ABSTRACT
        This is a report from the Project on Individually
Guided Mathematics, Phase 2 Analysis of Mathematics Instruction. The
report outlines some of the characteristics of probability
measurement procedures for scoring objective tests, discusses
hypothesized advantages and disadvantages of the methods, and reports
the results of three experiments designed to learn more about the
technique and compare it with standard procedures of scoring
objective tests. The procedure used required the students to specify
a degree of belief probability for each of the given alternatives to
a question. The students were given a multiple-choice item and asked
to specify what they believed to be the probability of correctness of
each choice. The initial intent of these experiments was to see if a
non-standard test-taking and scoring procedure would provide useful,
reliable information for such tests. The studies indicated that the
problem of getting useful, reliable information on difficult tests
has not been solved. (Author/FL)

THREE EXPERIMENTS INVOLVING PROBABILITY MEASUREMENT PROCEDURES WITH MATHEMATICS TEST ITEMS

WISCONSIN RESEARCH AND DEVELOPMENT

CENTER FOR COGNITIVE LEARNING

RESEARCH
COGNITIVE
LEARNING
DEVELOPMENT
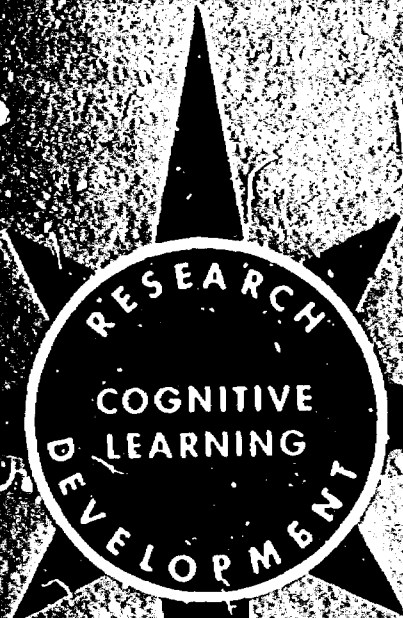
Technical Report No. 129

THREE EXPERIMENTS INVOLVING PROBABILITY
ME/.SUREMENT PROCEDURES WITH
MATHEMATICS TEST ITEMS

by

Thomas A. Romberg, Jack L. Shepler,[1] and James W. Wilson[2]

Report from the Project on Individually Guided Mathematics,
Phase 2 Analysis of Mathematics Instruction

Thomas A. Romberg and John G. Harvey,
Principal Investigators

[1]Indiana University of Pennsylvania

[2]University of Georgia

Wisconsin Research and Development
Center for Cognitive Learning
The University of Wisconsin
Madison, Wisconsin

June 1970

NATIONAL EVALUATION COMMITTEE

Samuel Brownell
Professor of Urban Education
Graduate School
Yale University

Henry Chauncey
President
Educational Testing Service

Elizabeth Koontz
Wage and Labor Standards
Administration, U.S.
Department of Labor,
Washington

Patrick Suppes
Professor
Department of Mathematics
Stanford University

Launor F. Carter
Senior Vice President on
Technology and Development
System Development Corporation

Martin Deutsch
Director, Institute for
Developmental Studies
New York Medical College

Roderick McPhee
President
Punahou School, Honolulu

***Benton J. Underwood**
Professor
Department of Psychology
Northwestern University

Francis S. Chase
Professor
Department of Education
University of Chicago

Jack Edling
Director, Teaching Research
Division
Oregon State System of Higher
Education

G. Wesley Sowards
Director, Elementary Education
Florida State University

RESEARCH AND DEVELOPMENT CENTER POLICY REVIEW BOARD

Leonard Berkowitz
Chairman
Department of Psychology

Russell J. Hosler
Professor, Curriculum
and Instruction

Stephen C. Kleene
Dean, College of
Letters and Science

B. Robert Tabachnick
Chairman, Department
of Curriculum and
Instruction

Archie A. Buchmiller
Deputy State Superintendent
Department of Public Instruction

Clauston Jenkins
Assistant Director
Coordinating Committee for
Higher Education

Donald J. McCarty
Dean
School of Education

Henry C. Weinlick
Executive Secretary
Wisconsin Education Association

Robert E. Grinder
Chairman
Department of Educational
Psychology

Herbert J. Klausmeier
Director, R & D Center
Professor of Educational
Psychology

Ira Sharkansky
Associate Professor of Political
Science

M. Crawford Young
Associate Dean
The Graduate School

EXECUTIVE COMMITTEE

Edgar F. Borgatta
Brittingham Professor of
Sociology

Robert E. Davidson
Assistant Professor,
Educational Psychology

Russell J. Hosler
Professor of Curriculum and
Instruction and of Business

Wayne Otto
Professor of Curriculum and
Instruction (Reading)

Anne E. Buchanan
Project Specialist
R & D Center

Frank H. Farley
Associate Professor,
Educational Psychology

***Herbert J. Klausmeier**
Director, R & D Center
Professor of Educational
Psychology

Robert G. Petzold
Associate Dean of the School
of Education
Professor of Curriculum and
Instruction and of Music

Robin S. Chapman
Research Associate
R & D Center

FACULTY OF PRINCIPAL INVESTIGATORS

Vernon L. Allen
Professor of Psychology

Frank H. Farley
Associate Professor of Educational
Psychology

James Moser
Assistant Professor of Mathematics
Education; Visiting Scholar

Richard L. Venezky
Assistant Professor of English
and of Computer Sciences

Ted Czajkowski
Assistant Professor of Curriculum
and Instruction

Lester S. Golub
Lecturer in Curriculum and
Instruction and in English

Wayne Otto
Professor of Curriculum and
Instruction (Reading)

Alan Voelker
Assistant Professor of Curriculum
and Instruction

Robert E. Davidson
Assistant Professor of
Educational Psychology

John G. Harvey
Associate Professor of
Mathematics and of Curriculum
and Instruction

Milton O. Pella
Professor of Curriculum and
Instruction (Science)

Larry Wilder
Assistant Professor of Curriculum
and Instruction

Gary A. Davis
Associate Professor of
Educational Psychology

Herbert J. Klausmeier
Director, R & D Center
Professor of Educational
Psychology

Thomas A. Romberg
Associate Director, R & D Center
Professor of Mathematics and of
Curriculum and Instruction

Peter Wolff
Assistant Professor of Educational
Psychology

M. Vere DeVault
Professor of Curriculum and
Instruction (Mathematics)

Donald Lange
Assistant Professor of Curriculum
and Instruction

B. Robert Tabachnick
Chairman, Department
of Curriculum and
Instruction

MANAGEMENT COUNCIL

Herbert J. Klausmeier
Director, R & D Center
V.A.C. Henmon Professor of
Educational Psychology

Thomas A. Romberg
Associate Director

James Walter
Director
Dissemination Program

Dan G. Woolpert
Director
Operations and Business

Mary R. Quilling
Director
Technical Development Program

* COMMITTEE CHAIRMAN

ii

STATEMENT OF FOCUS

The Wisconsin Research and Development Center for Cognitive Learning focuses on contributing to a better understanding of cognitive learning by children and youth and to the improvement of related educational practices. The strategy for research and development is comprehensive. it includes basic research to generate new knowledge about the conditions and processes of learning and about the processes of instruction, and the subsequent development of research-based instructional materials, many of which are designed for use by teachers and others for use by students. These materials are tested and refined in school settings. Throughout these operations behavioral scientists, curriculum experts, academic scholars, and school people interact, insuring that the results of Center activities are based soundly on knowledge of subject matter and cognitive learning and that they are applied to the improvement of educational practice.

