DOCUMENT RESUME

ED 191 006 CS 005 603

TITLE	Independent Classroom Problem-Solving Model. A Teacher's Manual for Solving Reading Problems in the Classroom.
INSTITUTION	Office of Education (DHEW), Washington, D.C. Right to Read Program.
PUB DATE	80
NOTE	95p.: A number of pages may be marginally legible.
AVAILABLE FROM	Superintendent of Documents, U. S. Government Printing Office, Washington, DC 20402
EDRS PRICE	MF01 Plus Postage. PC Not Available from EDRS.
DESCRIPTORS	*Classroom Research: Elementary Secondary Education: *Reading Achievement: *Reading Research: Reading Teachers: *Research Methodology: Sampling: Statistical Analysis: Teaching Guides; Teaching Methods

ABSTRACT

This guide presents a classroom problem solving model designed to help teachers conduct their own classroom research. It suggests developing a procedure for identifying the instructional problems influencing reading achievement. The model is presented in steps that can be used independently or in concert with other steps. Practice activities are presented to reinforce elements of each step. The steps in the model are: (1) identifying the instructional problem, including formulating and evaluating the questions to be answered; (2) determining the hypothesis to be tested: (3) carefully defining and recording relevant characteristics of students or subjects: (4) listing types of sampling procedures; (5) selecting a study design: (6) outlining and listing all procedures to be employed in the study; (7) collecting, processing and interpreting the data: and (8) determining whether the hypothesis was supported or not and planning for the next course of action. Extensive tables, statistical equations, sample studies, and a glossary are appended. (MKM)

Independent Classroom Problem-Solving Model

A Teacher's Manual For Solving Reading Problems in the Classroom

Fall 1980

Department of Education Basic Skills Program Office

Donohoe Building/400 Maryland Ave. S.W./Washington, D.C. 20202

This handbook was originally prepared pursuant to a grant from the Office of Education, U.S. Department of Health, Education, and Welfare. However, the opinions expressed do not necessarily reflect the position or policy of the Office of Education, and no official endorsement by the Office of Education should be inferred.

For sale by the Superintendent of Documents, U.S. Government Printing Office
Washington, D.C. 20402

Foreword

The National Right to Read Office within the U.S. Office of Education (USOE) has the responsibility of administering Title VII (The National Reading Improvement Program) of Public Law 93-380 as amended by Public Law 94-194. In Public Law 94-194, a congressional charge is given to the Commissioner of Education "to carry out, either directly or through grants or contracts:

1. Innovation and development projects and activities of national significance which will show promise of having a substantial impact in overcoming reading deficiencies in children, youth, and adults through incorporation into ongoing state and local educational systems throughout the Nation, and
2. Dissemination of information related to such programs."

After a careful review of the "research to classroom practice" type of efforts, the Basic Skills Program Office determined that the most effective vehicle for moving research into classroom practice is the classroom teacher. Classroom teachers must be guided into solving their own instructional problems, utilizing a systematic process. The turn-around time for classroom improvement must be short if it is to be of use to the students currently needing assistance.

To this end, Dr. Alvin Loving, then serving as a Right to Read technical assistant and I, as Program Development Branch Chief, organized a steering committee in December 1-4, 1975, to examine the problem and a solution.

Right to Read Premises

Every teacher of reading must be the catalyst for effective learning of all children in that classroom.

Individualized and group instruction which use problem-solving techniques to find effective methods and materials must become a part of the teaching style of every teacher. The target of effective instruction should be children who are reading at grade level. The grade level norm could be based on national testing norms or criterion-referenced norms developed by the teacher or system reading staff. If the child learns to read effectively as indicated by either of the two measurements, success has been achieved.

The Steering Committee agreed that:

(1) the focus of the problem-solving model should be on teaching practice as contrasted with reading theory or tradition.

(2) the crux of the problems of ineffective readers is associated with the nature of the reading instruction children receive.

(3) more effort should be directed to aiding the classroom teacher in his/her effort to teach reading

(4) reading for all children should be targeted at grade level achievement

(5) problems dealing with instructional effectiveness and management should become the focus of attention of any problem-solving model.

(6) teachers should be guided in assessing systematically the impact of specific materials and methods with different kinds of children.

The Basic Skills Program Office accepted these conclusions and based on them and Right to Read's program rationale, had a model developed with accompanying instruments to enable the reading teacher to resolve his/her problems by:

1. identifying the problem.
2. assessing pupils and developing a diagnostic profile.
3. selecting from possible solutions those best suited to the students' needs.
4. implementing and analyzing the results and their effect on student achievement.

The model provides a procedural outline for conducting active classroom problem-solving in reading. It represents the first concentrated attempt to get classroom teachers to examine the effectiveness of their instructional practices in a systematic manner.

Shirley A. Jackson
Basic Skills Director

Key to Statistical Models in the ICPS Models

Model	Description	Decision Tree	Equation	Sample Study	Tables
Binomial	p. 28	p. 40	p. 53	p. 58	T, p. 81, D, p. 69
Chi Square One Sample	p. 28	p. 40	p. 53		C, p. 69
Chi Square-Two Sample	p. 29	p. 40	p. 53	p. 53	C, p. 69
McNemar	p. 28	p. 40	p. 53		C, p. 69
Mann-Whitney U Test	p. 29	p. 40	p. 55	p. 63	A, p. 68, J, p. 76, K, p. 76
Kruskall-Wallis	p. 30	p. 41	p. 56	p. 62	
Spearman Rank	p. 29	p. 41	p. 56	p. 64	P, p. 80
Difference of Correlations	p. 29				
Kendall Coefficient of Concordance: W	p. 30	p. 41	p. 57	p. 61	R, p. 80
Kolmogorov-Smirnov		p. 40	p. 53	p. 61	E, p. 69
Sign		p. 40	p. 54	p. 63	D, p. 69
Randomization/Matched Pairs		p. 40	p. 54	p. 63	A, p. 68
Fisher's Exact		p. 40	p. 54	p. 60	I, p. 71
Randomization/two independent samples		p. 40	p. 55		T, p. 81, B, p. 68
Cochran's Q		p. 40	p. 55	p. 59	C, p. 69
Friedman Two-Way		p. 40	p. 55	p. 60	C. p. 69, N, 78
Chi Square-K independent sample		p 41	p 55		C, p. 69
Contingency Coefficient		p. 41	p. 56	p. 59	C, p. 69
Kendall's Partial Rank		p 41	p. 56		Q, p. 80
Wilcoxon	p. 43		p 54	p. 65	G, p. 70
Runs				p. 53	F, p. 70

Contents

Introduction

In recent years, the public schools have been characterized by change and innovation resulting from changing demographic factors, new curricular offerings, and varied instructional methods and technology. Even though much of this change was intended to exert a positive influence on student performance and attitudes, the schools have experienced a steady decline in measured achievement performance for many of the nation's children. This situation has encouraged a persistent search for solutions that will aid in the reduction of these negative trends. One such solution treats instructional methods as a major influence on the achievement performance of children. Our efforts here will pursue aspects of instructional method necessary for efficient and effective learning.

Any instructional method must provide teachers with skills to use new materials and techniques for selected children experiencing different kinds of circumstances In order to employ new materials and techniques, teachers must learn to make appropriate selection and to execute competently. Many teachers have never been provided with specific training that enables them to develop and implant strategies for selecting and using instructional resources effectively.

In the area of reading, the selection and execution problem takes on special significance because of the vast array of materials and approaches that have to be chosen and matched with different learners. Teachers are asked to evaluate print and nonprint materials such as workbooks, practice exercises, filmstrips, textbooks, slides, and tapes dealing with various aspects of the reading program. In addition, teachers are expected to employ different approaches such as individualized reading instruction, learning stations, peer tutoring, and skill development techniques for children with different abilities, needs and support resources. It becomes apparent that . . . the problem we face is how to foster better selection of materials for use in executing reading instruction To accomplish this, more attention must be directed toward systematic approaches to problem identification, problem analysis, application of treatments and resource utilization. The mastery of these approaches can lead to a process of problem solving that utilizes relevant resources for meeting the needs of reading teachers and learners alike. Opportunities to learn problem solving and data utilization processes should provide teachers with a workable pattern for planning, managing and evaluating learning experiences more efficiently and effectively.

To assist reading teachers in data utilization and problem solving, we have developed this model to present the reading teacher with the necessary technical and practical components required for successful selection and execution of instructional experiences in reading. We decided to name the model Independent Classroom Problem Solving (ICPS) to reflect the activity that teachers will be doing.

The ICPS Model is intended to aid reading teachers in the development of competencies in the

1. Identification and definition of a problem area.
2. Selection of a specific problem for study.
3. Selection of a working hypothesis or prediction that guides the problem solving processes.
4. Careful recording of action taken.
5. Analysis of evidence regarding relations between actions and desired objectives.
6. Pursuit of an instructional procedure based on specific data generated on the children to be taught.
7. Accumulation and summary of evidence gained from instruction.
8. Continuous evaluation and retesting of actions pursued in instruction.

In order to aid teachers in the learning and practice of these problem-solving steps, the model was employed to integrate the steps and focus them on specific problems in the teaching of reading. Such a model must enable reading teachers to handle and understand each step individually and to assess their impact and value on practice outcomes collectively. The model employed should help reading teachers to understand both the nature of the instructional process and the resulting changes in the achievement pattern of those being taught The model must also emerge as a viable approach for understanding the nature and effects of teaching practice across a variety of content areas and environmental settings.

When considering the development of a model that helps reading teachers solve classroom instructional problems, one should attempt to elicit the views of other professional educators with differing experience backgrounds. The systematic and organized approach to problem-solving provides a common communication mode for professional educators to share practice and intellectual experiences Without such a common means for comparing the results of instructional approachs across a variety of environmental conditions, improved practice resulting from classroom teaching is unlikely to occur to the extent that it makes a differance. Therefore, the model is likely to evolve and change as practitioners from around the country provide feedback on the actual effects of the model when employed in classrooms under different conditions It is crucial to the meaningful improvement of the model that information concerning its use be incorporated into the formal presentation for widespread adoption

The intent of this model is to use a step-by-step approach to help reading teachers solve problems in reading instruction. As a result of this experience we expect teachers to be able to employ a systematic process for making key decisions involving instructional practice. Our procedure should enable teachers to apply key elements of the scientific method as an approach to problem solving and formulation of practice applications or to the reporting of results to others with similar problems. The general purpose of this model will be achieved if it will help teachers in their efforts to identify and isolate problems, devise a plan for approaching the problem systematically, collect and interpret relevant data, design reasonable activities and procedures for implementing a solution based on observed results, evaluate the performance of participants . . . and evaluate the effectiveness of the instructional and problem-solving processes employed. As teachers are provided with systematic practice in the use of this model, it is expected that they will adopt and adapt a realistic systematic process for guiding their instructional practice in reading. While the instructional area of focus is reading, this model is not limited to finding solutions to problems in the reading area only. Any subject matter area at any grade level can be the focus of the ICPS Model. The ICPS Model can be employed to deal with problems associated with learning content information, use of reference materials, attitudes toward others and self, attitudes toward learning, comprehension in the content areas, special characteristics of subject matter, approaches to teaching subject matter concepts, effects of organizational patterns on learning outcomes, effects of different types of instructional media on learning, and the influence of demographic sectors on learning subject matter in a variety of settings. In short the ICPS Model can be employed to find reasonable solutions to most instructional problems encountered by classroom teachers.

To achieve the purposes outlined earlier, the development of a problem solving model is essential so that teachers can solve instructional problems and improve teaching practice. *This process is based on the premise that improvement in student achievement is the direct result of improved teaching practice, and improvement in teaching practice is a personal and individual process that is the end product of self-evaluation, and from our point of view, self-evaluation that leads to instructional improvement is the end product of a systematic process of problem-solving.* Therefore, any attempt to improve teaching practice must provide a means for teachers to do systematic self-evaluation in the context of the activity they want to improve. It is this notion that shapes the specific objectives that will guide the continued development and implementation of ICPS.

Before there is a statement of specific objectives, it is necessary to indicate some of the beliefs that provide the basic foundation of our approach to building and using the model. These beliefs are:

1. The focus of the model is on teaching practice as contrasted with reading theory or tradition.
2. The crux of the problems of most ineffective readers is associated with the nature of the reading instruction children are receiving. In order to effectively improve the reading instruction, the children's strengths and needs must be assessed, a program to meet these needs must be implemented, and the effects of the treatment evaluated
3. More effort should be directed toward aiding the classroom teachers in their efforts to teach reading, to improve instructional methods, and to establish selection criteria for evaluating the program's effectiveness
4. There is an achievement level in reading most children should be expected to reach in judging whether the program implemented is effective. For instance, the expectation or criterion of program success may be that all children make one year's progress in reading as measured by a particular test or achieve some measurable standard of performance.
5. This criterion for instructional effectiveness and management should become the focus of any effort to improve reading achievement. If the criterion is not being met, it is important to establish the cause of the problem
6. Teachers should select and utilize specific materials and methods for different kin' f children with different types of learning problems. The ction and use should be based on observed outcomes resulting from teaching practice

Some methods that can be used to deal with problem solving are:

Problem	Materials	Method
Motivation	A number of different programmed lessons. Independent Activity packets.	Give children more choices in selecting the activities they would like to do.
Placement of the children	Standardized Diagnostic Reading Test Informal Assessment Procedures	Use this test for children achieving below the average ability because this test is designed to assess specific skill abilities for selected groups accurately.
Children not following instructions accurately	Tests for determining. (1) ability to follow instructions, (2) knowledge of vocabulary used and ability to follow written directions	Try verbal instruction, written instruction and a combination of the two to find which is more effective in improving children's ability to follow instructions.

It should be apparent that our understanding and beliefs will have a significant influence on the statement of specific objectives, the methods of developing the model, the settings for testing the model, and the evaluation of the worth of the model for improving practice

The specific objectives of the ICPS Model are

1. To help teachers develop a procedure for identifying the instructional problems influencing achievement.
2. To help teachers select, collect and organize data required to implement the ICPS Model.
3. To help teachers analyze data and interpret findings
4. To help teachers select solutions based on available data.
5. To help teachers apply data within the context of classroom instrumentation.

6. To help teachers determine the kinds of data they have on hand and the best method for interpreting and using the data for instructional purposes.
7. To help teachers evaluate the model's effectiveness for solving instructional problems by determining what insight they have gained by use of the problem-solving model.
8. To help teachers analyze the realities of the classroom and become more aware of how these realities affect teaching practices. This will be accomplished through the use of the Classroom Data Profile Sheet chart method which will help the teacher construct a data profile of each child and discover trends and relationships between a group of traits, skills or backgrounds of children in a given classroom
9. To help the teachers conduct a systematic study of the instructional process of reading by providing an outline of specific procedures that can be employed to evaluate the effects of different instructional methods.
10. To help teachers determine how implementation patterns might vary across various problem situations by determining which children benefit most from what kind of instructional approach.
11. To evaluate the effectiveness of the model

Questions Most Often Asked by Teachers about The ICPS Model

Rationale

1. What is the ICPS?
ICPS stands for Independent Classroom Problem-Solving It is a self-help model for reading teachers The purposes of the model are threefold.
- To introduce and teach reading teachers a systematic problem-solving process.
- To help reading teachers improve their instructional practice by the use of a systematic problem-solving process.
- To teach them how to select, collect and report data relevant to their instructional needs in reading

2. Who can use this model?
The model is designed to be used by classroom teachers with access to their students' achievement scores It can be used by teachers of any level, teaching any subject

3. In what situation should this model be used?
It is best utilized to help a teacher look at relationships between different scores that represent performance and characteristics that represent expected student behavior

4. What are the benefits accrued by using this model?
The users are provided with an approach to problem solving. The ICPS Model helps them approach instructional problems more systematically. With practice the teachers can develop a problem-solving mind set, refining their observational skills to that they can pick out the relevant variables to look at in any problem situation

5. Does the complete model have to be implemented?
No, after becoming familiar with the different parts of this problem-solving method, the teachers are encouraged to use the steps of the process most useful to them Once the problem-solving mind set is established many of the steps can be modified or skipped entirely
In fact, the first step (formulating the problem) is the key step. *Often the solution of the problem is contained in the proper statement of the problem* It is in the statement of the problem that the relevant variables of the problem and their relationships become evident.

6. Do I have to read the whole manual from cover to cover to learn this problem-solving technique?
No, you may already have many of the skills needed to complete the problem-solving method. Refer to "Steps in the Problem-Solving Procedure" in Step 8 to determine which steps you have already accomplished and which steps you will need more information to complete. Then go directly to those steps. It might be helpful to cover the entire manual to make sure something important is not being left out.

7. What kind of problems is this problem-solving method used for?
This method is used in determing the relationships between or among different characteristics, abilities and skills as measured by test scores. It can be used to help teachers to compare the effectiveness of various teaching techniques as measured by achievement tests. The method can also be used to determine which of the various characteristics are associated with various skills or abilities.

8. What kind of problems can *not* be answered with this method?
Questions which are so general that there are several parts to the questions and which have more than two or three variables involved cannot be answered using this method unless each specific part and two variables can be isolated and the question broken down into component parts.
Another type of question which may not be answerable is one wherein the variables cannot be measured. The information needed must be quantifiable in order to use this problem-solving model.
A third type of question which is unanswerable using this problem-solving method is one that calls for value judgments. This method simply allows teachers to generate, expand and organize factual information with which to make the judgments concerning instructional practices

9. Is the ICPS Model designed to develop new theories or to help teachers in practice?
Although the results of these studies may add to existing theory, the main emphasis of the ICPS Model is to help teachers improve instruction in Reading.

10. How long will it take to complete a project?
A project can be completed in about two weeks. The amount of time used is dependent upon the kinds of questions guiding the project.

11. What is the procedure for using the problem-solving method in the ICPS Model?
Step 1: The problem is identified. This entails formulating a question and determining if answering it will be worth while.
Step 2. The hypothesis is formulated. A decision on how to measure variables is made.

Step 3: The characteristics of the test population are defined.
Step 4: Methods of sampling are described.
Step 5: A study design is selected.
Step 6: All procedures are listed.
Step 7: Data is processed. Information and scores are charted. The statistical analysis most appropriate to the data is determined and the results calculated
Step 8: From the results, problem-solving decisions are made. The answer to these questions should be determined

a. Was the hypothesis supported?
b. Is there a significant difference?
c. Is there a significant relationship?
d. What was the answer to my questions?
e. Was the problem solved?

On the basis of the answers to *a.–e.*, decide the next course of action.

Defining the Problem

12. How does one formulate a measurable problem?
The requirements of a problem-solving question are

1. The problem be of manageable size (i e , it does not lead the teacher into excess expenditures of time, resources, and commitments).
2. The question must be concrete and explicit (i e , what is being questioned, and who is involved in the questions must be answered).
3. The question must be measurable.

In order to formulate a measurable question, the problem solver must be aware of the different techniques which are available to test the implications of the hypothesis.

13. What kind of test should be used to measure the variables?
First, the variables (i.e., the characteristics or performance) must be isolated. Then, the test best suited to measuring that trait for the particular age group of the study must be found. Most tests, test characteristics and test publishing companies are listed in Mental Measurement Yearbook by Buros. Teachers can also develop tests for the variables, though it is more difficult to establish a valid and reliable test than to utilize a published test already in use

Selecting a Design

14. Given a specific problem, how should one go about designing a study?
After defining the problem, it must be matched with a generic problem type and that format should be followed The number of subjects in a group, whether there is a pre test or a post-test, and what method was used for selecting subjects, all affect the design selection Consult "Statistical Model. Study Designs" in Step 5 which refer to the most appropriate statistical model in Appendix 1

15. What are experimental and control groups?
These are the groups that are formed when two or more different treatments will be given. The experimental group will receive the special treatment and the control group will receive the normal treatment, so as to ensure that changes in the relevant characteristics aren't due to other variables

Processing the Data

16. What knowledge of statistics is needed in order to interpret the data?
All essential information is provided in the manual. To determine which statistical equation will be needed for the problem, consult Step 5, Appendices 1 and 2. The procedure for ranking scores is a prerequisite to many statistical equations, and is shown in Step 7.

Results

17. What can be done with the results of the study?
Depending on the results and type of problem solved, the results can be used to: make the teacher more sensitive to relationships of certain traits and abilities, provide decision-makers with evidence supporting needed changes, share the results with other teachers, make decisions about continuing present curricular practices, adding to it, or changing it completely, and add to existing theory.

18. Which studies will go into the manual and why?
The studies which are most useful to other teachers, or which have a unique emphasis will be used in the manual.

Evaluation

19. What are the limitations of the model?
The most obvious limitations are the time constraints and the fact that some results may not be generalizable.

The ICPS Model

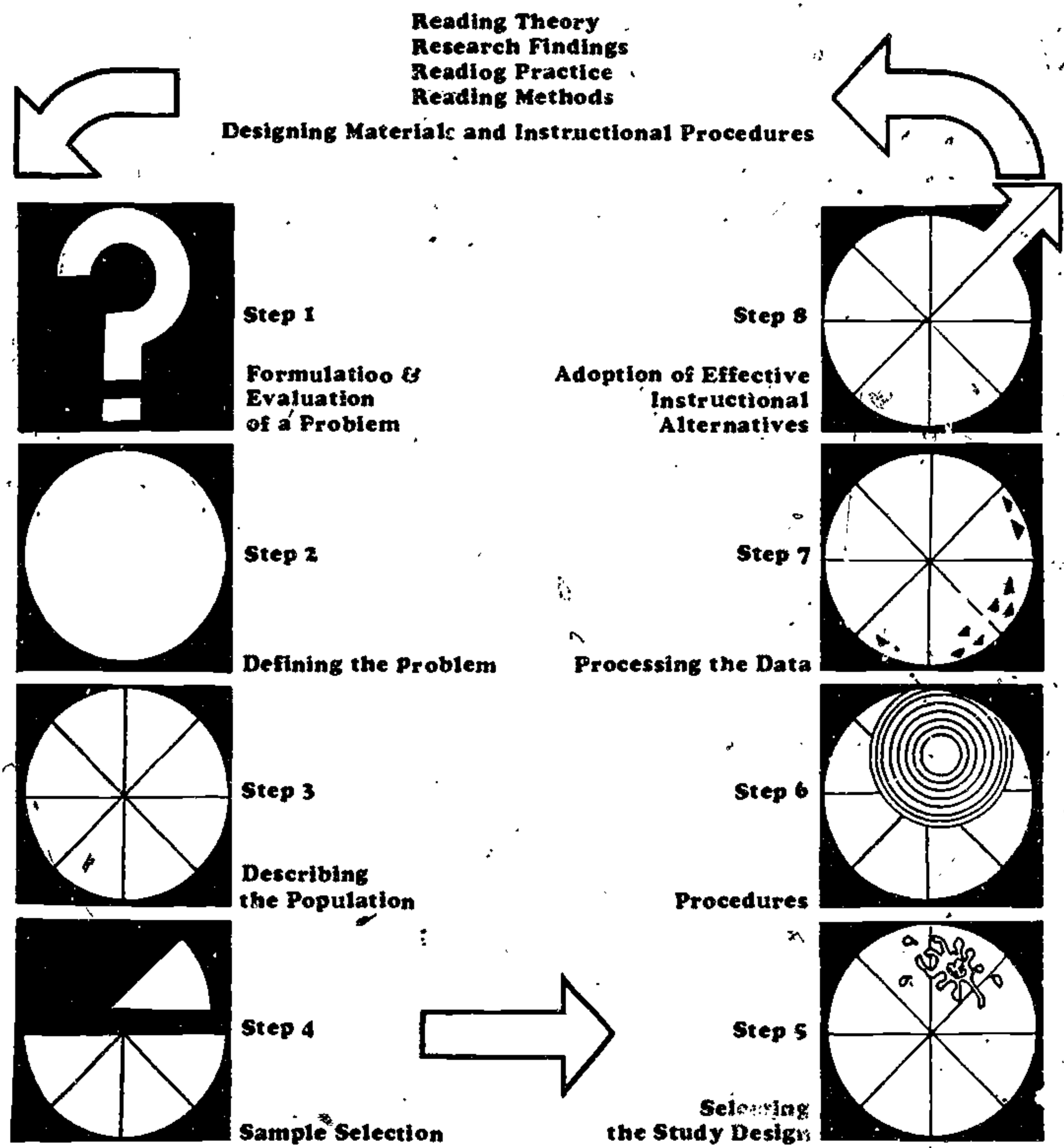
The Independent Classroom Problem Solving (ICPS) Model will be presented in steps that can be studied and used independently or in concert with other steps. It is important for the user to understand each step so that the relationship between the steps are useful and meaningful. The practice activities are designed to reinforce elements of each step while focusing attention on how each step contributes to the total process of problem-solving. • The first step in the ICPS Model is the identification of the problem. Problem identification includes formulating and evaluating the questions to be answered. • The second step is concerned with determining the hypothesis to be tested This step also entails choosing the best alternative solution and deciding how to measure the variables selected for study. • The third step involves a careful definition and recording of relevant characteristics of students or subjects. • The fourth step lists types of sampling procedures • In the fifth step a study design is selected • The sixth step requires the outline and listing of all procedures employed in the study. ● Collecting, processing and interpreting the data comprise the seventh step. Decisions and analysis of the data are made at this point. ● Step eight determines whether the hypothesis was supported or not, and gives plans for the next course of action.

