN indicate a discrepancy between confidence and one's behavior in distributing his responses.

Means and standard deviations of the 19 variables under the relaxed and the incentive conditions are shown in Tables 6 and 7.

The reliability of the test under the incentive condition was 0.261. Under the relaxed condition it was 0.493. Although the mean scores were significantly higher under the incentive condition, the reliability of these scores was consistently lower.

Ss reported a slightly more favorable attitude toward the test under the incentive condition, although the average liking in both groups was indicative of mild displeasure. In the incentive group, there was a significantly greater amount of confidence than there was in the relaxed group, along with a significantly higher propensity to gamble. This was perhaps due to the greater desire of these Ss to improve their scores, for the confidence expressed in the incentive group was more congruent with their distribution of preference than was the confidence expressed by the relaxed group. Confidence was most appropriate on the easy item, and was least appropriate on the item of moderate difficulty.

In her master's paper, Grozelier (1970) made a further breakdown by sex and social class. The following interpretation of these data is taken directly from her paper.

> Sex. Examination of the data indicates that girls were slightly more sensitive to the incentive effect than were boys (the increase in the girls PLN mean is significant at .02). With regard to the level of risk taking thereare no important differences coming out but at the item 13 where boys appear rather conservative and girls high risk minded. It follows from our basic assumption that the motive to achieve success would be stronger among boys whereas girls would rather be failure avoidance oriented.

Obviously this assertion is not absolutely reliable; an independent and simultaneous measure of the achievement motive would have been sounder. This observation might be accounted for by cultural factors, in their education boys are subject to stronger emphasis to reinforce success achievement trends than are girls.

On item 6, higher class subjects appear to be more conservative than the other two class students. This is particularly conspicuous under the incentive condition (PLN_i mean = 5.1 for the higher class, PLN_i mean = 5.9 for the middle class, PLN_i mean = 6.1 for the lower class). Unfortunately some looseness due to uncertainness in the sample does not allow to estimate the difference between middle class and lower class as significant. But if we assume this difference as significant we would be confronted with a trichotomous situation, middle class appearing success achievement oriented and the two other classes as fear of failure motivated. It should also be noted that at this item which represents the average level of difficulty of the two other classes, were not sensible to incentive effect whereas the lower class felt it strongly (PLN_i increase of 1.0). Suprisingly the neutral condition determined a slight increase in the middle class level of risk taking.

In short, middle class subjects appear as moderate risk takers and therefore qualify as motivated to achieve success whereas lower class. . .fear of failure oriented. These assertions are supported by the results but the upper class pattern is not so clearly defined. On the whole test and on the particular item 6 upper class appeared conservative in risk taking but in the neutral condition and on item 13 they qualified as fear of failure oriented.

It follows that middle class students fared the best in the confidence testing situation, their PLN indexes stand at the intermediate level and they get scores slightly higher than the two other classes (though it was not statistically significant). On the other hand they tend to display a motivation to achieve success. . . It seems that the cultural and educational training which characterizes middle class youngsters would give a priviligiated preparation to face this confidence testing situation. Strodtbeck (1958) in his investigation of childhood training and achievement described the emphasis put upon early mastery by children. Youngsters are trained to be able to adapt themselves quickly to a new situation and be able to get the best of it; they are ready to work longer to succeed in a task and, as they are rather ego-oriented, personal realization represents more value to them than a good grade.

Lower class students fared the worst on this test. They were more risky minded and tended to get the lowest scores. This finding can be interpreted as an issue of confidence testing procedure which emphasizes the importance of using partial knowledge and uncertainty; guessing is penalized but acknowledgement of partial

TABLE 8

GROUP MEANS

		TOTAL	SEX M	SEX F	GRADE SO	GRADE JU	SOCIAL CLASSES I	SOCIAL CLASSES II	SOCIAL CLASSES III
PLN MEAN	NEUTRAL CONDITION	6.5	6.4	6.7	6.4	6.6	6.2	6.6	6.9
	INCENTIVE CONDITION	7.0	6.9	7.2	7.1	7.0	7.1	7.0	7.8
	TOTAL	6.8	6.6	6.9	6.7	6.8	6.8	6.8	7.3
PLN ITEM 7 (Easy)	NEUTRAL CONDITION	8.5	8.6	8.4	8.7	8.4	8.7	8.6	8.2
	INCENTIVE CONDITION	8.9	8.9	8.9	8.8	8.9	8.9	9.0	8.3
	TOTAL	8.7	8.7	8.7	8.7	8.7	8.8	8.7	8.3
PLN ITEM 6 (Average)	NEUTRAL CONDITION	5.5	5.6	5.3	6.3	5.1	5.2	5.5	5.1
	INCENTIVE CONDITION	5.8	5.0	6.3	5.1	6.1	5.1	5.9	6.1
	TOTAL	5.6	5.4	5.9	5.7	5.6	5.1	5.7	5.6
PLN ITEM 13 (Difficult)	NEUTRAL CONDITION	5.4	4.5	6.3	5.8	5.0	4.6	5.5	5.4
	INCENTIVE CONDITION	6.1	5.4	6.6	6.4	6.0	6.9	6.0	5.9
	TOTAL	5.7	4.9	6.4	6.1	5.5	6.1	5.7	5.7

knowledge is rewarded. This test emphasizes adjustment and compromise but lower class people take higher risk because they precisely do not have the ability to compromise. This hypothesis has already been suggested by Bourdieu (1964) who pointed out this particular incapacity, affecting lower class people, to stand on an intermediate position. He showed how lower class children in the class room have only two extermes of attitude at their disposal, to recognize coarsely either their ignorance or their knowledge without any continuum in their attitude as the student lives in a two pole world limited to right and wrong. These students never learned how to scale a lack of knowledge by a skillful utilization of their available information because this would require a higher command of language which they do not learn at school. Hence one may conclude that many lower class students are not equipped to cope with this new kind of testing. They do not have the necessary skills.

Higher and lower class students could be labled as fear of failure motivated (taking extreme risks). . . the fear of failure motivation acts as an obstacle which keeps them from fitting into the new system. But the reason why this motive is working this way probably brings us back to some sociological factors. Nevertheless higher class people might get the advantage from their family milieu, being used to a more intellectual environment could help them to understand the complex presentation of the test. Of course this works just the opposite for lower class people who appeared particularly disadvantaged at this test; they tend to get the lowest scores on the whole test but the discrepancy becomes really striking when considering the neutral situation where they get a score mean of 1.9 against a score mean of 3.2 among the higher class and a score mean of 3.6 among the middle class.

Correlations were computed for each of the two samples for all 19 variables, The correlation matrices are shown in Tables 9 and 10. Correlations larger than r = .195 will be examined. For a single pair of variables, a correlation of 0.195 indicates a significant departure from 0.0 at the 0.025 level with 100 degrees of freedom. (Walker and Lev, 1953). Comparing significant correlations in the two matrices, it can be seen that there was a significant relationship between sex and attitude toward the test with the girls liking it better than the boys. This sex difference was accentuated under the incentive condition. The males were more Autonomous and less Succorant in both groups. This should be expected because the personality test was not involved in the incentive instructions. Finally, only the difficult item provided a significant correlation with appropriateness of judgment of confidence and the propensity to gamble with the females showing a greater willingness to make extreme choices, and also exhibiting greater congruence between their feelings of certainty and their behavior in responding to the items. That is, the females were more inclined to choose single responses, but they also felt more certain about their choices than did the males. Confidence was significantly related to score under both conditions, though the relationship was higher under the relaxed condition. That is, subjects were more willing to take extreme positions under the relaxed condition.

TABLE 9

CORRELATION COEFFICIENTS - RELAXED CONDITION

	VARIABLE	1	2	3	4	5	6	7	8	9	10
1	Sex	1.000									
2	Year	0.022	1.000								
3	Score	0.096	-0.039	1.000							
4	Attitude	0.278	-0.036	0.110	1.000						
5	Confidence	0.194	0.062	0.209	0.230	1.000					
6	Autonomy	-0.283	-0.211	-0.039	-0.183	-0.189	1.000				
7	Harm Avoidance	-0.090	0.074	-0.102	-0.004	-0.028	-0.423	1.000			
8	Impulsivity	0.066	0.101	0.044	-0.041	-0.002	0.168	-0.357	1.000		
9	Order	0.085	-0.018	-0.133	-0.016	0.099	-0.306	0.314	-0.543	1.000	
10	Succorance	0.396	0.066	0.103	0.365	0.218	0.539	0.241	-0.058	0.089	1.000
11	Social Class	0.076	0.194	-0.058	0.156	0.023	-0.105	-0.016	-0.065	-0.039	0.070
12	Approp. Confidence	-0.033	-0.190	-0.170	-0.188	-0.325	0.089	0.029	-0.089	0.034	-0.065
13	Propensity to Gamble	0.134	0.085	0.221	0.137	0.180	-0.071	0.008	-0.016	0.045	0.047
14	Approp. on Med. Diff.	-0.006	0.059	-0.028	-0.078	-0.205	0.053	0.018	0.133	-0.080	0.059
15	Gamble on Med. Diff.	0.076	-0.235	0.082	0.125	0.122	0.192	-0.023	-0.196	0.057	-0.070
16	Approp. on Easy Item	-0.119	-0.154	-0.114	-0.206	-0.547	0.095	0.122	-0.105	0.045	-0.098
17	Gamble on Easy Item	0.001	-0.117	0.219	0.137	0.293	-0.133	0.021	-0.080	0.043	0.228
18	Approp. on Hard Item	-0.202	-0.119	-0.036	-0.089	-0.254	0.069	0.037	0.077	0.012	-0.045
19	Gamble on Hard Item	0.270	-0.122	0.178	0.199	0.123	-0.030	0.031	-0.112	0.067	0.076

		11	12	13	14	15	16	17	18	19
11	Social Class	1.000								
12	Approp. Confidence	0.036	1.000							
13	Propensity to Gamble	0.176	-0.481	1.000						
14	Approp. on Med. Diff.	-0.060	0.370	-0.251	1.000					
15	Gamble on Med. Diff.	0.046	0.026	0.218	-0.695	1.000				
16	Approp. on Easy Item	-0.070	0.597	-0.178	0.278	0.093	1.000			
17	Gamble on Easy Item	-0.159	-0.036	-0.009	0.080	-0.051	-0.153	1.000		
18	Approp. on Hard Item	-0.063	0.343	-0.217	0.268	-0.012	0.342	0.167	1.000	
19	Gamble on Hard Item	0.036	0.034	0.297	-0.013	0.194	0.080	-0.065	-0.458	1.000

TABLE 10

CORRELATION COEFFICIENTS - INCENTIVE CONDITION

	VARIABLE	1	2	3	4	5	6	7	8	9	10
1	Sex	1.000									
2	Year	-0.148	1.000								
3	Score	0.150	-0.219	1.000							
4	Attitude	0.334	0.188	0.167	1.000						
5	Confidence	0.100	-0.159	0.178	0.192	1.000					
6	Autonomy	-0.284	0.029	-0.060	-0.173	-0.045	1.000				
7	Harm Avoidance	0.116	-0.007	0.032	0.058	-0.080	-0.206	1.000			
8	Impulsivity	0.142	-0.007	-0.041	0.065	-0.007	0.056	-0.382	1.000		
9	Order	0.024	0.115	0.093	-0.019	-0.046	-0.080	0.260	-0.374	1.000	
10	Succorance	0.429	-0.018	0.143	0.167	-0.053	-0.678	0.358	0.027	0.072	1.000
11	Social Class	0.033	0.097	0.093	0.139	0.073	0.090	0.115	-0.057	0.124	0.051
12	Approp. Confidence	-0.095	0.013	-0.074	-0.030	-0.398	-0.009	0.150	-0.077	0.178	0.081
13	Propensity to Gamble	0.158	0.024	0.203	0.215	0.473	-0.066	-0.191	0.152	-0.158	-0.013
14	Approp. on Med. Diff.	-0.199	-0.110	-0.034	-0.152	0.013	-0.039	-0.012	-0.139	0.036	-0.034
15	Gamble on Med. Diff.	0.160	0.159	0.087	0.174	0.174	0.120	0.024	0.106	-0.049	-0.045
16	Approp. on Easy Item	-0.074	-0.074	0.097	-0.050	0.108	0.178	0.137	-0.102	-0.081	-0.209
17	Gamble on Easy Item	-0.003	0.141	-0.050	-0.007	-0.138	-0.061	-0.114	0.080	0.079	0.128
18	Approp. on Hard Item	-0.244	0.105	0.166	-0.080	-0.300	0.077	0.155	-0.056	0.219	0.040
19	Gamble on Hard Item	0.253	-0.104	-0.273	0.063	0.033	0.001	-0.211	0.096	-0.108	-0.039

11	Social Class	1.000								
12	Approp. Confidence	-0.059	1.000							
13	Propensity to Gamble	0.136	-0.707	1.000						
14	Approp. on Med. Diff.	-0.023	0.138	-0.157	1.000					
15	Gamble on Med. Diff.	0.139	-0.165	0.304	-0.807	1.000				
16	Approp. on Easy Item	0.174	-0.053	0.047	0.247	-0.049	1.000			
17	Gamble on Easy Item	-0.133	0.209	-0.136	-0.101	0.091	-0.710	1.000		
18	Approp. on Hard Item	0.131	0.443	-0.226	0.063	-0.076	0.067	0.027	1.000	
19	Gamble on Hard Item	-0.099	-0.189	0.214	-0.037	0.008	-0.121	0.060	-0.567	1.000

Thus the subjects seemed to be more motivated by fear of failure than by potential reward. It is of additional interest to note that there was no relationship between score and the gamble score on the easy items in the incentive condition, while the significant relationship was on the hard item under the incentive condition. In fact, the gamble hard score - score correlation changed sign going from the relaxed condition to the incentive condition. That is, for the high scoring Ss, there was a tendancy to assume extreme positions on the hard items under the relaxed condition, but an unwillingness to do so under the incentive condition. That is, where grades were at stake, the high scoring Ss played the cautious role. S's attitude toward the test was related primarily to his confidence, although there was also a significant relationship with the gamble score in the incentive condition. Confidence was significantly related to inappropriatness of judgment and to willingness to take extreme positions under the incentive condition. That is, under the incentive, subjects who were confident about their responses were more willing to take extreme positions in responding. However, these extreme positions did not match their degrees of confidence very well. Several other of the item scores were related to confidence in the relaxed condition, while the gamble score became less important. The personality variables showed substantial intercorrelations as did the cluster of gamble and appropriatness scores. The significant negative correlations between the gamble and the appropriatness scores is due to the fact that these two scores are not independent of one another. The negative sign becomes obvious when one examines the means of computation of the appropriatness score (WPLN) from gamble score (PLN).