This Technical Report is from Phase 2 of the Project on Prototypic Instructional Systems in Elementary Mathematics in Program 2. General objectives of the Program are to establish rationale and strategy for developing instructional systems, to identify sequences of concepts and cognitive skills, to develop assessment procedures for these concepts and skills, to identify or develop instructional materials associated with the concepts and cognitive skills, and to generate new knowledge about instructional procedures. Contributing to the Program objectives, the Mathematics Project, Phase 1, is developing and testing a televised course in arithmetic for Grades 1-7 which provides not only a complete program of instruction for the pupils but also inservice training for teachers. Phase 2 has a long-term goal of providing an individually guided instructional program in elementary mathematics. Preliminary activities include identifying instructional objectives, student activities, teacher activities, materials, and assessment procedures for integration into a total mathematics curriculum. The third phase focuses on the development of a computer system for managing individually guided instruction in mathematics and on a later extension of the system's applicability.

CONTENTS

LIST OF FIGURES AND TABLES

ABSTRACT

This Technical Report presents the results of three experiments designed to study the utility of probability measurement procedures with mathematics test items. In each experiment it was hypothesized that:

1. The use of a probability measurement procedure introduces a test-taking style which changes the performance being measured.

2. Probability measurement procedures will yield a higher reliability coefficient than standard scoring procedures.

3. The mean score obtained by probability measurement procedures for the same students scored in the standard way which, in turn, will be greater than the means of students in the control group who take the test under standard conditions.

4. The reliability coefficients will be ordered in the same manner as the means.

In each study these hypotheses were not confirmed. The first two studies used test items measuring high level cognitive abilities with Eleventh and Twelfth Grade students. The third used information items measuring low cognitive abilities with Eighth Grade students.

INTRODUCTION

This paper outlines some characteristics of probability measurement procedures for scoring objective tests, discusses hypothesized advantages and disadvantages of the methods, and reports the results of three experiments designed to learn more about the technique and compare it with standard procedures of scoring objective tests.

In many testing situations a student is presented a multiple-choice item in which he is asked to decide which of the given alternatives is correct, or the best. The item is scored 1 or 0 depending on whether his answer corresponds to that on the key or not, regardless of the student's confidence in his response. Tests comprised of difficult items such as tests constructed to measure problem solving, insightful, or creative cognitive behaviors generally produce low reliabilities using the standard test-taking and scoring procedures. The initial purpose of the studies reported here was to see if a non-standard test-taking and scoring procedure would provide useful, reliable information for such a test.

The test-taking procedure used asks the student to specify a degree of belief probability for each of the given alternatives. That is, the student is presented a multiple-choice item, with five choices, and asked to specify what he believes to be the probability of correctness of each choice. The total of the probabilities for the five choices should be 1.

This procedure was proposed in an article by Shuford (1965) who called it an "admissible scoring procedure" and claimed it to be a more sensitive instrument to partial knowledge.

Any admissible probability measurement procedure has a scoring system which guarantees that any student, at whatever level of knowledge or skill, can maximize his expected score if and only if he follows instructions and honestly reflects his

'degree-of-belief probability' as to the correctness of a possible answer to the test item. [Shuford defines testing procedures which utilize such scoring systems as admissible probability measurement procedures.]

These degrees-of-belief probabilities contain all the information that can be made available about the student's knowledge structure as a consequence of asking the particular question under consideration. By way of contrast, multiple-choice and constructed-response procedures can yield only partial information as to whether or not these probabilities exceed certain values or lie within a very broad range. (Shuford, 1965, p. 2)

The notion of using degree-of-belief probabilities is not new in educational literature. However, little seems to have been done except to periodically re-discover it and postulate its utility until the Italian probabilist De Finetti (1965) reopened the topic with a comprehensive theoretical treatment. This was quickly followed by a careful treatment of scoring procedures associated with degree-of-belief test-taking (Albert, Massengill, & Shuford, 1966).

In the meantime several empirical studies have been reported. Ahlgren (1969) summarizes the results of recent research in this area and reports that in 26 out of 31 studies an increase in reliability was obtained by using confidence scoring studies rather than standard scoring. However, other than the studies reported here, none dealt with mathematics items.

Wilson (1965) observed that attempts to measure "insightful mathematical ability" were rather unfruitful in spite of considerable feeling among mathematicians that this is an important mathematical ability. Instruments developed for the National Longitudinal Study of Mathematical Abilities (NLSMA) to detect insightfulness were considered to be poor. One possible reason for this was that the tests were too

insensitive. Yet, the mathematicians responsible for developing the tests for NLSMA still wanted insightful scales to be included in the Longitudinal Study. Three scales totalling 31 items were administered to Eleventh Year students in Spring 1964 (NLSMA, 1968). Upon analyses of the data, the scale reliabilities were quite low. At that time, Wilson (1965) hypothesized that using admissible scoring procedures on this type of test would yield higher means and higher reliability coefficients.

The advantages of such procedures stem from the fact that degree-of-belief probabilities contain all of the information that can be made available about this student's knowledge structure as a consequence of asking the particular question under consideration. Specific advantages would include:

(1) Higher reliabilities. For example, Shuford (1965) reported increases in split-half reliabilities from .6 or .7 to around .9 when probability measurement procedures were used rather than standard scoring procedures. This could be expected since the probability measurement procedure would produce scores with a smaller fraction of chance behavior than the standard scoring procedure. Shuford also argued that increased reliabilities would be found in almost all testing situations encountered in practice if one used an "admissible probability measurement procedure."

(2) Better prediction and higher validity. These could be expected since correlations and validities are limited by test reliabilities.

(3) More sensitive item analysis. An item-analysis technique based on the examination of the patterns of probabilities assigned to a given item by a population should be very sensitive.

The most obvious disadvantage for the use of a probability measurement procedure is that students must be trained, or instructed, to follow the probability assignment procedure and convinced that maximum score can be expected if, and only if, it is followed. Another disadvantage is the greater cost in time and materials. It takes longer for the student to assign probabilities to each of five possible choices than to pick one choice as the best.

In addition to the different test-taking characteristics, various scoring procedures are possible.

Four scoring methods were used in the experiments reported in this paper. For the control groups:

(1) Standard scoring (0, 1) and summing the correct choices were used. For the treatment groups:

(2) Summing the probability weights on the correct choices,

(3) Transforming the data by a spherical scoring function, and

(4) Transforming the data by a logarithmic transformation were used.

The last two scoring procedures are examples of what Albert, et al. (1966, p. 127), have called reproducing scoring systems. These two transformations are scoring systems which are a part of test procedures which have been referred to as "admissible probability measurement procedures."

The spherical scoring function applied to each item is:

$$f(r_1, r_2, r_3, r_4, r_5) = \frac{r_k}{\left(\sum_{j=1}^{5} r_j^2\right)^{1/2}}$$

where r_j is the probability weight assigned to the jth alternative and r_k is the correct choice for the item. What this transformation does to a score on an item where the choice (a) is correct is illustrated in Figure 1.

The score for an item is strictly determined by the probability assigned to the correct answer and the way in which the student's uncertainty is distributed over the other answers (i.e., the relative magnitudes of the other assigned probabilities). The order of distributing these weights is of no importance. For instance, Subjects (3) and (4) have the same transformed score (.29) (since the magnitudes of the other four alternatives are the same).