Personal Questions Concerning ICPS Model:

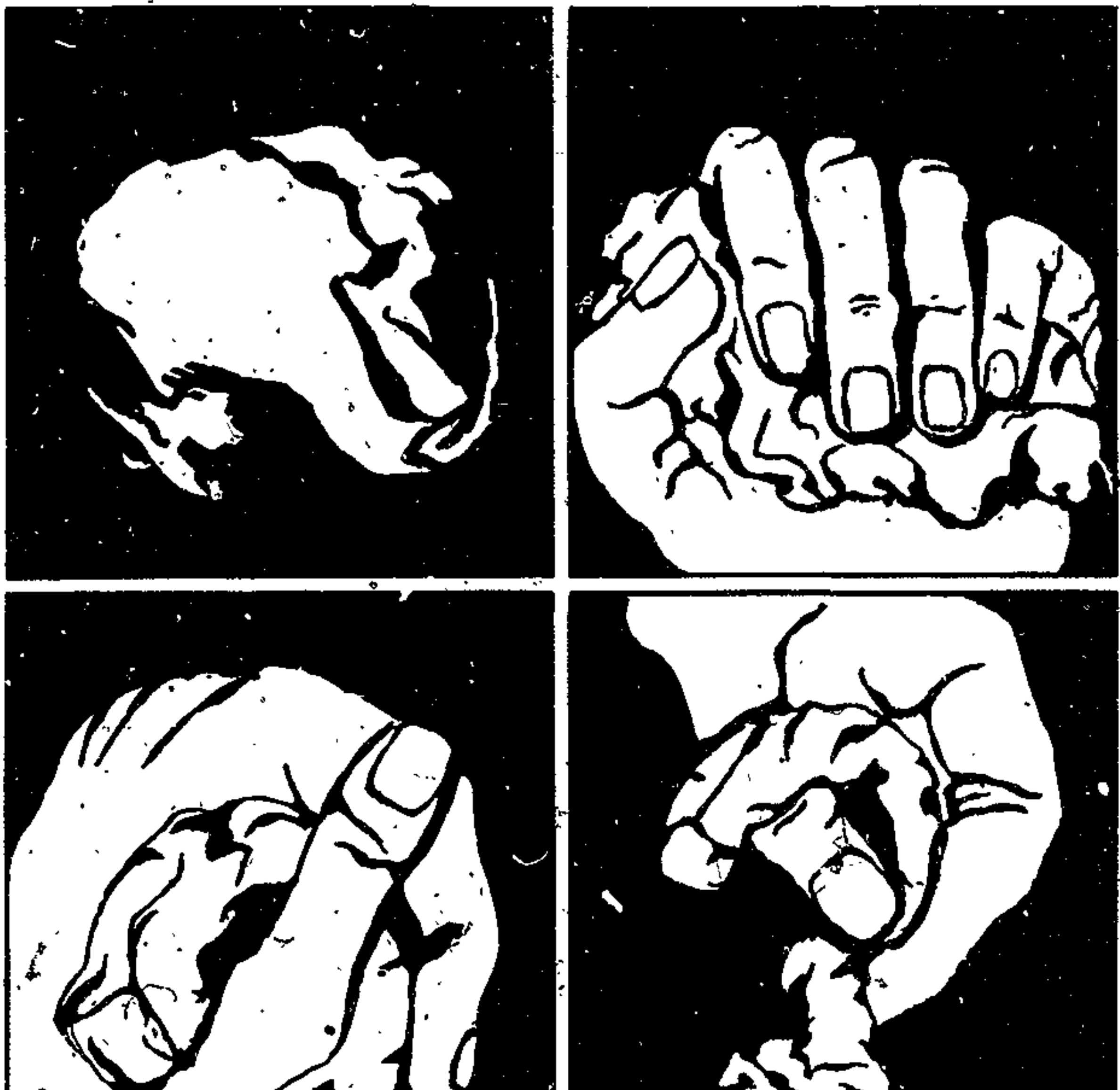
1.
2.
3.
4.
5.

NOTES:

The Independent Classroom Problem-Solving Model



Step 1:
The Question



Formulation of the Problem

"The purpose of problem-solving is to discover answers to questions through the application of scientific processes. To be answerable by problem-solving, questions must be asked in such a way that experimentation or observation in the real world will yield the needed information. Scientific questions, of course, are not haphazard or sheerly speculative. They are prompted by reflections on the relationships of different forces and influences as we know them. Problem-solving processes allow us to test our hunches about how certain forces and influences affect one another (P 23, Barnes)

The requirements of a problem-solving question are: (1) that the problems be of manageable size, that is, it should not lead the teacher into expenditures of time, resources, and commitments, (2) the question must be concrete and explicit. It must clarify who is involved in the question, what is to be determined, and why it is significant; (3) the question must be measurable, and the results must be quantifiable terms, symbols, determinations, or expressions of facts rather than values, inferences, or suppositions. The following exemplifies a good problem-solving question "Is oral presentation more effective than written presentation in improving students' performance on tests within a classroom?"

This problem is manageable in several ways. First, it would involve a simple comparison between two types of instruction, both of which the teacher uses or has used in his/her classroom instruction. The teacher has all necessary resources available to answer the question. (S)he has the subjects, her/his class, and her/his own instructional techniques to use. Little commitment is involved except the decision as to when to begin the study and when to end it, in other words, how long the study will last.

The problem is concrete and explicit. Those involved are the students and their performances on tests, as well as the teacher and the two different styles of instruction. What is to be determined is whether oral instruction or written instruction is more effective. And finally, the reason the problem is significant is because one type of instruction versus another may improve students performance on test scores.

The problem is measurable because the results can be quantified in terms of test scores and the improvement or lack of improvement measured. Decisions will have to be made as to how the results will be measured. For example, will the teacher use two groups of students? Will (s)he use different academic subjects for different types of instruction? Will (s)he use the entire class and give oral instructions for a designated period of time and then give written instructions for an equal period of time? Will (s)he use a post-test only or will (s)he use both a pretest and post-test? These decisions will necessarily have to be made to eliminate items which might make the results unclear; however, the problem is measurable, and many of these aforementioned decisions will be clearer in a given classroom because they are logical.

In determining whether a question can be answered using the problem-solving method, the following questions must be answered:

1. Do you already know what the answer is?
2. Can the question be answered by use of a test or other quantifiable data?
3. Is the question precise enough to direct efforts toward efficient and effective ways to find meaningful answers?
4. Does the question clarify what is to be determined and who is involved in the question?

If you can answer yes to each of the preceding questions for a given problem, then the problem is researchable. It is the objective of this section to teach the user how to identify and define a problem in instruction

Some Illustrative Examples of Researchable Problems

Problem 1. How does the *amount of time spent* in *individualized instruction* affect *reading performance* for sixth grade students?

Activity 1:1

Following is a list of problems. Put an A in front of each appropriate problem. Put an I in front of each inappropriate problem.

______	1. Is *oral presentation* more effective than *written presentation* in improving *student performance* on tests?
______	2. Is there greater improvement in *reading comprehension* when an *individualized approach* is used instead of the *traditional reading group approach*?
______	3. Does *homogeneous ability grouping* produce more positive attitudes toward learning than does *heterogeneous grouping?*
______	4. Is there a relationship between *self-concept* and *reading achievement?*
______	5. What is the relationship between *oral language skills* and *written language skills* as determined by Selected Standardized Test?
______	6. Is there a relationship between *Socioeconomic Status* and *reading achievement scores?*
______	7. This study seeks to prove that *bad attitudes* cause *bad grades.*
______	8. What are the *factors* causing *low achievement* in reading?
______	9. What *changes* should I make in my *curriculum?*
______	10. *Should the government* mandate force *busing?*

See next page for answers

Problem Evaluation

The answers that should have been given for activity 1:1 are as follows:

1. *A* 2. *A* 3. *A* 4. *A* 5. *A*
6. *A* 7. *I* 8. *I* 9. *I* 10. *I*

NOTES.

The explanation and evaluation of these answers are discussed further in the paragraphs below.

The first three questions (1–3) are "applied" problem-solving questions; they were formulated to find a specific way to improve reading instruction. The next three (4–6) are more "basic" problem-solving questions in that they were formulated for the purpose of learning more about specific relationships. Both types of questions can be useful to the reading teacher in solving problems in his/her classroom and improving his/her methods of teaching. These are examples of good problems because they pinpoint the relationship to be studied and observable results can be obtained to answer the question asked.

The last four sample questions (7–10) are not examples of appropriate problem-solving questions. Number 7 is not even stated as a question. The statement reflects the problem-solver's bias and points to the possibility of subjective, unscientific interpretation of the results. Questions 8 and 9 could be made researchable by grouping them into smaller, more specific questions. For instance, in question 8, one particular possible factor could be isolated, like motivation, and the strength of the relationship between motivation and high achievement could be measured. For question 9, two different methods or materials could be implemented at different times, or with different students and the results of the two programs could be compared. Number 10 could not be answered using this problem-solving model because it involves value judgment and there is no way to test these values.

Personal questions on the Formulation of a Problem:

1.

2.

3.

ACTIVITY 1:2

Activity 1.2 is designed to assist you in clarifying problems and evaluating them in relation to the factors that must be considered in formulating a problem for systematic study.

Practice Activity

In the chart below, write the problem area which concerns *you* then check the appropriate lines below to help define *your* problem. Then formulate *your* problem solving question and write it below. More than one letter may be checked in each section.

Additional Readings

Easy: Barnes, p. 12, 27; Ferguson, p. 1
Hard: Slakter, p. 255

Area of concern: Time allotment in instruction and reading performance.
Your area of concern:

Yes	Couldn't Determine	
		1. Who was affected by the problem?
____	____	a. learner
____	____	b. teacher
____	____	c. group of learners
____	____	d. administrators
____	____	e. parents
____	____	f. other specify ____________
		2. What is the nature of the problem?
____	____	a. motivation
____	____	b. materials
____	____	c. instructional methods
____	____	d. behaviors
____	____	e. methods of teaching
____	____	f. evalua' on
____	____	g. other specify ____________
		3. Who can bring about necessary change?
____	____	a. learner
____	____	b. teacher
____	____	c. group of learners
____	____	d. parents
____	____	e administrators
____	____	f. other specify ____________

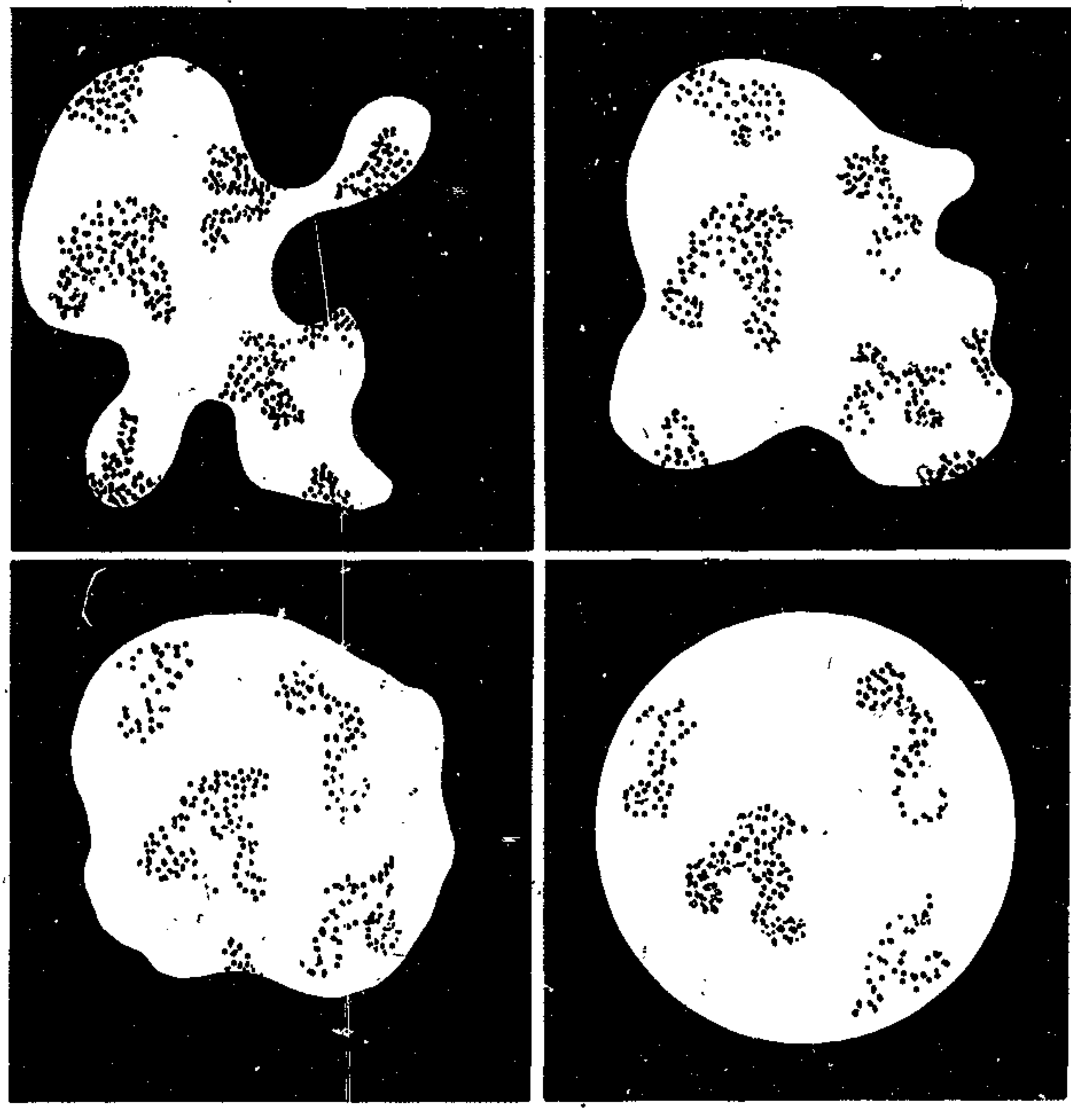
Yes	Couldn't Determine	
		4. What is the goal?
____	____	a. more effective teaching methods
____	____	b. more effective evaluation and placement
____	____	c. alleviating inadequate performance
____	____	d. alleviating inadequate behavior
____	____	e. discovering relationships
____	____	f. other specify ____________
		5. Is this problem worthwhile to pursue?
____	____	a. Will the results help me to teach better?
____	____	b. Will the results help me make decisions?
		Explain

Statement of the problem solving question:

How does the amount of time spent in individualized instruction affect reading performance for sixth grade students.

Your personal statement of an adequate problem-solving question.

Step 2:
Defining the Problem



Hypothesis Formulation

To apply the scientific method in solving a problem one must formulate a hypothesis or possible solution to the problem. *The hypothesis is a tentative proposition suggested as a solution to a problem or as an explanation of some observed state of affairs.* It presents a simple statement of the problem-solver's expectations relative to a relationship between variables within the problem. The hypothesis is then tested with the aid of a systematic problem-solving procedure. One must decide whether the hypothesis should be based on specific theory or predict what the outcome will be from surveying current research or personal experience.

A hypothesis should consist of a theoretical definition and an operational definition of terms and/or concepts. *For example, a theoretical definition of intelligence might be the ability to adapt to one's environment, but the operational definition of intelligence could be those behaviors that are measured on the Stanford-Binet Test.* Having both an operational and theoretical definition has three advantages for the problem-solver: (1) interpretation of results will be more realistic; (2) the two definitions will support the techniques of observation and measurement used; and (3) the problem-solving process should be more meaningful.

A different example is aggressive behavior. A theoretical definition could be the intent to injure, hurt, or destroy. In this case, an operational definition is crucial for the term "intent" and to whom it is intended. On the one hand, it may be the hitting, kicking, punching, and verbal insults from one child to another. On the other hand, it may be the same behavior measured in doll play. *Obviously, an operational definition is extremely important for the measurement of results, whereas a theoretical definition adds more meaning to hypothesis.*

In formulating hypotheses, one must look at all avenues for testing the statement, choose the most feasible solution to the problem, and find the standardized or appropriate test which will provide the most precise data needed to determine the solution. There are a variety of tests available for measuring many diverse traits and abilities. For the full spectrum of tests, and a critique of each test's uses and validity, consult Buro's *Yearbook of Tests and Measurement*. A brief list of tests used by previous participants of a problem-solving model workshop is included in Appendix 2 p. 58.

Some Illustrative Examples of Researchable Problems:

Theoretical Hypothesis: Time on task by students has a direct influence on their learning performance.

Operational Hypothesis: Students who spend more time working independently with instructional materials will make greater gains in measured reading performance than students who spend more time receiving direct instruction from the teacher.

Personal questions about the formulation and statement of hypothesis:

1.

2.

3.

4.

5.

NOTES:

Activity 2:1

Put an O in front of hypotheses that are operationally defined, a T in front of hypotheses that are theoretically defined, and a B if they are both.

________	1. There will be significantly more improvement in reading comprehension, as measured by the Stanford Diagnostic Test, when an individualized reading program is utilized than when a traditional group reading program is used.
________	2. Children in a racially integrated classroom will be more academically comfortable. Academically comfortable means more learning will take place.
________	3. Self concept will be improved by learning to read better. Self-concept means feeling positively or negatively about oneself.
________	4. There will be a significantly positive relationship between self-concept and reading achievement. Self-concept means a positive or negative feeling about oneself. Reading achievement means how well one reads. Self-concept will be measured by the Coopersmith Self Esteem Inventory, and the Gates-MacGinite Reading Test Survey "E" will be used to test reading achievement.
________	5. Children in a class where a regular teacher is absent will not perform as well on the SRA Reading Comprehension Test as peers of equal ability who are in a class with one teacher of regular attendance. Ability is defined as scores received on SRA Reading Comprehension, which is a machine-scored achievement battery administered throughout the district.

Answers to Activity 2:1 1. O 2. T 3. B 4. O 5. O

Write your personal operational hypothesis:

Null Hypothesis

The first step in initiating a systematic process in problem-solving is to formulate and state the null hypothesis (Ho). The *null hypothesis* establishes a standard against which a problem-solver tests data statistically. For example, if a problem solver wants to test a hunch that a certain method of teaching vocabulary development will result in greater gains for students than a conventional method of teaching vocabulary development, it suggests that there is a need to advance a statement that enables one to test that hunch. In order to test the hunch, the problem-solver advances the *null hypothesis* that there will be no (null) difference in measured reading comprehension between the two groups regardless of the method of instruction employed. The problem-solver teaches one group of children by the conventional method outlined in the test and another group by the special method of vocabulary development. After completing a procedure of systematic problem-solving like the one outlined in the ICPS Model, vocabulary data are collected and analyzed for each of the groups taught. The measured results obtained by each group are compared statistically to determine the extent the special method of vocabulary development resulted in a significant difference ($p < .05$) test results obtained by the group taught by the conventional method. If the problem-solver determines that a "statistically significant difference" exists between two groups, it can be assumed that the observed difference resulted from the difference in the method of teaching used. It is at this point that the problem-solver will make extensive use of the null hypothesis procedure to make a reasonable decision to guide instructional objectives and approaches. While the assumption may be entirely justified and the method of instruction employed to teach vocabulary development may have resulted in increased vocabulary, it is also possible that some undetected or unrecognized external factor may have produced the observed difference in vocabulary performance. Therefore, any conclusion drawn on the basis of the null hypothesis must be weighed against the possibility of a valid assumption proving to be incorrect.

The problem-solver must keep in mind the fact that the testing of the null hypothesis only shows that the results indicate a direction for further study and future discovery. It is not enough to determine that one method made a difference over another method. The teacher as problem-solver would not be satisfied with this finding as the sole basis for determining all future actions toward teaching vocabulary development. As the person responsible for outlining instructional directions for many learners, the teacher as problem-solver would want to know precisely what influenced the observed difference in vocabulary development in terms associated more with quality of learning rather than quantity alone. In this sense the null hypothesis serves to direct the application of problem-solving methods to assist the teacher in determining which real differences occur when different teaching methods are employed. Even though the null hypothesis does aid the teacher in making a decision about whether there is a difference between two methods of teaching, it does not reveal all of the underlying factors that influenced the results. Testing the null hypothesis is just a single step in helping the teacher as a problem-solver to understand the nature and impact of instructional decisions and approaches.

Some Illustrative Examples of Researchable Problems.

Hypothesis: Students who spend more time working independently with instructional materials will make greater gains in measured reading performance than students who spend more time receiving direct instruction from the teacher.

Null Hypothesis: There is no difference in the measured reading performance gains for students who spend more time working independently than for students who spend more time receiving direct instruction from the teacher

Activity 2:2

Write the null nypothesis for each hypothesis listed below.

Hypothesis	Null Hypothesis
1. There is a significant correlation between the number of books read outside of class and reading achievement in class.	
2. There is a significant difference in comprehension between children reading material only orally and those who read it only silently.	
3. Children will significantly increase their retention of word meanings if they study the words' etymologies.	
4. There is a significant positive correlation between children's enjoyment of reading and number of books in the home.	

Activity 2:3

Personal Questions about writing Null Hypothesis:

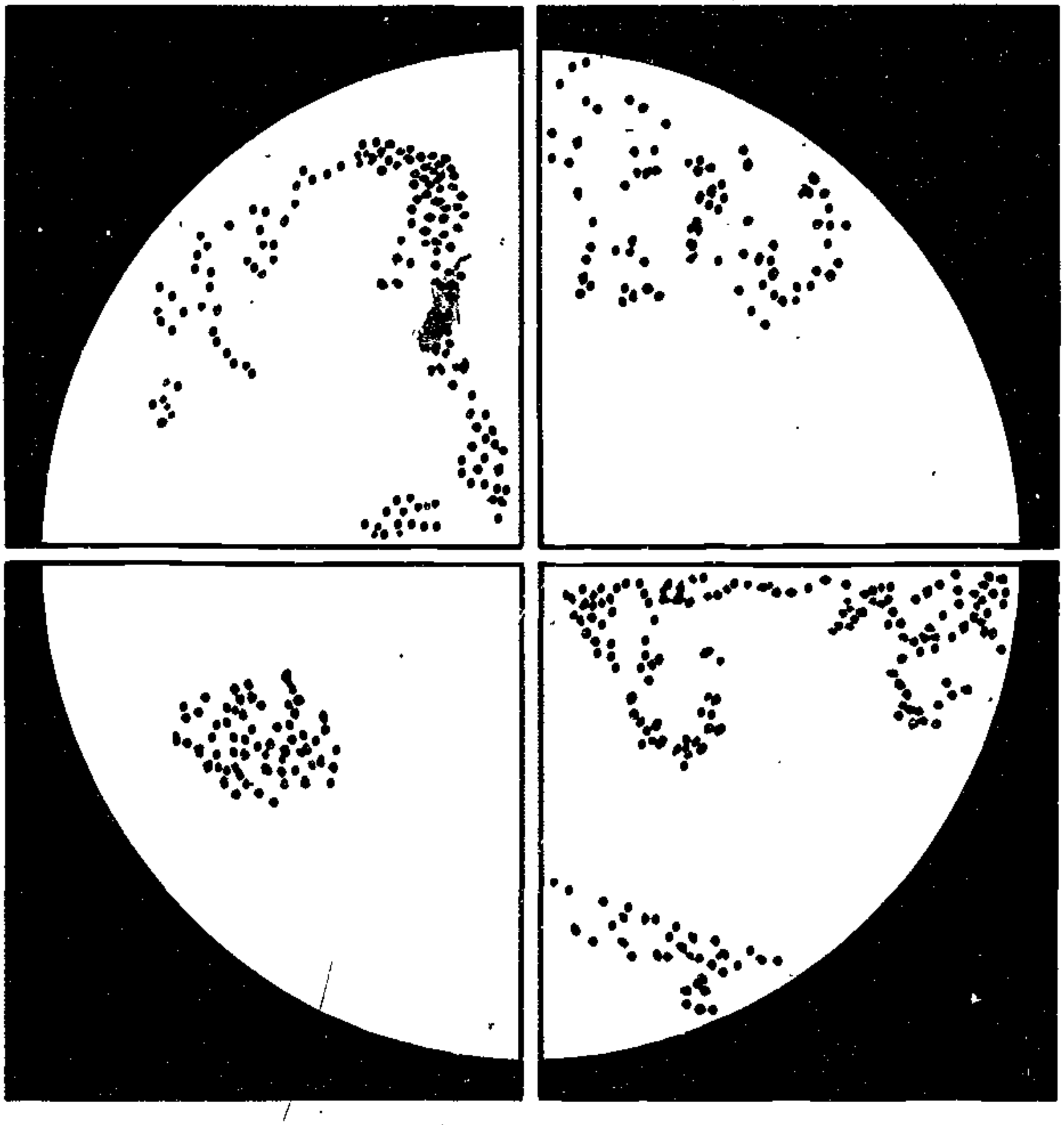
1.
2.
3.

NOTES:

Write a hypothesis and null hypothesis fo. each of the following problem-solving questions.

1. Do children understand oral or written instructions better?
2. Do children understand and retain a story better if they read it or listen to it?
3. Do children who have laterality problems also have more difficulty in reading?
4. Do children understand and learn the material better if they can choose what they read?