In order to better understand what variables contributed to S's expression of confidence, a regression analysis was performed. No significant regression held between confidence and any other variables, although high succorance and low harm avoidance did contribute a small amount to the prediction of confidence in the relaxed condition only.

Seventeen of the scores were factor analyzed. The PLN and WPLN variables for the item of medium difficulty were left out since they did not seem to provide much information. A principal components analysis was first performed. Then the principal components were rotated according to the following specifications: A maximum of nine factors were to be extracted, the lower limit of eigenroots was set at 1.00 and no factors were to have loadings of less than .30 for at least one variable. According to these specifications, seven factors were rotated. Ten rotations were required in the incentive condition. Thirteen were required in the relaxed condition. The factor matrix is shown in Tables 11and 12 Loadings in excess of 0.30 are underlined.

In interpreting these results, it should be recalled that a low numerical score on the Appropriate variable means that a person's responses were congruent with his confidence. The factor analysis did not reveal much about confidence, except to underline the fact that there is a dependence between it and the gamble and appropriate measures. This is illustrated in Factor 1 in both conditions. Factor 2 is made up of sex and several personality variables. Attitude is also a relevant variable

TABLE 11

FACTOR MATRIX--INCENTIVE CONDITION

No.	Name	Communality	1	2	3	4	5	6	7
	Factor Number		1	2	3	4	5	6	7
	Factor Number (before Rotation)		1	2	3	6	5	4	7
	Sum of Squares		2.245	2.069	1.847	1.845	1.687	1.621	1.302
	PerCent 17 Factors		12.5	24.0	34.2	44.5	53.8	62.9	70.1
	7 Factors		17.8	34.2	48.8	63.5	76.8	89.7	100.0
1	Sex	0.741	-0.033	0.463	-0.006	0.393	-0.070	0.485	0.364
2	Year	0.802	0.000	-0.002	-0.125	-0.159	0.034	0.278	-0.826
3	Score	0.733	0.250	0.087	-0.025	-0.461	0.049	0.230	0.629
4	Attitude	0.511	0.081	0.260	-0.004	0.069	-0.071	0.652	-0.023
5	Confidence	0.572	0.699	-0.025	0.054	0.057	0.071	0.113	0.244
6	Autonomy	0.806	-0.107	-0.877	0.093	-0.005	-0.061	0.113	0.025
7	Harm Avoidance	0.604	-0.195	0.390	0.257	-0.088	-0.569	0.124	-0.022
8	Impulsivity	0.738	-0.028	0.022	-0.100	0.004	-0.837	0.157	0.022
9	Order	0.619	-0.101	0.005	-0.176	-0.114	0.743	0.108	0.023
10	Succorance	0.825	-0.073	0.888	-0.104	-0.051	0.039	0.112	0.062
11	Social Class	0.419	0.044	-0.080	0.254	-0.165	0.179	0.527	-0.096
12	Approp. Confidence	0.772	-0.840	0.008	-0.104	-0.198	0.097	0.004	0.082
13	Propensity to Gamble	0.793	0.833	0.042	0.057	0.057	-0.173	0.245	-0.022
14	Approp. on Easy Item	0.827	0.022	-0.156	0.888	-0.091	0.023	0.031	0.060
15	Gamble on Easy Item	0.830	-0.164	0.012	-0.894	-0.002	0.003	0.028	-0.049
16	Approp. on Hard Item	0.746	-0.365	-0.030	0.019	-0.778	0.065	0.043	-0.028
17	Gamble on Hard Item	0.762	0.063	-0.046	-0.086	0.860	-0.086	0.039	0.009

TABLE 12

FACTOR MATRIX--RELAXED CONDITION

No.	Name	Communality	1	2	3	4	5	6	7
	Factor Number		1	2	3	4	5	6	7
	Factor Number (before rotation)		1	5	2	6	3	7	4
	Sum of Squares		2.233	2.128	2.055	1.623	1.545	1.415	1.313
	PerCent 17 Factors		12.4	24.2	35.6	44.7	53.2	61.1	68.4
	7 Factors		18.1	35.4	52.1	65.3	77.8	89.3	100.0
1	Sex	0.622	0.051	0.671	-0.115	0.381	0.072	-0.050	-0.056
2	Year	0.610	0.124	0.019	-0.010	-0.014	-0.705	0.073	-0.302
3	Score	0.651	0.068	0.172	-0.179	0.034	0.101	0.706	0.273
4	Attitude	0.562	0.257	0.578	-0.051	0.009	0.287	0.072	-0.269
5	Confidence	0.633	0.719	0.285	0.059	-0.048	0.097	0.063	0.124
6	Autonomy	0.726	-0.040	0.608	-0.439	-0.032	0.389	-0.094	0.022
7	Harm Avoidance	0.638	0.179	0.133	0.686	0.063	-0.306	0.129	0.049
8	Impulsivity	0.719	0.001	0.038	-0.802	0.037	-0.262	0.032	0.060
9	Order	0.664	0.072	0.047	0.790	-0.059	0.078	-0.145	0.034
10	Succorance	0.722	0.045	0.815	0.141	0.059	0.161	0.073	0.033
11	Social Class	0.697	0.011	0.159	-0.011	0.037	0.024	0.005	-0.818
12	Approp. Confidence	0.736	0.710	0.101	0.015	0.101	0.227	-0.391	0.086
13	Propensity to Gamble	0.727	0.237	-0.034	0.081	-0.236	0.016	0.716	-0.307
14	Approp. on Easy Item	0.812	0.886	0.066	0.097	0.028	0.107	0.007	0.044
15	Gamble on Easy Item	0.645	0.276	0.300	0.082	0.461	0.204	0.171	0.434
16	Approp. on Hard Item	0.775	-0.387	-0.013	-0.034	0 783	0.104	-0.011	0.008
17	Gamble on Hard Item	0.764	-0.111	0.212	0.070	-0.748	0.253	0.271	0.071

in the relaxed condition, but the importance of attitude in this factor is much reduced in the incentive condition. Factor 3 in the relaxed condition and Factor 5 in the incentive condition are quite similar and are made up entirely of the personality factors. Factor 4 is perhaps the only one of much interest. It shows a relationship among sex and the way in which Ss deal with the difficult items. This, however, only confirms what has been previously said about sex differences with respect to making dogmatic choices on items.

Conclusions

The findings may be summarized as follow:

1. The reliability is proportional to the deviation of scoring functions from simple methods which subjects anticipate.

2. The student is unfairly penalized by the scoring function which assigns him a score equal to the probability assigned to the right answer. Therefore, this function should only be used when minimal or no rewards are attached to subject performance. Otherwise the Euclidean function will not penalize the student, and will give high reliabilities in a no-feedback situation. In the event that feedback through tables, computers, or scoring aids is available, the logarithmic function is recommended.

3. On items not having unique correct responses, the weighted Euclidean function is only slightly superior to the unweighted with respect to reliability.

4. Under incentive conditions, scores on confidence tests are higher, and reliability significantly lower.

5. Females have a greater tendency toward taking extreme positions than males, especially in the incentive condition.

6. Subjects in the incentive group liked the test better, had more of a tendency to take extreme positions, and made more appropriate estimates of their confidence.

7. Middle SES subjects, compared to both upper and lower SES subjects, made higher scores and more appropriate estimates of confidence. They seemed to be motivated more by desire for success than fear of failure.

8. High scoring subjects gambled more on difficult items under the relaxed condition, but gambled less on difficult items in the incentive condition.

9. Liking of tests was directly related to confidence.

10. There was no significant regression between confidence and the battery of personality variables, although high succorance and low harm avoidance made small contributions to prediction.

Much work remains to be done in studying confidence testing. Although it is clear that technical improvements may be made in the reliability and validity of tests through confidence scores, it is also clear that subjects do not handle their confidence uniformly. What is confidence to one may be hazard to another. As Wang and Stanley state, (1970) The derivation of optimum response strategies in multiple choice testing represents an application of mathematical decision theory which underscores the decision process inherent in such tests. The success of testing procedures which attempt to control the decision process will be critically dependent on the ability of subjects to effectively use optimal strategies. It is not certain that all subjects are equally capable of learning to use such strategies.

The question of optimal strategies is likely to be perhaps the most significant outcome of further research on confidence testing. Although Bruner (1956) pointed out two basic differences in the way subjects use their confidences - the sentry condition and accuracy condition, and demonstrated empirical evidence of these two modes of behavior, there are other complex conditions which intervene between a subjective probability and a decision or action. Since it is possible, although not guaranteed that one may assess subjective probabilities accurately by means of reproducing scoring functions, two basic steps are needed. First, subjects need more experience in utilizing reproducing scoring functions. It takes a while to learn to respond intelligently to the rules of that game. Once it is possible to be confidence of measures of subjective probability on a set of subjects, further study may be made of the use of optimal strategies by subjects in problematic situations. Such strategies would perhaps start with what is known about optimal search procedures in polychotomic trees (Watanabe, 1969). Although the ability to utilize optimal strategies, and the ability to make appropriate assessments of one's subjective probabilities is of value in its own right, it would perhaps be a useful next step to begin to apply information about subjective probabilities to the study of the structure of subject matter. This could be done through an analysis of the associative networks of highly trained subjects terms utilizing a system of analysis similar to that of Quillian (1968), substituting subjective probability in place of his all or nothing at all lines of association. Further development of such techniques, and further gathering of data on sophisticated human subjects may lead to the uncovering of most of the appropriate parameters involved in guiding decision making in problematic situations.

This goal is an ambitious one. Perhaps at a more realistic level would be the goal of increasing emphasis on the ways students react to problematic situations. Are students able to assess their state of information and respond intelligently to it? Do our teaching and testing practices make them aware that there are differences among the ways we use our information?