Albert, et al. (1966), refer to the truncated logarithmic scoring system as not being strictly a reproducing scoring system, but having the reproducing property for values of p between .027 and .973. They recommend this procedure be followed for practical purposes, since it is likely that the effect resulting from the truncation at $p = .01$ is quite acceptable. The truncated logarithmic scoring function is:

Figure 1						
Spherical Scoring Function Applied to the Responses of Six Subjects to the Same Item						
Subjects			Choices			
	a	b	c	d	e	$f(r_1, r_2, r_3, r_4, r_5)$
(1)	.2	.2	.2	.2	.2	.45
(2)	.5	.5	0	0	0	.71
(3)	.2	.7	.1	0	0	.29
(4)	.2	.1	0	.7	0	.29
(5)	.7	.3	0	0	0	.92
(6)	.3	.7	0	0	0	.40

$$f(r_k) = \begin{cases} 1 + \log r_k & \text{for } .01 < r_k \leq 1 \\ -1 & \text{for } 0 \leq r_k \leq .01 \end{cases}$$

where r_k is the probability weight assigned to the correct choice. This is the only reproducing scoring system that depends only on the probability weight that the subject assigns to the correct choice. The range of scores assigned to an item is between -1 and 1. This transformation is particularly hard on misinformation in that one receives a score of -1 on an item for assigning 0 to the correct choice. Figure 2 illustrates what the logarithmic transformation does to the weights the subject places on the correct choice.

Figure 2		
Logarithmic Scoring Function Applied to the Responses of Four Subjects to the Same Item		
Subject	r_k	Truncated Logarithmic Score
(1)	0	-1
(2)	.1	0
(3)	.4	.602
(4)	.7	.845

From this background the following hypotheses were proposed:

Hypothesis 1. The use of a probability measurement procedure introduces a test-taking, style which changes the performance being measured.

It was decided to examine this hypothesis by examining the percentage of responses in three categories: (1,0) or right-wrong responses, (.2, .2, .2, .2, .2) or guessing responses, and other responses. If subjects are using degree-of-belief probabilities the percentage of other responses should be large in comparison to the other categories.

Hypothesis 2. Probability measurement procedures will yield a higher reliability coefficient than standard scoring procedures.

This hypothesis was to be examined by putting 90% confidence intervals around the coefficient (Hoyt, 1941) and seeing if the intervals overlap (Feldt, 1965).

Hypothesis 3. The mean score obtained by probability measurement procedures for the treatment group will be greater than the mean score for the same students scored in the standard way which, in turn, will be greater than the means of students in the control group who take the test under standard conditions.

This hypothesis was to be tested by simply ordering the means and rejecting the hypothesis if the means are not ordered as hypothesized. The spherical transformation on the treatment group scores should produce higher means than the original means; and the logarithmic transformation, lower means.

Hypothesis 4. The reliability coefficients will be ordered in the same manner as the means.

The four coefficients will be examined and the hypothesis will be rejected if the ordering is not as specified by the hypothesis.

In order to examine the plausibility of these hypotheses, three experiments were conducted using students from James Madison Memorial High School in Madison, Wisconsin. The first involved Twelfth Graders; the second,

3

Eleventh Graders; and the third, Eighth Graders. The first two studies used a test derived from selected items from the NLSMA "insightful scales." The third study used a geometry information test, also derived from the NLSMA battery.

STUDY NO. 1

The first experiment involving students taking Twelfth Year mathematics was conducted in Fall 1967. Using a stratified random assignment procedure, 32 subjects were assigned to the treatment group and 32 subjects to the control group. Blocking was done on grade, sex, previous mathematics, grade, and I.Q. The subjects assigned to the treatment group met immediately before the test for 15 minutes to learn the probability scoring procedure. Using an overhead projector, the students in this session were presented sample multiple-choice items with five alternatives (Appendix B). For each item they were asked to specify their beliefs as to the probability of correctness of each alternative where the sum of the probabilities for the five choices is 1. The students were instructed that they could maximize their scores if they honestly reflected their degree-of-belief probabilities as to the correctness of each of the choices for an item. The control group was instructed to take this test in the usual manner. The testing time for both groups on a 15-item test was 49 minutes. The items were selected from insightful items included in the NLSMA battery (Appendix A). The results of this study are summarized in Table 1.

The first hypothesis was only partially substantiated since students in the Twelfth-Grade treatment group used (1, 0) scoring 50% of the time and guessing (.2, .2, .2, .2, .2) 14% of the time. Hence, the students used a different strategy on only 36% of the questions.

Hypotheses 2, 3, and 4 are not supported by the data. The differences between sum scores for the treatment and control groups were negligible. The magnitudes of the means and the reliabilities are very similar. So similar, in fact, that no confidence intervals were calculated for the reliabilities. However, the variance was reduced for the treatment group.

The transformed data for the treatment group produce conflicting information with the hypotheses. As expected, the spherical transformation produced a higher mean. However, the transformation had the opposite effect from what was expected concerning reliabilities and variances. The logarithmic transformation produces a dramatically lower mean and reliability, but a larger variance.

Why the hypotheses were not confirmed is a matter of conjecture. One plausible explanation was that the items proved not to be as difficult as had been anticipated. Thus, it was decided to repeat the experiment.

STUDY NO. 2

The second study, also conducted in Fall 1967, used Eleventh Graders. For this study it was decided to increase the length of the test to 17 items (Appendix A), to decrease the testing time to 40 minutes, and to increase the instruction for the treatment group to 40 minutes by including practice in using the procedure on difficult mathematical items. Because of schedule difficulties, a matched,

Table 1

Results of Experiment with Twelfth Grade Students

Control	Treatment		
(Sum)	(Sum)	(Spherical)	(Logarithmic)
$\bar{X} = 6.75$	$\bar{X} = 6.80$	$\bar{X} = 8.16$	$\bar{X} = 3.45$
$r = .638$	$r = .624$	$r = .51$	$r = .43$
$s^2 = 8.25$	$s^2 = 5.63$	$s^2 = 4.81$	$s^2 = 13.77$
$N = 32$	$N = 32$	$N = 32$	$N = 32$
$K = 15$	$K = 15$	$K = 15$	$K = 15$

$\bar{X}$ = mean; r = reliability (Hoyt); s^2 = variance; N = subjects; K = items

Table 2

Results of Experiment with Eleventh Grade Students

Control	Treatment		
(Sum)	(Sum)	(Spherical)	(Logarithmic)
$\bar{X}$ = 4.00	$\bar{X}$ = 4.25	$\bar{X}$ = 6.33	$\bar{X}$ = -1.78
r = .185	r = .10	r = -.02	r = -.56
s^2 = 3.36	s^2 = 1.75	s^2 = 1.56	s^2 = 5.60
N = 25	N = 25	N = 25	N = 25
K = 17	K = 17	K = 17	K = 17

$\bar{X}$ = mean; r = reliability (Hoyt); s^2 = variance; N = subjects; K = items

rather than a random, sample was taken, blocking on the same variables as before. The results of this study are summarized in Table 2.

Again, the first hypothesis was only partially substantiated. For the Eleventh Grade treatment group the subjects used (1, 0) scoring 35% of the time and guessing (.2, .2, .2, .2, .2) 33% of the time. Or, students were using a different strategy only 32% of the time.

The second, third, and fourth hypotheses were again not supported by the data. As in Study No. 1, the differences between the sum scores for the treatment and control groups were negligible.

SUMMARY OF STUDIES NO. 1 AND NO. 2

Why the hypotheses were not confirmed is not clear. One possibility is that the test instrument was not suitable to probability scoring. Even for these types of difficult items, students apparently attempt to arrive at answers by mathematical techniques and are willing to bet that their responses are correct even though the techniques used often lead to wrong answers.

For example, the typical way many students in the Eleventh and Twelfth Grades found a wrong answer to Problem 3 in Appendix A was to use, in solving a difficult problem, the technique of first simplifying the algebraic expression. Thus, the equation became

$$(x + 1)(x + 2)(x + 3) = (x + 2)(x + 3)(x + 4)$$
$$x + 1 = x + 4$$

which has no roots (response (e)). Since the answer was reached using a mathematical method, and the response is one of the multiple choices, the subject is certain that his answer is correct. In line with this, if the answer found is not one of the five alternatives, the student resorts to guessing. The data related to the first hypothesis somewhat substantiated this conjecture.