Step 3: Describing the Population



Establishing a Classroom Data Base

One of the key tools required for a classroom teacher interested in problem-solving is a classroom data base for all learners. The classroom data base (CDB) provides foundational information on each child that reveals demographic and personal characteristics that might influence the way each child reacts to instruction in a group setting. Many times a well-conceived and constructed classroom data base will reveal important patterns that define the nature of the group to be taught. Sometimes these classroom patterns suggest meaningful approaches to potential problem areas or provide a realistic base for interpretation of observations and instructional planning.

The first step in the establishment of a classroom data base (CDB) is the construction of a Classroom Data Profile Sheet (CDPS). The CDPS is a form that lists key descriptive and numerical information on each child utilizing format that aids the teacher in spotting key relationships or significant demographic patterns. A sample CDPS pictured below shows the information that should be collected on all children routinely and some items that might be good to have on hand when there is a need to find answers to certain instructional qu 'ions. Sometimes solutions are indicated by the nature or the display in the CDPS and no further study procedures would be called for to de·' with a specific problem. When additional data are required to answer certain questions, the teacher can associate newly collected data with existing information and be in a better position to make appropriate instructional decisions.

Instructions for Using CDPS Chart

Regardless of the study, the CDPS Chart should include the whole class when filling out Sections 1-26. Only students involved in the study should be included when filling out Sections 27-31.

Name—the students' (subjects) names are entered first. This is for your benefit only in collecting data. If the study is published or in any way shared, Federal law requires that the name be deleted from the study.

ID#—the identification number is also only for your benefit. This can be the number of each child's test file or one that you give the students to ensure confidentiality. This may or may not be deleted from the final study.

Age in Months—this is calculated using the following formula
$y(12) + f = a$
y = years
f = full months (over fifteen days)
a = age in months
This calculation is useful when the study requires more exact measurement of age than years.

Sex—M is male, F is female. This may be used to discover if there is a relationship between sex and another variable, such as number of books read, achievement or verbal ability.

Race—W = white B = black O = other
Here is another variable which may have bearing on the students' performance.

Grade Point Average—4.0 = A 3.0 = B 2.0 = C 1.0 = D 0 = F

The formula is $\frac{\text{sum of grades}}{\text{number of courses}}$

Title I—does this child participate in the title I program?

Special Education—Is the child enrolled in a special education program?

Special Reading Programs—Is the student enrolled in any special reading programs in the school? List programs.

Bilingual Program—Is the student enrolled in foreign language programs?

Word knowledge—*comprehension and total work study skills.* These are all tests the teacher can give to determine more precisely what the child's ability really is in reading.

Number of Books read by Student—give number of books students have read during the current year.

Responses on reading interest inventory—List major interests revealed in interest inventory.

Latest available achievement scores—these will be available in the child's file. These scores will help you to get a clearer, more complete picture of your students' strengths and weaknesses.

Scores collected for problem solving—these scores will be gathered from the tests you as the teacher will give the children. The data received from those will be used to support or refute your hypothesis.

RS = Raw Score
SS = Standard Score

Classroom Data Profile Sheet

1. Name												
2 ID#												
3 Age in Months												
4 Sex												
5 Race												
6 Grade-Point Average												
7 Number of Days in Attendance												
8 Major Health Problems												
General Background Factors												
9 Title I												
10 Special Education												
11 Special Reading Programs												
12 Bilingual Programs												
Latest Available Reading Scores												
13 World Knowledge Vocab	RS											
	SS											
14 Comprehension	RS											
	SS											
15 Total	RS											
16 Work Study Skills	RS											
	SS											
Background Factors Educational Latest Available Achievement												
17 No of Books read by Student												
18 Scores on Reading Interest Inventories												
19 Social Studies	RS											
	SS											
20 Math	RS											
	SS											
21 Science	RS											
	SS											
22 Library Skills	RS											
	SS											
23 Language Arts	RS											
	SS											

Activity 3:1

Criterion Checklist for Classroom Data Base

Yes	No	N.A.	
______	______	______	1. Did it provide foundational information that might influence the way each child reacts to information in a group setting?
______	______	______	2. Did it reveal important patterns that define the nature of the group to be taught?
______	______	______	3. Did it suggest potential problem areas?
______	______	______	4 Did it provide meaningful approaches to potential problem areas?
______	______	______	5. Does it provide a realistic base for interpretation of observations and instructional planning?
______	______	______	6. Does it indicate that further problem-solving procedures will be necessary?
______	______	______	7. Will more data have to be collected?

Personal Questions about Classroom Data Profile Sheet

1.

2.

3.

4.

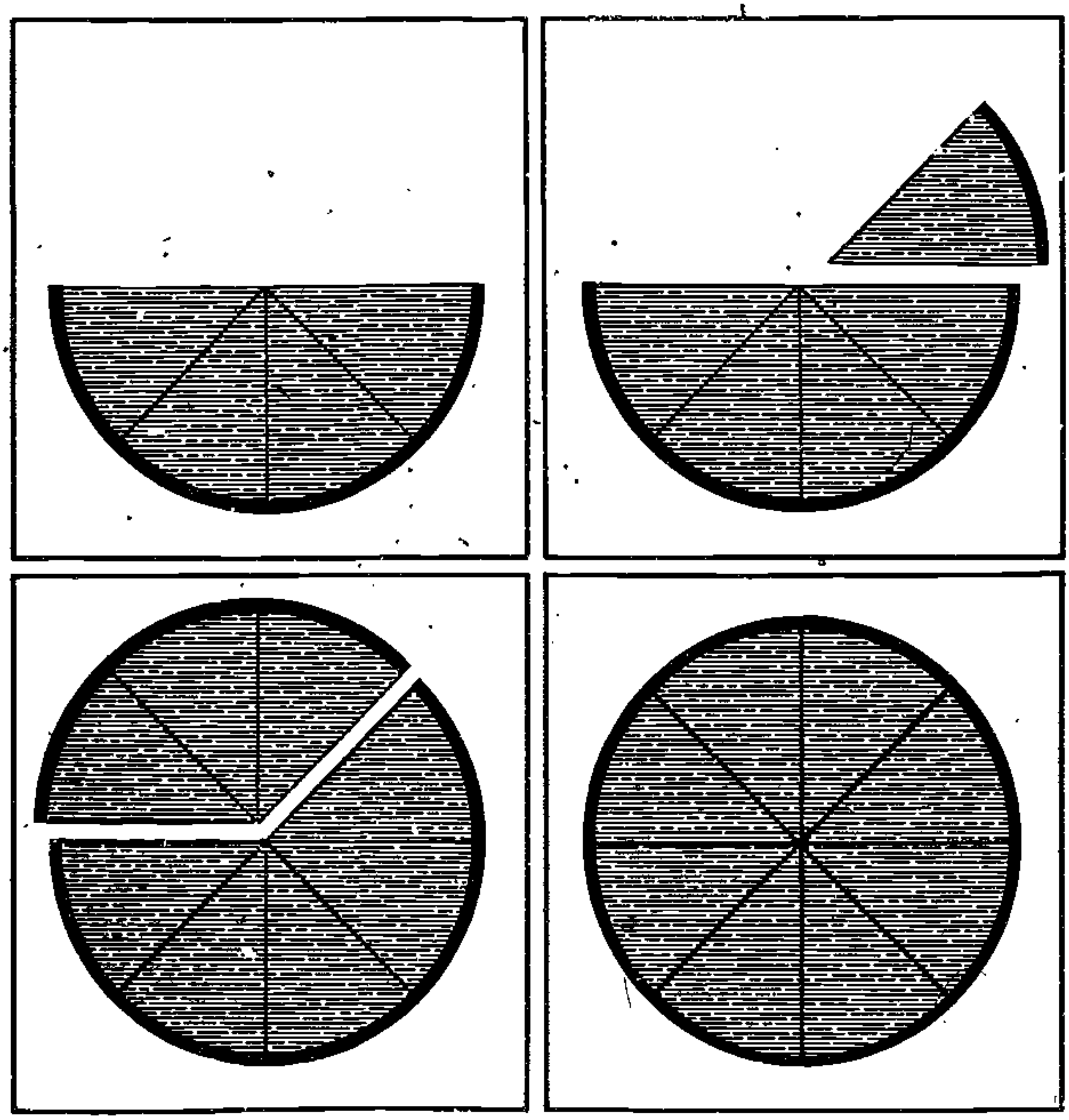
List types of data you require for your study:

NOTES:

Additional Readings

Easy: Barnes, p. 33, Dixon, p. 31
Hard: Walles and Roberts, p. 100

Step 4: Types of Sampling Procedures



Selecting and Describing the Samples

When teachers attempt to gather data on students, they are usually interested in making observations and then drawing conclusions from these observations. If teachers can observe all instances of a population they can, with a degree of confidence, base conclusions about the population on these observations. This is, usually the case, when an individual teacher with a single class of students is involved. In other situations where teachers want to observe only some instances of a population, they can do no more than infer that these observations are representative of the population as a whole. Basically, this is the concept of sampling, which involves taking a portion of the population, making observations on this smaller group and generalizing the findings to the large population. Sampling assists the problem solver in studying a portion (sample) of the population rather than the entire population. *The small group that is observed is called a sample and the larger group about which the generalization is made is called a population. A population comprises all the members of any well-defined class of people, events, or objects. A sample is simply a portion of the population.*

In order for your study to be of use to other teachers, you must also define the important characteristics of your subjects that may affect the experiment's outcome (see Classroom Data Profile Sheet in previous section). (Other teachers may be able to determine how successful your treatment will be for their students, by comparing the two groups of students on salient characteristics like reading habits, attendance patterns, whether the child has access to a library, and reading interests.) For instance, if your students are mostly white, middle class, children attending a rural public school, the treatment that may work successfully for them may not work for children in a affluent suburban private school, or an urban racially mixed public school. Other characteristics to be considered are suggested in the Classroom Data Profile Sheet shown in previous chapter.

How your subjects were selected is another important factor to note, as this will help determine the statistical analysis used to process the data. There are types of samples you can select: matched related samples, random independent samples.

An independent sample is present when the experimental and the control groups have been drawn randomly from the population. Every subject has an equal chance of being selected for the sample. When the two groups are drawn in this way it is assumed that they are essentially equal. Then, the experimental group is given the special treatment and the control group continues to receive regular treatment and any difference in the performance is assumed to be due to the difference in treatment. However, if by the nature of your study you cannot assume the two groups are equal, as is often the case in educational problem solving, then all subjects must receive the same treatment, and a related or matched sampling must be used

In matched sampling the two groups are divided individually into pairs that have equivalent characteristics that might affect performance. Each member of the pair is then assigned randomly to one group or the other, or the subjects are tested twice and each subject serves as his own control. When matching, it is necessary to make sure the pairs are matched over every possible relevant characteristic. Otherwise, you will get irrelevant or possibly insignificant results.

In many instances, a teacher may choose to try a new treatment with the whole class to see if this method produces greater gains than the class as a group made previously or greater than normally expected at their age and grade level. This type of study does not require the sampling procedure as outlined. On the other hand if a teacher wants to divide the class into two or more groups to test different methods, these sampling procedures will work quite well.

At times it is not feasible to acquire a group that matches the characteristics of the group receiving the treatment being used by the problem-solver. In those instances, it is possible to determine whether or not there are significant differences between groups receiving treatment and those receiving no treatment. If there are significant differences between the performance of groups receiving the treatment and the performance of groups not receiving treatments, it can be assumed that the treatment made the difference in the observed performance changes Most problems in reading require teachers to work with groups that are not matched. The procedure for determining the extent to which groups differ on a specific performance should be followed.

NOTES:

Personal Questions about selecting and describing samples

1

2.

3.

4.

Some Illustrative Examples of Selecting and Describing Samples:

Example One: A teacher is interested in dividing the class into three groups that are equivalent in performance on a common measure of reading. In order to select students for each of the three groups, a random method will be employed. Arrange all students in the class in some orderly sequence: alphabetically by surname, by birth date (youngest to oldest or the reverse), any other systematic arrangement. Then assign each student a number. Write the number corresponding to each student on a small piece of paper and place all the numbers in a box. Shake up the numbers thoroughly. Select a number without looking in the box. Record the first number selected under Group 1. Place the number in the box and shake before selecting another number. Record the second number selected to Group 2. Replace the number shake and select a third number for recording under Group 3. This procedure should be repeated until all students are assigned to one of the three groups. Once students have been assigned to one of the three groups, the teacher is ready to begin the treatment phase of the problem-solving process.

Example Two: Students can be divided into sample groups on the basis of sex. Males students are assigned to one group and female students are assigned to another group.

Example Three: Approximately half of the students in a classroom were given a specialized program in reading the previous year and the teacher is interested in determining if there has been a cumulative effect on these children's reading performance under certain conditions. All students who had the specialized program are placed in one sample group. The remaining members in the class are put in the other sample group.

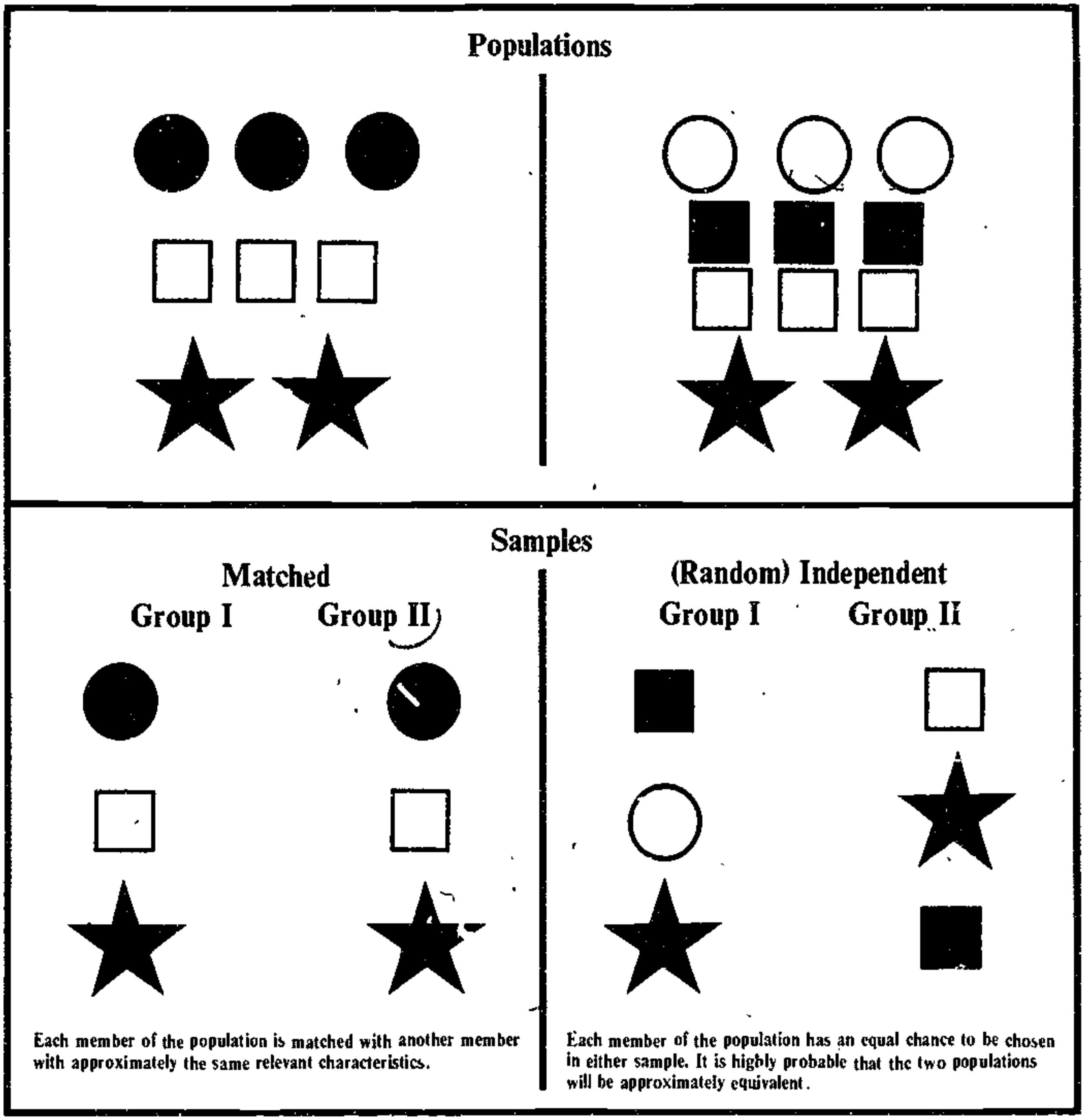
Example Four: In a school that has assigned students to different classrooms at the same grade level on a heterogeneous basis. For example, a fifth grade teacher might be interested in employing a different approach to vocabulary development that is different from the approach suggested in the regular instructional materials. In order to determine whether or not this different approach to vocabulary development will have lead to significant improvements in performance, two different fifth grade classes can be thought of as sample groups. Salient characteristics of each classroom should be described and a determination of how each group compare prior to the administration of a treatment to the experimental group.

Example of Sample Description—Data on Fifth Grade, Month of October

Students	Age in Mos	Sex	Race	Raw Score in Reading
1	125	F	W	87
2	132	F	B	73
3	127	F	W	48
4	134	M	B	68
5	126	M	B	39
6	129	F	B	89
7	130	M	W	63
8	135	F	W	78
9	129	M	B	59
10	136	M	W	85
	1303	5F, 5M	5B, 5W	689
Mean	130.3			68.9
Range	125–136			39–87

The average age of the group is 130 3 months with the youngest 125 months and the oldest being 136 months Five members of the sample are female and five members are male. Half of the sample is comprised of black students (2 females and 3 males) and half are white students (3 females and 2 males). The average raw score on a common reading test for members of the sample is 68 9 with a range of 39–87. The average raw score of 68.9 translates into a 5.4 reading grade level score. In general, this is a cooperative group of students who work up to their ability on most occasions There are few if any discipline problems and students are highly motivated to complete assigned class work. Reading interests of group members are varied and reflect expert patterns of students from similar backgrounds. These students come from average working class backgrounds with parents who are generally supportive of their children's school activities.

Matched vs. Independent (Random) Sampling



Activity 4:1

Checklist for Determining Types of Samples

Mrs. X is a fourth grade teacher who knows she will be ill for a month and a half. She decides she would like to know if her reading class will be affected negatively by her absence and with the presence of a substitute teacher for that time. She decides to measure her group against a group comprised of other fourth grade reading students, from a variety of different schools, but each of whom will have his/her regular teacher. The students were all nine years old and of the same race.

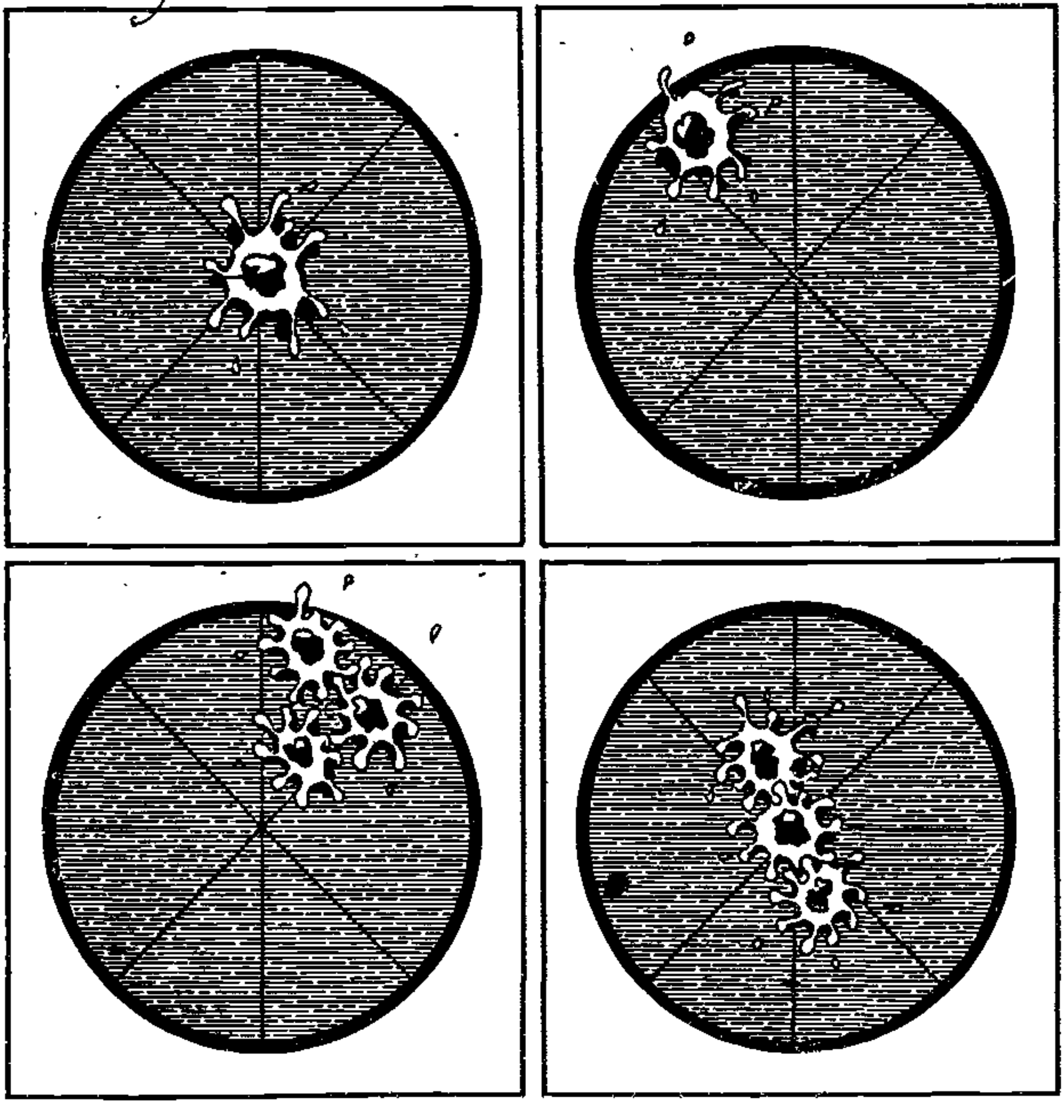
Yes	No	N.A.	
______	______	______	1. Are the two groups equal?
______	______	______	2. Is it an independent sample?
______	______	______	3. Are there other effects besides the absence which might affect performance?
______	______	______	4. Is there a control group?
______	______	______	5. Is it a matched sample?
______	______	______	6. Is there a special treatment?
______	______	______	7. Is a pretest necessary?
______	______	______	8. Is a post-test necessary?

Discuss this example in a small group, and then consider your own study for consideration.

Additional Readings

Easy: Dixon, p. 32; Barnes, p. 33; Ferguson, p. 112
Hard: Wallis, p. 100; Galfo and Miller, p. 25

Step 5: Selecting the Study Design



Selecting the Study Design

When selecting the design, the main consideration is the type of question asked. The question may involve a simple companson between two measurements, i.e., test scores. It may, also deal, with the effects a treatment may have when applied to one/two/three groups. In this case, a pretest and post-test will be necessary to examine, precisely, the effects of the special treatment. The following summary tabulates eleven representative study designs. A more complete list of basic statistical models occurs in Chapter 7 after a discussion of nominal, ordinal, and interval level measurement, which is a necessary determinant for using the correct model.

Study Designs: Statistical Models

1. One Group—Treatment—Post-test

E ⟶ Treatment ⟶ Post-test

Assumption: neither pretests nor uses a control goup Introduces a treatment and evaluates the post-test, attributing any assumed change to the treatment. Generally a poor model

Example: a teacher instructs a class on fractions, tests for understanding, and evaluates to some assumed standard

Appropriate statistic: normally. none.

2. Two Groups—Treatment—Post-test

E1 ⟶ Treatment ⟶ Post-test
E2 ⟶ Treatment ⟶ Post-test

Assumption: two groups assumed to be equal before exposure to the same treatment. Tests for difference between the two groups on the post-test. Attributes differences to the ability of the groups to relate to the task.

Example: a teacher assumes boys and girls do equally well on math tests, teaches a math lesson using only sports examples and then tests to see if boys did significantly better than girls.

Appropriate statistic: Binomial Test or Chi Square for one sample (*Equations, p. 53. Sample Studies, p. 58*)

3. Two Groups—Two Treatments—Post-tests

E1 ⟶ Treatment 1 ⟶ Post-test
E2 ⟶ Treatment 2 ⟶ Post-test

Assumption: the two groups are assumed to be equal before the treatments. Treatment 2 may either be a different treatment from 1 or *no treatment at all*. If no treatment is given to one group, it is called the Control Group (C), as opposed to the Expenmental Group (E).

Example: a teacher may want to take two groups of children of equal ability to test the effectiveness of two separate methods of teaching a particular reading skill One group is taught by method A, the other by method B. Both groups receive the same post-test after the treatment. Any difference is *assumed* attributable to the difference in the effectiveness of one method over the other Or the teacher may select groups of assumed equal ability and teach only one a new learning task and post-test to see if the new method produced a significant difference in performance between the two groups.

Appropriate statistic: Binomial Test (*Equation. p. 53. Sample Study. p. 58*).

4. One Group—Pretest/Post-test

E ⟶ Pretest ⟶ Treatment ⟶ Post-test

Assumption: the only change between the pre- and post-test is the treatment. Thus any difference between the pre- and post-test scores is attributed to the treatment. Note that this model only differs from Model 1 in gathering pretest information for comparison with the post-test. It is the availability of two sets of scores for companson which makes this model stronger. Its results can be tested.