APPENDIX A

TEST SCORING PROGRAMS

```
CC
C
C
C
C         PROBABILISTIC TEST SCORING PROGRAM R.RIPPEY MARCH 1971
C        THIS PROGRAM ALLOWS THE SCORING BY ALL FUNCTIONS IN ONE RUN
C        DECK SETUP ------
C        CARD 1 IS A TITLE CARD (COLUMN 2-72)
C        CARD 2 IS A CONTROL CARD
C              COL 1 STANDARD DEVIATION OPTION
C                    1 = YES
C                    2 = NO
C              COL 4 PUNCHED OUTPUT OPTION
C                    1 = YES
C              BLANK OTHERWISE
C              COL 6 LOG TRANSFORM OF EUCLIDEAN FUNCTION
C                    1 = YES
C                    2 = NO
C                    NOT APPLICABLE IF EUCLIDEAN FUNCTION IS NOT REQUESTED
C        CARD 3 PARAMETER CARD
C              COL 1 DATA TYPE
C                    1 = REGULAR DATA
C                    2 = PORTA PUNCH
C              COL 2-3 TOTAL NUMBER OF ITEMS ON TEST
C              COL 4 RESPONSES PER ITEM -- MUST BE SIX OR FEWER
C              COL 5-6 NUMBER OF CARDS PER SUBJECT -- MUST BE TEN OR FEWER
C              COL 7-8 NUMBER OF ITEMS PER CARD
C        CARD 4 FUNCTION SELECT CARD
C              A 1 IS TO BE PUNCHED IN THE COLUMN CORRESPONDING TO THE
C              FUNCTIONS TO BE SELECTED
C                    COL 1 -- EUCLIDEAN FUNCTION FOR SCORING INTRINSIC ITEMS
C                    COL 2 -- SPHERICAL FUNCTION
C                    COL 3 -- TRUNCATED LOGARITHMIC
C                    COL 4 -- INFERRED CHOICE
C                    COL 5 -- EUCLIDEAN FUNCTION FOR SCORING ITEMS WITH
C                          UNIQUE CRRECT ANSWERS
C                    COL 6 -- SIMPLE PROBABILITY ASSIGNED TO THE CORRECT AN
C                    COL 7 -- ENTROPIC
C                    COL 8 --
C                    COL 9 --
C                    COL 10--
C        KEY CARD(S)
C              FOR FUNCTIONS 1 AND 7 -----
C                    PASS ONE OUTPUT
C              FOR FUNCTIONS 2, 3, 4, 5, ND 6 -----
C                    A SINGLE KEY CARD CONTAINING THE NUMBER OF THE CORRECT
C                    RESPONSES IN THE COLUMN CORRESPONDING TO THE NUMBER OF
C                    THE ITEM.  FOR EXAMPLE, IF THE CORRECT RESPONSE TO
C                    ITEM SEVEN WOULD BE RESPONSE TWO, THEN A 2 SHOULD BE
C                    PUNCHED IN COLUMN 7.
C        RESPONSE CARDS
C        LAST CARD OF EACH JOB SHOULD BE BLANK
C        JOBS MAY BE STACKED
C        LAST CARD OF LAST JOB SHOULD CONTAIN A -1 PUNCHED IN CLS 1-2.
C
C
C
C
C
```

```
C
      COMMON MA,MC,MD,KF,KI,KN,KK,KIC,Z,A,B
      DIMENSION AVANS(72,6), ANSPR(72,6), SI(72,10), ID(10), ANS(72,6),
     1    SS(72), S2(72), Z(72), AJOB(12), KS(10), A(72),B(72),SST(10),
     2    ST(10), SSP(10),REM(10),VAREM(10),VARP(10),DEFR(10), REL(10),
     3    CF(10),SSI(10),JFUNC(10),SM(72),D(6)
C
C
C
C
C
      JOB = 0
  987  JOB = JOB + 1
      WRITE OUTPUT TAPE 6, 986, JOB
     READ INPUT TAPE 5,2235, (AJOB(I),I=1,12)
     WRITE OUTPUT TAPE 6,2235,(AJOB(I),I=1,12)
C
C
C
C
C
      DO 4000 I = 1,72
      SS(I) = 0
      S2(I) = 0
      Z(I) = 0
      A(I) = 0
      B(I) = 0
      DO 4001 J = 1,6
      AVANS(I,J) = 0
      ANSPR(I,J) = 0
 4001 ANS(I,J) = 0
      DO 4000 J = 1,10
 4000 SI(I,J) = 0
      DO 4002 I = 1,10
      SSI(I) = 0
      SST(I) = 0
      ST(I) = 0
      SSP(I) = 0
      REM(I) = 0
      DEFR(I) = 0
      VAREM(I) = 0
      VARP(I) = 0
      REL(I) = 0
      CF(I) = 0
 4002 JFUNC(I) = 0
      EN = 0
      BZRK = 0
C
C
C
C
C
      READ INPUT TAPE 5,1,MA,MB,MC,MD
      IF(MA)940,940,241
  241 READ INPUT TAPE 5,1,                     KF,KI,KN,KK,KIC
      WRITE OUTPUT TAPE 6,930,KI,KN
      READ INPUT TAPE 5,4050, (JFUNC(I),I=1,10)
C
```

```
C
C
C
C
      IF (JFUNC(1) + JFUNC(7) + JFUNC(8)) 4020,4020,4021
 4020 IF (JFUNC(2)+JFUNC(3)+JFUNC(4)+JFUNC(5)+JFUNC(6)) 4022,4022,4023
 4021 IF (JFUNC(2)+JFUNC(3)+JFUNC(4)+JFUNC(5)+JFUNC(6)) 4024,4024,4025
 4022 WRITE OUTPUT TAPE 6,4030
      CALL SEARCH
      GO TO 987
 4023 GO TO 4
 4024 GO TO 3
 4025 WRITE OUTPUT TAPE 6,4031
      CALL SEARCH
 4040 GO TO 987
    3 KL=KI KN
      DO82 K=1,KL
      READ INPUT TAPE 5 , 120,I,J,AK
   82 AVANS(I,J)=AK
      GO TO 6
    4 READ INPUT TAPE 5,994,(Z(I),I=1,KI)
      DO 5 I=1,KI
      M=Z (I)
    5 AVANS(I,M)=1.0
    6 CONTINUE
C
C
C
C
C
      IF (JFUNC(8)) 15,15,14
   14 DO 8 J = 1,KI
    8 B(J) = 0
      DO 7 I = 1,KI
      DO 9 K = 1,6
    9 D(K) = 0
      DO 10 K =1,6
      DO 11 J = 1,6
      COMP = AVANS(I,J) ** 2
      IF (K - J) 13,12,13
   12 COMP = (1. - AVANS(I,J)) ** 2
   13 D(K) = D(K) + COMP
   11 CONTINUE
   10 D(K) = SQRTF(D(K))
      B(I) = D(1)
      DO 7 K = 2,KN
      IF (D(K) - B(I)) 7,7,18
   18 B(I) = D(K)
    7 CONTINUE
C
C
C
C
C
      WRITE OUTPUT TAPE 6, 5002
 5002 FORMAT (21H1MAXIMUM DISTANCE FOR)
      DO 5000  I = 1,KI
 5000 WRITE OUTPUT TAPE 6, 5001, I, (AVANS(I,J),J=1,3), B(I)
```

```
5001  FORMAT (5H ITEM, I5, 3H---, 3F10.4, 4H IS , F10.4)
C
C
C
C
C
  15  CONTINUE
C
C
C
C
C
 211 GO TO  (101,102),KF
 101 JP=1
      K=1
      IF (KK-1) 32,32,999
 999 DO 121 K=1,KK
      JL=K*KIC
      IF (JL-KI)42,42,22
  32 JL=KI
  42 CONTINUE
      READ INPUT TAPE 5 ,111,ID(K),KS(K),((ANS(I,J),J=1,KN),I=JP,JL)
         IF(ID(K))940,240,121
 121 JP=JL+1
         GO TO 37
C
C
C
C
C
 102 JP=1
      K=1
      IF (KK-1) 532,532,998
 998 DO 172 K=1,KK
      JL=K*KIC
      IF(JL-KI)533,533,532
 532 JL=KI
 533 CONTINUE
      READ INPUT TAPE 5,112,ID(K),KS(K),((ANS(I,J),J=1,KN),I=JP,JL)
         IF(ID(K))940,240,172
 172 JP=JL+1
C
C
C
C
C
  37 G=ID(1)
      EN = EN+1.
      N=EN
      IF(G)940,240,41
  41 CONTINUE
C
C
C
C
C
      IF(KK-1)  36,36,81
  81 CONTINUE
```

```
      DO 36 K=2,KK
      LA=ID(K)
      LB=ID(K-1)
      IF(LA-LB)33,80,33
   80 LC=KS(K)
       LD=KS(K-1   )
      IF(LC-LD) 35,35,36
   33 WRITE OUTPUT TAPE 6,113,ID(K)
       CALL SEARCH
       GO TO 987
   35  WRITE OUTPUT TAPE 6,114,ID(K)
       CALL SEARCH
       GO TO 987
   36 CONTINUE
C
C
C
C
C
       WRITE OUTPUT TAPE 6, 4003
       WRITE OUTPUT TAPE 6,4004, ID(1)
C
C
C
C
C
C      PROGRAM NOW WEIGHTS AND  COMPUTES PROBABILITIES
       DO 367 I=1,KI
       SM(I)=0
       DO 366 J=1,KN
  366 SM(I) = SM(I) + ANS(I,J)
       BZRK= SM(I)
       IF(BZRK)5327,5327,5328
 5327  ANSPR(I,J) = 0.
       GO TO 367
 5328  CONTINUE
       DO 367 J=1,KN
      ANSPR(I,J)= ANS(I,J) / SM(I)
  367 CONTINUE
      WRITE OUTPUT TAPE 6,124, ((ANSPR(I,J),J=1,3),I=1,KI)
C
C
C
C
C
 4061  IF(JFUNC(1)) 4062,4062,4071
 4062  IF(JFUNC(2)) 4063,4063,4072
 4063  IF(JFUNC(3)) 4064,4064,4073
 4064  IF(JFUNC(4)) 4065,4065,4074
 4065  IF(JFUNC(5)) 4066,4066,4075
 4066  IF(JFUNC(6)) 4067,4067,4076
 4067  IF(JFUNC(7)) 4068,4068,4077
 4068  IF(JFUNC(8)) 4069,4069,4078
 4069  IF(JFUNC(9)) 4070,4070,4079
 4070  IF(JFUNC(10)) 4090,4090,4080
 4071  CALL FUNCA(AVANS,ANSPR,SSP,SI,SST,ST,SS,S2)
       GO TO 4062
 4072  CALL FUNCB(AVANS,ANSPR,SSP,SI,SST,ST)
```

```
      GO TO 4063
 4073 CALL FUNCC(AVANS,ANSPR,SSP,SI,SST,ST)
      GO TO 4064
 4074 CALL FUNCD(AVANS,ANSPR,SSP,SI,SST,ST)
      GO TO 4065
 4075 CALL FUNCE(AVANS,ANSPR,SSP,SI,SST,ST)
      GO TO 4066
 4076 CALL FUNCF(AVANS,ANSPR,SSP,SI,SST,ST)
      GO TO 4067
 4077 CALL FUNCG(AVANS,ANSPR,SSP,SI,SST,ST)
      GO TO 4068
 4078 CALL FUNCH(AVANS,ANSPR,SSP,SI,SST,ST)
      GO TO 4069
 4079 CALL FUNCI(AVANS,ANSPR,SSP,SI,SST,ST)
      GO TO 4070
 4080 CALL FUNCJ(AVANS,ANSPR,SSP,SI,SST,ST)
 4090 GO TO 211
C
C
C
C
C
C
  240 CONTINUE
C
C
C
C
C
      WRITE OUTPUT TAPE 6,4005
C
C
C
C
C
      EKI = KI
      DO 4006 J = 1,10
      MB = J
      IF (JFUNC(MB)) 4006,4006,3007
C
C
 3007 GO TO (921,922,923,924,921,2002,2004),MB
  921 WRITE OUTPUT TAPE 6,926
      GO TO 925
  922 WRITE OUTPUT TAPE 6,927
      GO TO 925
  923 WRITE OUTPUT TAPE 6, 928
      GO TO 925
  924 WRITE OUTPUT TAPE 6,3002
      GO TO 925
 2002 WRITE OUTPUT TAPE 6,2003
      GO TO 925
 2004 WRITE OUTPUT TAPE 6,2005
  925 CONTINUE
C
C
      CF(J) = (ST(J) * ST(J)) / (EKI * EN)
      SST(J) = SST(J) - CF(J)
```

```
      DO 702 I = 1,KI
  702 SSI(J) = SSI(J) + SI(I,J) * SI(I,J)
      SSI(J) = ((SSI(J) / EN) - CF(J))
      SSP(J) = ((SSP(J)) / EKI) - CF(J)
      DO 705 I = 1,KI
  705 SI(I,J) = SI(I,J) / EN
      WRITE OUTPUT TAPE 6,706
  706 FORMAT(18HOITEM DIFFICULTIES)
      DO 636 I=1,KI
  636 WRITE OUTPUT TAPE 6,704, I, SI(I,J)
      REM(J) = SST(J) - SSP(J) - SSI(J)
      DEFR(J) = (EKI - 1.) * (EN - 1.)
      VAREM(J) = REM(J) / DEFR(J)
      VARP(J) = SSP(J) / (EN - 1.)
      REL(J) = 1. - (VAREM(J) / VARP(J))
      WRITE OUTPUT TAPE 6,742, REL(J), N
      WRITE OUTPUT TAPE 6,776, VARP(J),VAREM(J),REM(J),SSI(J),SSP(J),
     1 SST(J),CF(J),ST(J)
 4006 CONTINUE
C
C
C
C
C
      GO TO (987 , 773 ) , MA
  773    EN = EN-1.
      WRITE OUTPUT TAPE 6,628
      DO 25 I=1,KI
      SS(I)=SS(I)*SS(I)/EN
      SS(I)= SQRTF((S2(I)-SS(I))/(EN-1.))
   25 WRITE OUTPUT TAPE 6,125,I,SS(I)
      GO TO 987
  940 CALL EXIT
C
C
C
C
C
    1 FORMAT(I1,I2,I1,I2,I2)
  111  FORMAT (I5, I1, 16(3F1.0     ))
  112 FORMAT( I5,I1,36F2.0)
  113 FORMAT(35H ERROR IN ID CODES FOR SUBJECT NO. ,I5)
  114 FORMAT(39H CARDS OUT OF SEQUENCE FOR SUBJECT NO. ,I5)
  116 FORMAT(6(1X,I2,1H-,I2,1X,F5.2) )
  124 FORMAT(12(1X,F5.2))
  118 FORMAT(14H TOTAL SCORE =,F7.2,5X,I5)
  119 FORMAT (/16H ITEM SCORES FOR,2X,I5)
  120 FORMAT(1X,I2,1X,I2,1X,F5.2)
  122 FORMAT( 1X,F2.0,1X,F2.0,1X,F5.2)
  125 FORMAT(I2,2X,F7.2)
  628 FORMAT(    50H CRITERION GROUP STANDARD DEVIATIONS FOR EACH ITEM)
  704  FORMAT (2X, I2, 3X, F5.2)
  742  FORMAT (23HORELIABILITY OF TEST = ,F6.3,2X,5H FOR ,I4,2X,4H SUB)
  776  FORMAT (9(1X,F7.2))
  926 FORMAT( 27H1EUCLIDEAN SCORING FUNCTION)
  927 FORMAT( 27H1SPHERICAL SCORING FUNCTION )
  928 FORMAT( 27H1LOGARITHM SCORING FUNCTION)
  929 FORMAT( 39H1REPRODUCING EUCLIDEAN SCORING FUNCTION)
```