Other possibilities are that these types of mathematical items do not lend themselves to easy elimination of alternatives, or that the treatment was not strong enough to convince students to use probability scoring more than they did. It may also be of importance to demonstrate in detail to the treatment group the admissible scoring transformation to be used. A better understanding of what the transformation will do to the weights assigned could influence the way a subject scores the items.

In conclusion, the problem of how one gets useful, reliable information on difficult tests measuring high level cognitive abilities had not been solved.

STUDY NO. 3

In the two preceding experiments, a probability measurement procedure was employed with a test consisting of very difficult, complex items which were designed to measure "insightful mathematical ability." However, the probability measurement procedure failed to yield a higher reliability coefficient. As a further examination of the usefulness of probability measurement procedures, a third

5

study was designed using a test which measures a low cognitive ability level. The purpose of this experiment was to investigate whether probability scoring used with a test, measuring knowledge of specific facts, would yield more reliable information than conventional scoring procedures.

An achievement test consisting of 30 multiple-choice items was constructed from a pool of 74 items from a NLSMA battery of geometry items (NLSMA, 1968, Report 2).

In case the treatment for learning the probability scoring procedure had not been strong enough to produce the desired results in the two previous experiments, the treatment was strengthened. The treatment period was lengthened to 50 minutes of instruction and practice the day before the testing followed the next day by fifteen minutes of review and practice immediately before the 30-minute testing period.

In the training period, as with the two previous experiments, a pamphlet, "Training for Probability Scoring" was handed out and discussed (Appendix B). An overlay similar to that used for the first two experiments was employed to demonstrate how to score items using the method (Appendix B). Also, two practice tests were used, one involving analogies from Lorge-Thorndike Intelligence Test, Verbal Battery (1954); the other, geometrical concepts not measured by the test used in the experiment (Appendix B). It was hypothesized that by having students score their own practice tests using the spherical scoring procedure (Appendix B) that they would be more prone to be convinced to use probability scoring procedures, rather than resort to their usual test-taking strategies. It was also decided to use a practice test consisting of items very similar to the test to be given.

Thus it was anticipated that the practice session would be similar to the testing session.

For this study, four Eighth Grade classes taught by the same teacher at Madison Memorial Junior High School in Madison, Wisconsin, were used. The teacher identified two of the four classes as being high-mathematical achieving classes and two as low-mathematical achieving classes. By flipping a coin, one class from each of the above two pairs was assigned to the control group (67 students) and the others to the treatment group (58 students). With respect to previous math grades the control group had 14 in the A to B+ range, 31 in the B to C+ range, and 22 in the C to F range. The treatment group had 11 in the A to B+ range, 26 in the B to C+ range, and 21 in the C to F range. The average IQ for the control group was 115.4 and for the treatment group, 120.5. The classroom teacher administered the training session for the two classes in the treatment group and also the testing sessions. The results of this study are summarized in Table 3.

Again, the results do not support the hypotheses. For Hypothesis 1, there was some change in the test-taking style for the treatment group—61% using (1, 0) scoring, 5% guessing (.2, .2, .2, .2, .2) scoring, and 34% using some other scoring scheme. The scoring using simple summing of the weights placed on the correct alternative yielded almost exactly the same mean and reliability coefficient as the control group.

With respect to Hypothesis 2, the probability transformation measures applied to the treatment group yielded a lower reliability than the control group or the treatment group under simple summing.

For Hypothesis 4, again the means are not ordered in the direction hypothesized, nor are the reliabilities ordered in the same manner as the means.

Table 3

Results of Experiment with Eighth Grade Students

Control (Sum)	Treatment		
	(Sum)	(Spherical)	(Logarithmic)
$\bar{X}$ = 20.66	$\bar{X}$ = 20.58	$\bar{X}$ = 22.06	$\bar{X}$ = 16.97
r = .83	r = .84	r = .80	r = .75
s^2 = 26.02	s^2 = 9.59	s^2 = 14.82	s^2 = 40.71
N = 67	N = 58	N = 58	N = 58
K = 30	K = 30	K = 30	K = 30

$\bar{X}$ = mean; r = reliability (Hoyt); s^2 = variance; N = subjects; K = items

SUMMARY

It was felt at the conclusion of Study No. 1 that the reasons that probabilit/ scoring did not increase the reliability coefficient vas caused by the items not being difficult enough and the training in probability scoring not strong enough. However in Study No. 2, when the training period was lengthened, practice given in scoring difficult mathematical items included, and the test made more difficult for the subjects, these changes still did not increase the reliability coefficients.

At that time, the following possible explanations were raised:

(1) The problem-solving set students employ when trying to solve difficult mathematical problems does not allow probability scoring procedures to be effective.
(2) The training procedure was not effective.
(3) Probability scoring procedures may not necessarily increase the reliability coefficient of a test.

It was then decided to design a third experiment using mathematical items designed to test recall of information at a lower cognitive level. In the previous experiments the students had not been told the/admissible scoring procedure being used. For Study No. 3 the training was lengthened to include teaching the subjects to use a spherical admissible scoring procedure. However, again the reliability coefficient was not increased. It was anticipated that Experiment 3 would clarify the utility of the procedure. However, the reliability coefficient for the control group was quite high (.83). Thus, the test reliability may have been too high to expect much of

an increase by employing an admissible scoring procedure. [One should note that the |treatment group's reliability did decrease from .84 under simple summing to .80 under the spherical scoring procedure.]

While these studies have not eliminated any of the three alternative explanations of the results, increasing the time of the training would seem questionable in light of the cost-effectiveness factor in putting the probability scoring procedure into practice. One alternative would be to use the commercial materials of Massengill and Shuford (1968). These materials employ a device which calculates the logarithmic scoring function. However, it is the opinion of these authors that using this device would, again, probably not appreciably increase the reliability of the tests used in these studies.

Although the probability scoring procedures have not produced greater reliabilities in the studies reported here, the method certainly had definite assets, particularly concerning information about an individual's score. Immediately, if the subject reflects his true degree of belief, a teacher can tell if a student is misinformed (0 on correct alternative) or whether he is guessing (.2, .2, .2, .2, .2) or whether he is correctly informed (1, or a number close to 1 on the correct alternative) on any particular item. This certainly is better than the traditional method of employing (1, 0) scoring.

In conclusion, the three studies indicate that the problem of how one gets useful, reliable information on difficult tests has not been solved.

REFERENCES

Albert, Arthur, Massengill, H. Edward, & Shuford, Emir H., "Admissible Probability Measurement Procedures," Psychometrika, 1966, 31, 125-145.

Ahlgren, Andrew, "Reliability, Predictive Validity, and Personality Bias of Confidence-Weighted Scores," Paper presented at the 1969 meeting of the AERA and NCME "Confidence on Achievement Tests—Theory, Applications," Los Angeles, Calif.

De Finetti, Bruno, "Methods for Discriminating Levels of Partial Knowledge Concerning a Test Item," The British Journal of Mathematical and Statistical Psychology, 1965, 18, 87-123.

Feldt, Leonard S. The approximate sampling distribution of Kuder-Richardson reliability coefficient twenty. Psychometrika, 1965, 31, 357-369.

"Grade 10, Y-Population, Geometry Information 1." In NLSMA Reports, No. 2, Wilson, J.W., Cahen, L. S., & Begle, E. G. (Eds.). The Board of Trustees of the Leland Stanford Junior University, 1968, Stanford, California.

"Grade 11, Z-Population, Year 2 Test Battery." In NLSMA Reports, No. 3, Wilson, J.W., Cahen, L.S., & Begle, E. G. (Eds.). The Board of Trustees of the Leland Stanford Junior University, 1968, Stanford, California.