Example: a teacher may want to know if removal of fear of failure would increase spelling grades. The pretest would be a spelling test under normal conditions, the treatment would be an announcement that anyone who tries will receive a passing grade and the post-test would be another spelling test after removal of treatment. Any difference in performance would be attributed to the change in test conditions *assuming* the test to be exactly the same or equally difficult.

Appropriate statistic: McNemar Test for Significance of Change (*Equation, p. 53*).

5. Two Groups—Pretest/Post-test

E1 ⟶ Pretest ⟶ Treatment ⟶ Post-test
E2 ⟶ Pretest ⟶ Treatment ⟶ Post-test

Assumption: two groups are assumed equal in the pretest, but different on the post-test following the same treatment. Any difference in performance would be attributed to the difference in the ability of the groups to relate to the task. Note that this model only differs from Model 2 because it includes a pretest. The presence of a common pretest score for the two groups allows the problem solver to test whether the assumption of the two groups being equal is valid. It is a stronger model for this reason because it operates on one less assumption.

Example: same as Model 2 except that a teacher pretests the boys and girls to establish their equal ability before the treatment rather than assume it.

Appropriate statistic: Binomial or Chi Square for one sample (*Equations, p. 53. Sample Studies, p. 58*).

6. Two Groups—Two Treatments—Pretest/Post-Test

E1 → Pretest → Treatment 1 → Post-test
E2 → Pretest → Treatment 2 → Post-test

E → Pretest → Treatment → Post-test
C → Pretest → No Treatment → Post-test

Assumption: the two groups are assumed to be equal in the pretest, but different in the post-test after different treatments. The difference would be attributable to the differing treatments. Note that Treatment 2 may be *no treatment at all*. No treatment groups are called *Control Groups*.

Example: same as Model 3 except that a teacher would pretest the two groups of children to establish equal ability instead of assuming equal ability before the treatment. Thus, it is a stronger model than Model 3 because it operates on one less assumption.

Appropriate statistic: Mann Whitney *U* Test or *T* Test (*Equation, p. 55, Sample Study, p. 63*)

7. Classification of Relationships

Reading Ability

	Low	Middle	High
Male			
Female			

Assumption: the distribution of boys and girls among low, middle, and high readers, if reading ability were divided into three sets, would be either relatively equal or significantly different.

Example: the teacher believes that girls generally read better than boys. She divides reading scores into low, middle, and high and notes the number of boys and girls in the higher classifications would be greater. In other words, there should be a significant difference in the distribution of girls and boys among high, middle, and low ability readers. The significance of the difference in distributions can then be tested. An alternative would be to assume no difference in the relative distribution of boys and girls among reading ability groups and then test.

Appropriate statistic: Chi Square (*Equation, p. 55, Sample Study, p. 58*).

8. Relationship Between Two Variables For One Group

Subject #	(X) Reading	Rank	(Y) Math	Rank
1	70	5	84	1
2	75	4	83	2
3	84	1	73	5
4	80	2.5	79	3.5
4	80	2.5	79	3.5
"	"		"	
N =	XN		YN	

Assumption: reading and math tests are independent.

Example: the teacher wants to know whether reading and math performance are predictive of each other, that is, whether there is a significant relationship between reading and math scores for her his students. (S)he then lists a reading and math score for each subject, ranks the scores and the total number of students, and uses a formula to test the relationship.

Appropriate statistic: Spearman Rank Correlation (*Equation, p. 56, Sample Study, p. 64*).

9. Difference Between Relationships

Boys Reading	(R)	Boys Spelling	(R)	Girls Reading	(R)	Girls Spelling	(R)
85	1.5	82	3	80	3	85	1
70	5	59	6	86	1	79	3
63	6	61	5	73	5	68	6
85	1.5	75	4	62	6	73	5
75	3.5	85	1.5	80	3	79	3
75	3.5	85	1.5	80	3	79	3
"		"		"		"	

Boys: Reading, Spelling → Correlation; Girls: Reading, Spelling → Correlation (*Spearman Rank Correlation)

Correlation → Difference of Correlations ← Correlation

Assumption: there either is or is not a relationship between the relationships of reading and spelling for boys and girls. Tests for a significant relationship across groups.

Example: the teacher wants to know if the relationship between reading and spelling performance is the same for boys and girls. In other words, do girls and boys differ in the degree to which scores on one test predict the other. The relationship between reading and spelling is calculated for each group and the relationship between the correlations is tested.

Appropriate statistic: $z = \frac{zr_1 - zr_2}{1/(N_1-3) + 1/(N_2-3)}$

Significance of the difference between two correlation coefficients.

10. More Than Two Groups On One Variable.

Reading Scores

Group	A	B	C	D	Etc.
	40	56	51	60	
	65	49	61	63	
	55	49	61	63	
	57	46	58	51	

Assumption: the groups are not different. thus the test is for differences between group means.

Example: the teacher designates three reading groups based on *assumed* differences in reading ability. This model allows the teacher to actually test whether assumed differences are real. based on scores from a reading test.

Appropriate statistic: Kruskal-Wallis One Way Analysis of Variance (*Equation, p. 56, Sample Study, p. 62*).

11. Relationship Among Three or More Variables for One Group

Subject #	Self-esteem	Fear of Failure	Achievement
1	60	70	60
2	82	80	75
3	70	75	65
4	75	70	75
5	65	65	65
N =	XN	YN	ZN

Assumption: self-esteem, fear of failure, and achievement are all independent (have no relationship with one another).

Example: a teacher wants to find out if self-esteem, fear of failure, and achievement are predictive of one another or if there is a significantly strong relationship among these three variables.

Appropriate statistic: Kendall Coefficient of Concordance: W (*Equation, p. 57 Sample Study, p. 61*).

To assist in selecting the sample/design, the following table compares a significant difference study and a correlational study. A correlational study basically seeks to find a relationship between two ideas, concepts, or topics. A significant difference study seeks to find differences between two measurements; i.e., performances and scores. By using the comparison, one can answer questions about sample, design, and treatment using the checklist in Activity 5:1.

Personal Questions about Selecting the Study Design:

1.

2.

3.

4.

5.

NOTES:

Activity 5:1

Format of Problem-Solving Procedure and Design Selection

Example A: Correlational Study

Problem
The relationship of motivation and achievement

Hypothesis
There is a relationship between motivation as measured by the Brown-Holtzman Survey of Study Habits and Attitudes and Achievement Tests.

Null Hypothesis
There is no relationship between motivation and achievement.

Procedure
Give all the students both tests at approximately the same time under the same set of circumstances

Results
The scores obtained on the achievement test and the motivation test are ranked for each subject and the two scores are compared to see if the subject scores high on one test whether he will score high on the other test also. Compare the two scores using a correlational technique.

Conclusions
If a strong relationship is established, then we have obtained support for our hypothesis. We may reject the Null Hypothesis. We then find there is a strong relationship between motivation and reading achievement. This could lead you to look for better ways to motivate your students. If you found that there wasn t a strong relationship between the teaching techniques and performance or days tardy and performance then problem-solving process will begin again.

Example B: Significant Differences Study

Problem
Is oral or written presentation of an instruction more effective for my class?

Hypothesis
The subjects will perform better on a test if oral rather than written instructions are given

Null Hypothesis
There will be no difference (significant) in the performance of different groups on the two tests

Procedure
Two tests on the same topic of comparable difficulty will be given to students within a week to a month apart One will be given to half the class orally; one will be given with written instructions to the other half of the class The other test will be given in reverse order, the half receiving oral instructions the first time will receive written instruction this time and vice-versa

Results
The difference between scores and which set of scores was higher will be found and the average difference calculated using the Wilcoxon-Sign rank test

Conclusion
If the performance on the orally presented test is significantly better than the performance on the written instructions test, then we reject the Null Hypothesis and find support for the hypothesis that orally presented instructions are more effective than written instructions. Based on that data, you would present your instructions orally, and/or train your students to handle written instructions better. If there were no difference, this would indicate both modes should be used or that the instruction mode has no important effect on the results.

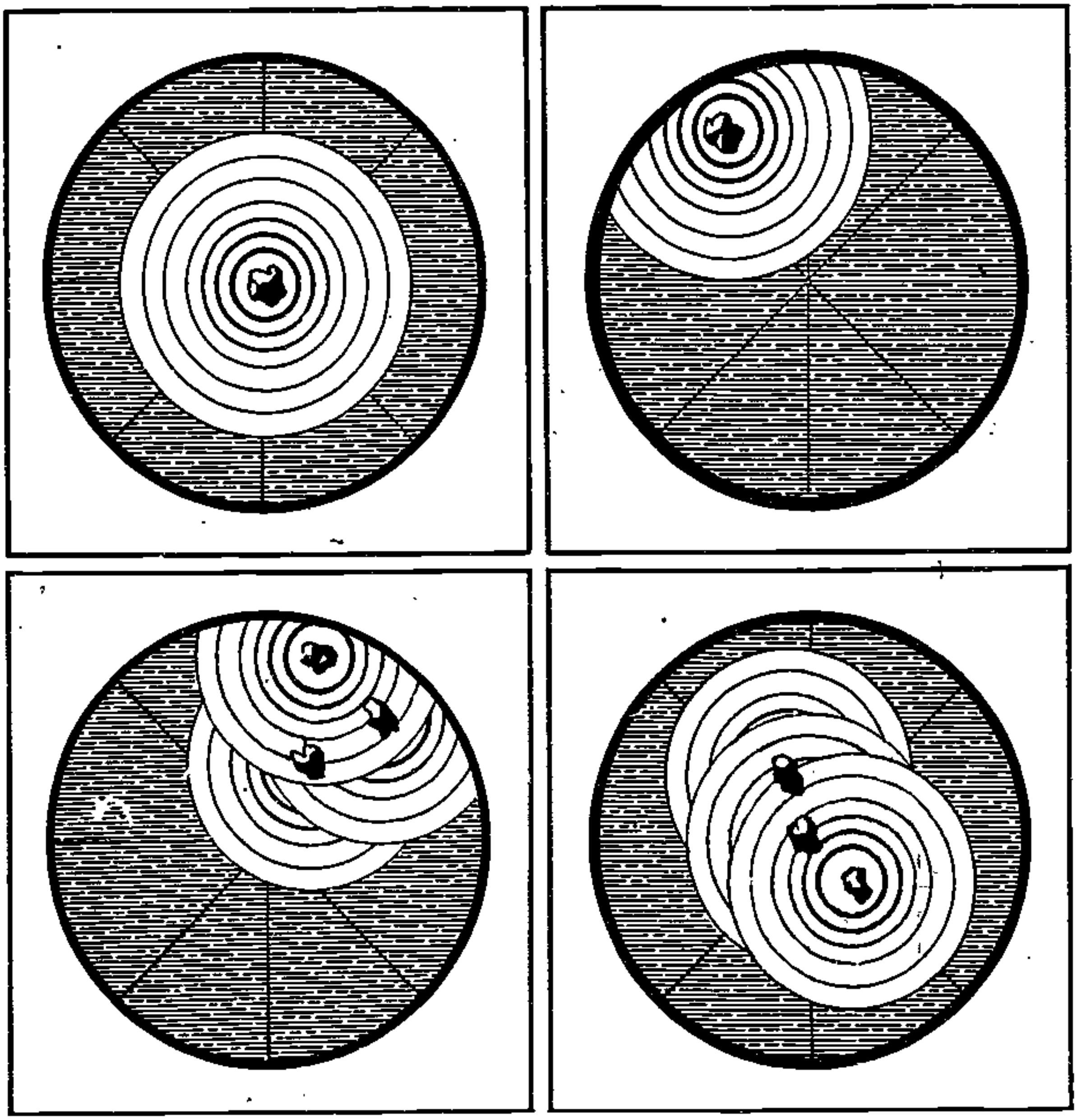
Criterion Checklist for Format of Problem-Solving Design/Procedure

Example A			Example B			Your Own Study			
Yes	No	N.A.	Yes	No	N.A.	Yes	No	N.A.	
___	___	___	___	___	___	___	___	___	1. Am I trying to find if there is a relationship between any two or more traits, scores, or behaviors?
___	___	___	___	___	___	___	___	___	2. Am I trying to find if there is a significant difference between two sets of scores for one group of subjects?
___	___	___	___	___	___	___	___	___	3. Am I comparing the performances of two groups of subjects?
___	___	___	___	___	___	___	___	___	4. Did I test the two groups of subjects before the treatment began (i.e. pretest)?
___	___	___	___	___	___	___	___	___	5. Am I trying to determine whether one group of subjects fall into a category more often than another group?
___	___	___	___	___	___	___	___	___	6. Am I comparing more than two groups of subjects?

Additional Readings

Easy: Dixon, p. 35; Barnes, p. 52
Hard: Wallis and Roberts, p. 211; Galfo and Miller, p. 153

Step 6: Procedures



Procedures

The procedures of a study includes time and circumstances of all activities involved in the problem-solving process. This section outlines what data are to be collected and recorded, how data are to be collected, the nature and extent of the treatments to be employed, the instructional techniques and materials required, and the time required to complete all activities. A description of any treatment, intervention approach or material used in the study is given in the procedures section.

In the procedure section the problem-solver should list all activities in sequencial order with as much detail as necessary to guide the study process in the direction intended. This is the point where the problem-solver should specify the specific details of what should be done, the nature of the organization, operationalize the schedule of events, determine data requirements for collection, recording, analysis, and reporting. Listing all procedures will lend credibility to your results. It will also assist teachers who are interested in replicating your study or at the very least make use of your treatment in their classroom instruction.

The real value of the procedures stage is in the organization and evaluation of plans for conducting a given study. It helps the problem-solver to perfect the formation of ideas and to help localize and pinpoint the problem that is to be studied. Because a careful articulation of procedures provides assistance in improving the formation of ideas that refine and further define the problem to be solved. It is our assumption that each improvement in idea formation and problem definition reveals new insights and observations that often yields new data to improve judgment and instructional practice. Given this important contribution, the problem-solving process that a well-reasoned statement of procedures can make, this section should be developed with care.

Some Illustrative Examples of Procedures

Example One: In order to determine how the nature of the time spent by students affects their measured reading performance, the class will be divided into three groups that are equated with regard to present measured reading performance. Each group will be taught for 30 minutes per day for ten consecutive school days The 30 minutes periods will be divided thusly:

			Total No. Minutes Spent	
Groups	No. of Minutes	Treatment	Within Treatment	Between Treatment
#1	5	Teachers demonstrate and explain lesson	50	
	20	Student working independently with instructional materials	200	300
	5	Student/teacher interaction	50	
#2	20	Teacher demonstrate and explain lesson	200	
	5	Student working independently with instructional materials	50	300
	5	Student/teacher interaction	50	
#3	5	Teacher demonstrate and explain lesson	50	
	5	Student working independently with instructional materials	50	300
	20	Student/teacher interaction	200	

Students working independently with instructional materials in Group #1 will receive the greatest amount of instructional time, students taught by teacher demonstration and explanation in Group #2 will receive the greatest amount of instructional time and students taught primarily through the use of Student/Teacher Interaction, Group #3 will receive the greatest amount of instructional time. At the end of the treatment for each group a post-test is administered to measure changes in measured, reading performance. A statistical test will be applied to determine if observed differences are significant.

Activity 6:1

Criterion Checklist for Problem-Solving Procedure

Below is an example of a list of procedures. Write all information and determine if all needed information is listed in the following activity. Examples of steps taken in implementing an experimental procedure

1. I compiled all the data for the CDPS chart.
2. I administered the Ginn Test on October 1, 1977 in the afternoon during reading class to all students
3. I then implemented my individualized reading programs by evaluating Ginn results and determining the appropriate level for each child. Then I compiled a work packet for each child and distributed them. Every Friday I had a conference with each individual student to assess his progress and answer any questions.
4. May 1, I administered the Ginn in the same way as I did on October 1.
5. I compared scores received in October and May to find if there was significant improvement

Pretest

What was the date of the pretest? ____________________
What time of day was the pretest administered? ____________________
Was it administered to the whole class? ____________________
Was it administered to a small group? ____________________
Who administered the pretest? ____________________
How much time was used to administer the pretest? ____________________

Treatment

What were the specific activities of the treatment? ____________________
What preparation was made by the teacher for the treatment? ____________________
How long was the treatment period? ____________________
What materials are required for each treatment? ____________________

Post-test

What was the date of the post-test? ____________________
What time of day was the post-test administered? ____________________
Was it administered to the whole class? ____________________
Was it administered to a small group? ____________________
How much time was used to administer the post-test? ____________________

This is a more specific look at the problem-solving procedure. After completing the criterion checklist for the above example, apply the checklist to your own study and discuss in groups any problems.

Your Personal Questions about Procedures:

1.

2.

3.

4.

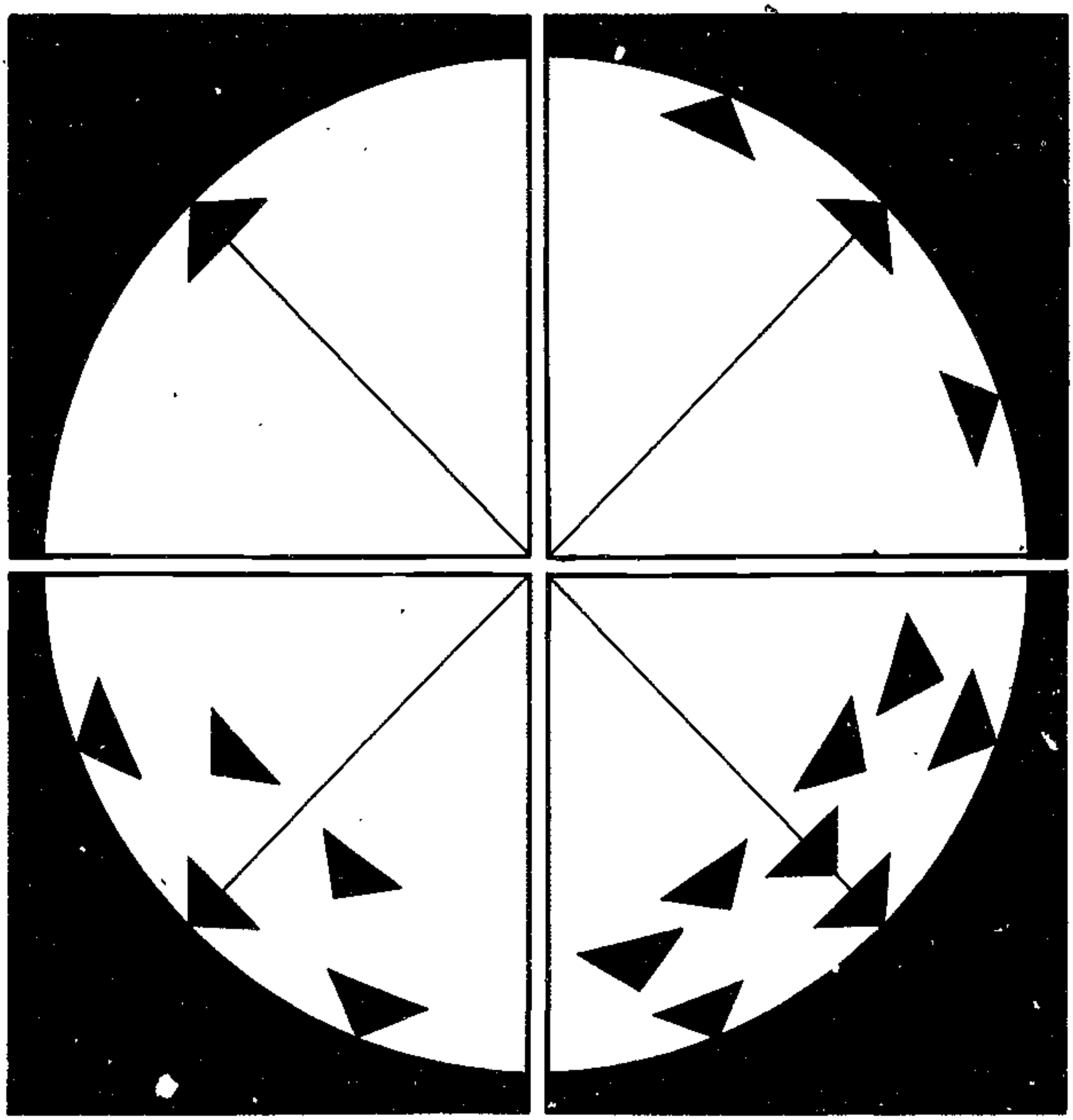
5.

NOTES.

Additional Readings

Hard: Wallis and Roberts, p. 142

Step 7:
Processing the Data



Processing the Data

When you process data for solving an instructional problem in reading, you basically will either want to know, "what is the relationship between these two traits?" or "is there a significant difference between two defined groups?" If you wish to find a relationship you will use a correlation technique like the Spearman Rho and others which are listed in the appendix. The type of correlation method you use will depend on the level of measurement of your data and whether you wish to find the relationship between two groups or among more than two groups. If you are looking for a significant difference, that is, a difference between two groups that is large enough that it probably cannot be reasonably attributed to chance, you will use the Mann-Whitney U or another test of this type listed in the appendix. Which test of significant difference that you will use depends on several factors: the type of sample used, the design of the study, the level of measurement of your data, whether your groups are randomly sampled or matched sampled, the number of subjects in each group and the number of groups tested. See the chart in Appendix I, p 52 summary of the types of statistical methods, types of questions answered by each method, the level of measurement, and the equation

Levels of Measurement

The level of measurement refers to how precisely the data are measured, and there are four levels. The least precise level of measurement is the nominal scale which is observations of unordered categories. In the ordinal scale the data can be ranked. The interval level of measurement requires that the difference in two consecutive values be the same as between any two other consecutive values. And the most precise measurement ratio requires an absolute zero so one scale value could be said to be twice as large as another value. Below are examples of each level of measurement.

Nominal level—boys, girls, Indian, Black, White and Oriental. These groups cannot be compared and ordered. All items or subjects are either in one group or the other. They cannot be in more than one group.

Ordinal level—Attitude Toward Reading Scale, Disruptive Behavior Rating. The values of the different data can be ranked from the highest to the lowest, but the differences between consecutive scores are not equivalent. In other words, if $A = 31$ and $B = 33$ and $C = 35$, you can say $A < B < C$, but cannot say $B - A = C - B$

Interval level—Some I Q tests, a spelling test, a vocabulary test. The difference between scores are equivalent so you can say $B - A = C - B$

Ratio Level—Yard stick. Very few things in the behavioral sciences can be measured so precisely. Using this level of measurement one can say $X = 2y$ or Abby is two times as tall as Ben.

In choosing the best technique for processing your data, you will want to choose the most powerful test. The most powerful test is the one that most often rejects the null hypothesis, there is no difference between groups due to the treatment when it is false.

Preparing the Data for Processing

Before inferential statistical analysis can be performed on the data, however, it must be organized and summarized. Descriptive statistics are used for this purpose. Descriptive statistics may indicate the data's central tendency or average, its variability or diversity, the location or place of its scores in relationship to one another and the correlation. The type of descriptive statistic used will vary depending on what information is needed and what level of measurement the data is. The following chart summarizes the types and specifies under what circumstances they are used.

In addition, most inferential statistics require that the raw scores be ranked. The ranking procedure is as follows:

1 Place all the first set of scores on a chart, arranged from highest score to lowest score, keeping all scores for one individual on the same line.

2 Number the scores, the highest score being numbered 1, the lowest the number equal to the number of scores.

3 If there are any ties, they are all assigned the rank which represents the average of the number of ranks the score would occupy.