SUBROUTINE SEARCH

```
      SUBROUTINE SEARCH
    2 READ INPUT TAPE 5,1, ID
    1 FORMAT (I6)
      IF(ID) 999,3,2
    3 RETURN
  999 CALL EXIT
      END
```

```
C
C
C
C
      SUBROUTINE OUT(FT)
      COMMON MA,MC,MD,KF,KI,KN,KK,KIC,Z,SF,A
      DIMENSION Z(72), SF(72), A(72)
      EKN=KN
      EKI=KI
      A=EKI/12.
      KAB=KI/12
      EKAB=KAB
      IF(EKAB-A)206,207,206
  206 KIP=KAB+1
      GO TO 208
  207 KIP=KAB
  208 CONTINUE
      JL=1
      DO 209 IK=1,KIP
      JM=12*IK
      IF(MC)232,232,233
  233 PUNCH 124,(Z(I),I=JL,JM)
  232 WRITE OUTPUT TAPE 6,124,(Z(I),I=JL,JM)
  209 JL=12*IK  +1
      IF (MC)234,234,235
  235 PUNCH 118, FT
  234 WRITE OUTPUT TAPE 6,118, FT
  124 FORMAT(12(1X,F5.2))
  118 FORMAT(14H TOTAL SCORE =,F7.2,5X,I5)
      RETURN
      END
```

```
  930 FORMAT(2X,I3,1X,6H ITEMS,1X,I1,1X,19H RESPONSES PER ITEM)
  933 FORMAT(1X,I1,I2,I1,I2,I2,40X,  F6.0)
  986 FORMAT(11H1JOB NUMBER,1X,I3)
  994 FORMAT(72F1.0)
 2003 FORMAT(17H1PROBABILITY ONLY)
 2005 FORMAT(18H1ENTROPIC FUNCTION)
 2235 FORMAT(12A6)
 3002 FORMAT(16H1INFERRED CHOICE)
 4003  FORMAT (1H1/50X,20H-- STUDENT SCORES --)
 4004  FORMAT (15H0STUDENT NUMBER, 4X, I5)
 4005  FORMAT (1H1/39X,42H -- ITEM DIFFICULTIES AND RELIABILITIES --)
 4030  FORMAT (30H0NO SCORING FUNCTION REQUESTED)
 4031  FORMAT (41H0INCOMPATIBLE SCORING FUNCTIONS REQUESTED)
 4050  FORMAT (10I1)
C
C
C
C
C
       END
```

```
C
      SUBROUTINE FUNCA(AVANS,ANSPR,SSP,SI,SST,ST,SS,S2)
      COMMON MA,MC,MD,KF,KI,KN,KK,KIC,Z,SF,A
      DIMENSION AVANS(72,6), ANSPR(72,6), SI(72,10), SST(10), ST(10),
     1 SS(72), S2(72), SSP(10), Z(72), SF(72), A(72)
  201 DO 203 I=1,KI
      SF(I)=0
      FT=0
  203 Z(I)=0
      DO 204 I=1,KI
      DO 295 J=1,KN
  295 SF(I)=  ( AVANS(I,J) - ANSPR(I,J)) **2 + SF(I)
      Z(I)= 1.- (SQRTF(SF(I))) / 1.414
      IF (MD-1)1112,1112,1111
 1111   IF(Z(I) - .01) 1113,1113,1114
 1113 Z(I)=0.
       GO TO 1112
 1114  Z(I)= (2. + (LOGF(Z(I))) / 2.3026) / 2.
 1112  SI(I,1) = SI(I,1) + Z(I)
      SST(1) = SST(1) + Z(I) * Z(I)
      ST(1) = ST(1) + Z(I)
      GO TO ( 204, 9),MA
    9 SS(I)= SS(I)+ Z(I)
      S2(I)= S2(I)+ Z(I)*Z(I)
  204 FT=FT+Z(I)
      WRITE OUTPUT TAPE 6,1
    1 FORMAT (//10HOEUCLIDEAN)
      CALL OUT(FT)
      SSP(1) = SSP(1) + FT * FT
      RETURN
      END
C
C
C
      SUBROUTINE FUNCB(AVANS,ANSPR,SSP,SI,SST,ST)
      COMMON MA,MC,MD,KF,KI,KN,KK,KIC,Z,DEN2,ANUM
      DIMENSION AVANS(72,6), ANSPR(72,6), SI(72,10), SST(10), ST(10),
     1 SSP(10), Z(72), DEN2(72), ANUM(72)
  202 DO 214 I=1,KI
      Z(I)=0
      DEN2(I)=0
      ANUM(I)=0
  214 FT=0
  217 DO 215 I=1,KI
      DO 216 J=1,KN
      DEN2(I)=DEN2(I) + (ANSPR(I,J))**2
  216 ANUM(I)=(AVANS(I,J))*   ANSPR(I,J)+ ANUM(I)
      DEN2(I)= SQRTF(DEN2(I))
      Z(I)=ANUM(I)/DEN2(I)
      SI(I,2) = SI(I,2) + Z(I)
      SST(2) = SST(2) + Z(I) * Z(I)
      ST(2) = ST(2) + Z(I)
  215 FT=FT+Z(I)
      SSP(2) = SSP(2) + FT * FT
      WRITE OUTPUT TAPE 6,1
    1 FORMAT (//10HOSPHERICAL)
      CALL OUT(FT)
      RETURN
      END
```

```
C
C
C
C
      SUBROUTINE FUNCC(AVANS,ANSPR,SSP,SI,SST,ST)
      COMMON MA,MC,MD,KF,KI,KN,KK,KIC,Z,SF,A
      DIMENSION AVANS(72,6), ANSPR(72,6), SI(72,10), SST(10), ST(10),
     1 SSP(10),Z(72),SF(72),A(72)
  205 FT=0
      DO 614 I=1,KI
      SF(I)=0
  614 Z(I)=0
      DO 615 I=1,KI
      DO 616 J=1,KN
  616 SF(I)= AVANS(I,J) *ANSPR(I,J)  +SF(I)
      IF (SF(I)-.01) 617,617,618
  617 Z(I)=0
      GO TO 621
  618 Z(I)= (2.+(LOGF(SF(I))/2.3026)/ 2.
  621 CONTINUE
      SI(I,3) = SI(I,3) + Z(I)
      SST(3) = SST(3) + Z(I) * Z(I)
      ST(3) = ST(3) + Z(I)
  615 FT= FT+ Z(I)
      SSP(3) = SSP(3) + FT * FT
      WRITE OUTPUT TAPE 6,1
    1 FORMAT (//12HOLOGARITHMIC)
      CALL OUT(FT)
      RETURN
      END
```

```
C
C
C
C
      SUBROUTINE FUNCD(AVANS,ANSPR,SSP,SI,SST,ST)
      COMMON MA,MC,MD,KF,KI,KN,KK,KIC,Z,A,B
      DIMENSION AVANS(72,6), ANSPR(72,6), SI(72,10), SST(10), ST(10),
     1 SSP(10),Z(72),A(72),B(72)
  271 FT=0
      DO 3010  I=1,KI
 3010 Z(I)=0
      DO 3041 I=1,KI
      K=0
      DO 3020 J=1,KN
      IF (AVANS(I,J)) 3020,3020,3021
 3021 K=J
 3020 CONTINUE
      AL=0
      DO 3022 J=1,KN
      IF (ANSPR   (I,J) -AL ) 3022,3029,3024
 3024 AL=ANSPR(I,J)
      M=J
      GO TO 3022
 3029 M=0
 3022 CONTINUE
      IF(M-K)3025,3026, 3025
 3025 Z(I) =0
      GO TO 3040
 3026 Z(I) = 1.0
 3040 CONTINUE
      SI(I,4) = SI(I,4) + Z(I)
      SST(4) = SST(4) + Z(I) * Z(I)
      ST(4) = ST(4) + Z(I)
 3041  FT=FT+Z(I)
      SSP(4) = SSP(4) + FT * FT
      WRITE OUTPUT TAPE 6,1
    1 FORMAT (//16HOINFERRED CHOICE)
      CALL OUT(FT)
      RETURN
      END
```

```
C
C
C
C
      SUBROUTINE FUNCF(AVANS,ANSPR,SSP,SI,SST,ST)
      COMMON MA,MC,MD,KF,KI,KN,KK,KIC,Z,SF,A
      DIMENSION AVANS(72,6), ANSPR(72,6), SI(72,10), SST(10), ST(10),
     1 SSP(10), Z(72), SF(72), A(72)
 2001 FT=0
      DO 2074 I=1,KI
      SF(I)=0
 2074 Z(I)=0
      DO 2075 I=1,KI
      DO 2076 J=1,KN
 2076  SF(I)= AVANS(I,J)*ANSPR(I,J)  + SF(I)
      Z(I)=SF(I)
       SI(I,6) = SI(I,6) + Z(I)
       SST(6) = SST(6) + Z(I) * Z(I)
       ST(6) = ST(6) + Z(I)
 2075  FT=Z(I) + FT
       SSP(6) = SSP(6) + FT * FT
       WRITE OUTPUT TAPE 6,1
    1  FORMAT (//43HOPROBABILITY ASSIGNED TO THE CORRECT ANSWER)
       CALL OUT(FT)
       RETURN
       END
C
C
C
C
       SUBROUTINE FUNCE(AVANS,ANSPR,SSP,SI,SST,ST)
       COMMON MA,MC,MD,KF,KI,KN,KK,KIC,Z,SF,A
       DIMENSION AVANS(72,6), ANSPR(72,6), SI(72,10), SST(10), ST(10),
     1  SSP(10) , Z(72), SF(72), A(72)
       DO 203 I = 1,KI
       SF(I) = 0
       FT = 0
  203  Z(I) = 0
       DO 204 I = 1,KI
       DO 295 J = 1,KN
  295  SF(I) = (AVANS(I,J) - ANSPR(I,J))**2 + SF(I)
       Z(I) = 1. - (SQRTF(SF(I))) / 1.414
       IF (MD - 1) 1112,1112,1111
 1111  IF (Z(I) - .01) 1113,1113,1114
 1113  Z(I) = 0
       GO TO 1112
 1114  Z(I) = (2. + (LOGF(Z(I))) / 2.3026) / 2.
 1112  CONTINUE
       SI(I,5) = SI(I,5) + Z(I)
       SST(5) = SST(5) + Z(I) * Z(I)
       ST(5) = ST(5) + Z(I)
  204  FT = FT + Z(I)
       SSP(5) = SSP(5) + FT * FT
       WRITE OUTPUT TAPE 6,1
    1  FORMAT (//17HOEUCLIDEAN UNIQUE)
       CALL OUT(FT)
       RETURN
       END
```

```
C
C
C
C
      SUBROUTINE FUNCG(AVANS,ANSPR,SSP,SI,SST,ST)
      COMMON MA,MC,MD,KF,KI,KN,KK,KIC,Z,BL,A
      DIMENSION AVANS(72,6), ANSPR(72,6), SI(72,10), SST(10), ST(10),
     1 SSP(10), Z(72), BL(72), A(72)
 2006 FT = 0
      DO 2007 I = 1,KI
      Z(I) = 0
 2007 BL(I) = 0
      DO 2017 I = 1,KI
      DO 2009 J = 1,KN
      IF (ANSPR(I,J)) 2010,2010,2012
 2010 BL(I) = -9999.
      GO TO 2013
 2012 BL(I) = LOGF(ANSPR(I,J)/2.3026)
 2013 BL(I) = AVANS(I,J) * BL(I)
 2009 Z(I) = Z(I) + BL(I)
      Z(I) = 10. + Z(I)
      IF (Z(I)) 2016,2016,2018
 2016 Z(I) = 0
 2018 CONTINUE
      SI(I,7) = SI(I,7) + Z(I)
      SST(7) = SST(7) + Z(I) * Z(I)
      ST(7) = ST(7) + Z(I)
 2017 FT = FT + Z(I)
      SSP(7) = SSP(7) + FT * FT
      WRITE OUTPUT TAPE 6,1
    1 FORMAT (//9HOENTROPIC)
      CALL OUT(FT)
      RETURN
      END
C
C
C
      SUBROUTINE FUNCH (AVANS,ANSPR,SSP,SI,SST,ST)
      COMMON MA,MC,MD,KF,KI,KN,KK,KIC,Z,SF,DMAX
      DIMENSION  AVANS(72,6), ANSPR(72,6), SI(72,10), SST(10), ST(10),
     1 SSP(10), Z(72), SF(72), DMAX(72)
  201 DO 203 I = 1,KI
      SF(I) = 0
      FT = 0
  203 Z(I) = 0
      DO 204 I = 1,KI
      DO 295 J = 1,KN
  295 SF(I) = (AVANS(I,J) - ANSPR(I,J)) **2 + SF(I)
      SF(I) = SQRTF(SF(I))
      Z(I) = 1. - ((SF(I)/DMAX(I)) * 2.0)
       SI(I,8) = SI(I,8) + Z(I)
      SST(8) = SST(8) + Z(I) * Z(I)
      ST(8) = ST(8) + Z(I)
  204 FT = FT + Z(I)
      WRITE OUTPUT TAPE 6, 202
  202 FORMAT (//23HOEUCLIDEAN PROPORTIONAL)
      CALL OUT(FT)
      SSP(8) = SSP(8) + FT * FT
      RETURN
      END
```

```
C
C
C
C
      SUBROUTINE FUNCI(AVANS,ANSPR,SSP,SI,SST,ST)
      COMMON MA,MC,MD,KF,KI,KN,KK,KIC,Z,A,B
      DIMENSION AVANS(72,6), ANSPR(72,6), SI(72,10), SST(10), ST(10),
     1 SSP(10), Z(72), A(72), B(72)
      RETURN
      END
```

```
C
C
C
C
      SUBROUTINE FUNCJ(AVANS,ANSPR,SSP,SI,SST,ST)
      COMMON MA,MC,MD,KF,KI,KN,KK,KIC,Z,A,B
      DIMENSION AVANS(72,6), ANSPR(72,6), SI(72,10), SST(10), ST(10),
     1 SSP(10), Z(72), A(72), B(72)
      RETURN
      END
```

```
*      XEQ
*      FORTRAN
*      BINARY
C      WEIGHTED EUCLIDEAN SCORING. ROBERT RIPPEY   JUNE 1970 32 ITEM
C      THIS PROGRAM SCORES TESTS USING THE WEIGHTED OR UNWEIGHTED
C      EUCLIDEAN FUNCTION .32 ITEMS MAX. ,3 RESPONSES PER ITEM AND A