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

Lorge, Irving & Thorndike, Robert L. The Lorge-Thorndike Intelligence Tests. Boston: Houghton Mifflin Company, 1954.

Massengill, H. Edward, & Shuford, Emir H., Jr. "A report on the effect of 'degree-of-confidence' in student testing." Technical Report SMC R-11, February 1968, The Shuford Massengill Corporation, Contract AF 49(638)-1744, Air Force Office of Scientific Research.

Shuford, Emir H., Cybernetic Testing. A paper presented at the National Society for Programmed Instruction Convention in Philadephia, Pa., during May 5-8, 1965.

Wilson, James W., Probability Measurement Procedures Experiment. A paper presented at the SMSG Mathematics Education Seminar on June 7, 1965.
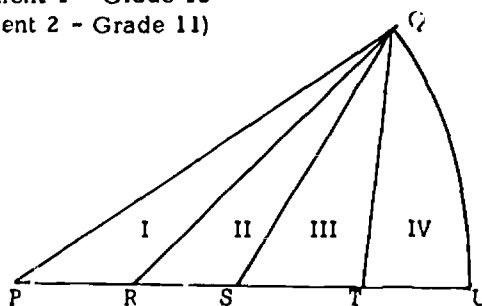
8

APPENDIX A

Tests for Experiments 1, 2, and 3

(Items 1-15 used for Experiment 1 - Grade 12
Items 1-17 used for Experiment 2 - Grade 11)

1. In the figure at the right, QU is an arc
of a circle with center at P, and PR =
RS = ST = TU. Which of the four regions
designated by Roman Numerals has the
greatest area ?

   (A)   I
   (B)   II
   (C)   III
   (D)   IV
   (E)   It cannot be determined from the information given above

2. Four interior angles of a convex polygon are each right angles. Which of the following state-
ments applies to this polygon ?

   (A) Some of the interior angles must be acute
   (B) The polygon must be regular
   (C) The sum of the measures of all the interior angles may be arbitrarily large
   (D) The polygon must be a rectangle
   (E) None of the above

3. Solve the equations: $(x + 1)(x + 2)(x + 3) = (x + 2)(x + 3)(x + 4)$

   (A) $-1, -2, -3$
   (B) $-2, -3$
   (C) $-2, -3, -4$
   (D) $-1, -2, -3, -4$
   (E) The equation has no roots

4. A club of 18 boys had a baseball team (9 players) and a football team (11 players). Five boys
were on neither team. How many were on both of the teams ?

   (A) 2
   (B) 5
   (C) 7
   (D) 9
   (E) You cannot tell from the information given

5. The numbers x for which $10 - x$, 10, and $10 + x$ are the lengths of the sides of a triangle
are exactly the numbers x such that

   (A) $|x| < 5$
   (B) $|x| > 5$
   (C) $|x| < 10$
   (D) $|x| > 10$
   (E) $|x| < 20$

6. The equation $2x^{10} + 5x - 1 = 0$ has a root near zero. Of the following, which best approximates
this root ?

   (A) -0.5
   (B) -0.2
   (C) 0.1
   (D) 0.2
   (E) 0.5

7. What is the greatest possible distance between a point in the plane and a nearest point with integer coordinates?

(A) 1

(B) $\dfrac{\sqrt{2}}{2}$

(C) $\sqrt{2}$

(D) $\sqrt{3}$

(E) 2

8. Find the **largest** value of x which satisfies the equation: $2(8^x) + 4(8^{-x}) - 9 = 0$.

(A) $-\dfrac{1}{3}$

(B) $\dfrac{1}{2}$

(C) $\dfrac{2}{3}$

(D) $\dfrac{3}{2}$

(E) 4

9. The solution of $\begin{cases} x + 5y = 17 \\ 1.5x + 7.501y = 25.503 \end{cases}$ is exactly $(2, 3)$ but the solution of

$\begin{cases} x + 5y = 17 \\ 1.5x + 7.501y = 25.5 \end{cases}$ is exactly $(17, 0)$. The best explanation of why the above happens is:

(A) the constants have different degrees of accuracy
(B) the graphs of the equations are nearly parallel lines
(C) zero has many peculiar properties
(D) one should never round off
(E) a regular 17-sided polygon is constructible

10. The graphs of the equations $y^2 = x$ and $x = y + 3$ split the plane into _five_ areas. (See diagram.) Which of these areas represent the points which satisfy both of the inequalities
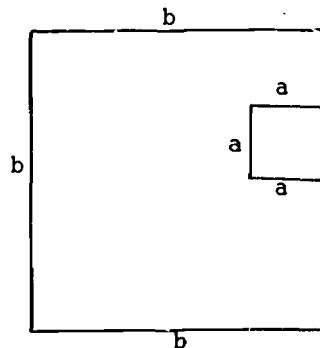
$$x - y - 3 > 0$$

and

$$y^2 - x > 0 ?$$

(A) I
(B) II
(C) I and II
(D) III and IV
(E) III and V
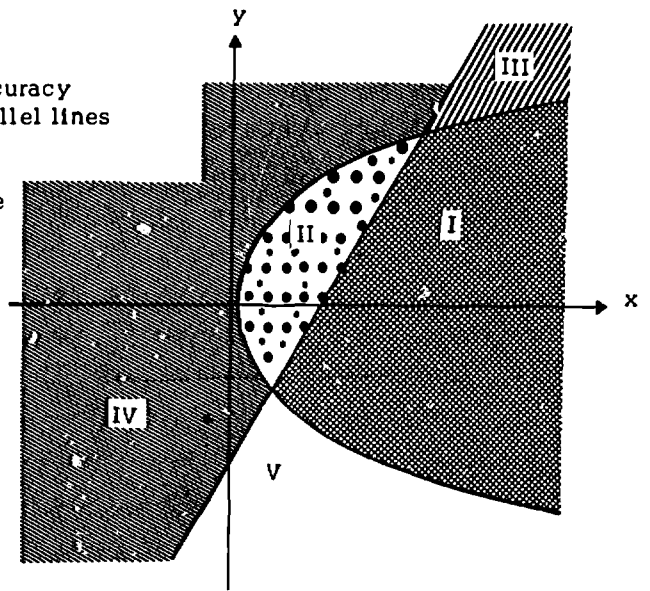
11. The diagram at the right is not necessarily drawn to scale. The line segments at each vertex are perpendicular. Both a and b are whole numbers. The area of the figure is 13 square inches. What is the perimeter of the figure?
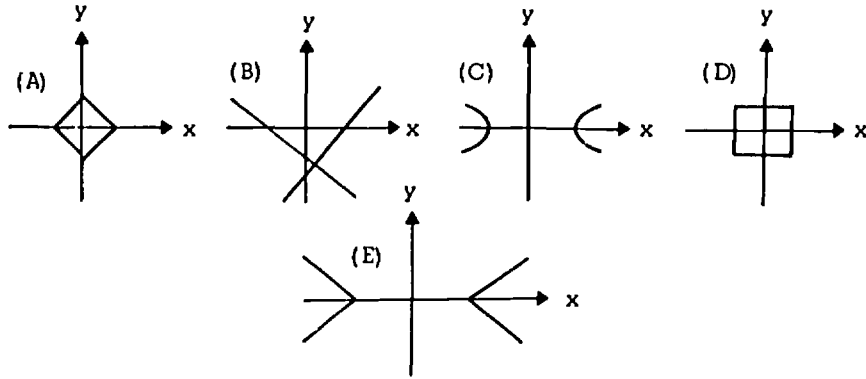
(A) 18 inches   (D) 40 inches
(B) 20 inches   (E) 42 inches
(C) 34 inches

12

12. Solve the inequality $\frac{1}{x} > \frac{1}{x+1}$.

   (A) All real numbers satisfy the inequality.
   (B) $x > -1$
   (C) $x < -1$
   (D) $x > 0$
   (E) $x < -1$ or $x > 0$

13. Which of the following is a sketch of the graph of $|x| = |y| + 1$?

14. Which of the following expressions is equivalent to: $(49)^3 \times (64)^3 \times (56)^{-3}$?

   (A) $(49 \times 64 \times 56)^{-27}$

   (B) $(49 \times 64 \times 56)^3$

   (C) $(56)^3$

   (D) $-(49 \times 64 \times 56)^3$

   (E) $(56)^9$

15. Let $\sqrt{a} = x$ and $\sqrt{b} = x + 1$. Which one of the following is equal to $2x + 1$?

   (A) $\sqrt{a + b}$

   (B) $\frac{a + b}{2}$

   (C) $a + b$

   (D) $b - a$

   (E) $\sqrt{a^2 + b^2}$

Problems 16 and 17 used for Experiment 2 only

16. Find all integers n such that $\frac{2n + 1}{3} < \frac{4n + 1}{5} < \frac{3n + 2}{4}$. The sum of these integers is (?).