Example

Raw Scores		
38	1 5 = $\frac{(1+2)}{2}$	*ties are assigned ranks by averaging the number of ranks the scores would occupy.*
38	1 5	
35	3	
30	5 $\frac{(4+5+6)}{3}$	
30	5	
30	5	*The last rank will always equal the number of scores*
28		

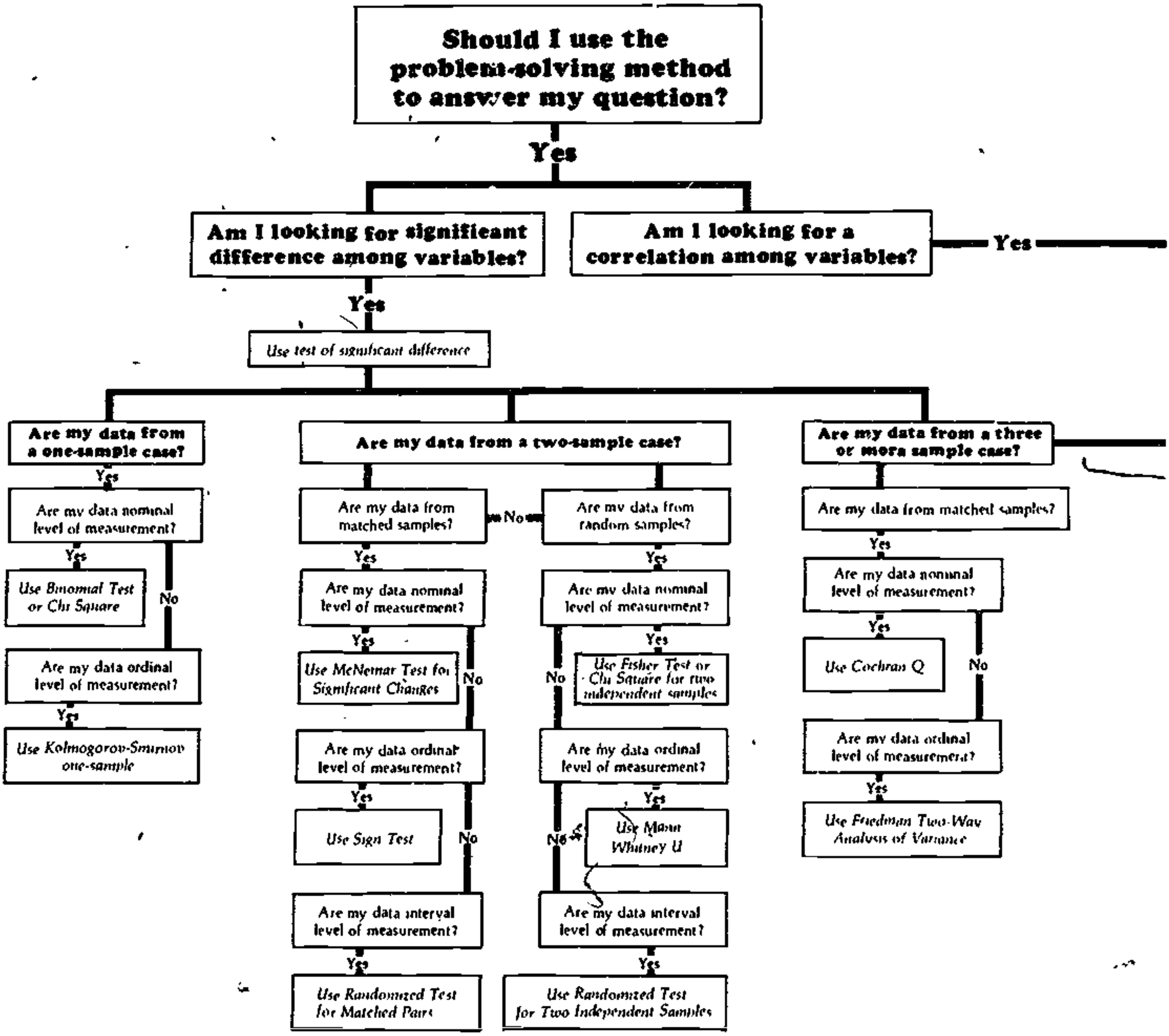
Types of Measures	Levels of Measurements		
	Nominal	Ordinal	Interval
Indices of central tendency	Mode (most freqoent score)	Median (middle score)	Mean (average mid valoe)
Indices of variability	Range (no. of scores between highest and lowest)	Qoartile deviation (no of scores between the bottom ¼ and top ¼	Variance and Standard deviation (the average difference of a score from the mean?)
Indices of location	Label or classification (categories)	Percentile rank (percentage of scores = to or less than that score)	Z-score, Z-score and other standard scores (distance a score from the mean)
Correlation		Spearman Rank (degree of relationship between scores)	Pearson r (closeness of agreement of two samples)

Your Personal Questions about processing the data:

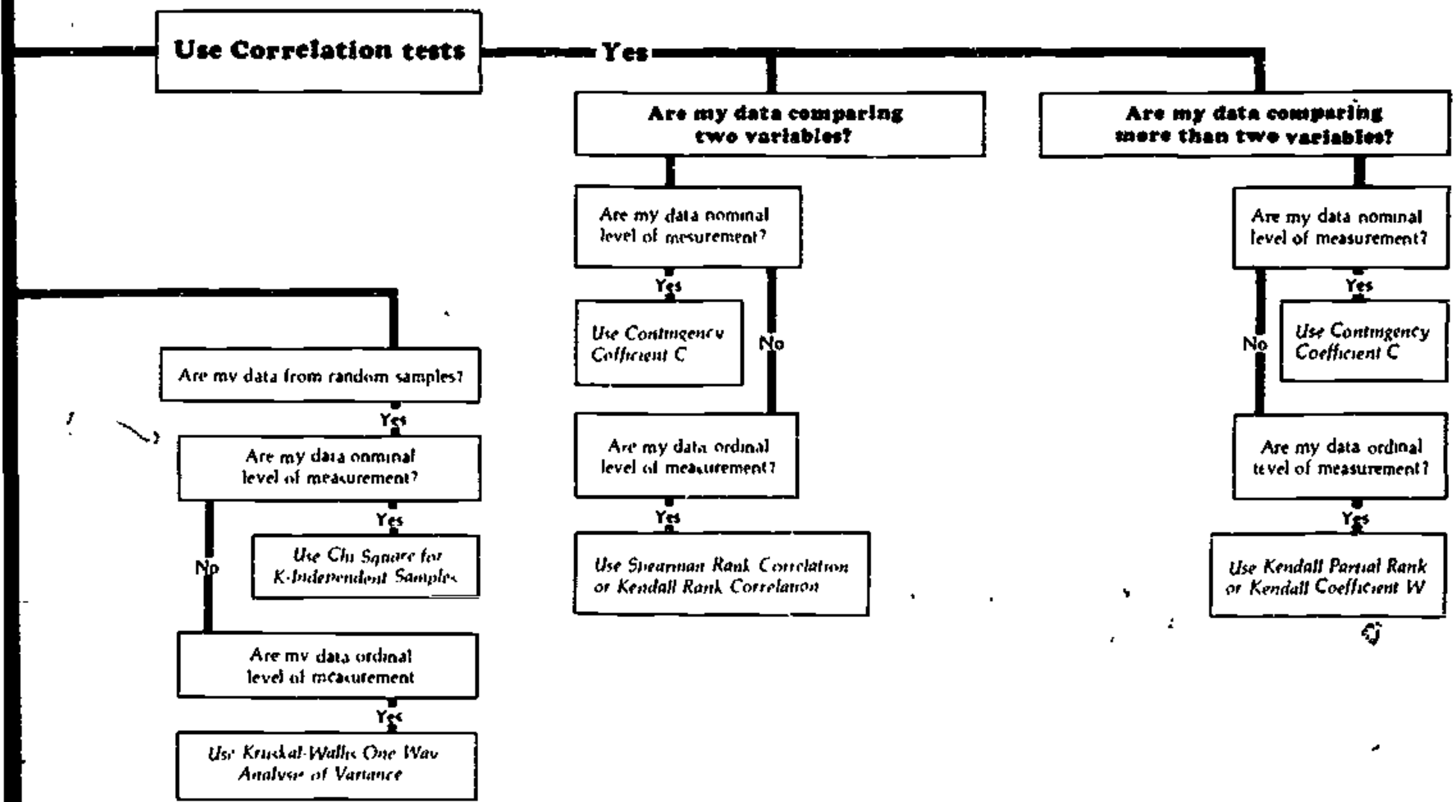
1.
2.
3.
4.
5.

NOTES:

[See Sample Studies for Illustrative Examples on pages 58–65]


Should I use the problem-solving method to answer my question?
Yes
Am I looking for significant difference among variables?
Am I looking for a correlation among variables?
Yes
Yes
Use test of significant difference
Are my data from a one-sample case?
Are my data from a two-sample case?
Are my data from a three or more sample case?
Yes
Are my data nominal level of measurement?
Yes
Use Binomial Test or Chi Square
No
Are my data ordinal level of measurement?
Yes
Use Kolmogorov-Smirnov one-sample
Are my data from matched samples?
No
Are my data from random samples?
Yes
Are my data nominal level of measurement?
Yes
Use McNemar Test for Significant Changes
No
Are my data ordinal level of measurement?
Yes
Use Sign Test
No
Are my data interval level of measurement?
Yes
Use Randomized Test for Matched Pairs
Yes
Are my data nominal level of measurement?
Yes
Use Fisher Test or Chi Square for two independent samples
No
Are my data ordinal level of measurement?
Yes
Use Mann Whitney U
No
Are my data interval level of measurement?
Yes
Use Randomized Test for Two Independent Samples
Are my data from matched samples?
Yes
Are my data nominal level of measurement?
Yes
Use Cochran Q
No
Are my data ordinal level of measurement?
Yes
Use Friedman Two-Way Analysis of Variance

Correlation Tests



Parametric versus Nonparametric Test

A parameter is a characteristic element of a population. Some examples of the parameters in which we are interested are the distribution or how the number of members of the population are distributed in relation to one another, the level of measurement or how precisely measured the population is, and the variance or dispersion of the population of scores.

In order to use a parametric test certain conditions about these parameters of the population must be met. Since these conditions are not ordinarily tested they are assumed to hold. The meaningfulness of the results of the parametric test are as follows:

1. The population of scores must be interval level of measurement.
2. The population must have normal distribution. A normal distribution is represented by a bell shaped curve  in which few scores fall at either extreme of the distribution, with a majority falling in the central portion of the distribution.
3. All samples must come from either the same population or populations with the same distribution.
4. All the scores tested must come from either the same population or populations with the same variance or dispersion of scores.

A nonparametric test on the other hand does not specify these conditions for the parameters of the population from which the sample is drawn. Although one must assume when using most nonparametric tests that the observations (scores) are independent from one another and that the variables being measured have underlying continuity, these assumptions are few and weak when compared to the parametric test. The nonparametric test can also be used with ordinal and sometimes nominal level data thus not requiring such precise measurement (Siegal, p. 31).

The parametric test is the more powerful test (see the following chart). That is, it will reject the null hypothesis when it is false more often than a nonparametric test. The reason for this is that because you make more assumptions about the parameters of the population in using the test so that you have more information on which to base your decision. There is less a chance of making a wrong decision, and so you are free to reject the null hypothesis with a smaller significant difference between the experimental and control measures. However, if you use the parametric test without making all the assumptions needed, then your results will be meaningless.

The nonparametric test, although not as powerful, is useful when one cannot make the assumptions necessary for using parametric tests. This is often the case when working with small groups in a classroom situation. Often the population of scores are only ordinal or nominal level of data and this too requires the use of nonparametric rather than parametric test. In addition, nonparametric methods are more understandable and usually much easier to apply in the classroom (Barnes, pp. 75–78).

For the reasons stated in the preceding paragraph, most of the tests found in this manual are nonparametric.

Choose the most powerful method that is possible, given your study design and the level of measurement of your data. The characteristics of each test, level of measurement required, type of question answered, sample problem and equation are listed on the chart on the next page.

If your results are significant at the .05 level, then you find support for your hypothesis and may reject the null hypothesis. This means that there is only a 5 percent chance that your results were due to error or chance.

Assumptions of Parametric and Nonparametric Test Compared

Parametric tests.	Nonparametric tests.
Interval level data.	Nominal, ordinal or interval level data depending on the test.
Normal distribution of the population.	No particular shape of the distribution needed.
All samples must come from the same population or population with the same shape distribution	Samples may come from differently shaped distribution

Matching the Reading Questions with Appropriate Statistical Model and Methods

Q. Is there a relationship between reading achievement and self-concept?
H. There will be a significant relationship between reading achievement as measured by the SRA Reading Achievement test and self-concept as measured by the Coopersmith Self-Concept Scale
Model: Model 8 (Chapter 5) is chosen to find the relationship between two scores for the same individual.
Method: The Spearman Rank Method (Appendix 1,2) is used because the scores (data) will be measured at the ordinal level of measurement and you want to find the relationship between these two scores.

Q. Is there a relationship among reading achievement, self-concept and fear of failure?
H. There will be a significant relationship among reading achievement as measured by the Metropolitan Primary Reading Achievement test, self-concept as measured by Bookover Academic Self-esteem test and fear of failure as measured by the Fear of Failure test.
Model: Model 11 (Chapter 5) is chosen to find the relationship among three or more scores.
Method: Kendall's Coefficient of Concordance (Appendix 1,2) is used because the data is ordinal level and you are comparing three or more scores.

Q. Do children with dominance problems do worse in reading than children with normal dominance?
H. Children with mixed-dominance as assessed on the Peripheral Dominance test will have lower scores on the Sloss Oral Reading test than children with lateral dominance.
Model: Model 3 (Chapter 5) is used because you are trying to determine if there is a significant difference between two groups which were not matched with only a post-test score
Method: X^2 (chi-square) for two independent samples (Appendix 1,2) is used because your sample was selected by independent (or random) selection and you are trying to find if there is significant difference between two groups' scores.

Q. Do my students learn more using an individualized spelling program (EX: Pretest—Treatment—Post-test)?
H. The students using an individual spelling program will improve their spelling scores significantly in a six-month period.
Model: Model 4 (Chapter 5) is used to determine if there is a significant difference in your students' scores.
Method: The Wilcoxon matched-pairs, sign-ranks test (Appendix 1,2) is the best statistical method for ordinal level data for related samples. (The scores compared are both from the same student so your sample would be considered a matched sample)

Q. Will the higher-ability students or the lower-ability students benefit more from participating in an Individualized Reading Program?
H. Students of lower ability will improve their performance on the SRA reading comprehension test significantly more than the higher-ability group.
Model: Model 5 (Chapter 5) is used to determine if there is a significant difference in performance between two different groups receiving the same treatment
Method: The Kolmogorov-Smirnov 2-sample test (Appendix 1, 2) is used because the level of measurement is ordinal and you are trying to determine if the score improvement of the lower-ability group is significantly more than the higher-ability group.

Q. Is an individual reading program more effective than a traditional group reading program?
H. There will be significantly more improvement in reading comprehension, as measured by the Stanford Diagnostic Test when an individualized reading program is used than when a traditional group reading program is used.
Model: Model 6 (Chapter 5) was used to test two groups receiving different treatments to find the significant difference.
Method: The Mann-Whitney U test (Appendix 1, 2) is used for ordinal level data to find significant difference between two groups' scores.

Your Personal Questions about matching Reading questions with appropriate Statistical Methods.

1.
2.
3.
4
5

Activity 7:1

Processing the Data Checklist

Yes	No	N.A.	
______	______	______	Have I ranked all scores?
______	______	______	Have I determined the level of measurement of my data?
______	______	______	Nominal
______	______	______	Ordinal
______	______	______	Interval
______	______	______	Ratio
______	______	______	Are my groups of subjects related (matched)?
______	______	______	Are my groups of subjects independent (randomly selected)?
______	______	______	Check the design of my study? (see Statistical Model chart)
______	______	______	Determined what statistical analysis I should use given my study design, level of measurement, and type of groups—related or unrelated, and question asked? (see charts)
______	______	______	Calculate my results

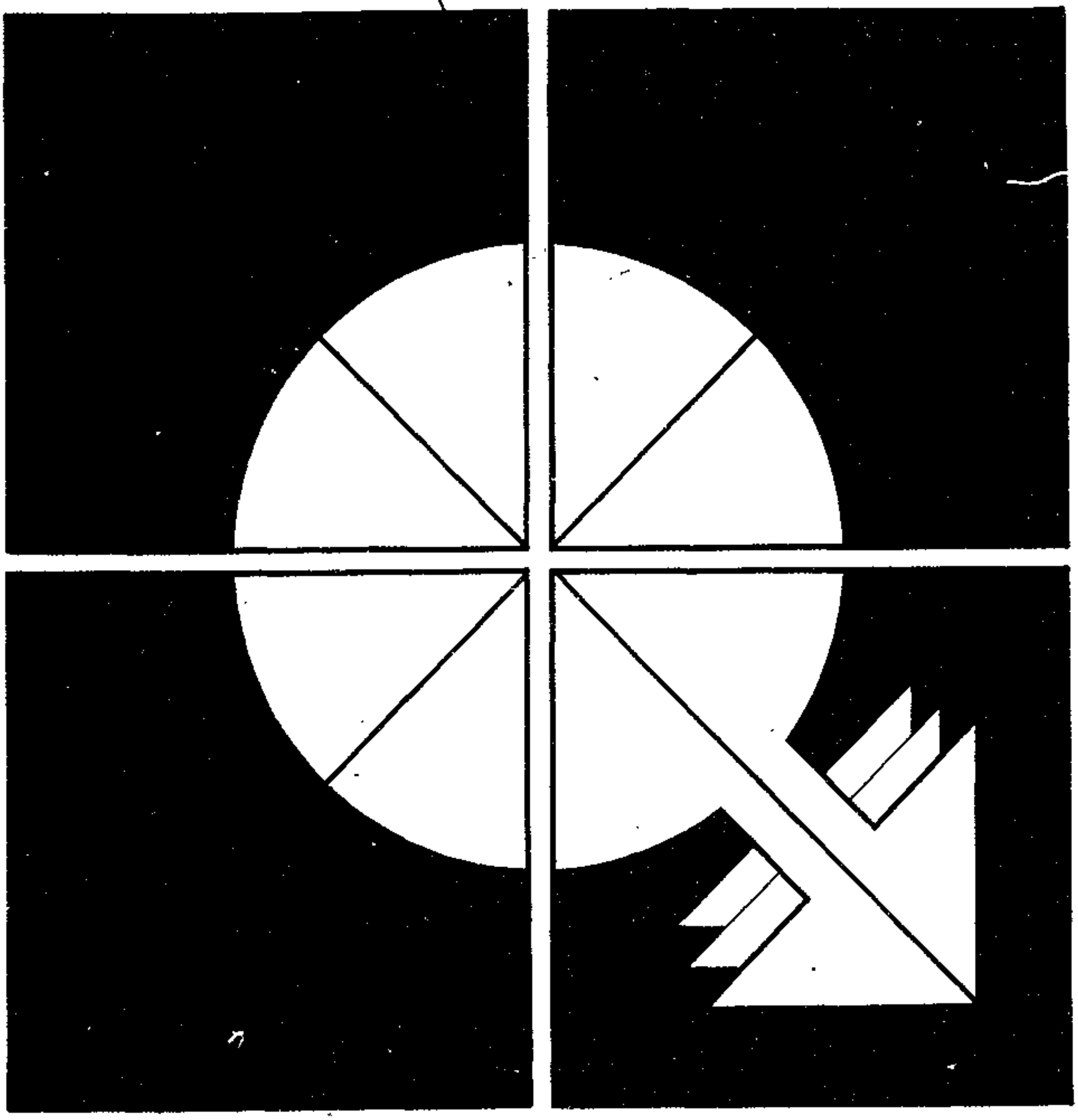
NOTES:

Additional Readings

Easy: Dixon, p. 70, 89, 93, 299; Barnes, p. 46, 75, 78, 85; Runyon, p. 3, 37, 63, 109, 124, 207, 307, 339; Ferguson, p. 1, 131, 179, 264, 275.

Hard: Siegal, p. 7, 20, 21, 30; McNeinar, p. 14, 16, 19, 39; Hayes, p. 81, 215; Andrews et al., p. 3; Vallis, p. 213, 384; Galfo, p. 103; Slakter, p. 255, 265, 382.

Step 8: Action Alternatives



Action Alternatives

One of the expected outcomes of the ICPS model is to assist the teacher in making decisions based on available or intentionally collected information. The completion of a study should have guided a teacher over the collection, organization, analysis and interpretation of information related to the problem to be solved. The only task remaining is to select among a varied array of action alternatives. In short, how are the results to be used to guide instructional activity?

It should be recognized that the utilization of information is, in the final judgment, a personal process Individual teachers acquire data in relation to their own needs, conceptions and perceptions. Since the generation of information is to a large extent a personal process, teachers as data users differ in terms of any variable one would care to suggest. Therefore, the personal utilization process is enhanced by watching the teachers' perspectives as problem-solvers to those action requirements guiding instructional decision making in the classroom. Taking into account the personal nature of data utilization, there are four general action alternatives that a teacher might consider at the conclusion of a problem-solving procedure.

Conduct additional study on the problem or related aspects of the problem.

If the results of the study are insignificant or reveal significance in relationship between variables, the teacher may be interested in doing additional study, exploring related questions, or generating additional hypotheses for the same questions. The teacher may want to change the type of study from one of studying the relationship between variables to one of finding a significant difference between them. Additional study might also be directed toward delineating the source or sources of the problem, defining and identifying the variables more precisely, or establishing the specific interaction effects between selected variables. The results of a problem-solving experience should increase the probability of teachers seeking additional answers to significant answers to critical instructional problems.

Review related research papers or articles on the problem being studied.

Often the results of problem-solving help the teacher to realize how complex a problem might be and how little (s)he knows about the problem. At that point, the teacher might turn to others who have been working on similar problems. Professional journals, information systems, libraries and resource people can be used to expand the knowledge of the teacher on selected topics related to the problems of immediate interest. It is possible that the problem solver might find additional evidence that might support or refute her/his results or suggest that other approaches to solving the problem might be more productive and useful. Reviews of related studies provide the teacher with additional knowledge to isolate and solve instructional problems more effectively.

Study different instructional approaches.

Along these same lines, the teacher sometimes finds reading in general or subject matter curriculum books that (s)he may glean some knowledge as to the various techniques involved in dealing with the problem. The teacher may find different types of methodologies or procedures that can be adapted to the requirements of teachers working under different circumstances. This alternative could help the teacher generate other hypotheses and pursue new teaching strategies and improve general ability to teach.

Share your findings with others.

The fourth type of action alternative is to disseminate the findings to others. This may include sharing information about the successful strategies for dealing with selected problems. The problem-solver might accomplish this by developing materials, attending professional meetings, talking with co-workers, or writing articles. In a more general way, the teacher can also put forth the problem-solving model as an effective way of matching problems with solutions or even show a more efficient modification of this model. Both types of dissemination require that the teacher present information in a format that is comprehensible and concise.

Instructional Changes Resulting From the Study

The fifth type of action alternative is to implement specific instructional changes This might involve refinement of instructional materials, and techniques of presentation As the problem-solver records the results of his instructional innovations, it is possible, to determine with greater accuracy the impact of an approach to teaching on the measured performance of selected students As a teacher makes incremental improvements in current teaching practice, there is likely to be a decided expansion of professional skills that has greater potency in affecting the achievement of children.

Activity 8:1

The following are summaries of hypothetical studies. What action alternative would you use in the given situations?

Case I (see p. 58)

Question

Tom Reed is doing a study to determine whether the disruptive children in his class like the token economy system he established in his classroom. Will his students like it?

Hypothesis: The disruptive students will like the token economy.

Results: Using the Binomial Test he discovered that there were significantly more disruptive students who like the token economy than who didn't like it.

Conclusion:

Case II (See p. 64)

Question:

John Stock wishes to find out if the two tests he gives his class at the beginning and end of every year measure the same aptitudes.

Hypothesis. There will be a significantly strong positive relationship between the scores obtained on the Woodstock test and the Metropolitan Primary II.

Results: Using the Spearman Rho, Mr. Stock supported his hypothesis that there is a significantly strong relationship between the two tests.

Conclusion:

Activity 8:2

Steps In The Problem-Solving Procedure

Dates Completed	Steps	Data Needed
	Identifying the problem Formulation	Who was affected by the problem? Who caused it? What was the nature of the problem? What is the goal for improvement?
	Evaluation	How will solving this problem help me teach better? How can I use these results to make decisions?
	Formulating the hypothesis Problem-solving design decided	Am I trying to discover whether a relationship exists between two traits or scores? Am I trying to discover significant difference between scores? Comparing performance of two groups of subjects? Tests two groups of subjects before the treatment, pretest Determine whether one group of subjects fall into one category more often than another group Compare more than two groups of subjects
	Proper test to quantify	What traits do I want to measure? Is there a standardized test for that trait? Does this test measure this trait effectively at my pupils' level? Is it available to me? Is it too costly in terms of time and money?
	Define the quantity	Obtain the information in your records to complete the CDPS Summarize the results (e.g. list the percentage of male-female, on free lunch program-not on it, black-white mean age, etc.)
	List all procedures	When and under what circumstances were the pretest and post-test given? Describe in detail what the treatment entailed
	Data processing	Complete the CDPS chart
	Determine the appropriate statistical test	Rank all the scores Determine the level of measurement: Nominal, Ordinal, Interval and Ratio Are subject groups matched? Are subject groups randomly selected? What is the statistical design of my study?
	Calculate results	Look up equations for the appropriate test Plug in values from your data Check the appropriate chart to determine significance
	Problem-solving decisions	Did I support my hypothesis?
	What is the solution to my problem?	Did I answer my questions?
	What should my next course of action be?	Should I maintain my program? Should I modify my program? How? Should I implement a new program? What? Should I set up a new hypothesis and implement another study?
	Evaluating the Model	
	Did I have trouble using the model?	No ___ Yes ___ List specific problems
	Could I use the model for another study?	No ___ Yes ___ List specific reasons
	Would I use the model for another problem?	No ___ Yes ___ Why not?
	Do I find the model a feasible way to improve the quality of instruction to my pupils?	In what way? ___ Planning ___ Selecting materials ___ Selecting methods ___ Selecting activities ___ Identifying specific instructional problems
	Would I change the format of the problem solving model?	In what way? ___ Have more workshops ___ Have less workshops ___ Include more information in the manual ___ Simplify the manual ___ Include more examples in the workshop and manual ___ Send out literature and sample questions ahead of time so the teachers can start formulation of problem before the workshop ___ Start earlier in the school year ___ Have more feedback from the workshop supervisors ___ Simplify the problem-solving method ___ Others

Activity 8:3

Steps Completed Using The Problem-Solving Model

Actions Taken: Com.	Not. Com.	N.A.	Steps
			1. Identify the problem.
____	____	____	a. Formulation.
____	____	____	b. Evaluation.
			2. Formulate the hypothesis.
____	____	____	a. Appropriate test or method for quantifying results
			3. Defining populatioı
____	____	____	a. Obtain all information for the C D.P S
____	____	____	4. Sampling procedure selected.
____	____	____	5. Problem-solving design decided.
____	____	____	6. List d all procedes.
			7. Data processing.
____	____	____	a. All scores and information are charted
____	____	____	b. Determine statistical analysis to be used
____	____	____	c. Calculate results.
			8. Problem-solving decisions.
____	____	____	a. Did I support my hypothesis— Is there a significant relationship?
____	____	____	b. Did I find the answer to my question?
____	____	____	c. Did I solve my problem?
____	____	____	d. What should be my next course of action?
			9. Evaluation and feedback.
____	____	____	a Did I have trouble using the model?
____	____	____	b. Could I use the model on another problem?
____	____	____	c. Would I use the model on another problem?
____	____	____	d. Did I find the model valuable in helping provide quality instruction to my pupils?