C       CONFIDENCE MEASURE.     TWO KINDS OF KEYS MAY BE USED
C       THE FIRST CARD CONTAINS THE FOLLOWING PARAMETERS.
C       COL 1   KK  KEYING OPTION  1  REGULAR, 2  PHASE 1 OUTPUT.
C       COL2 PUNCHED OUTPUT IF 1, OTHERWISE BLANK.
C       COL 3   WEIGHTING OPTION 1  NO WEIGHTING, OTHERWISE BLANK.
C       COL 4 = KC,    IF 2 IN COL.4, THERE ARE MORE THAN 16 ITEMS.
C       COL 5-6 =KI, NUMBER OF ITEMS, SOME NUMBER FROM 1-32
C       OPTION 1 CONTAINS KEYED ANSWERS IN APPROPRIATE COLUMNS
C       OPTION 2 UTILIZES THE OUPUT OF PHASE 1
C       CARD ORDER  CONTROL,KEYS,TITLE, RESPONSES,BLANK, REPEATS,-1
C       RESPONSES
C      COL 1-6 I.D. COL 7-70 RESPONSES UP TO 16 ITEMS IN GROUPS OF 4
C      THE FOURTH RESPONSE  REPRESENTS CONFIDENCE. 0= NO CONFIDENCE.
C      9 = ABSOLUTE CERTAINTY. COLUM 71 REPRESENTS TEST ATTITUDE.
C      0 = HATE THE TEST. 9 = LOVE THE TEST.
C      THE SECOND CARD CONTAINS RESPONSES FOR ITEMS 17 ON
C      TITLE CARD CONTAINS TEXT IN COLS 1-72
C      THE FORMAT OF THE SECOND CARD PER SUBJECT IS THE SAME AS CARD
C      END OF JOB  BLANK CARD
C      END OF RUN  -1 IN COLS 1-2
     1DIMENSION          ATITL(12), AKEY(32,3),PKEY(32,3),ANS(32,3),
     2PANS(32,3),CON (32), IT(32), DMAX(32), D(32) , S(32),AT(32)
     3 , ZV(32,300)  ,Z(32),  SS(32),  SP(300),  SI(32)
      EQUIVALENCE (DEN2,S2),(ANUM,SS), (Z,S) ,(SK,FT)
  999 READ(5,993)KK,MPUN,MWT,  KC,KI
  993 FORMAT(4I1,I2)
       GO TO (992,991),KK
  991 KL=3*KI
       DO 82 K=1,KL
       READ(5,120),I,J,AK
   82 AKEY(I,J)=AK
       GO TO 707
  120 FORMAT(1X,I2,1X,I2,1X,F5.2)
  992 READ (5,1) TID,((AKEY(I,J),J= 1,3 ),I=1,16)
   52 FORMAT(4(2X,F6.2))
    1 FORMAT( F6.0,16(3F1.0,1X))
       IF(KC-1)800,800,801
  801 READ (5,1) TID,((AKEY(I,J),J= 1,3 ),I=  17,KI )
  800 CONTINUE
  707 CONTINUE
       IF (TID) 2,999,995
  995 READ(5,5) (ATITL(I),I=1,12)
    5 FORMAT(12A6)
      D=0
      DA=0
      DB=0
       SOB=0
      DC=0
       EN=0
       TLIK = 0
      EKI=KI
      SSP=0
      CF=0
      ST=0
```

```
      SSI=0
      SSP=0
      REM=0
      DEFR=0
       VAREM =0
      VARP=0
      REL =0
      FT=0
      DO 8 I=1,KI
       SI(I)=0
      IT(I)=0
      CON (I)=0
      S(I)=0
       DO 555 N=1,300
555    ZV(I,N)=0
        DO 8 J=1,3
       PKEY(I,J) = 0
      ANS(I,J)=0
      PANS(I,J)=0
    8 CONTINUE
      WRITE (6,5) (ATITL (I),I=1,12)
      WRITE (6,6) TID
    6 FORMAT(13H TEST NUMBER ,F6.0)
      DO 7 I=1,KI
       DO 7 J=1,3
    7 AT(I) = AKEY(I,J) + AT(I)
       IF (AT(I)) 500,500,501
  500 WRITE (6,502)
  502 FORMAT(10H KEY ERROR)
       CALL EXIT
  501  DO 9 I=1,KI
       DO 9 J=1,3
    9 PKEY(I,J) = AKEY(I,J) / AT(I)
      DO 16 I=1,KI
      CALL MAX(PKEY(I,1),PKEY(I,2),PKEY(I,3),DMAX(I))
       WRITE (6,510),PKEY(I,1),PKEY(I,2), PKEY (I,3), DMAX(I)
  510  FORMAT(4(F6.3 ,2X))
   16 CONTINUE
   30 READ (5,20) SID, ((ANS(I,J),J=1,3),CON (I),I=1,16),ALIK
   20 FORMAT(F6.0 ,65F1.0)
       IF (KC-1)803,803,804
  804 READ (5,20) SOB, ((ANS(I,J),J=1,3),CON (I),I=17,KI)
  803  CONTINUE
       TLIK = TLIK + ALIK
      SK=0
       IF(SID) 2,240,842
  842 EN=EN+1.0
       N=EN
      DO 21 I=1,KI
      D(I)=0
   21 AT(I)=0
      DO506 I=1,KI
      DO 22 J=1,3
   22 AT(I) = ANS(I,J) + AT(I)
       IF(AT(I)) 503,503,504
  504 CONTINUE
      DO 23 J=1,3
   23 PANS(I,J) = ANS(I,J) / AT(I)
      D(I) =0
      DO 25 J=1,3
```

```
 25 D(I) = (PKEY(I,J) - PANS(I,J))**2 + D(I)
    D(I) = SQRTF(D(I))
     S(I)= 1.0-((D(I)/DMAX(I)) *2.0)
    IF (MWT)24,24,50
 24 S(I) = S(I) * CON(I)
 50 SK = S(I) + SK
    SCON = SCON + CON  (I)
     ZV(I,N)= Z(I)
    GO TO 506
503 S(I)=0
     ZV(I,N) =0
506 CONTINUE
      SSP=SSP + FT*FT
    AVCON = SCON / EKI
    WRITE(6,27)SID,SK,ALIK
 27 FORMAT(8H STUDENT,F8.0,11H   SCORE = ,F7.3,7H ATT = ,F6.3)
     WRITE (6,67) AVCON
 67 FORMAT(22H AVERAGE CONFIDENCE = ,F5.2)
     SCON = 0
 26  FORMAT(16(1X,F5.2))
     WRITE (6,26),(CON(I), I=1,KI)
 32 FORMAT(1H )
    WRITE (6,26) ,(S(I),I=1,KI)
    IF (MPUN) 60,60,31
 60 WRITE (6,32)
     GO TO 3
 31 PUNCH 52,SID,SK,ALIK,AVCON
    PUNCH 51,     (S(I),I=1,16)
     IF(KC-1)810,810,811
811 PUNCH 51,     (S(I),I=17,32)
810  CONTINUE
 51 FORMAT(  16F5.2)
  3 CONTINUE
     GO TO 30
240 DO 701 I=1,KI
     DO 701 J=1,N
     SI(I)= SI(I) + ZV(I,J)
     SST= SST + ZV(I,J) * ZV(I,J)
701 ST= ST + ZV(I,J)
     CF = (ST*ST)/ (EKI*E N)
     SST=SST-CF
     DO 702 I=1,KI
702 SSI=SSI + SI(I)*SI(I)
     SSI=((SSI)/EN)-CF
     SSP = ((SSP)/EKI)-CF
     DO 705 I=1,KI
705 SI(I)=SI(I)/EN
     AVLIK=TLIK/EN
     WRITE ( 6,537) AVLIK
537 FORMAT(18H AVERAGE LIKING = ,F5.2)
     WRITE OUTPUT TAPE 6,706
706 FORMAT(18HOITEM DIFFICULTIES)
     DO 636 I=1,KI
636   WRITE OUTPUT TAPE 6,704,I,SI(I)
704 FORMAT (2X,I2,3X, F5.2)
     REM = SST - SSP - SSI
     DEFR = (EKI-1.) * (EN - 1.)
     VAREM = REM/DEFR
     VARP = SSP/(EN - 1.)
     REL = 1.-(VAREM / VARP)
```

```
      WRITE OUTPUT TAPE 6,742,REL ,N
  742 FORMAT( 23HORELIABILITY OF TEST = ,F6.3,2X,5H FOR ,I4,2X, 4H SUB)
      WRITE OUTPUT TAPE 6,776,VARP,VAREM,      REM,SSI,SSP,SST,CF,ST
  776 FORMAT(8(1X, F9.1))
      IF(SOB) 2,999,999
  125 FORMAT(I2,2X,F7.2)
    2 CALL EXIT
      END
      SUBROUTINE MAX(PA,PB,PC,DMAX )
      DMAX = 0
      DA=0
      DB=0
      DC=0
      DA = SQRTF((1.0-PA)**2 + PB**2 + PC**2 )
      DB = SQRTF(PA**2 + (1.0-PB)**2 + PC**2)
      DC= SQRTF (PA**2+PB**2+ (1.0-PC)**2)
      IF (DA-DB)10,11,11
   11 DMAX = DA
      GO TO 12
   10 DMAX = DB
   12 IF (DMAX - DC ) 13,14,14
   13 DMAX = DC
      GO TO 15
   14 DMAX = DMAX
   15 CONTINUE
      RETURN
      END
```

```
*        XEQ
*        FORTRAN
*        BINARY
C         RISK TAKING INDEXES  ANNE MARIE GROZELIER  MARCH 70
         DIMENSION A(15),B(15),C(15),D(15),PLN(15),RT(15)
  0099   TPLN=0
         TRT=0
         TC=0
         READ(5,1)ID,(A(I),B(I),C(I),D(I),I=1,15)
       1 FORMAT(I6,60F1.0)
         IF(ID)98,98,5
       5 DO 2 I=1,15
         PLN(I)=((A(I)-3.0)**2 +(B(I)-3.0)**2+(C(I)-3.0)**2)/6.0
         RT(I)= ABSF(PLN(I)-D(I))
         TRT=TRT+RT(I)
         TPLN=TPLN+PLN(I)
         TC=TC+D(I)
       2 CONTINUE
         TRT=TRT/15.0
         TPLN=TPLN/15.0
         TC=TC/15.0
         WRITE (6,3) ID,TRT,TPLN,TC
         PUNCH 13,ID,TRT,TPLN,TC
       3 FORMAT(1X,I6,2X,3(F5.1,1X))
      13 FORMAT(I6,2X,3(F5.1,1X))
         GO TO 99
      98 CALL EXIT
         END
```

```
*      XEQ
*       FORTRAN
         PROBABILITY TEST KEY PREPARATION
C       FIRST CARD CONTAINS CONTROL PARAMETERS AND TITLE
C       COL 1      1 = STANDARD FORMAT
C       COL 2-3     = NUMBER OF ITEMS COL 4 = NUMBER OF RESPONSES
C       COL 5-6 = NUMBER OF CARDS   COL 7-8 = NUMBER OF ITEMS PER CARD
C       COL 9-49 = TITLE
C       DATA CARDS CONTAIN ID IN COL 1-5 A CARD SEQUENCE NUMBER IN COL 6
C       AND RESPONSES BY CRITERION GROUP MEMBERS IN COL 7-78
C       LAST CARD IN JOB IS BLANK
C       LAST CARD IN RUN IS A SECOND BLANK CARD
      DIMENSION ANS(72,6), ID (10), KS(10),ANSPR(72,6), SANSP(72,6),
     1 AVANS(72,6), TITL(10), SM(72)
   99 EN=0
      DO 71 I=1,72
      DO 71 J=1,6
      AVANS (I,J)=0
      ANS(I,J)=0
      ANSPR(I,J)=0
   71 SANSP (I,J) =0
      READ (5,1) KF, KI, KN, KK, KIC, (TITL(I), I=1,10)
      KG=KF+1
      GO TO (202,101), KG
  101 JP=1
      K=1
      IF (KK-1) 32,32,999
  999 DO 121 K=1,KK
      JK=K*KIC
      IF (JL-KI) 42,42,32
   32 JL=KI
   42 CONTINUE
      READ(5,111)ID(K),KS(K), ((ANS(I,J),J=1,KN),I=JP,JL)
  121 JP= JL + 1
      GO TO 37
   37 G=ID(I)
      IF (G) 50,50,41
   41 CONTINUE
   36 CONTINUE
C       PROGRAM NOW WEIGHTS AND COMPUTES PROBABILITIES
      DO 367 I=1,KI
      SM(I)=0
      DO 366 J=1,KN
  366 SM(I) = SM(I) + ANS(I,J)
      IF (SM(I)) 202,202,203
  203 CONTINUE
      DO 367 J=1,KN
      ANSPR(I,J) = ANS(I,J) / SM(I)
  367 SANSP (I,J) = SANSP (I,J) + ANSPR(I,J)
      EN=EN+1
      GO TO 101
   50 DO 51 I=1,KI
      DO 51 J=1,KN
   51 AVANS(I,J)=SANSP(I,J)/EN
      DO 49 I=1, KI
      DO 49 J=1,KN
   49 WRITE (6,117) I,J, AVANS(I,J)
      PUNCH 115,KF, KI, KN, KK, KIC, (TITL(I),I=1,10)
      DO 60 I=1,KI
      DO 60 J=1, KN
   60 PUNCH 116, I,J,AVANS(I,J)
      GO TO 99
  202 CALL EXIT
    1 FORMAT (I1, I2, I1, I2, I2, 10A4)
```

```
115 FORMAT (I1, I2, I1, I2, I2, 2X, 12HCONTROL FOR , 10A4)
111 FORMAT (I5, I1, 72F1.0)
112 FORMAT (I5, I1, 72F1.0)
116 FORMAT (1X, I2, 1H-, I2, 1X, F5.2)
117 FORMAT (1X, I2, 2X, 1H-, 2X, I2, 2X, 1H-, F5.2)
    END
```