   (A) 12
   (B) 14
   (C) 15
   (D) 18
   (E) 22

17. Which of the following values of x satisfies the equation

$$ax^2 + bx + c = 0 \quad \text{when} \quad a + b + c = 0?$$

(A) $\dfrac{b}{a}$

(B) $\dfrac{c}{a}$

(C) $\dfrac{a + c}{b}$

(D) $-\dfrac{b}{a}$

(E) $-\dfrac{c}{a}$

GEOMETRY TEST INSTRUCTIONS

In this test you will be asked questions about different topics in geometry. Do not become discouraged if there are some questions you cannot answer. No one is expected to know about every topic.

Although there will be some very hard questions, there are also some very easy ones that you will certainly be able to answer correctly, and these are mixed in among the others. Read every question!

Here is a sample question to show you how you should mark your answer.

Example 0. If one angle of a triangle contains 90°, the triangle is called:

_____(A) acute   _____(D) isosceles
__✓_(B) right   _____(E) equilateral
_____(C) obtuse

The answer is B. See how letter B has been checked for Example 0.

You are to answer as many questions as you can. Do not spend too much time n any one question. You should guess only if you can rule out some of the choices. Do not guess wildly.

*For these problems, you will mark each of your answers by checking one of the letters A, B, C, D, or E. You may use any space on the page for scratchwork.

You will have 30 minutes to answer 30 questions. DO NOT TURN THIS PAGE UNTIL YOU ARE TOLD TO DO SO.

*The treatment group was told to ignore this. They were instructed to place a probability weight in the blank reflecting their degree of belief as to the correctness of the alternative.

1. The diagonals of a parallelogram must be

_____(A) mutually perpendicular
_____(B) parallel
_____(C) equal in length
_____(D) bisectors of each other
_____(E) oblique

2. The geometric shape suggested by a can or a drinking straw is called a

_____(A) sphere
_____(B) cone
_____(C) pyramid
_____(D) cylinder
_____(E) cube

3. If the intersection of two different planes is not empty, then the intersection is

_____(A) a point
_____(B) two different points
_____(C) a line
_____(D) two different lines
_____(E) a plane

4. How many vertices has the above polygon?

_____(A) 3
_____(B) 6
_____(C) 9
_____(D) 15
_____(E) 24

5. An equilateral triangle is

_____(A) obtuse
_____(B) scalene
_____(C) right
_____(D) hyperbolic
_____(E) equiangular

6. If two lines are in the same plane, a line which intersects them in two different points is called

_____(A) a ray
_____(B) an oblique line
_____(C) a transversal
_____(D) a skew line
_____(E) a transit

7. Which of the following is true for this figure?

_____(A) $l \parallel m$
_____(B) $l = m$
_____(C) $l \sim m$
_____(D) $l \cong m$
_____(E) $l \perp m$

8. The following figure illustrates a

_____(A) prism
_____(B) cube
_____(C) cone
_____(D) pyramid
_____(E) cylinder
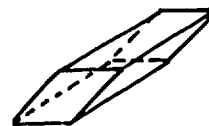
9. If two parallel lines are cut by a transversal, the alternate interior angles are

_____(A) supplementary
_____(B) complementary
_____(C) acute
_____(D) obtuse
_____(E) congruent

10. In the figure below, if $\overline{XY} \cong \overline{YZ}$ and $\overline{XZ}$ is not congruent to $\overline{YZ}$, then $\triangle XYZ$ is
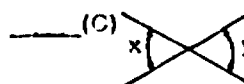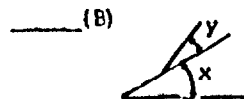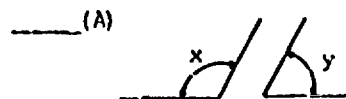

_____(A) equiangular
_____(B) scalene triangle
_____(C) a right triangle
_____(D) an equilateral triangle
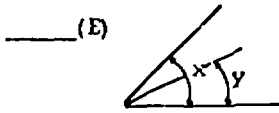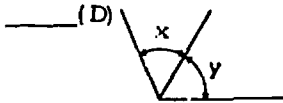_____(E) an isosceles triangle

11. If the adjacent angles formed by two intersecting lines have equal measures, the lines are

_____(A) parallel
_____(B) oblique
_____(C) perpendicular
_____(D) horizontal
_____(E) vertical

12. In which of the following figures are angles x and y adjacent?

_____(A)

_____(B)

_____(C)

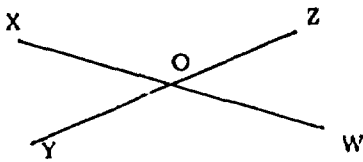16

12. (cont.)

_____(D)

_____(E)

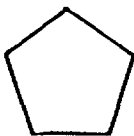13. Which of the figures below are parallelograms?

I.

II.

III.

IV.

_____(A) II and III only
_____(B) I and II only
_____(Ç) I, II,and III only
_____(D) II only
_____(E) I, II, III,and IV

14. In the figure below, which angle is supplementary to ∠XOZ?

_____(A) I only
_____(B) II only
_____(C) III only
_____(D) I and II only
_____(E) II and III only

15. The following figure represents a

_____(A) quadrilateral
_____(B) pentagon
_____(C) rectangle
_____(D) hexagon
_____(E) decagon

16. Which of the following are true?

I. A square is a rectangle
II. A square is a rhombus
III. A square is a parallelogram

_____(A) I and III only
_____(B) II and III only
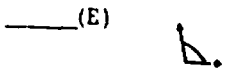_____(C) III only
_____(D) I and II only
_____(E) I, II,and III

17. Which of the following figures represents ℓ ∥ m?

_____(A)

_____(B)

_____(C)

_____(D)

_____(E)

18. Which of the angles below is the largest?

_____(A)

_____(B)

_____(C)

_____(D)

17

18. (cont.)

_____(E)

19. The geometric shape suggested by a tennis ball or a globe is called a

_____(A) sphere
_____(B) cone
_____(C) pyramid
_____(D) cylinder
_____(E) cube

20. How many points has a straight line?

_____(A) 1
_____(B) 2
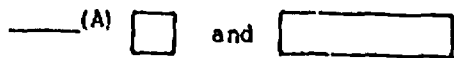_____(C) 5
_____(D) 17
_____(E) More than can be counted

21. Which one of the following has a different number of diagonals than the others listed?

_____(A) Rectangle
_____(B) Rhombus
_____(C) Trapezoid
_____(D) Hexagon
_____(E) Parallelogram
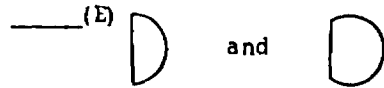
22. All squares are

_____(A) congruent
_____(B) equal
_____(C) similar
_____(D) collateral
_____(E) isoperimetric
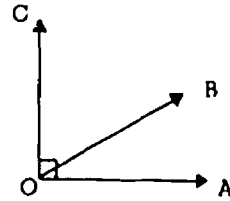
23. Which of the following pairs of figures appears to be similar?