Additional Readings

Easy: Dixon, p. 2, Barnes, p. 52, 108
Hard: Wallis, p. 23, Galfo, p. 299; Slakter, p. 382.

This concludes the ICPS Model as a process for making systematic instructional decisions. The material shown in the Appendix is intended to assist teachers who are planning to conduct a formal study and would like to apply statistical tests to decide whether observed differences are significant. It is entirely possible that teachers will need some assistance in using some of the information in the Appendix. However, the information is within reach of most teachers.

Appendix 1: Statistical Equations*

GLOSSARY OF SYMBOLS

A	Upper left-hand cell in a 2 × 3 table, number of cases observed in that cell
α	Alpha. Level of significance = probability of a Type I error
B	Upper right-hand cell in a 2 × 2 table, number of cases observed in that cell.
β	Beta Power of the test = probability of a Type II error
C	Lower left-hand cell in a 2 × 2 table, number of cases observed in that cell
C	Contingency coefficient.
Chi square	A random variable which follows the chi-square distribution, certain values of which are shown in Table C of the Appendix.
χ^2	A statistic whose value is computed from observed data
χr^2	The statistic in the Friedman two-way analysis of variance by ranks.
d_i	A difference score, used in the case of matched pairs, obtained for any pair by subtracting the score of one member from that of the other
df	Degrees of freedom.
D	Lower right-hand cell in a 2 × 2 table, number of cases observed in that cell.
D	The maximum difference between the two cumulative distributions in the Kolmogorov-Smirnow test
E_{ij}	Under H_0, the expected number of cases in ith row and the jth column in a χ^2 test
F	Frequency, i.e., number of cases
F	The F test, the parametric analysis of variance
$F_0(X)$	Under H_0, the proportion of cases in the population whose scores are equal to or less than X. This is a statistic in the Kolmogorov-Smirnov test
g	In the Moses test, the amount by which an observed value of s_h exceeds $nc - 2h$, where $nc - 2h$ is the minimum span of the ranks of the control cases
G_j	In the Cochran Q test, the total number of "successes" in the jth column (sample)
h	In the Moses test, the predetermined number of extreme control ranks which are dropped from each end of the span of control ranks before s_h is determined
H	The statistic used in the Kruskal-Wallis one-way analysis of variance by ranks
H_0	The null hypothesis
H_1	The alternative hypothesis, the operational statement of the research hypothesis.
i	A variable subscript, usually denoting rows
j	A variable subscript, usually denoting columns
K	In the Kolmogorov-Smirnov test, the number of observations which are equal to or less than X
K_D	In the Kolmogorov-Smirnov test, the numerator of D
L_i	In the Cochran Q test, the total number of "successes" in the ith row.
μ	Mu The population mean
μ_0	The population mean under H_0
μ_1	The population mean under H_1
n	The number of independently drawn cases in a single sample
N	The total number of independently drawn cases used in a statistical test
O_{ij}	The observed number of cases in the ith row and the jth column in a χ^2 test
p	Probability associated with the occurrence under H_0 of a value as extreme as or more extreme than the observed value.
P	In the binomial test, the proportion of "successes".
Q	In the binomial test, $1 - P$
Q	The statistic used in the Cochran test
r	The number of runs.
r	The Pearson product moment correlation coefficient
r	The number of rows in a $k \times r$ table.
R_j	The sum of the ranks in the jth column or sample
r_s	The Spearman rank correlation coefficient.
$\bar{r}_s$	The mean of several rs's
s	In the Kendall W, the sum of the squares of the deviations of the R_j from the mean value of R_j
s'	In the Moses test, the span or range of the ranks of the control cases
s_h	In the Moses test, the span or range of the ranks of the control cases after h cases have been dropped from each extreme of that range
S	A statistic in the Kendall τ.
$S_N(X)$	In the Kolmogorov-Smirnov teste, the observed cumulative step function of a random sample of N observations.
σ	Sigma The standard deviation of the population When a subscript is given, standard error of a sampling distribution, for example, σU - the standard error of the sampling distribution of U
σ^2	The variance of the population
Σ	Summation of.
t	Student's t test, a parametric test
t	The number of observations in any tied group
T	In the Wilcoxon test, the smaller of the sums of like-signed ranks
T	A correction factor for ties
τ	Tau The Kendall rank correlation coefficient
$\tau_{xy.z}$	The Kendall partial rank correlation coefficient
U	The statistic in the Mann-Whitney test
U'	$U = n_1n_2 - U'$, a transformation in the Mann-Whitney test
W	The Kendall coefficient of concordance
x	In the binomial test, the number of cases in one of the groups
X	Any observed score
$\bar{X}$	The mean of a sample of observations
z	Deviation of the observed value from μ, when $\sigma = 1$. z is normally distributed. Probabilities associated under H_0 with the occurrence of values as extreme as various z's are given in Table A of the Appendix
$\binom{a}{b}$	The binomial coefficient $\binom{a}{b} = \frac{a!}{b!(a-b)!}$ Table T of the Appendix gives binomial coefficients for N from 1 to 20
!	Factorial $N! = N(N-1)(N-2)\ldots 1$ For example $5! = (5)(4)(3)(2)(1) = 120$. $0! = 1$ Tables S of the Appendix gives factorials for N from 1 to 20

*All tests were adapted from Siegel, Sidney, *NonParametric Statistics for the Behavioral Sciences*, McGraw-Hill Book Co., New York, 1956

$\lvert X - Y \rvert$	The absolute value of the difference between X and Y That is, the numerical value of the difference regardless of sign. For example, $\lvert 5 - 3 \rvert = \lvert 3 - 5 \rvert = 2$
$X > Y$	X is greater than Y
$X < Y$	X is less than Y
$X = Y$	X is equal to Y
$X \geq Y$	X is equal to or greater than Y
$X \leq Y$	X is equal to or less than Y
$X \neq Y$	X is not equal to Y

One-Sample Cases

Test & Design Models	Level of Measurement	Type of Question Answered	Table	Characteristics of the Test
Binomial Test E Treatment Post-test C	Nominal	Proportion of P and O observed in sample d awn f a population having the P value specified in hypothesis? EX Do boys do better than girls in Math, when lessons use sports examples?	T for factorials $\binom{N}{X}$ D for P(X) Multiply by 2 for a 2-tailed test)	Used nominal level data with discrete categories Power 95% when N = 6 63% when N
Chi Square X^2 Test E_1 E_2 Treatment Post-test E_3	Nominal	Is there a significant difference between expected group frequencies and observed group frequencies? EX—Which level of questioning is used most frequently in class? Is it used significantly more than others?	C for X^2 when df = K-1 K = number of categories The greater X^2 is, the smaller the probability that observed frequency [illegible] from the expected frequency population	Used several categories, 20% of catagories have a frequency of 5 or more. If frequency is lower, categories must be combined It is a goodness-of-fit-test
Kolmogorov-Smirnov One-sample Test E Treatment Post-test	Norminal and Ordinal	Is the divergence between a theoretical and an observed distribution due to chance or are they from different populations? EX Does a child from a high, medium, or low economic status have a significantly higher drop-out rate?	E for D (for a 2-tailed test)	The test can be used for categories having very low frequency (few subjects in each category) It is a more powerful test than chi square.
One Sample Run Test	Nominal	Is distribution randomly selected? How many runs (or identical scores appearing consecutively) are in sample?	F_I and F_{II} for the limits between which r must fall. If r falls above or below limits, then sample is biased	This test is used to determine whether a sample is really randomly selected If r falls above F_{II} or below F_I then the sample is biased

Two Related Samples

Test & Design Models	Level of Measurement	Type of Question Answered	Table	Characteristics of the Test
McNear Test E Pre-test Treatment Post-test	Nominal	Is there is a significant difference between pretest and post-test scores? EX Is there a significa[illegible] improvement in test score[illegible] en oral instructions instead of written instructions are given?	C For X^2 Degrees of freedom = 1	If the expected frequencies of the sample are small use Binomial test instead of McNemar 4 + D must equal 6 when vA + D = 6 power = 95% ASA + D gets larger the power declines to 63%

Sign Test E Pre-test Treatment Post-test	Ordinal within pairs	Are there significantly more or fewer plus signs then minus signs? *EX* Will children's scores on spelling tests increase or decrease if they have time to study their words independently during class?	D for P (2-tailed test) A for Z score (for large samples) (1-tailed test)	Use this test when you can not rank all the scores but you can judge the greater score of a pair This test measures the direction of the difference within the pairs measured. When N = 6 the power of the test is 95%, but as N is larger the power of the test decreases to 63%
Wilcoxon Matched-Pairs, Sign Rank Test E Pre-test Treatment Post-test	Ordinal	Is there a significantly greater increase in the frequency and the magnitude of the positive values in relation to the negative values? *EX.* Are children who attend nursery school more socially perceptive than their matched peers who do not attend nursery school	G for T (1 and 2-tailed test)	For this test you need a continuous distribution. Since it utilizes more information, the Wilcoxon is a more powerful test than the sign test. Its power equals 95%
Randomization Test E Pre-test Treatment Post-test	Interval	What is the exact probability under the Null Hypothesis that our observed data has occurred? *EX* What is the probability that 20 students in 50 would have I Q 's of 120 or above ?	When N > 12 A for Z	A continuous distribution is necessary to perform this test, but you need not assume that you have a normal distribution. This is one of the most powerful nonparametric tests. It is better to use the *Wilcoxom Test* if N > 12.

Two Independent Samples

Test	Level of Measurement	Type of Question Answered	Table	Characteristics of the Test
Fisher Exact Probability Test E Pre- Treatment I Post- C test Treatment II test	Nominal or Ordinal	Do group I and group II differ significantly? *EX* Do children in a segregated classroom for EMH differ in achievement from children also classified EMH who are mainstreamed?	S for factorial (N!) I for significance of p	This test is the most powerful one for Nominal or Ordinal level data Use it when N is small (A + B 15 C + D 14) and when the data fits into two discrete categories

Chi Square test for Independent Samples	Nominal	Do two groups differ with respect to one characteristic? *EX* Do bright or dull students show more improvement academically in a highly structured program or in an open classroom?	C for X^2 If X^2 - C value that it is significant df = (r-1)(k-1) r = no of classifications k = no of groups	Use Chi square when 20 N 40 if expected frequencies are 5 or more, use the Fisher test if N 20 or the expected frequency is less than 5.
Mann-Whitney U Test E_1 Pre- Treatment I Post- E_2 test Treatment II test	Ordinal	Have two independent groups been drawn from the same population? *EX* Do children with dominance problems have more trouble reading than children with lateral dominance?	J for $3<n^2<8$ K for $9<n^2<20$ n^2 = largest group to find U When $n^2>20$ use A for Z	This is a very powerful test. Use it as an alternative to the t-test when the data is only ordinal level or when the groups are unequal.
Randomization Test for Two Independent Samples E_1 Pre- Treatment I Post- E_2 test Treatment II test	Interval	What is the significant difference between the means of two independent samples when n_1 and n_2 are small? *EX* Are group A's scores from a different population then group B's scores?	T for $(\frac{n_1+}{n_2})$ B for T-test if $\frac{1}{5} \leq \frac{n_1}{n_2} \leq 5$	This is the most powerful non-parametric test of central tendency. You do not have to assume a normal distribution or homogeneity of variance when n_1 and n_2 are large, the Mann-Whitney U test would be more effective than the randomized test and should be used

K-Related Samples

Test	Level of Measurement	Type of Question Answered	Table	Characteristics of the Test
Cochran Q Test E_1 Treatment 1 Pre- Post- E_2 Treatment 2 test test E_3 Treatment 3	Nominal or Dichotomized Ordinal	Do K-related samples differ significantly among themselves? *EX.* Is there a difference between three matched groups which receive three different types of instruction?	C for Q df = K-1 If C- Q then Q is significant df-K-1 K = number of categones.	Use this test only for nominal or dichotomous data. If N is too small the results will no longer have the distribution of Chi square.
Friedman Two-Way Analysis of Varience? E_1 Treatment 1 Pre- Post- E_2 Treatment 2 test test E_3 Treatment 3	Ordinal	Were K samples drawn from same population? Do rank totals vary significantly? *EX* Do three groups of students' achievement scores differ with three different types of teaching techniques?	C for Xr^2 df = K-1 If Xr^2 C then it is significant N for Xr^2 when K = 3n = 2-9 k = 4 n = 2-4	This test is very powerful. It should be used in preference to the Cochran Q whenever the data is appropriate

K-Independent Samples

Test	Level of Measurement	Type of Questions Answered	Table	Characteristics of the Test
X^2 for K Independent Samples E_1 Treatment 1 Pre- Post- E_2 Treatment 2 test test C Treatment	Nominal or Ordinal	Are samples from same population? *EX* Will children with parents who have different levels of education read the same number of books outside of class for enjoyment?	C for X^2 df = (K-1)(r-1) K = number of columns r = number of rows	Limit the number of groups. When there is more than 1 group, the Chi square test is insensitive to order. K or R must be greater than 1. And 80% of the categories must have an expected frequency of 5 or more
Kruskal-Wallis One-Way Analysis of Varience E_1 Treatment 1 ... Post- E_2 Treatment 2 test test C Treatment 3	Ordinal	Are K samples from the same population with respect to averages? *EX* Do children from different socioeconomic classes achieve significantly different grades in school?	O for H df = K-1 For large N use C for H	This test requires a continuous distribution. This test has 95% of the power of F test (the equivalent of the parametric test).

Correlations

Test	Level of Measurement	Type of Question Answered	Table	Characteristics of the Test
Contingency Coefficient C Sub-ject # (X) Read'g Score (y) Math Score 1 70 84 2 75 83 3 84 73 4 80 79 5 81 83 N =	Nominal	What is extent of association between two attributes? *EX.* How much relationship is there between oral ability and writing ability?	C for C df = (K-1)(r-1)	This test has the same limitations as X^2, in that it is not very powerful, but sometimes it is the only alternative. Also you can't get a perfect correction (can approach it though with a larger N)
Spearman Rank Correlation Coefficient Rs(Rho)	Ordinal	Are individuals ranked similarly on two characteristics? *EX* Do children who score high in Math also get high scores in Reading?	P for rs (1-tailed) $4 \leq N \leq 30$	This test is 90% as powerful as Pearson Product, the comparable parametric test.
Kendall Partial Rank Correlation Coefficient: tau xyz Sub-ject # Corre-lation x y (Z) trait 1 90 93 91 2 85 90 86 3 89 91 95	Ordinal	Is there a third variable causing correlation between x and y? *EX* Can level of parental education influence relationship between achievement and socioeconomic status?	Q for tau xyz if $N \geq 10$ use A for Z	There is no test of significance for this test because the distribution is unknown. Use this test to control for variance when you cannot control for outside factors in the study

Kendall Coefficient of Concordance W

Subject #	Trait Score (x)	Trait Score (y)	Trait score Z
1	15	20	23
2	21	13	16
3	14	17	18
N =			

Ordinal

What is degree of variance among several (K) rankings of N objects or individuals? EX How close is the interjudge agreement on ranking children in a class according to disruptive behavior? Order entities according to consensus and clusters of variables

R for W
$3 \leq K \leq 20$
$3 \leq n \leq 7$
When N > 7 C for X^2 df = N = 1

This test measures civergence of the actual agreement with perfect agreement. A significant W means Judges are using the same criteriato judge This does not sell whether the criteria is the right one with which to judge subjects.

Appendix 2: Sample Studies

Binominal Test

The Binomial is a goodness-of-fit test It is useful in determining whether it is reasonable to believe that the portions (or frequencies) within a given sample could have been drawn from a population having a specified value of P. P—is the proportion of cases in one class within a population. The data must be nominal level and separated into two discrete categories such as. male and female, in-school and out-of-school, member and non-member. There is no comparable parametric test which is applicble to data measured at the nominal level. However, when the relative power of the Binomial test was measured, the results showed it to have 95% of the power of a paramenc test to reject the null hypothesis when N–6. As N grew power went down to 63%

Tom Reed is doing a study to determine whether the disruptive children in his class like the token economy system he established in the class. (see p. 47)

i Null Hypothesis: There will be no difference between the number of disruptive students who like the token economy and the number of those who don't.
Hypothesis: The disruptive student will like the token economy
Statistical Test:

ii The binominal test is chosen because the data are in two discrete categories (disruptive nondisruptive) and the design is of the one sample type (he only asks the pupil preference once).

iii Significance Level: Let probability - .05. N = the number of disruptive students = 10

iv *Data*

Names	Disruptive	Non-Disruptives
John	✓	
Susie		✓
Fred	✓	
Larry	✓	
Mary	✓	
Tina		✓
Tom		✓
Kathy	✓	
Dan	✓	
Pam	✓	
Bill		✓
Terry	✓	
Jean	✓	
Joan		✓
Rob	✓	

Disruptive student = N 10 N = 10
P = the expected number of disruptive students who like the token economy occurring by chance.
Q = 10 disruptive students - P

v. The Equation $\sum_{i=0}^{N} (\;) P^i Q^{N-i}$

Table D gives the probabilities associated with the occurrence under the Null Hypothesis.

vi. Rejection region: He will find support for his hypothesis if the number of disruptive students who don't like the token economy is so small that the probability of that number of students occurring under the Null Hypothesis is equal to or less than P = .05

vii Decision. Only one of the 10 did not like the token economy. Looking up in Table D with N = 10 and X = 1, we find P = .011 Since .011 < .05 the hypothesis is supported Mr Reed's students did enjoy using the token system.

Chi-Square Two-Sample Case

This test is used to compare the number of scores in two or more categories with the expected number of scores in those categories, based on the null hypothesis. This goodness-of-fit-test can be used with nominal level data. There is no information in the literature as to the power of Chi-square to reject the false null hypothesis. However, it is suggested that if the statistician's data is ordinal level or higher, the Kolmogorov-Smirnov or some other test be used because they use more information and therefore would have more power to reject the null hypothesis.

Example

Helen McBride felt that her second graders who were read to by their parents enjoyed reading more than those whose parents did not read to them. Some of her students's parents did not read to them but had allowed their children to attend Story Hour at the library. Helen decided to find out if there was a difference between the three groups preference for reading

i Hypothesis: The students whose parents read to them will have a higher score on the reading enjoyment test than those whose parents do not.
Null Hypothesis: There will be no difference in reading perference among the three groups

ii. Statistical Test: The Chi Square test is chosen because the data is categorical or nominal level of measurement and because Mrs. McBride wants to compare a theoretical distribution with the distribution of scores which actually occurred.

iii Level of Significance: Let a (alpha) = .05. N = 30 students.

iv. Computation Data

	Parent read to children	Children attend Library hour	Children not read to	Total
like to read	Expected 5 (under the null hypothesis) Observed 9	5	5	15
don't like to read	Expected 5 (under null)Observed 1	5 4	5 8	15
Total	10	10	10	30

Equation

1. $$\chi^2 = \sum^{r}\sum^{k} \frac{(0 - E)^2}{E}$$

χ^2 = chi square
O = observed scores
E = expected scores
$\sum\sum$ = add the sum of each cell

2. $$\chi^2 = \frac{(9-5)^2}{5} + \frac{(6-5)}{5} + \frac{(2-5)^2}{5} + \frac{(1-5)^2}{5} + \frac{(4-5)^2}{5} + \frac{(8-5)^2}{5}$$

$$\chi^2 = \frac{16}{5} + \frac{1}{5} + \frac{9}{5} + \frac{16}{5} + \frac{1}{5} + \frac{9}{5} = \frac{52}{5}$$

$$\chi^2 = 10.4 \quad df = (r - 1)(k - 1) = (2 - 1)(3 - 1) = 2$$

3. Consult Table C. df = 3 5.99 = critical value of Chi-Square

v. **Region of Rejection:** In order to reject the null hypothesis, Mrs. McBride must have 5.99 or more.

vi. **Decision:** Chi-Square = 10.4 which is greater than 5.99. Therefore, Ms. McBride may reject the null hypothesis at the .05 level. She finds support for her hypothesis that the children who have someone read to them will enjoy reading more. She decided to start reading aloud to her class.

Cochran Q Test

The Cochran Q tests for significance difference among three or more matched sets of data that are at the nominal or dichotomized ordinal level of measurement. The test can be used to compare the same subjects under three or more conditions, or three or more different sets of subjects under the same condition. The power of this test is not known, however the fact that it should be used with nominal or ordinal level data makes comparison between the Cochran Q and parametric tests meaningless. Parametric test must have interval level of measurement. The statistician is cautioned, however, not to use the Cochran Q if the data is interval level of measurement, as some of the information may be wasted

Example

Mary Stewart was interested in finding out what teaching method would increase reading enjoyment for the children in her 6th grade Title I reading program. The three methods she employed were as follows: 1) a traditional round-robin oral reading approach; 2) educational comic books were used in class as supplementary reading materials, and 3) time was provided during class to go to the library for selection of books. Mary would like to know if there is any difference in enjoyment for the three methods of instruction

The following is an outline of Ms. Stewart's study:

1. **Hypothesis:** There will be a significant difference of reading enjoyment for the three different methods of instruction
Null Hypothesis: There will be no difference in the reading enjoyment for the three methods.

ii. **Statistical Test:** The Cochran Q is chosen because the data are for three related groups and are dichotomized as yes and no reading enjoyment.

ii. **Significance Level:** Let a (alpha) = .01, N = 14 (students)

vi. **Data** 0 = no enjoyment 1 = enjoyment

Names	Traditional	Educational Comics	Library Time	L t	L 2
Tod	0	0	0	0	0
Anna	0	1	1	2	4
Sue	0	1	0	1	1
Kenny	0	0	0	0	0
Steven	0	1	0	1	1
Marie	0	0	1	1	1
Mike	0	0	1	1	1
Barry	0	1	1	2	4
Lenore	1	1	1	3	9
Kathy	0	0	1	1	1
Craig	0	0	1	1	1
Terry	1	0	1	2	4
John	0	1	0	1	1
Joey	0	1	1	2	4
	$G_1 = 2$	$G_1 = 7$	$G_1 = 9$	$L_1 = 18$	$L_1^2 = 33$

v. **Computation**

1. $$Q = \frac{(k-1)[k\Sigma G_i^2 - (\Sigma G_i)^2]}{k\Sigma L_i - \Sigma L_i^2}$$

2. $$Q = \frac{(3-1)\,3\,(4 + 49 + 81) - (18)}{(3)(18) - 33}$$

$$Q = \frac{2\,3\,(134) - 18}{54 - 33} \qquad Q = \frac{768}{21}$$

$$Q = 36.57$$

3. Consult Table C, with df = k–1 = 3 – 1 = 2
The figure under .05 column in row 2 = 5.99

vi. **Rejection region:** One may reject the null hypothesis when the value of Q is larger than the figure listed for the significance level of .05.

vi. **Decision:** Since Ms. Stewart's data yielded a Q = 3657 which is larger than the necessary 5.99 for significance at the .05 level, she can reject the null hypothesis. She finds support for her hypothesis that there is a significant difference among methods of teaching for increasing reading enjoyment.

Contingency Coefficient: C

The Contingency Coefficient C is a measure of relationship or the extent of association between two sets of attributes It is the only test of its kind for data which is nominal scale, that is of an unordered series of frequencies. This test is not powerful in rejecting the false null hypothesis, but is uniquely useful with its ease of computation and freedom from restrictive assumptions about the population.

Example

John Stock felt that there was a relationship between low achievement and disruptive behavior He wanted to determine the strength of this relationship in his fifth grade students This part of his study follows:

i. **Hypothesis:** There will be a significantly strong positive relationship between the low scores obtained on the achievement test and ratings of disruptiveness.
Null Hypothesis: There will be no relationship between low achievement and disruptiveness.

ii. **Statistical Test:** The Contingency Coefficient C was chosen because the data of one of the variables was categorical (disruptive/nondisruptive) and also Mr Stock was seeking to discover the strength of the relationship.

iii. Level of Significance—Let a (alpha) = .05, N = 15 students
iv. Computation

Data

Disruptive	Below 50 Percentile		Above 50 Percentile		Total
	5	8	5	2	10
Nondisruptive	3	2	3	4	6
Total		10		6	16

Equations

1. $x^2 = \Sigma\Sigma \frac{(O_i - E_i)^2}{E_i} \Sigma\Sigma$ = add the sum of differences of each cell.
O_i = observed number of cases in a cell.
E_i = expected number of cases in a cell, all chances being equal

$$x^2 = \frac{(8-5)}{5} + \frac{(2-5)}{5} + \frac{(2-3)}{3} + \frac{(4-3)}{3}$$

$$x^2 = 9/5 + 9/5 + 1/3 + 1/3 = \frac{27}{15} + \frac{27}{15} + \frac{5}{15} + \frac{5}{15} = \frac{64}{15}$$

$$= 4.26$$

2. $C = \sqrt{\frac{x^2}{N + x^2}}$ C = contingency coefficient
x^2 = value of x^2 for data
N = number of subjects

$$C = \sqrt{\frac{4.26}{16 + 4.26}} = \sqrt{\frac{4.26}{20.26}} = \sqrt{n} = 458$$

46 = 5

3. Consult Table C. .46 is significant at .5 level
v. Rejection region: The null hypothesis will be rejected C = .05 significance or less
vi. Decision: Since C was not significant at the 05 level, Mr Stock can *not* reject the null hypothesis Therefore, Mr. Stock does not find support for his hypothesis that there is a significant relationship between disruptiveness and low achievement.