APPENDIX B

SAMPLE ITEMS FROM TESTS EMPLOYING
EUCLIDEAN CONFIDENCE SCORING

The student is given a description of an experiment to read (Bandura, 1968). Then he is given the following:

a. Select the correct answer with certainty on the items which have unique correct responses.

b. Express your honest degrees of confidence on those items which do not have unique correct answers.

c. Do not choose your numbers randomly. If you have no preferences for any of the responses, give each response the same numerical value, i.e., 1,1,1.

If you like all of the answers equally well, mark the answer sheet as follows:

a. 3
b. 3
c. 3

If you like none of the answers, mark the answer sheet as follows:

a. 0
b. 0
c. 0

If you like just the second answer and none of the others, mark:

a. 0
b. 9
c. 0

If you like some answers better than others, respond with a digit from 0-9 which will express your odds on that answer. For example, if you like answer c three times as much as answer b, mark:

a. 0
b. 3
c. 6

A. Answer the following questions according to the instructions. The questions refer to the report of the experiment by Bandura and McDonald.

1. The purpose of the experiment by Bandura and McDonald was to:
   a. illustrate the fact that strong dominant repertoires limit the opportunities for reinforcing subordinate responses.
   b. to demonstrate the superiority of modeling over operant conditioning.
   c. to test the validity of the stage theory of moral development.

2. It is important to research the relative efficiency of modeling over operant conditioning because:
   a. operant conditioning can be automated.
   b. many desired complex behaviors seldom occur as random behaviors.
   c. the excessive emphasis on modeling in Spartan education resulted in an increase in homosexuality.

3. It is important to research the stage theory of moral development because:
   a. the morals of today's young people need improvement.
   b. an understanding of the different reactions of subjective and objective children may suggest a basis for making instructional decisions.
   c. the stage theory has an elegant simplicity.

4. The factor which differentiates between subjective and objective children is:
   a. deviancy.
   b. invariance.
   c. intention.

5. How well described are the following? State your relative preference.
   a. the scheme for classifying subjects.
   b. the treatment.
   c. the dependent variable.

6. What additional information would you like?
   a. the sex of the subjects.
   b. the age of the subjects.
   c. the IQs of the subjects.

7. What do you like about this experiment?
   a. the random assignment of subjects to treatments.
   b. its pursuit of a new or novel line of thought.
   c. the statistical analysis.

8. If you were to design a similar experiment you would:
   a. modify the design.
   b. modify the statistics.
   c. modify the treatments.

9. Relative threats to validity in this experiment are:
   a. history.
   b. regression effect.
   c. instrument decay.

10. In applying these findings to the classroom, one must be concerned with:
   a. the economic cost of modeling.
   b. the external validity of the findings.
   c. the difficulties inherent in applying the results of animal experiments to human behavior.

Sample Items From
TEST OF NURSE ACTIONS
Leona Peterson

This test is designed to learn what you think nurses should do in certain situations and how much certainty you have when you choose these actions. The information you give will be confidential. It will not be used to determine any grade for a course.

There are 7 pages. Please check that you have all 7 pages plus an Answer Sheet.

This is a timed test. Please wait at the bottom of those pages where it asks you to wait, until the examiner says to go on.

The test consists of 14 incidents. Sets of statements or answers follow these incidents. There may or may not be a single correct answer. The precise response which you might desire may not be listed. Information that you might like to have may be missing. However, for each set of statements, you are to mark:

1) the strength of your preference for each statement "a, b, and c". This is done by placing a number from zero to nine (0 to 9) on the lines following "a, b, and c" on the Answer Sheet. A zero indicates your rejection of the statement. A nine indicates that you prefer that statement over the others.
2) how certain you are of the set of responses you have made. This is done by placing a number from zero to nine (0 to 9) on the line following "d" on the Answer Sheet. Zero = completely unsure. Nine = absolutely certain.

For example, if you like only response "a", you would mark:

a. 9
b. 0
c. 0

If you are partially certain of this set of responses, you would mark: d. 5

If you like "a" not at all, "b" most, and "c" a lesser amount, you could mark:

a. 0	a. 0	a. 0
b. 5	b. 9	b. 7
c. 1	c. 3	c. 4
d. 1	d. 2	d. 7

In these last examples, certainty, marked in "d", varied from very little to quite a bit.

Here is an example for you to work:

The nurse who wants to suggest good sources of Vitamin C will include:
a. citrus fruits.
b. broccoli.
c. apples.

ANSWER SHEET
a._____
b._____
c._____
d._____

The more certain you are, the higher your score will be if your responses are close to the mean of answers given by expert nurses. However, if your responses "a, b, and c" are far from the mean of answers given by the experts, a low certainty, as marked in "d" will give you a lower score.

The numbers to the right of the lines on the Answer Sheet are for data processing only.

PLEASE WAIT UNTIL THE EXAMINER TELLS YOU TO START

INCIDENT A: Mr. Frank Buttry, a 78 year old retired machinist, was helping his wife cart weeds to the garbage can when he suddenly slipped and fell. He was brought to the hospital by ambulance, and over the past day, he has been comatose. His respirations are stertorous, and the right side of his mouth blows outward with each expiration. His right leg has a tendency to fall into external rotation. A lumbar puncture shows bloody cerebrospinal fluid. His wife says he has been treated by Dr. Adams for hypertension for 20 years. His tentative diagnosis: CVA. His admission chest x-ray shows a marked hypertrophy of the left ventricle.
Dr. Adams' orders include: Mercuhydrin 1 ml. IM q.o.d.
500 ml. 5% D/W
1000 ml. Lactated Ringers IV

1. Essential in the care of Mr. Buttry at this time is (are):
   a. regulating IV fluids for rapid replenishment of electrolytes and of glucose to the brain.
   b. leaving a padded tongue blade at the bedside.
   c. suctioning oropharyngeal secretions.

2. During Mr. Buttry's first few days in the hospital, a small elevation of his temperature, pulse, and respirations, may signal
   a. infection or dehydration.
   b. progressive deterioration of the centers of the brain.
   c. a normal course.

3. Preventing deformity is an important part of the nurse's responsibility to Mr. Buttry. A full range-of-motion exercise program should be started as soon as
   a. Mr. Buttry awakens and starts moving in bed.
   b. Mr. Buttry is ordered to be "up in chair."
   c. Mr. Buttry's blood pressure, pulse, and respirations stabilize.

INCIDENT B: Mr. Buttry opened his eyes during the third day of his hospitalization. Although unable to speak intelligibly, he grasped his wife's arm with recognition. He responded to instructions to move, albeit sluggishly. He is able to swallow sips of water.

4. At this point in his recovery, the nurses should assume that self-help training should include
   a. turning from side to side.
   b. sitting up.
   c. feeding himself.

5. The nurses should also
   a. phone Dr. Adams.
   b. encourage Mr. Buttry to talk.
   c. refrain from asking Mr. Buttry questions.

INCIDENT C: On the 6th day, a half hour after Mr. Buttry's bed had been rolled down for the night, Mr. Buttry started coughing. Although his cough was infrequent, it was bothersome enough to keep him awake. The night nurse found him still awake at 11:30 p.m. He had received Nembutal gr iii at 10 p.m. In addition to the order for Nembutal at h.s., there is a further order written: "May give Nembutal gr iss once during the night for sleep p.r.n."

6. The night nurse should
   a. phone Dr. Adams.
   b. give a warm drink of milk and another pillow, and assure him that he will be given another sleeping pill as soon as it is all right to do so.
   c. give him 2 tsp. of a non-narcotic cough syrup.

GO ON TO THE NEXT PAGE

INCIDENT D: Mr. Buttry's evening routine and medication were soon under control. A few days later, however, Mr. B's respirations sounded gurgling. He looked pale and was orthopneic. Blood chemistry test results (drawn the evening before) showed:

Na (normal: 135-145 mEq/l.)...130 mEq/l.
Cl (normal: 95-110 mEq/l.).... 90 m Eq/l.
K (normal: 3.5-5 mEq/l.)..... 2.5 mEq/l.
CO_2 (normal: 22-30 mEq/l.).... 45 mEq/l.
BUN (mormal: 8-28 mg%)........ 20 mg%
pH.......................... 7.8

A chest x-ray indicated increased vascularity of the lung fields. Dr. Adams ordered Morphine SO_4 15 mg. 'H' given stat. He then gave Ouabain 0.5 mg IV and ordered:

Stop Mercuhydrin
Aldactone 25 mg. q.i.d.
Digitoxin 0.6 mg. stat and in 6 hours
then Digitoxin 0.2 mg. q 6 hrs.
1 Gm. Sodium diet.
Rotating tourniquets on 45 min., off 15 min (four extremities)
Oxygen per cannula at 2 to 7 liters/min. p.r.n.

7. The nurse should assume that Dr. Adams must be notified by phone if
   a. Mr. Buttry becomes nauseated or has abdominal cramps.
   b. Mr. Buttry's extremities become uncomfortable and discolored.
   c. Mr. Buttry's face becomes pink and flushed.

8. The nurse should tell Mr. Buttry that
   a. coffee and tea should be avoided.
   b. meat, fish, and poultry will be limited on his diet tray.
   c. if he does not like some food on his try, she will get a substitute.