_____(A) ☐ and ▭

_____(B) △ and ◺

_____(C) ⬠ and ⬠

_____(D) ◿ and ◺

23. (cont.)

_____(E) and

24. In the figure below, ∠AOB and ∠BOC are

_____(A) supplementary angles
_____(B) complementary angles
_____(C) both right angles
_____(D) congruent angles
_____(E) both obtuse angles

25. In a trapezoid, one pair of sides must be

_____(A) parallel
_____(B) vertical
_____(C) supplementary
_____(D) congruent
_____(E) isosceles

26. The sum of the measures in degrees of the angles of a triangle

_____(A) is between 30 and 180
_____(B) is 180
_____(C) is between 180 and 360
_____(D) is 360
_____(E) depends upon the sizes of the angles

27. Which of the following figures represents a simple closed curve?

_____(A)

_____(B)

_____(C)

_____(D)

_____(E)
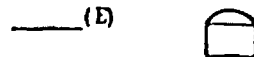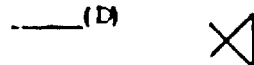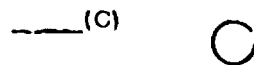
28. Two planes perpendicular to the same
line are

_____(A) perpendicular
_____(B) oblique
_____(C) intersecting
_____(D) parallel
_____(E) skew

29.
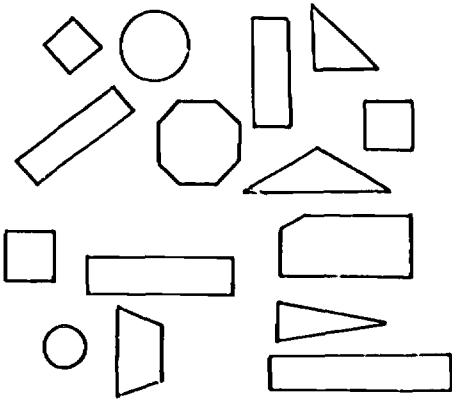
How many rectangles are shown above?

_____(A) 2
_____(B) 3
_____(C) 4
_____(D) 7
_____(E) 8

30. Which of the following is the measure
in degrees of an obtuse angle?

_____(A) 45
_____(B) 90
_____(C) 135
_____(D) 225
_____(E) Both C and D

APPENDIX B

Training for Probability Scoring,

Experiments 1, 2, and 3

TRAINING FOR PROBABILITY SCORING
(Used in all three experiments.)

The use of a probability scoring technique is somewhat different from the usual test taking strategy.

Instead of choosing one of five alternatives in a multiple choice item, one puts probability weights between 0 and 1 on each alternative. A person is guaranteed a maximum score if he follows instructions and honestly reflects his degree of belief as to the correctness of a possible answer to the test item. Your score is to be determined by summing the weights you assign to the correct answers.*

Strategy:

(1) If possible, work the problem using mathematical methods.

(2) If you arrive at what you believe to be the correct answer, assign 1 to the correct answer and 0 to the other choices.

(3) If you are not definite as to which of the alternatives is the correct choice, try to eliminate those which are definitely wrong. Assign 0 probability weights to these. Of the alternatives that could possibly be right, assign weights to these with regard to your belief in their correctness. The weights should sum to one.

Strategies for assigning weights:

(1) Do not waste a lot of time figuring out probabilities that add to 1. Keep the weights as simple as possible. Use .1, .2, .3, etc., as much as possible [don't use .64, .16, .10, .07, and .03, for example].

(2) Assign 0 to definitely wrong alternatives and 1.00 to a definitely right one. If you have no idea which alternatives are correct or incorrect and your choice is a random guess, give each alternative a weight of .2.

(3) If 1 out of 5 alternatives is definitely wrong and the other four seem equally likely to be right, assign .25, .25, .25, .25 to these alternatives (0, of course, to the wrong one). If 2 of 5 are definitely wrong and the other 3 seemingly equally likely to be right, assign .33, .33, .33 to each of these, etc.

(4) If one alternative seems more correct than another, be sure this is reflected by assigning a higher probability weight to it.

*Except for Experiment 3.

OVERLAY PRACTICE SHEET
(Experiments 1, 2, and 3)

1. The President of the U.S. is:
    a) Rusk   b) Johnson   c) Nixon   d) Humphrey   e) Romney

2. The Governor of North Dakota is:
    a) George Wallace   b) Nils Boe   c) William Guy   d) James Rhodes
    e) Ronald Reagan

3. Solve the equation: $5/n - 3/n = 1/4$
    a) 8   b) 4   c) 2   d) 1/2   e) 1/8

4. The Premier of Israel is:
    a) David Ben Gurion   b) Dayan   c) Nassar   d) Abba Eban
    e) Levi Eshkol

5. The Prime Minister of Canada is:
    a) John Diefenbaker   b) John Smith   c) Lester Pearson
    d) John D. Rockefeller   e) Sir Walter Thomson

6. The number of points common to a straight line and the sides of a triangle cannot be:
    a) 0   b) 1   c) 2   d) 3   e) infinite

PRACTICE ITEMS
Experiment 2 - 11th Grade

1. If $n + 20$ is a multiple of 8, then when (if ever) is $n + 10$ a multiple of 4 ?

    (A) never
    (B) always
    (C) whenever n is even
    (D) whenever n is a multiple of 4
    (E) whenever n is a multiple of 8

2. Which of the following equations has no rational root ?

    (A) $x - \frac{1}{x} = 0$

    (B) $x^2 - 1 = 0$

    (C) $2x + 3x = 5x$

    (D) $x^2 + x = 1$

    (E) $x^3 = \frac{8}{27}$

3. Arrange the areas P, Q, and R of the following shaded regions in increasing order.

24
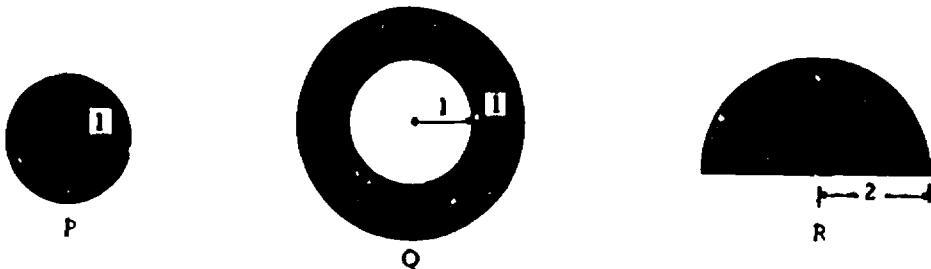
(A) P < Q < R
(B) Q < P < R
(C) P < R < Q
(D) R < P < Q
(E) Q < R < P

4. If the shaded region of the square pic-
   tured has an area between 60 and 70
   square inches and the unshaded area is
   between 75 and 85 square inches, the
   best estimate below of the length of a
   diagonal of the square is:
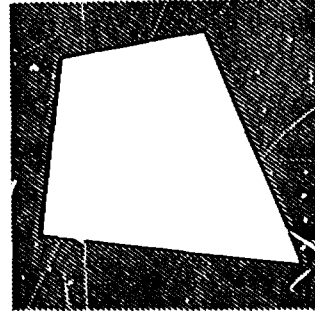
   (A) 12 inches
   (B) 17 inches
   (C) 23 inches
   (D) 29 inches
   (E) 35 inches

5. If $x \log_b 5 = \log_b 25$, then $x = (?)$.

   (A) 5

   (B) $\dfrac{1}{5}$

   (C) 2

   (D) log 20

   (E) the base b must be known before x can be determined.