Fisher Exact Probability Test

This test is used to determine if one of two independent samples has significantly more scores in one of the two mutually exclusive categories than the other The two groups must be small in size and be measurable at the nominal or dichotomial ordinal level The Fisher test is the most powerful one a teacher can use on data that meets these specifications

Example

Joan Mitchell noticed that the children in her first grade class who seemed least well adjusted were also the youngest children chronologically. She wanted to discover if there was a significant difference between the adjustment ratings between the group of children whose birthdates fell between October 1 and December 1 and those born in other months An outline of her study follows:

Hypothesis: The children born between October 1 and December 1, 1972 will be rated as having adjustment problems significantly more frequently than the other group

i. Null Hypothesis: There will be no difference in adjustment ratings between the two groups.
ii. Statistical Test: The Fisher Exact Probability Test was selected because the catagories—adjusted and maladjusted—are mutually exclusive, and since the N is small.
iii. Data

	Adjusted		Maladjusted		
youngest children	A	1	B	8	9
older children	C	6	D	0	6
		7		8	15

v. Computation
1 Refer to Table I
2 A + B = 9 C + D = 6, go to this row under Totals in the Right Hand Margin
3. B = 8, therefore go to 8 under the B (or A) column
4. D = 0, therefore go across the 0 to find the level of significance, which is 0.1

vi. Rejection Region: The null hypothesis will be rejected if D = a .05

Decision: Since D = 0 which is less than the 2 needed for significance at the 05 level, Mrs. Mitchell's hypothesis is supported. She can reject the null hypothesis. She can therefore conclude that there was a significant difference in adjustment between the two age groups.

Friedman Two-Way analysis of Variance by Ranks

The Friedman test is to be used to determine if three or more samples differ significantly from one another and thus come from different populations. The samples must be matched and be measurable at the ordinal level of data. This matching may be achieved by studying the same group of subjects under three or more conditions or by assigning matched groups to the different conditions Although the exact power of the Friedman test is not reported, Friedman compared the results of his test with the results of the F test, a parametric test, and the comparison shows the Friedman statistical test to be equally as powerful as the F test.

Example

John McGee was interested in discovering which of three methods of teaching was most effective in helping his fifth graders to learn vocabulary words. The first method he tried was to have each child look up the definitions of the words only. In the second method, he divided the children into groups with each group giving a report on the etymology of each word The third type was to have the students write stones using the words. His study follows:

1. Hypothesis: There will be significant difference among the group's performance under different conditions
Null Hypotehsis: There will be no significant difference among the treatments.
ii Statistical Test: The Friedman Two-Way Analysis of Variance was used in this case because the data was ordinal level in measurement with matched samples
iii Significance Level: Let a (alpha) = 05. N = 9 (the number of students in Mr McGee's reading class)
iv Data

Names	Scores					
	Definitions	Ranks	Etymology	Ranks	Stones	Ranks
Mary	5	1	8	2	9	3
Greg	8	2	9	3	7	1
Tom	7	2	6	1	9	3
Renee	6	2	7	3	5	1
Tony	8	2	9	3	7	1
Ray	8	1	10	3	9	2
Ann	9	3	8	2	7	1
Rick	5	1	9	3	8	2
Martha	10	3	9	2	7	1
	R_1 =	17	R_2 =	20	R_3 =	15

v. Computation

1 $Xr^2 = \frac{12}{Nk(k+1)} \Sigma (r_j)^2 - 3N(k+1)$

N = number of rows
k = number of columns
R_j = sum of ranks in each individual column
Σ = directs one to sum the squres of the sums of ranks over all three teaching methods

2 $Xr^2 = \frac{12}{9(3)(3+1)} (289 + 400 + 225) - 3(9)(3+1)$

$Xr^2 = \frac{12}{108} 918 - 108$

$Xr^2 = .11\ (806)$

$Xr^2 = 88.66$

3. Consult Table N where K = 3 N = 9
Significant at the .05 level = 6 22 = Xr^2

vi. Rejection region: The rejection region for the Null Hypothesis of all values of Xr^2 which are so large that the probability associated with their occurrence under the null hypothesis is equal to or less than a (alpha) = 05 (6.222)

vii. Decisions: Since the Xr^2 = 88.66 is considerably larger than the 6.222 to considerably larger than the 6.222 to find significance at the .05 level, Mr. McGee finds support his hypothesis. He can reject Null hypothesis, that all the treatments help his students equally.

Kendall Coefficient of Concordance: W

The Concordance W is used to measure correlation or strength of relationship among three or more sets of scores.* Such a measure may be particularly useful in studies of inter-test or inter-judge reliability or in studies of clusters of variables. A high or significant value of W may be interpreted as meaning that the tests or judges are applying the same standards in ranking the variables under study. Clusters of variables are studied in the example below

Example

Mrs Dema Jane Martin had the idea that some of her students felt that the work she assigned was too difficult to perform. therefore they didn't even try to listen or make their best efforts She wants to find out if there is a relationship among the variables lack of self-esteem, reading comprehension scores and fear of failure scores.

i. Hypothesis: There will be a significant strong relationship among the three variables. self-esteem (as measured by the Self-Esteem questionnaire), fear of failure (as measured by the test of that same name) and achievement (as measured by the SRA Reading Comprehension Test).
Null Hypothesis: There will be no significant relationship among the variables.

ii. Statistical Test: The Kendall Concordance W was used in this study to find the strength of relationship among three variables and the data is ordinal level in measurement

iii. Significance Level: Let a (alpha) = 05. N = 15.

iv. Computation
Data

Student	SRA Rank	Self-Esteem Rank	Fear of Failure Rank	R_j	$R_j - \frac{R_j}{N}$	$(R_j - \frac{R_j}{N})^2$
Debbie	12	14	12	38	13 73	188.51
Linda	3	4	4	11	– 13 27	176 09
Patrick	2	6	3	11	– 13 27	176 09
Allen	11	10	8	29	4 73	22 37
Steve	13	12	14	39	14 73	216 97
Alice	1	3	5	9	15 27	233 17
William	7	6	7	20	4 27	18 23
Pam	8	9	10	27	2 73	7 45
Terry	10	8	9	27	2 73	7 45
Joan	4	1	2	7	17 27	298.25
Mike	14	13	13	40	15 73	247 43
Fred	5	2	1	8	16 27	264.71
Marty	6	7	6	19	5 27	27 77
Sue	9	15	15	39	14 73	216 97
Kevin	15	13	12	40	15 73	247 43

$\frac{R_j}{N} = 24\ 2$ $\quad \frac{(R_j - R_j)^2}{N}$

= 2348 89

Equation $\quad W = \frac{s}{1/12k^2(N^3 - N)}$

$s = \sum (R_j - \frac{\Sigma R_j}{N})^2$ sum of observed deviations from the mean of R_j = 2348.89

k = number of sets of rankings = 3
N = number of subjects (students) = 15

1. $W = \frac{2348\ 89}{1/12(3)^2(15^3 - 15)}$

$W = \frac{2348\ 89}{1/12(9)(3360)} = \frac{2348\ 89}{2520}$

W = .932 Since N 7 find x^2 (chi square)

2 $x^2 = K(N-1)W$
$x^2 = 3(15 - 1)\ .932$
$x^2 = 39.144$

3. df = N − 1 = 15 − 1

4. Consult Table C. (If N 7 consult Table R) df = 14
Critical value for .05 = 23.68

v. Rejection region: In order to reject the null hypothesis x^2 must be equal to or greater than 23.68

vi Decision: 39.144 is greater than 23.68, therefore Mrs Martin may reject the null hypothesis at the .05 level of significance. She finds a significant relationship among the three variables of achievement, self-esteem and fear of failure

*The data must be ordinal or ordered in level of measurement

Kolmogorov-Smirnov One-Sample Test

This is a goodness-of-fit test It is used to determine the degree of agreement between the distribution of a set of sample values (observed scores) and some specified theoretical distribution The Kolmogorov-Smirnov test is useful with ordinal or ordered level of data, particularly if the number of scores in each category is small. This test is more powerful than its counterpart the chi-square test in rejecting the null hypothesis when it is false, so the statistician has a better chance of getting significant results by using the Kolmogorov-Smirnov test than by using the chi-square test

Example

John Stock's 5th grade students are classified as having the lowest language arts ability of all fifth graders in the school. Most of these students have been in this low grouping throughout grade school. Many of them have also been discipline problems throughout their grade school days. Consequently, John as instituted a point system to try and improve the discipline in his room. He wants to find out if the class likes the point system, so he administered a questionnaire to his students. His study follows.

i. Hypothesis: The students will rate the point system significantly highly as measured by the questionnaire.
   Null Hypothesis: The number of students who rate the point system high will equal those who rate it low.
ii. Statistical Test: The statistical test which is appropriate to test this hypothesis is the Kolmogorov-Smirnov test because the data is of ordinal scale and it is compared with a theoretical distribution.
iii. Significance Level: Let a (alpha) = .01. N = 20. number of students.
iv. Computation
   Data

	Score on Questionnaire (10 = high preference)									
	1	2	3	4	5	6	7	8	9	10
f = number of subjects choosing rank	0	0	1	0	2	4	3	2	4	4
$f_o(X)$ = theoretical cumulative distribution under H_o	$\frac{1}{10}$	$\frac{2}{10}$	$\frac{3}{10}$	$\frac{4}{10}$	$\frac{5}{10}$	$\frac{6}{10}$	$\frac{7}{10}$	$\frac{8}{10}$	$\frac{9}{10}$	$\frac{10}{10}$
S25(X) = observed cumulative distribution	$\frac{0}{20}$	$\frac{0}{20}$	$\frac{1}{20}$	$\frac{1}{20}$	$\frac{3}{20}$	$\frac{7}{20}$	$\frac{10}{20}$	$\frac{12}{20}$	$\frac{16}{20}$	$\frac{20}{20}$
$f_o(X) - S_{20}(X)$ =	$\frac{2}{20}$	$\frac{4}{20}$	$\frac{5}{20}$	$\frac{7}{20}$	$\frac{7}{20}$	$\frac{5}{20}$	$\frac{4}{20}$	$\frac{4}{20}$	$\frac{2}{20}$	0

highest $F_o(X) - S_{20}(X)$

Equation

1. $F_o(X) - S_{20}(X)$ = D Highest D = 7/20 = 3.5/10 = .35
2. Consult Table E .01 probability, N = 20, Significant D = .356

v. Rejection region: To reject the null hypothesis D .56.
vi. Decision: John's data shows .35 which is less than .356. He can reject the null hypothesis and finds support for his hypothesis that his students like the point system.

Kruskal-Wallis One-Way Analysis of Variance by Ranks

This test is used to determine whether two or more independent samples comes from the same population or whether they can be said to be significantly different from one another. The test assumes that the variable under study be at least ordinal level of measurement and have an underlying continuous distribution. This means that the variable may have any value in a certain interval, not restricted to isolated values. The Kruskal-Wallis test has a relative power of 95.5% compared to the parametric F test.

Example

Joan Rollinson taught twenty-seven third graders, who had various levels of abilities. She is trying to find the most efficient mode of giving Instructions for the children in her class.

i. Hypothesis: There will be significant difference in performance on the test among the three groups receiving oral instructions, written instructions and both.
   Null Hypothesis. The performance of the different groups will be no different.
ii. Statistical Test: The Kruskal Wallis test is chosen because Joan is looking for statistical significance among two or more samples and the data is ordinal level in measurement.
iii. Significance Level: Let a (alpha) = 05. N = 30 number of students.
iv. Computation
   Data

Oral	Rank	Written	Rank	Both	Rank
17	7 5	10	15	20	1.5
18	5 5	14	12	19	3 5
15	1.1	13	13	20	1.5
17	7.5	16	9.5	18	5.5
16	9.5	11	14	19	3 5
R_1 =	41.0	R_2 =	63 5	R_3 =	13.5

Equation

$$H = \frac{\frac{12}{N(N+1)} \sum \frac{R_j^2}{N_j} - 3(N+1)}{1 - \frac{\Sigma T}{N^3 - N}}$$

N = number of scores
R_j = sum of the ranks of a column squared
N_j = number of scores in a column
T - t^3 - t (when t is the number of tied scores in a tied group of scores

Note. $\frac{1 - T}{N_3 - N}$ is only used to correct for ties. If there are no ties delete this part from the equation

1 $$H = \frac{\frac{12}{15(15+1)} \frac{(41)^2}{5} + \frac{(63.5)^2}{5} + \frac{(13.2)^2}{5} - 3(15+1)}{\frac{6+6+6+6+6}{(15)^3 - 15}}$$

$$H = \frac{\frac{12}{240} \frac{1681}{5} + \frac{3969}{5} + \frac{172.25}{\cdot 5} - 48}{\frac{30}{3360} \quad 1 -}$$

$$H = \frac{\frac{1}{20} 336.2 + 793.8 + 34.45 - 48}{\frac{1}{112} \quad 1 -}$$

H = 10.12

2. Consult Table O. sample size = 5555, p = .049 = 5 78

v. Rejection region: Ms. Rollinson can reject the null hypothesis if the H derived from her data is greater than the H on the chart (5.78)
vi. Decision: Since her H = 10.12 which is larger than 5.78 she may reject the null hypothesis. Ms. Rollinson does find support for her hypothesis that there is significant difference in performance with different type of instruction. She would therefore, use both oral and written instructions to facilitate her pupils performance.

Mann-Whitney U Test

This test should be used to determine whether two independent groups have been drawn from different populations, and whether they are significantly different. The data must be at least ordinal level of measurement. Mann-Whitency U is one of the most powerful nonparametric tests. When compared to the parametric T test, it had 95.5% of its power to reject the null hypothesis. Mann-Whitney U is a useful alternative to the T-test, then, when the assumptions of a parametric test cannot be made

Example

Marlene Piscitelli noticed that her students with dominance problems in the remedial reading clinic seemed to have more problems with reading comprehension than the other students. She decided to find out if there was a significant difference between the two groups. Her study follows:

i. Hypothesis: There will be a significant difference in achievement between the students who have lateral dominance problems and those who do not.
Null Hypothesis: The two groups were drawn from the same population; there will be no significant difference in performance.

ii. Statistical Test: The Mann-Whitney U test is chosen because the data is at the ordinal level of measurement and Marlene is looking for significant difference between two independent groups.

iii Significance Level: Let a (alpha) = .05. N = 2. students with dominance problems. N^2 = 6 students with no dominance problems

iv Computation
Data

N 1	41	45				
N 2	78	75	82	90	51	68

1	90	82	78	75	68	51	45	41
	N_2	N_2	N_2	N_2	N_2	N_2	N	N_1

U = No. of N's preceding N_2's.
U = 0

2. Table J N_2 = 6, U = 0 N_1 = 2. Significant at .036.

v. Rejection Region: The U has to be so small that there would only be .05 chance or less that it could have occurred by chance. In other words the significant level must be .05 or less.

vi. Decision: Since the significance level for Ms Piscetteli's data is equal to .036 which is less than .05, she finds support for her hypothesis that students with laterality problems also have difficulty in reading. She may reject the null hypothesis

Randomization Test for Two Matched Samples

This test is a useful and powerful nonparametric technique for testing the significant difference between the means of two matched samples when the two samples are small. It requires at least interval measurement of the variable being studied. There is no assumption of normal distributions or homogeneity of variance (which means the two samples don't have to scores equidistant from one another) like the comparable T test assumes. The Randomization test efficiently uses all information and therefore its power to reject the false null hypothesis is essentially 100%. An example is shown below. However, if the samples are large, the Wilcoxon test may be more efficient to use

Example

Mike Brown is concerned that some of his students aren't doing as well on his tests as they are capable of doing. He feels that they get too anxious and feel they will flunk the test. To verify this observation he administered the Fear of Failure test and found a high negative correlation between the scores on that test and his own reading tests. Mike therefore decided to give the students an option to retake the test as often as they want. He wants to find out if the students will do better on the tests if they feel less pressure. His study is shown below.

i. Hypothesis. There will be a significant increase in reading scores when the students are given an option to retake the tests.
Null Hypothesis: There will be no difference in the two scores

ii. Statistical Test: The Randomization Test for Matched Pairs is chosen because the interval level of measurement is used, the samples are matched and Mike is looking for significant difference

iii. Significance Level: Let a (alpha) = .05. N = number of pairs = 8.

iv. Computation
Data

Names	Differences between option/no option
Tina	+19
Ed	+27
Marty	-1
Larry	+6
Pam	+7
Michelle	+13
Amy	-4
Tom	+3

Six most extreme positive outcome (a (alpha) = .05)

	Outcomes								Ed.
1	+19	+27	+1	+6	+7	+13	+4	+3	80
2	+19	+27	-1	+6	+7	+13	+4	+3	78
3	+19	+27	+1	+6	+7	+13	+4	-3	74
4	+19	+27	+1	+6	+7	+13	-4	+3	72
5	+19	+27	-1	+6	+7	+13	+4	-3	72
6	+19	+17	-1	+6	+7	+13	-4	+3	70

v. Rejection region: Mike's data falls into one of three six extreme outcomes than his scores represent a significant difference of a (alpha) = .047.

vi. Decision: Since Mike's set of data did match one of these extreme outcomes, he achieved significant results at .05 probability (alpha). He can reject the null hypothesis and finds support for his hypothesis that giving a retake option on his tests will improve his students' scores.

Sign Test

The Sign test uses plus and minus signs rather than quantitive measures as its data. It is particularly useful with data which can not be quantitatively measured but in which it is possible to rank with respect to each other the two members of each pair. The only assumption is that the variable measured has a continuous distribution, that is, the data can take on any value. This test is to be used when the subjects are drawn from matched samples. This can be accomplished either by using the subjects as their own controls (test/retest) or by using different subjects which are matched in respect to the relevant variables. The power of the Sign test to reject the false null hypothesis is about 95% when N = 6, and as the sample increases its eventual power is 63%. A sample problem is shown below.

Example

Joe Risk's sixth graders having been labelled "troublemakers" early in their grade school careers were constantly disrupting his reading class. As Mr. Risk felt these disruptions were not conducive to learning either for the troublemakers or the other class members, he decided to try a token economy in the class to decrease the acting out behavior. He put the class on a point system, in which the class as a whole could earn points for participating in class and points were taken away for disruptive behavior. The points earned went toward the privilege of getting out of class early on Friday. Mr. Risk wants to find out if this point system is an effective way of controlling behavior. A synopsis of his study follows.

i. **Hypothesis:** The children will receive a higher behavioral rating after the use of the point system, where 1 = disruptive and 5 = well behaved.
   Null Hypotehsis: There will be no difference before and after the use of the point system.
ii. **Statistical Test:** The sign test was chosen because the measurement is ordinal and the samples matched.
iii. **Significance Level:** Let a (alpha) = .05, N = 17 students (this may be reduced if there are ties).
iv. **Computation**
   Data

Name	Pre-rating	Post-rating	Direction of Difference	Sign
Marth	5	5	5 = 5	0
Frank	2	4	2 < 4	+
Eddie	1	3	1 < 3	+
Mitch	3	4	3 < 4	+
Rita	4	4	4 = 4	0
Suzanne	4	5	4 < 5	+
Leta	4	5	4 < 5	•
Tom	2	3	2 < 3	+
Rick	4	2	4 > 2	–
Mike	4	3	4 > 3	–
James	3	4	3 < 4	+
Chris	3	5	3 < 5	+
Bob	2	4	2 < 4	+
Barb	3	4	3 < 4	+
Allan	1	3	1 < 3	+
Joe	2	3	2 < 3	+
Ellen	4	5	4 < 5	+

$N = 15$ X = number of fewer signs = 2

1. Consult Table D $N = 15$, $X = 2$, $p = .004$
v. **Rejection region:** Joe may reject the null hypothesis if $p < .05$.
vi. **Decision:** Since $p = .004$ which is less than .05, Joe can reject the null hypothesis and finds support for his hypothesis that the point system reduces disruptive behavior. He will continue to use the point system.

The Spearman Rank Correlation Coefficient: Rank

The [illegible]an test sometimes called rho is one of the best known non[illegible] [illegible]tics. It was developed to determine the amount of relationship between variables which must be measurable at the ordinal level. However, no relationships can be established using this test, only degree of relationship. The more the ranks within two sets of scores agree, the higher the positive correlation, the more they diverge from one another, the higher the negative correlation. If there is no discernable pattern to the variables' variance in relation to one another than there is said to be low correlation or little relationship between the scores. The Spearman Rank Correlation test has 91% of the power to reject the null hypothesis of the Pearson r, a comparable parametric test.

Example

John Stock wishes to find out if two tests he gives his class at the beginning and end of every year measure the same aptitudes. The two tests are the Woodstock Reading test and the reading section of the Metropolitan Primary II. He will determine this by finding the strength of relationship between the scores his fifth graders obtained on the two tests (see p. 47) A synopsis of this portion of his study follows.

i. **Hypothesis:** There will be a significantly strong positive relationship between the scores obtained on the Woodstock test and the Metropolitan Primary II.
   Null Hypothesis: There will be either no relationship or negative relationship between the scores obtained.
ii. **Statistical test:** The Spearman Rank test will be used because Mr. Stock wishes to find the strength of relationship and the data is ordinal level of measurement.
iii. **Significance Level:** Let a (alpha) = .05, N = 15 (students or sets of scores)
iv. **Computation**

Name	Woodstock test score Rank	Metropolitan Primary II score Rank	(difference) di	(difference) di^2
Jean	14	14	0	0
Tony	4	4	0	0
Rachel	13	11	2	4
Tim	5	3	2	4
Ann	10	12	-2	4
Lela	3	2	1	1
Tammy	1	1	0	0
Rick	7	9	2	4
Ruth	8	6	2	4
Mary	2	5	–3	9
Phillip	12	10	1	1
Clark	12	10	2	4
Kevin	6	7	–1	1
Mike	15	15	0	0
Pam	11	13	–2	4
				$di^2 = 40$

Equation

1.
$$r_s = 1 - \frac{6\Sigma di^2}{N^3 - N}$$
$$r_s = 1 - \frac{6(40)}{(15)^3 - (15)}$$
$$r_s = 1 - \frac{240}{3375 - 15}$$
$$r_s = 1 - \frac{240}{3360}$$
$$r_s = 1 - .07$$
$$r_s = .93$$

2. Consult Table P. N = 15 significance level .05, critical value of r_s = .425
v. **Rejection region.** If our score is larger than the critical value of r_s listed in P, then we can reject the null hypothesis.

vi Decision: Since Mr Stock's value 93 was larger than the critical value 425 he can reject the null hypothesis and find support for his hypothesis that there is a relationship between the Metropolitan Primary and Woodstock Reading tests

Wilcoxon Matched-Pairs Signed-Ranks Test

This test is used to determine if there is significant difference in direction and magnitude between pairs of scores For example if a teacher is interested in ascertaining whether more achievement scores increased or decreased *and* whether there was a bigger difference in the scores which increased or those which decreased, then (s)he would use this me''od

Teacher will use Statistical Models 4, 5, or 6 in conjunction with this test The data must be ordinal level, i e the teacher must be able to make a judgment of greater than between any pair's two performances

If all these criteria are met then use this method rather than the Sign test, as the Wilcoxon is more powerful than the Sign, that is, it uses more information than the Sign When compared to the t-Test (a parametric test), it is found to be about 95% as efficient This is powerful for a nonparametric test

Example

Kathleen Bergman is trying to determine if the PLATO programmed instruction technique improves children's learning

i **Null Hypothesis:** There will be no increase in the achievement scores obtained on the pre and post test given to the children before and after treatment

Hypothesis: There will be a significant increase in the achievement scores obtained on the pretests and post tests

ii **Statistical Test:** The Wilcoxon Matched-Pairs Signed-Rank Test is chosen because the study employs two related samples (two sets of scores for each child) and it yields different scores which may be ranked in order of absolute magnitude (absolute difference in the two scores)

iii **Significance Level:** L a (alpha) = 025 N = Number of pairs of scores or number of children minus any pairs of scores whose difference was zero

iv

Data	12 - 2 = 10 = N Pretest score	Post test score	Difference d	Rank of L	Rank with less frequent sign
John	63	82	19	8	
Suzi	42	69	27	9	
Fred	55	55	0		
Larry	74	73	-1	-1	(-) 1
Tim	37	43	6	5	
Mary	51	58	7	6	
Rob	43	56	13	7	
Terry	80	76	-4	-3	(-) 3
Joe	60	60	0		
Mac	79	82	3	2	
Lois	90	85	-5	-4	(-) 4
Ed	40	70	30	10	8 T

v Equation—none for $N \leq 25$

(For $N \geq 25$ the equation is $= z = \dfrac{T - \frac{N(N+1)}{4}}{\sqrt{\frac{N(N+1)(2N+1)}{24}}}$)