9. The nurse will further Mr. Buttry's recovery by
   a. using a massager on his posterior chest wall.
   b. suggesting he turn toward the left side when eating.
   c. giving oxygen at 7 liters on evidence of dyspnea.

10. The nurse notes that Mr. B's urine output suddenly increases substantially with this regime. She should
   a. expect that laboratory reports will show a further decrease in potassium.
   b. expect a greater likelihood of anorexia and gastrointestinal complaints.
   c. recognize this as a guarded prognostic sign.

STOP. PLEASE DO NOT TURN THIS PAGE UNTIL TOLD TO DO SO.

Sample Items From
Test of Troublesome Group Managment Situations
Sallie R. Churchill

Form II

This study is concerned with actions taken by social group workers in group meetings and with their reasons for these actions. Comparisons will be made among several schools of social work and between beginning and second year students.

There are two parts to this questionnaire:

PART I: You will be given a description of a boys' group and a description of a men's group. Following each group description you will be given a set of three incidents which occur during meetings of that group.

You will be asked to rate several actions that a group worker might select as possible responses to the incidents. You will also be asked to rate several reasons that a group worker might offer for his actions. In each instance, you will be asked to indicate how certain you are about your ratings.

In addition you will be asked to rate each set of worker actions and worker reasons in the way you think your first year field instructor would rate his preference and (2) the way you think your first year social group work method teacher would rate his preference. You will also be asked to indicate how certain you are about your estimates.

The answers to PART I will be recorded on a separate answer sheet.

PART II: You will be asked to give some information about yourself, your education and your work experience.

As a researcher I am interested in learning how groups of students respond to the items in the questionnaire, rather than how any specific individual responds. Therefore, I will not request that you put your name on the questionnaire. However, the information will be considered confidential and the responses of each of you will be available to the researcher only.

It is expected that it should take between one and two hours to complete this questionnaire.

Please turn to the Instructions on the Use of the Answer Sheet.

Form II

INSTRUCTIONS FOR USING THE ANSWER SHEET

1. Record your own preferences under Self Rating.

2. Record how you think your first year field instructor would respond under Field Instructor.

   a. If you had more than one field instructor in your first year, select either:

      The field instructor whom you had for the longest time; or if you had field instructors for equal lengths of time, select the field instructor whom you had during your last quarter or semester.

3. Record how you think your first year social group work method classroom teacher would respond under Method Teacher.

   a. If you had more than one group work method teacher during your first year of social group work education select either:

      The teacher whom you had the longest time;

      or

      The teacher who carried the major responsibility for your first year method class;

      or

      The teacher whom you had during the last semester or quarter in your first year.

4. Place a digit from 0 to 9 in the blank opposite A,B, or C. Each number represents the strength of your preference for each response. However, the sum of the three responses should add up to 9. Zero indicates rejection, 9 represents maximum preference.

If you liked only response A, you would answer A 9, B 0, C 0

If you liked all three answers equally, you would mark - A 3, B 3, C 3

If you liked answer B most, C a small amount and A not at all, you might mark -

A	0	or	0
B	7		8
C	2		1

The number you write down should be proportional to your preference.

THERE ARE NO CORRECT ANSWERS

After marking your preference for A, B, and C, mark in square D, a number from 0 to 9 which tells how certain you are of your distribution of preferences for A, B, and C.

0 = Completely unsure

9 = Absolutely certain

Turn to the FIRST GROUP DESCRIPTION.

BOYS' GROUP

Please read the following description of THE BOYS' GROUP and the GROUP MEMBERS. Spend as much time studying this material as you wish. When you are ready, turn to the group incidents. If you wish you may refer to the group and/or individual descriptions as you rate worker actions and/or worker reasons.

Setting: Children's Unit of a State Mental Hospital

Timing: The group has been meeting with a social group worker once a week for two months.

Group Description:

All group members, aged 7 to 9, were admitted to the hospital because of serious behavioral problems which prevented attendance at school, and/or retention in an open community.

When the group was formed all the boys had made considerable progress in the closed residential setting and the hospital was considering discharge. A group worker was assigned to these boys to assist in their resocialization to community living. The group service was planned to assist the boys to increase their attention span, to increase their impulse control and to learn appropriate social skills (including interactional and physical components of games). The boys want the group to serve as a means of having fun, getting off the ward, and getting away from the "crazy" kids.

The worker thinks the group members are benefitting from the group experience and usually the group meetings proceed in a constructive manner towards stated goals. In most ways the group can be considered a good group, with members and worker successfully working together.

Individuals

Aaron, 7½: Large for his age, poor coordination, easily loses his temper and destroys equipment. Steals other boys' possessions.

Barry, 9: Very passive; cries easily. Often angers other members by refusing to do something, such as take his turn in a game. May become irrationally angry and lose all control.

Danny, 7½: Subject to epileptic seizures. He threatens to have seizures to control group. Grimaces and acts "crazy" to get attention. Hyperactive.

Eddie, 8: Leader of the group; good athlete and very bright. He has little frustration tolerance and will bite anyone close to him when he is upset. Often tries to challenge adult authority. Upset by worker's attention to other group members.

Gary, 8½: Very immature; sucks thumb, frequently masturbates. He clings to worker while being verbally abusive to other members. Scapegoated frequently.

When ready, turn to INCIDENT ONE.

Form II

BOYS' GROUP: INCIDENT ONE

Please read the following incident:

> All of the members were playing a fairly rough, but controlled game of "King of the Mountain." Barry and Gary seemed to have teamed up to hold off the "attack." Eddie seemed quite surprised when the weaker members held up against him. Suddenly Eddie bit Gary. Gary gave up, fell off the mountain, began screaming, "I quit!" Eddie immeadiately began taunting, "Baby, baby, where's your bottle? Where is your diapers?" The sing-song chant took hold and all five boys shouted it loudly at Gary, apparently ignoring that he had been bitten.

RATE THE FIRST FOUR SETS OF STATEMENTS IN REFERENCE TO THE ABOVE INCIDENT IN THE BOYS' GROUP.

I. ON THE ANSWER SHEET, IN THE SELF RATING COLUMN, INDICATE THE STRENGTH OF YOUR PREFERENCE FOR EACH WORKER ACTION.

II. ON THE ANSWER SHEET, IN THE FIELD INSTRUCTOR COLUMN, RATE HOW YOU THINK YOUR FIRST YEAR FIELD INSTRUCTOR WOULD RATE THE STRENGTH OF HIS PREFERENCE FOR EACH WORKER ACTION.

III. ON THE ANSWER SHEET, IN THE METHOD TEACHER COLUMN, RATE HOW YOU THINK YOUR FIRST YEAR GROUP WORK METHOD TEACHER WOULD RATE THE STRENGTH OF HIS PREFERENCE FOR EACH WORKER ACTION.

IV. IN EACH INCIDENT INDICATE ON LINE D HOW CERTAIN YOU ARE.

---

1. A. The group worker helps the group quiet down and reviews the nature of the group interactions that have just occurred. He encourages the group members to discuss their feelings toward Gary.

   B. The group worker stops the game and gathers the boys together. He tells them that he will not permit the group to tease Gary for crying when he has been bitten, nor will he permit the boys to bite each other.

   C. The group worker comforts Gary.

---

2. A. The group worker sends Eddie out of the group for biting Gary. He discusses with the group that boys usually like to play rough but they must play fair and square.

   B. The group worker reminds the group of its contract, pointing out that Gary's crying, as well as Eddie's biting when he is upset, are behaviors which the group might help the boys change.

   C. The group worker reminds the group of its contract and encourages the group members to discuss what has taken place and how this may affect the functioning of the group.

If you wish to make any comments about the Worker Actions in response to this incident, write them below. (OPTIONAL)

I. ON THE ANSWER SHEET, IN THE SELF RATING COLUMN, INDICATE THE STRENGTH OF YOUR PREFERENCE FOR EACH REASON WHICH A GROUP WORKER MIGHT OFFER FOR HIS ACTION.

II. ON THE ANSWER SHEET, IN THE FIELD INSTRUCTOR COLUMN, RATE HOW YOU THINK YOUR FIRST YEAR FIELD INSTRUCTOR WOULD RATE THE STRENGTH OF HIS PREFERENCE FOR EACH REASON.

III. ON THE ANSWER SHEET, IN THE METHOD TEACHER COLUMN, RATE HOW YOU THINK YOUR FIRST YEAR GROUP WORK METHOD TEACHER WOULD RATE THE STRENGTH OF HIS PREFERENCE FOR EACH REASON.

IV. IN EACH INCIDENT, INDICATE ON LINE D HOW CERTAIN YOU ARE.

3.

A. The extent of aggression and/or hostility expressed in a group should be controlled by the group worker.

B. A group member who is excessively teased or hurt needs comforting and/or support.

C. A group may need help in identifying and dealing with obstacles which prevent the group from progressing towards the group goal.

4.

A. A group member who intentionally hurts another member physically or psychologically, should be disciplined, regardless of whether the hurt member had provoked the aggression or not.

B. The group worker should not attempt to retain control of the interactions between or among the members, except in incidents of possible physical danger to a member.

C. When it appears that group members have lost control of themselves, so that their behavior is neither helpful to themselves nor to the other group members, the worker should act so as to restore control.

If you wish to make any comments about the Reasons for Worker Actions in response to this incident, write them below. (OPTIONAL)

Go on to the next INCIDENT.

Form II

BOYS' GROUP: INCIDENT TWO

Please read the following incident:

All the members had been working on plastic airplane models. The group was sharing three tubes of glue. As the meeting time was up and the group began to clean up, the worker noticed one tube of glue was missing. This has happened frequently.

The worker asked the group members if they could locate the tube of glue. Immediately Danny said Aaron had swiped it. Aaron uneasily put the tube of glue back on the table. Danny yelled, "I was right, Aaron had it. He's a little thief?" Gary said, "Our group doesn't need any dirty thieves. Does anyone want him in the group?" Suddenly Eddie demanded silence. "It's our group; let's vote on it." All the boys shouted at once, "We don't want him." Eddie declared, "It's five against one . . . he's out."

RATE THE NEXT FOUR SETS OF STATEMENTS (5-8) IN REFERENCE TO THE ABOVE INCIDENT IN THE BOYS' GROUP.

I. ON THE ANSWER SHEET, IN THE SELF RATING COLUMN, INDICATE THE STRENGTH OF YOUR PREFERENCE FOR EACH WORKER ACTION.

II. ON THE ANSWER SHEET, IN THE FIELD INSTRUCTOR COLUMN, RATE HOW YOU THINK YOUR FIRST YEAR FIELD INSTRUCTOR WOULD RATE THE STRENGTH OF HIS PREFERENCE FOR EACH WORKER ACTION.

III. ON THE ANSWER SHEET, IN THE METHOD TEACHER COLUMN, RATE HOW YOU THINK YOUR FIRST YEAR GROUP WORK METHOD TEACHER WOULD RATE THE STRENGTH OF HIS PREFERENCE FOR EACH WORKER ACTION.

IV. IN EACH INCIDENT INDICATE ON LINE D HOW CERTAIN YOU ARE.

---

5. A. The group worker asks the boys to think about their quick vote. He adds that the boys seemed awfully angry and did not seem to be thinking about what they were doing. He asks if the group really wants to lose one of its members. He accepts the final group decision.

   B. The group worker says that the decision about who is a member of the group is the worker's and that nobody could be voted out of the group by group action.

   C. The worker asks Eddie if he really thinks that Aaron is a thief. He tells him that Aaron had returned the glue and that Aaron is not a thief.

6. A. The group worker helps the group members tell Aaron how they feel about his behavior, in this meeting and others. He helps Aaron respond to what the boys say.

   B. The group worker comments that all the boys have concerns about problems like stealing. The boys can talk about these concerns in the meeting but they may not kick anyone out of the group.

   C. The group worker tells Eddie that he is upsetting everyone by yelling for the group to vote a member out; he tells the boys to calm down and finish cleaning up.

Form II
Boys' Group
Incident Two

If you wish to make any comments about the Worker Actions in response to this incident, write them below. (OPTIONAL)

I. ON THE ANSWER SHEET, IN THE SELF RATING COLUMN INDICATE THE STRENGTH OF YOUR PREFERENCE FOR EACH REASON WHICH A GROUP WORKER MIGHT OFFER FOR HIS ACTION.

II. ON THE ANSWER SHEET, IN THE FIELD INSTRUCTOR COLUMN RATE HOW YOU THINK YOUR FIRST YEAR FIELD INSTRUCTOR WOULD RATE THE STRENGTH OF HIS PREFERENCE FOR EACH REASON.

III. ON THE ANSWER SHEET, IN THE METHOD TEACHER COLUMN, RATE HOW YOU THINK YOUR FIRST YEAR GROUP WORK METHOD TEACHER WOULD RATE THE STRENGTH OF HIS PREFERENCE FOR EACH REASON.

IV. IN EACH INCIDENT, INDICATE ON LINE D HOW CERTAIN YOUR ARE.

7.

A. The group is a mutual aid system which must remain free to make its own decisions.

B. The group is used as both means and context of the treatment of the individual; thus membership is determined by the worker in relation to individual needs.