FIRST PRACTICE TEST
Experiment 3 - 8th Grade

INSTRUCTIONS

Look at Sample Question 0.

0. ROSE   DAISY   VIOLET

    _____A red   _____B garden   _____C sweet   _____D grow   _____E lily

The words in question 0 are names of flowers. On the next line only _lily_ is the name of a
flower. The letter before _lily_ is E so we check that blank.

Now look at Question 00. Think in what way the words in Question 00 go together.   Then find the
word on the line below that belongs with them.

00. GO   RUN   WALK   MOVE

    _____A think   _____B dream   _____C march   _____D sing   _____E seem


The right answer is _march._

Wait for the signal to begin.

25

1. BENCH   SEAT   STOOL

    _____A table   _____B chair   _____C desk   _____D bed   _____E sit

2. POTATO   BEET   PEA

    _____A nut   _____B banana   _____C vegetable   _____D dinner   _____E carrot

3. BOOK   MAGAZINE   LETTER

    _____A movie   _____B newspaper   _____C radio   _____D lecture   _____E read

4. SHEEP   PIG   COW   HORSE

    _____A dog   _____B rabbit   _____C deer   _____D wolf   _____E beaver

5. PEEL   RIND   BARK   SHELL

    _____A corn   _____B orange   _____C tree   _____D husk   _____E box

6. DOLLAR   PESO   MARK   LIRA

    _____A change   _____B franc   _____C foreign   _____D purchase   _____E bank

7. MUSICIAN   ACTOR   HUMORIST   SINGER

    _____A ventriloquist   _____B professional   _____C amateur   _____D program
    _____E radio

8. ALLEY   ROAD   DRIVE   PATH

    _____A country   _____B glade   _____C passageway   _____D glen   _____E lane

9. STAIRWAY   LADDER   STAIRS   STAIRCASE

    _____A elevator   _____B climb   _____C hill   _____D escalator   _____E grade

10. HERD   FLOCK   SWARM   DROVE

    _____A lair   _____B den   _____C bunch   _____D pack   _____E insects

11. CAR   CAB   WAGON   CART

    _____A train   _____B carriage   _____C vehicle   _____D motor   _____E tandem

12. PIN   SAFETY PIN   HOOK AND EYE   ZIPPER

    _____A button   _____B belt   _____C strap   _____D suspenders   _____E garters

13. TIE   CRAVAT   STOCK   NECKCLOTH

    _____A bib   _____B collar   _____C scarf   _____D kirtle   _____E girdle

14. HONESTY   LOYALTY   SINCERITY   FAITHFULNESS

    _____A passivity   _____B servility   _____C devotion   _____D obsequiousness
    _____E compliance

15. PINE   SPRUCE   HEMLOCK

    _____A chestnut   _____B willow   _____C poplar   _____D fir   _____E maple

1. The accompanying figure shows a construction of a

_____(A) mean proportional of two segments

_____(B) perpendicular bi-sector of two segments

_____(C) median of a triangle

_____(D) diameter of a circle

_____(E) tangent to a circular arc
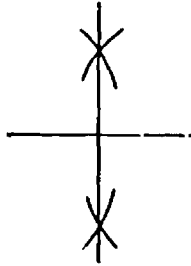
2. How many radii has a circle?

_____(A) 1
_____(B) 3
_____(C) 5
_____(D) 9
_____(E) More than can be counted.

3. The abbreviation m∠FGH means

_____(A) measure of angle FGH
_____(B) metric arc FGH
_____(C) m is the midpoint of FGH
_____(D) m is perpendicular to FGH
_____(E) minor angle FGH

4. Circles having the same center are called

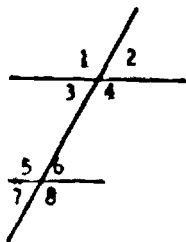_____(A) congruent
_____(B) asymmetric
_____(C) concentric
_____(D) corresponding
_____(E) coincident

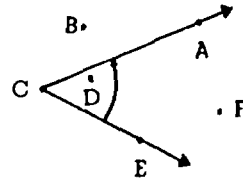5. In how many points do a circle and a line tangent to the circle intersect?

_____(A) None
_____(B) One
_____(C) Two
_____(D) At least two
_____(E) Infinitely many

6. In the figure below, which pair of angles are corresponding angles?

_____(A) 5, 8
_____(B) 2, 8
_____(C) 4, 5
_____(D) 1, 2
_____(E) 3, 7

7. Which of the points, in the figure below, are in the exterior of angle ACE?

_____(A) only B
_____(B) only D
_____(C) only B and F
_____(D) only D and F
_____(E) A, D, E, and F

8. Which of the following is part of a circle?
_____(A) Radius
_____(B) Center
_____(C) Arc
_____(D) Chord
_____(E) All of these

9. The total length of a closed curve is called its

_____(A) apothem
_____(B) area
_____(C) longitude
_____(D) slant height
_____(E) perimeter

10. The axis of the cone shown below is
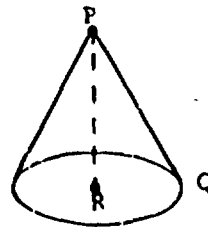
_____(A) point P
_____(B) point R
_____(C) segment PR
_____(D) segment PQ
_____(E) the circle with center R

Spheri:al Transformation Scoring Sheet
(Given to Students)

PROBABILITY WEIGHTS ON THE FIVE CHOICES

Correct Choice	A B C D or E	A B C D or E	A B C D or E	A B C D or E	Actual Score One Receives for Item
1	0	0	0	0	1.00
.5	.5	0	0	0	.71
.33	.33	.33	0	0	.58
.25	.25	.25	.25	0	.50
.2	.2	.2	.?	.2	.45
.9	.1				.99
.8	.2				.97
.8	.1	.1			.98
.7	.3				.92
.7	.2	.1			.96
.7	.1	.1	.1		.97
.6	.4				.83
.6	.3	.1			.88
.6	.2	.2			.90
.6	.2	.1	.1		.93
.6	.1	.1	.1	.1	.95
.5	.4	.1			.79
.5	.3	.2			.81
.5	.3	.1	.1		.84
.5	.2	.2	.1		.86
.5	.2	.1	.1	.1	.88
.4	.6				.56
.4	.5	.1			.62
.4	.4	.2			.67
.4	.4	.1	.1		.69

Equal Weighting

Spherical Transformation Scoring Sheet

PROBABILITY WEIGHTS ON THE FIVE CHOICES

Correct Choice	A B C D or E	A B C D or E	A B C D or E	A B C D or E	Actual Score One Receives for Item
.4	.3	.3	.0	.0	.69
.4	.3	.2	.1	.0	.73
.4	.2	.2	.2	.0	.76
.3	.7	.0	.0	.0	.40
.3	.6	.1			.59
.3	.4	.1	.1	.1	.57
.3	.4	.3			.52
.3	.3	.2	.2		.59
.3	.2	.2	.2	.1	.64
.2	.8				.24
.2	.7	.1			.29
.2	.6	.2			.30
.2	.6	.1	.1		.33
.2	.5	.2	.1		.35
.2	.4	.2	.2		.31
.2	.4	.2	.1	.1	.39
.2	.3	.3	.2		.39
.2	.3	.3	.1	.1	.47
.2	.3	.2	.2	.1	.43
.1	.9				.11
.1	.8	.1			.12
.1	.7	.2			.14
.1	.5	.2	.1	.1	.18
.1	.4	.4	.1		.18
.1	.6	.1	.1	.1	.16
.1	.7	.1	.1		.14
.1	.3	.3	.2	.1	.20
.1	.3	.3	.3		.19
.1	.4	.2	.2	.1	.20
.0	.a	.b	.c	.d	.00