C. The decision by the group members to exclude a member from the group must be based on reasons which seem valid to the worker.

8.

A. The responsibility for each member's welfare belongs to the group worker. Therefore, the group worker retains a veto power over all group decisions.

B. The group worker should correct the members when they express wrong ideas or feelings in order to maintain an accurate awareness of reality within the group situation.

C. The use of authority by the worker can become a central stumbling block in the helping relationship.

If you wish to make any comments about the Reasons for Worker Actions in response to this incident, write them below.

Go on the the next INCIDENT.

APPENDIX C

CRITERION RESPONSE DISTRIBUTION PLOTTING PROGRAM

Criterion Response Distribution Plotting Program

This program plots the distribution of responses assigned to each item on a test. For convenience, the probability assigned to each option is rounded to the nearest tenth. The response triangle can then be considered a set of 55 discrete points, with each point corresponding to a specific set of probabilities. The program calculates the probabilities assigned by each subject to each item (rounded to the nearest tenth) and performs a frequency count for each response set. Output is in the form of a plot of the response triangle.

Program limitations

The program will plot responses for up to seventy-two items. All items must have three options.

Card setup

Card 1 — Title card (cols 2-60)

Card 2

Cols.	
1	Punch a 1 in this column
2-3	Number of items
5-6	Number of cards for each subject (Not more than ten)
7-8	Number of responses per card

A sample of the output and the program listing follow.

only.

PLEASE WAIT UNTIL THE EXAMINER TELLS YOU TO START

Sample Output from Criterion Response Plotting Program

CRITERION RESPONSE FOR ITEM 1

```
                                   *10*   0   *0*
                               *9*   0     0   *1*
                           *8*   0     0     0   *2*
                       *7*   0     0     0     0   *3*
                   *6*   0     0     0     0     0   *4*
   A           *5*   0     1     0     0     0     0   *5*       B
           *4*   0     1     2     0     0     0     0   *6*
       *3*   0     4     6     1     0     0     0     0   *7*
   *2*   2     5    10     4     0     0     0     0     0   *8*
*1*  10     9     4     4     2     0     1     0     0     0   *9*
*0*  15    11     6     5     3     1     1     0     0     0     0  *10*
   *10*   *9*   *8*   *7*   *6*   *5*   *4*   *3*   *2*   *1*   *0*
                                  C
```

```
      XEQ
*     FORTRAN
       CRITERION RESPONSE DISTRIBUTIONS
       DIMENSION AJOB(12), ID(10), KS(10), ANS(72,3), ANSPR(72,3), SM(7
     1), JSTORE(72,11,11)
       JOB = 0
  987  JOB = JOB + 1
       WRITE OUTPUT TAPE 6, 986, JOB
      READ INPUT TAPE 5,2235, (AJOB(I),I=1,12)
      WRITE OUTPUT TAPE 6,2235,(AJOB(I),I=1,12)
       DO 4000 I = 1,72
       DO 4001 J = 1,3
       ANS(I,J) = 0
 4001  ANSPR(I,J) = 0
       DO 4000 JA = 1,11
       DO 4000 JB = 1,11
 4000 JSTORE(I,JA,JB) = 0
  241 READ INPUT TAPE 5,1,                KF,KI,KN,KK,KIC
       WRITE OUTPUT TAPE 6,930,KI
       IF(KF)940,940,211
  211 GO TO  (101,102),KF
  101 JP=1
       K=1
       IF (KK-1) 32,32,999
  999 DO 121 K=1,KK
       JL=K*KIC
       IF (JL-KI)42,42,32
   32 JL=KI
   42 CONTINUE
       READ INPUT TAPE 5 ,111,ID(K),KS(K),((ANS(I,J),J=1,KN),I=JP,JL)
          IF(ID(K))940,240,121
  121 JP=JL+1
          GO TO 37
  102 JP=1
       K=1
       IF (KK-1) 532,532,998
  998 DO 172 K=1,KK
       JL=K*KIC
       IF(JL-KI)533,533,532
  532 JL=KI
  533 CONTINUE
       READ INPUT TAPE 5,112,ID(K),KS(K),((ANS(I,J),J=1,KN),I=JP,JL)
          IF(ID(K))940,240,172
  172 JP=JL+1
   37 G=ID(1)
       EN = EN+1.
       N=EN
       IF(G)940,240,41
   41 CONTINUE
       IF(KK-1)  36,36,81
   81 CONTINUE
      DO 36 K=2,KK
      LA=ID(K)
      LB=ID(K-1)
      IF(LA-LB)33,80,33
   80 LC=KS(K)
       LD=KS(K-1  )
      IF(LC-LD) 35,35,36
   33 WRITE OUTPUT TAPE 6,113,ID(K)
       CALL EXIT
   35  WRITE OUTPUT TAPE 6,114,ID(K)
```

```
      CALL EXIT
   36 CONTINUE
C     PROGRAM NOW WEIGHTS AND  COMPUTES PROBABILITIES
      DO 367 I=1,KI
      SM(I)=0
      DO 366 J=1,KN
  366 SM(I) = SM(I) + ANS(I,J)
      BZRK= SM(I)
      IF(BZRK)5327,5327,5328
 5327  ANSPR(I,J) = 0.
      GO TO 367
 5328 CONTINUE
      DO 367 J=1,KN
     ANSPR(I,J)= ANS(I,J) / SM(I)
  367 CONTINUE
      DO 5013 I = 1,KI
     JA = ANSPR(I,1) * 10. + 1.
      JB = ANSPR(I,2) * 10. + 1.
 5013 JSTORE(I,JA,JB) = JSTORE(I,JA,JB) + 1
      GO TO 211
  240 CONTINUE
      I = 0
 5014 I = I + 1
      IF (I-KI) 5015,5015,987
 5015 WRITE OUTPUT TAPE 6, 5016, I
 5016 FORMAT (41H1CRITERION RESPONSE DISTRIBUTION FOR ITEM, I5)
      WRITE OUTPUT TAPE 6,  5001, JSTORE(I,11,1)
      WRITE OUTPUT TAPE 6,  5002, (JSTORE(I,10,J), J=1,2)
      WRITE OUTPUT TAPE 6,  5003, (JSTORE(I,9,J), J=1,3)
      WRITE OUTPUT TAPE 6,  5004, (JSTORE(I,8,J), J=1,4)
      WRITE OUTPUT TAPE 6,  5005, (JSTORE(I,7,J), J=1,5)
      WRITE OUTPUT TAPE 6,  5006, (JSTORE(I,6,J), J=1,6)
      WRITE OUTPUT TAPE 6,  5007, (JSTORE(I,5,J), J=1,7)
      WRITE OUTPUT TAPE 6,  5008, (JSTORE(I,4,J), J=1,8)
      WRITE OUTPUT TAPE 6,  5009, (JSTORE(I,3,J), J=1,9)
      WRITE OUTPUT TAPE 6,  5010, (JSTORE(I,2,J), J=1,10)
      WRITE OUTPUT TAPE 6,  5011, (JSTORE(I,1,J), J=1,11)
      WRITE OUTPUT TAPE 6,  5012
      GO TO 5014
  940 CALL EXIT
 5001 FORMAT( //60X, 4H*10*, 5X,115,5X,3H*0*)
 5002  FORMAT (//59X, 3H*9* , 5X, 2I5, 5X, 3H*1*)
 5003  FORMAT (//57X,3H*8*, 5X, 3I5, 5X, 3H* 2*)
 5004 FORMAT (//55X, 3H*7*, 5X, 4I5, 5X, 3H*3*)
 5005  FORMAT (//53X, 3H*6*, 5X, 5I5, 5X, 3H*4*)
 5006  FORMAT (// 41X, 1HA, 9X, 3H*5*, 5X, 6I5, 5X, 3H*5*, 9X, 1HB)
 5007  FORMAT (//49X, 3H*4*, 5X, 7I5, 5X, 3H*6*)
 5008  FORMAT (//47X, 3H*3*, 5X, 8I5, 5X, 3H* 7*)
 5009   FORMAT (//45X, 3H*2*, 5X, 9I5, 5X, 3H*8*)
 5010  FORMAT (//43X, 3H*1*, 5X, 10I5, 5X, 3H*9*)
 5011  FORMAT (//41X, 3H*0*, 5X, 11I5, 5X, 4H*10*)
 5012  FORMAT (//50X, 54H *10* *9*  *8*  *7*  *6*  *5*  *4*  *3*  *2*
    1I*  *0*///60X, 1HC)
  930 FORMAT(2X,I3,1X,6H ITEMS,1X,I1,1X,19H RESPONSES PER ITEM)
    1 FORMAT(I1,I2,I1,I2,I2)
  111  FORMAT (I5, I1, 72F1.0)
  112 FORMAT( I5,I1,36F2.0)
  113 FORMAT(35H ERROR IN ID CODES FOR SUBJECT NO. ,I5)
```

```
  114 FORMAT(39H CARDS OUT OF SEQUENCE FOR SUBJECT NO. ,I5)
  986 FORMAT(11H1JOB NUMBER,1X,I3)
 2235 FORMAT(12A6)
      END
```

REFERENCES

Ahlgren, A. Confidence marking and predicting retention. A paper read at the annual meeting of the American Educational Research Association, Chicago, Illinois, 1968 (mimeo).

Bandura, A. Behavioral modifications through modeling procedures. In L. Ullman & L. Krasner (Eds.), Research in behavior modification. New York: Holt, Rinehart & Winston, 1968. Pp. 314-320.

Bourdieu, P. & Passeron. Les heritiers, les etudiants, et la culture. Paris, ed. de Minuit, 1964.

Bruner, J.S., Goodnow, J.J. & Austin, G.A. A study of tinking. New York: Wiley & Sons, 1956.

Dressel, P.L. & Schmid, J. Some modifications of multiple choice items. Educational and Psychological Measurement, 1953, 13, 574-595.

DeFinetti, B. Methods for discriminating levels of partial knowledge concerning a test item, British Journal of Mathematical and Statistical Psychology, 1965, 18, 87-123.

Educational Testing Service, Manual for STEP writing Test, Level 1.

Ebel, R. Measuring Educational Achievement. New York: Prentice-Hall, 1965.

Ebel, R. Review of "Valid Confidence Testing Demonstration Kit." Journal of Educational Measurement, 1968, 5, 353-354.

Grozelier, A.M. The influence of social values on risk taking in a confidence testing situation. M.A. Dissertation, University of Chicago, Chicago, Illinois.

Guttman, L. The Cornell technique for scale and intensity analysis. Educational and Psychological Measurement, 1947, 7, 247-279.

Hambelton, R.K., Roberts, D.M., & Traub, R.E. A Comparison of the reliability and validity of two methods for assessing partial knowledge on a multiple choice test. Journal of Educational Measurement, 1970, 7, 75-90.

Hoyt, C. Test reliability estimated by analysis of variance, Psychometrika, 1941, 6, 153-160.

Kogan, N. & Wallach, M.A. Risk taking: A study in cognition and personality. New York: Holt, Rinehart, & Winston, 1964.

Lewy, A & McGuire, C. A study of alternative approaches to estimating the reliability of unconventional tests. Paper delivered to the American Education Research Association, Chicago, Illinois, 1966 (mimeo).

Plato, Republic. London: Macmillan, 1925.

Platt, J.R. Strong Inference. Science, 1963.

Quillian, M.R. Semantic memory. In Minsky, M. (Ed.) Semantic Information Processing. Cambridge, Mass.: The MIT Press, 1968.

Romberg, T. & Shepler, J. An experiment involving a probability measurement procedure. Paper presented to the annual meeting of the American Educational Research Association, Chicago, 1968 (mimeo).

Shuford, E.N. Cybernetic testing. Bedford, Mass.: Decision Sciences Laboratory, Air Force Systems Command, 1967.

Shuford, E., Albert, A. & Massengill, N.E. Admissable probability measurement procedures. Psychometrika, 1966, 31, 125-145.

Smith, E. & Tyler, E. Appraising and recording student progress. New York: Harper, 1942.

Strong, E.K. Vocational interests of men and women. Stanford, Cal.: Stanford University Press, 1943.

Swineford, F. Analysis of a personality trait. Journal of Educational Psychology, 1941, 32, 438-444.

Toda, M. Measurement of Subjective Probability Distributions.ESD TDR 63 407 Bedford, Mass.: Decision Sciences Laboratory, 1963.

Walker, H.M. & Lev, J. Statistical Inference. New York: Holt, Rinehart, and Winston, 1953.

Wang, M.W. & Stanley, J.C. Differential weighting: A review of methods and empirical studies, Review of Educational Research, 1970, 40, 663-705.

Ward, & Edwards, W. Behavioral Decision Theory, In Farnsworth, P.R., McNemar, O., McNemar, Q. (Eds.) Annual Review of Psychology, 1961, 12, 473-497.

Watanabe, S Knowing and guessing: A formal and quantitative study. New York: Wiley & Sons, 1969.

Williams, R. & Millman, J. Confidence Testing: A Reply, NCME Measurement News, July, 1970.