N in this study = 10 ≤ 25 Therefore go directly to Table G Find 025 under one tailed test Find 10 under N

vi **Rejection region:** Since the direction of the difference is predicted (i e the scores will increase), a one-tailed test is appropriate. The region of rejection consists of all values of T so small that the probability associated with their occurrence under the Null Hypothesis is less than or equal to a (alpha) = 025 for one-tailed test.

vii **Decision:** Only three children in the study regressed while using PLATO Looking at Table G with N = 10 and T = 8, so a T = 8 allows us to reject the Null Hypothesis at a (alpha) = 025 for a one-tailed test We find support for Kathleen's hypothesis the children do show significant increase in the achievement scores obtained on the pretest and post test

Appendix 3: Tables

List of Tables*

*All tables from
Siegel, Sidney *Nonparametric Statistics*, New York, N Y Publisher, McGraw-Hill Company, 1956, pages 312

Table A. Table of Probabilities Associated with Values as Extreme as Observed Values of z in the Normal Distribution

The body of the table gives one-tailed probabilities under H_0 of z. The left hand marginal column gives various values of z to one decimal place. The top row gives various values to the second decimal place. Thus, for example, the one-tailed p of z ≥ 11 or z ≤ − 11 is p = 4562

z	00	01	02	03	04	05	06	07	08	09
0	5000	4960	4920	4880	4840	4801	4761	4721	4681	4641
1	4602	4562	4522	4483	4443	4404	4364	4325	4286	4247
2	4207	4168	4129	4090	4052	4013	3974	3936	3897	3859
3	3821	3783	3745	3707	3669	3632	3594	3557	3520	3483
4	3446	3409	3372	3336	3330	3264	3228	3192	3156	3121
5	3085	3050	3015	2981	2946	2912	2877	2843	2810	2776
6	2743	2709	2676	2643	2611	2578	2546	2514	2483	2451
7	2420	2389	2358	2327	2296	2266	2236	2206	2177	2148
8	2119	2090	2061	2033	2005	1977	1949	1922	1894	1867
9	1841	1814	1788	1762	1736	1711	1685	1660	1635	1611
1 0	1587	1562	1539	1515	1492	1469	1446	1423	1401	1379
1 1	1357	1335	1314	1292	1271	1251	1230	1210	1190	1170
1 2	1151	1131	1112	1093	1075	1056	1038	1020	1003	0985
1 3	0968	0951	0934	0918	0901	0885	0869	0853	0838	0823
1 4	0808	0793	0778	0764	0749	0735	0721	0708	0694	0681
1 5	0668	0655	0643	0630	0618	0606	0594	0582	0571	0559
1 6	0548	0537	0526	0516	0505	0495	0485	0475	0465	0455
1 7	0446	0436	0427	0418	0409	0401	0392	0384	0375	0367
1 8	0359	0351	0344	0336	0329	0322	0314	0307	0301	0294
1 9	0287	0281	0274	0268	0262	0256	0250	0244	0239	0233
2 0	0228	0222	0217	0212	0207	0202	0197	0192	0188	0183
2 1	0179	0174	0170	0166	0162	0158	0154	0150	0146	0143
2 2	0139	0136	0132	0129	0125	0122	0119	0116	0113	0110
2 3	0107	0104	0102	0099	0096	0094	0091	0089	0087	0084
2 4	0082	0080	0078	0075	0073	0071	0069	0068	0066	0064
2 5	0062	0060	0059	0057	0055	0054	0052	0051	0049	0048
2 6	0047	0045	0044	0043	0041	0040	0039	0038	0037	0036
2 7	0035	0034	0033	0032	0031	0030	0029	0028	0027	0026
2 8	0026	0025	0024	0023	0023	0022	0021	0021	0020	0019
2 9	0019	0018	0018	0017	0016	0016	0015	0015	0014	0014
3 0	0013	0013	0013	0012	0012	0011	0011	0011	0010	0010
3 1	0010	0009	0009	0009	0008	0008	0008	0008	0007	0007
3 2	0007									
3 5	0005									
3 4	0003									
3 5	00023									
3 6	00016									
7	00011									
[illegible]	00007									
3 9	00005									
4 0	00003									

Table B. Table of Critical Values of t*

df	Level of significance for one-tailed test					
	10	05	025	01	005	0005
	Level of significance for two-tailed test					
	20	10	05	02	01	001
1	3 078	6 314	12 706	31 821	63 657	636 619
2	1 886	2 920	4 303	6 965	9 925	31 598
3	1 638	2 353	3 182	4 541	5 841	12 941
4	1 533	2 132	2 776	3 747	4 604	8 610
5	1 476	2 015	2 571	3 365	4 032	6 859
6	1 440	1 943	2 447	3 143	3 707	5 959
7	1 415	1 895	2 365	2 998	3 499	5 405
8	1 397	1 860	2 306	2 896	3 355	5 041
9	1 383	1 833	2 262	2 821	3 250	4 781
10	1 372	1 812	2 228	2 764	3 169	4 587
11	1 363	1 796	2 201	2 718	3 106	4 437
12	1 356	1 782	2 179	2 681	3 055	4 318
13	1 350	1 771	2 160	2 650	3 012	4 221
14	1 345	1 761	2 145	2 624	2 977	4 140
15	1 341	1 753	2 131	2 602	2 947	4 073
16	1 337	1 746	2 120	2 583	2 921	4 015
17	1 333	1 740	2 110	2 567	2 898	3 965
18	1 330	1 734	2 101	2 552	2 878	3 922
19	1 328	1 729	2 093	2 539	2 861	3 883
20	1 325	1 725	2 086	2 528	2 845	3 850
21	1 323	1 721	2 080	2 518	2 831	3 819
22	1 321	1 717	2 074	2 508	2 819	3 792
23	1 319	1 714	2 069	2 500	2 807	3 767
24	1 3[illegible]	1 711	2 064	2 492	2 797	3 745
25	[illegible]16	1 708	2 060	2 485	2 787	3 725
26	1 315	1 706	2 056	2 479	2 779	3 707
27	1 314	1 703	2 052	2 473	2 771	3 690
28	1 313	1 701	2 048	2 467	2 763	3 674
29	1 311	1 699	2 045	2 462	2 756	3 659
30	1 310	1 697	2 042	2 457	2 750	3 646
40	1 303	1 684	2 021	2 423	2 704	3 551
60	1 296	1 671	2 000	2 390	2 660	3 460
120	1 289	1 658	1 980	2 358	2 617	3 373
∞	1 282	1 645	1 960	2 326	2 576	3 291

* Table B is abridged from Table III of Fisher and Yates *Statistical tables for biological agricultural and medical research* published by Oliver and Boyd Ltd, Edinburgh, by permission of the authors and publishers

Table C. Table of Critical Values of Chi Square*

df	Probability under H_0 that $\chi^2 \geq$ chi square													
	.99	.98	.95	.90	.80	.70	.50	.30	.20	.10	.05	.02	.01	.001
1	.00016	.00063	.0039	.016	.064	.15	.46	1.07	1.64	2.71	3.84	5.41	6.64	10.83
2	.02	.04	.10	.21	.45	.71	1.39	2.41	3.22	4.60	5.99	7.82	9.21	13.82
3	.12	.18	.35	.58	1.00	1.42	2.37	3.66	4.64	6.25	7.82	9.84	11.34	16.27
4	.30	.43	.71	1.06	1.65	2.20	3.36	4.88	5.99	7.78	9.49	11.67	13.28	18.46
5	.55	.75	1.14	1.61	2.34	3.00	4.35	6.06	7.29	9.24	11.07	13.39	15.09	20.52
6	.87	1.13	1.64	2.20	3.07	3.83	5.35	7.23	8.56	10.64	12.59	15.03	16.81	22.46
7	1.24	1.56	2.17	2.83	3.82	4.67	6.35	8.38	9.80	12.02	14.07	16.62	18.48	24.32
8	1.65	2.03	2.73	3.49	4.59	5.53	7.34	9.52	11.03	13.36	15.51	18.17	20.09	26.12
9	2.09	2.53	3.32	4.17	5.38	6.39	8.34	10.66	12.24	14.68	16.92	19.68	21.67	27.88
10	2.56	3.06	3.94	4.86	6.18	7.27	9.34	11.78	13.44	15.99	18.31	21.16	23.21	29.59
11	3.05	3.61	4.58	5.58	6.99	8.15	10.34	12.90	14.63	17.28	19.68	22.62	24.72	31.26
12	3.57	4.18	5.23	6.30	7.81	9.03	11.34	14.01	15.81	18.55	21.03	24.05	26.22	32.91
13	4.11	4.76	5.89	7.04	8.63	9.93	12.34	15.12	16.98	19.81	22.36	25.47	27.69	34.53
14	4.66	5.37	6.57	7.79	9.47	10.82	13.34	16.22	18.15	21.06	23.68	26.87	29.14	36.12
15	5.23	5.98	7.26	8.55	10.31	11.72	14.34	17.32	19.31	22.31	25.00	28.26	30.58	37.70
16	5.81	6.61	7.96	9.31	11.15	12.62	15.34	18.42	20.46	23.54	26.30	29.63	32.00	39.29
17	6.41	7.26	8.67	10.08	12.00	13.53	16.34	19.51	21.62	24.77	27.59	31.00	33.41	40.75
18	7.02	7.91	9.39	10.86	12.86	14.44	17.34	20.60	22.76	25.99	28.87	32.35	34.80	42.31
19	7.63	8.57	10.12	11.65	13.72	15.35	18.34	21.69	23.90	27.20	30.14	33.69	36.19	43.82
20	8.26	9.24	10.85	12.44	14.58	16.27	19.34	22.78	25.04	28.41	31.41	35.02	37.57	45.32
21	8.90	9.92	11.59	13.24	15.44	17.18	20.34	23.86	26.17	29.62	32.67	36.34	38.93	46.80
22	9.54	10.60	12.34	14.04	16.31	18.10	21.24	24.94	27.30	30.81	33.92	37.66	40.29	48.27
23	10.20	11.29	13.09	14.85	17.19	19.02	22.34	26.02	28.43	32.01	35.17	38.97	41.64	49.73
24	10.86	11.99	13.85	15.66	18.06	19.94	23.34	27.10	29.55	33.20	36.42	40.27	42.98	51.18
25	11.52	12.70	14.61	16.47	18.94	20.87	24.34	28.17	30.68	34.38	37.65	41.57	44.31	52.62
26	12.20	13.41	15.38	17.29	19.82	21.79	25.34	29.25	31.80	35.56	38.88	42.86	45.64	54.05
27	12.88	14.12	16.15	18.11	20.70	22.72	26.34	30.32	32.91	36.74	40.11	44.14	46.96	55.48
28	13.56	14.85	16.93	18.94	21.59	23.65	27.34	31.39	34.03	37.92	41.34	45.42	48.28	56.89
29	14.26	15.57	17.71	19.77	22.48	24.58	28.34	32.46	35.14	39.09	42.56	46.69	49.59	58.30
30	14.95	16.31	18.49	20.60	23.36	25.51	29.34	33.53	36.25	40.26	43.77	47.96	50.89	59.70

*Table C is abridged from Table IV of Fisher and Yates, *Statistical tables for biological, agricultural, and medical research*, published by Oliver and Boyd Ltd., Edinburgh, by permission of the authors and publishers.

Table D. Table of Probabilities Associated with Values as Small as Observed Values of *x* in the Binomial Test*

Given in the body of this table are one-tailed probabilities under H_0 for the binomial test when $P = Q = \frac{1}{2}$. To save space, decimal points are omitted in the *p*'s.

N \ x	0	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
5	031	188	500	812	969	†										
6	016	109	344	656	891	984	†									
7	008	062	227	500	773	938	992	†								
8	004	035	145	363	637	855	965	996	†							
9	002	020	090	254	500	746	910	980	998	†						
10	001	011	055	172	377	623	828	945	989	999	†					
11		006	033	113	274	500	726	887	967	994	†					
12		003	019	073	194	387	613	806	927	981	997	†	†			
13		002	011	046	133	291	500	709	867	954	989	998	†	†		
14		001	006	029	090	212	395	605	788	910	971	994	999	†	†	
15			004	018	059	151	304	500	696	849	941	982	996	†	†	†
16			002	011	038	105	227	402	598	773	895	962	989	998	†	†
17			001	006	025	072	166	315	500	685	834	928	975	994	999	†
18			001	004	015	048	119	240	407	593	760	881	952	985	996	999
19				002	010	032	084	180	324	500	676	820	916	968	990	998
20				001	006	021	058	132	252	412	588	748	868	942	979	994
21				001	004	013	039	095	192	332	500	668	808	905	961	987
22					002	008	026	067	143	262	416	584	738	857	933	974
23					001	005	017	047	105	202	339	500	661	798	895	953
24					001	003	011	032	076	154	271	419	581	729	846	924
25						002	007	022	054	115	212	345	500	655	788	885

*Adapted from Table IV, B, of Walker, Helen, and Lev, J. 1953. *Statistical inference*. New York: Holt, p. 458, with the kind permission of the authors and publisher. †1.0 or approximately 1.0.

Table E. Table of Critical Values of *D* in the Kolmogorov-Smirnov One-Sample Test*

Sample size (N)	Level of significance for D = maximum $\lvert F_0(X) - S_N(X) \rvert$				
	.20	.15	.10	.05	.01
1	.900	.925	.950	.975	.995
2	.684	.726	.776	.842	.929
3	.565	.597	.642	.708	.828
4	.494	.525	.564	.624	.733
5	.446	.474	.510	.565	.669
6	.410	.436	.470	.521	.618
7	.381	.405	.438	.486	.577
8	.358	.381	.411	.457	.543
9	.339	.360	.388	.432	.514
10	.322	.342	.368	.410	.490
11	.307	.326	.352	.391	.468
12	.295	.313	.338	.375	.450
13	.284	.302	.325	.361	.433
14	.274	.292	.314	.349	.418
15	.266	.283	.304	.338	.404
16	.258	.274	.295	.328	.392
17	.250	.266	.286	.318	.381
18	.244	.259	.278	.309	.371
19	.237	.252	.272	.301	.363
20	.231	.246	.264	.294	.356
25	.21	.22	.24	.27	.32
30	.19	.20	.22	.24	.29
35	.18	.19	.21	.23	.27
Over 35	$\frac{1.07}{\sqrt{N}}$	$\frac{1.14}{\sqrt{N}}$	$\frac{1.22}{\sqrt{N}}$	$\frac{1.36}{\sqrt{N}}$	$\frac{1.63}{\sqrt{N}}$

*Adapted from Massey, F. J., Jr. 1951. The Kolmogorov-Smirnov test for goodness of fit. *J. Amer. Statist. Ass.*, 46:70, with the kind permission of the author and publisher.

Table F. Table of Critical Values of r in the Runs Test*

Given in the bodies of Table F_I and Table F_{II} are various critical values of r for various values of n_1 and n_2. For the one-sample runs test, any value of r which is equal to or smaller than that shown in Table F_I or equal to or larger than that shown in Table F_{II} is significant at the .05 level. For the Wald-Wolfowitz two-sample runs test, any value of r which is equal to or smaller than that shown in Table F_I is significant at the .05 level.

Table F_I

n_1 \ n_2	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20
2											2	2	2	2	2	2	2	2	2
3					2	2	2	2	3	2	2	2	2	3	3	3	3	3	3
4				2	2	2	3	3	3	3	3	3	3	3	4	4	4	4	4
5			2	2	3	3	3	3	4	4	4	4	4	4	4	4	5	5	5
6		2	2	3	3	3	3	4	5	4	4	5	5	5	5	5	5	6	6
7		2	2	3	3	3	4	4	5	5	5	5	5	6	6	6	6	6	6
8		2	3	3	3	4	4	5	5	5	6	6	6	6	6	7	7	7	7
9		2	3	3	4	4	5	5	6	6	6	6	7	7	7	7	8	8	8
10		2	3	3	4	5	5	5	6	6	7	7	7	7	8	8	8	8	9
11		2	3	4	4	5	5	6	7	7	7	7	8	8	8	9	9	9	9
12	2	2	3	4	4	5	6	6	7	7	7	8	8	8	9	9	9	10	10
13	2	2	3	4	5	5	6	6	7	7	8	8	9	9	9	10	10	10	10
14	2	2	3	4	5	5	6	7	7	8	8	9	9	9	10	10	10	11	11
15	2	3	3	4	5	6	6	7	8	8	8	9	9	10	10	11	11	11	12
16	2	3	4	4	5	6	6	7	8	8	9	9	10	10	11	11	11	12	12
17	2	3	4	4	5	6	7	7	8	9	9	10	10	11	11	11	12	12	13
18	2	3	4	5	5	6	7	8	8	9	9	10	10	11	11	12	12	13	13
19	2	3		5	6	6	7	8	9	9	10	10	11	11	12	12	13	13	13
20	2	3		5	6	6	7	8		9	10	10	11	12	12	13	13	13	14

Table F. Table of Critical Values of r in the Runs Test* (*Continued*)

Table F_{II}

n_1 \ n_2	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20
2																			
3																			
4				9	9														
5			9	10	10	11	11												
6			9	10	11	12	12	13	13	13	13								
7				11	12	13	13	14	14	14	14	15	15	15					
8				11	12	13	14	14	15	15	16	16	16	16	17	17	17	17	17
9					13	14	14	15	16	16	16	17	17	18	18	18	18	18	18
10					13	14	15	16	16	17	17	18	18	18	19	19	19	20	20
11					13	14	15	16	17	17	18	19	19	19	20	20	20	21	21
12					13	14	16	16	17	18	19	19	20	20	21	21	21	22	22
13						15	16	17	18	19	19	20	20	21	21	22	22	23	23
14						15	16	17	18	19	20	20	21	22	22	23	23	23	24
15						15	16	18	18	19	20	21	22	22	23	23	24	24	25
16							17	18	19	20	21	21	22	23	23	24	25	25	25
17							17	18	19	20	21	22	23	23	24	25	25	26	26
18							17	18	19	20	21	22	23	24	25	25	26	26	27
19							17	18	20	21	22	23	23	24	25	26	26	77	27
20							17	18	20	21	22	23	24	25	25	26	27	27	28

*Adapted from Swed, Frieda S., and Eisenhart, C. 1943. Tables for testing randomness of grouping in a sequence of alternatives. *Ann. Math. Statist.* 14, 83-86, with the kind permission of the authors and publisher.

Table G. Table of Critical Values of T in the Wilkcoxon Matched-pairs Signed-ranks Test*

N	Level of significance for one-tailed test		
	.025	.01	.005
	Level of significance for two-tailed test		
	.05	.02	.01
6	0		
7	2	0	
8	4	2	0
9	6	3	2
10	8	5	3
11	11	7	5
12	14	10	7
13	17	13	10
14	21	16	13
15	25	20	16
16	30	24	20
17	35	28	23
18	40	33	28
19	46	38	32
20	52	43	38
21	59	49	43
22	66	56	49
23	73	62	55
24	81	69	61
25	89	77	68

*Adapted from Table I of Wilcoxon, F. 1949. *Some rapid approximate statistical procedures.* New York: American Cyanamid Company, p. 13, with the kind permission of the author and publisher.

Table I. Table of Critical Values of *D* (or *C*) in the Fisher Test*,†

Totals in right column		*B* (or *A*)†	Level of significance			
			.05	.025	.01	.005
A + *B* = 3	*C* + *D* = 3	3	0	—	—	
A + *B* = 4	*C* + *D* = 4	4	0	0		
	C + *D* = 3	4	0			
A + *B* = 5	*C* + *D* = 5	5	1	1	0	0
		4	0	0		—
	C + *D* = 4	5	1	0	0	
		4	0	—	—	
	C + *D* = 3	5	0	0		—
	C + *D* = 2	5	0	—	—	
A + *B* = 6	*C* + *D* = 6	6	2	1	1	0
		5	1	0	0	
		4	0			—
	C + *D* = 5	6	1	0	0	0
		5	0	0		
		4	0	—	—	—
	C + *D* = 4	6	1	0	0	0
		5	0	0		
	C + *D* = 3	6	0	0		
		5	0	—	—	—
	C + *D* = 2	6	0	—		—
A + *B* = 7	*C* + *D* = 7	7	3	2	1	1
		6	1	1	0	0
		5	0	0	—	
		4	0			
	C + *D* = 6	7	2	2	1	1
		6	1	0	0	0
		5	0	0	—	
		4	0		—	—
	C + *D* = 5	7	2	1	0	0
		6	1	0	0	
		5	0			—
	C + *D* = 4	7	1	1	0	0
		6	0	0		
		5	0			
	C + *D* = 3	7	0	0	0	—
		6	0	—		
	C + *D* = 2	7	0			

*Adapted from Finney, D. J. 1948. The Fisher-Yates test of significance in 2 × 2 contingency tables. *Biometrika* 35, 149-154, with the kind permission of the author and the publisher.

†When *B* is entered in the middle column, the significance levels are for *D*. When *A* is used in place of *B*, the significance levels are for *C*.

(*Continued*)

Table I. Table of Critical Values of *D* (or *C*) in the Fisher Test*,† (*Continued*)

Totals in right margin		*B* (or *A*)†	Level of significance			
			.05	.025	.01	.005
A + *B* = 8	*C* + *D* = 8	8	4	3	2	2
		7	2	2	1	0
		6	1	1	0	0
		5	0	0		
		4	0			
	C + *D* = 7	8	3	2	2	1
		7	2	1	1	0
		6	1	0	0	
		5	0	0		—
	C + *D* = 6	8	2	2	1	1
		7	1	1	0	0
		6	0	0	—	
		5	0			—
	C + *D* = 5	8	2	1	1	0
		7	1	0	0	0
		6	0	0		—
		5	0			
	C + *D* = 4	8	1	1	0	0
		7	0	0		—
		6	0		—	—
	C + *D* = 3	8	0	0	0	
		7	0	0		—
	C + *D* = 2	8	0	0	—	—
A + *B* = 9	*C* + *D* = 9	9	5	4	3	3
		8	3	3	2	1
		7	2	2	1	0
		6	1	1	0	0
		5	0	0		—
		4	0	—	—	—
	C + *D* = 8	9	4	3	3	2
		8	3	2	1	1
		7	2	1	0	0
		6	1	0	0	—
		5	0	0		
	C + *D* = 7	9	3	3	2	2
		8	2	2	1	0
		7	1	1	0	0
		6	0	0	—	—
		5	0		—	
A + *B* = 9	*C* + *D* = 6	9	3	2	1	1
		8	2	1	0	0
		7	1	0	0	
		6	0	0		—
		5	0		—	
	C + *D* = 5	9	2	1	1	1
		8	1	1	0	0
		7	0	0	—	
		6	0			
	C + *D* = 4	9	1	1	0	0
		8	0	0	0	
		7	0	0		
		6	0			
	C + *D* = 3	9	1	0	0	0
		8	0	0		
		7	0			
	C + *D* = 2	9	0	0		

*Adapted from Finney, D. J. 1948. The Fisher-Yates test of significance in 2 × 2 contingency tables. *Biometrika* 35, 149-154, with the kind permission of the author and the publisher.

† When *B* is entered in the middle column, the significance levels are for *D*. When *A* is used in place of *B*, the significance levels are for *C*.

Table I. Table of Critical Values of *D* (or *C*) in the Fisher Test*,† (*Continued*)

Totals in right margin		*B* (or *A*)†	Level of significance			
			05	025	01	005
A + *B* = 10	*C* + *D* = 10	10	6	5	4	3
		9	4	3	3	2
		8	3	2	1	1
		7	2	1	1	0
		6	1	0	0	
		5	0	—	—	—
		4	0		—	—
	C + *D* = 9	10	5	4	3	3
		9	4	3	2	2
		8	2	2	1	1
		7	1	1	0	0
		6	1	0	0	—
		5	0	0	—	
	C + *D* = 8	10	4	4	3	2
		9	3	2	2	1
		8	2	1	1	0
		7	1	1	0	0
		6	0	0	—	—
		5	0	—	—	—
	C + *D* = 7	10	3	3	2	2
		9	2	2	1	1
		8	1	1	0	0
		7	1	0	0	—
		6	0	0	—	—
		5	0	—	—	—
A + *B* = 10	*C* + *D* = 6	10	3	2	2	1
		9	2	1	1	0
		8	1	1	0	0
		7	0	0	—	—
		6	0	—	—	—
	C + *D* = 5	10	2	2	1	1
		9	1	1	0	0
		8	1	0	0	—
		7	0	0	—	—
		6	0	—	—	—
	C + *D* = 4	10	1	1	0	0
		9	1	0	0	0
		8	0	0	—	—
		7	0	—	—	—
	C + *D* = 3	10	1	0	0	0
		9	0	0	—	—
		8	0	—	—	—
	C + *D* = 2	10	0	0	—	—
		9	0	—	—	—
A + *B* = 11	*C* + *D* = 11	11	7	6	5	4
		10	5	4	3	3
		9	4	3	2	2
		8	3	2	1	1
		7	2	1	0	0
		6	1	0	0	—
		5	0	0	—	—
		4	0	—	—	—
	C + *D* = 10	11	6	5	4	4
		10	4	4	3	2
		9	3	3	2	1
		8	2	2	1	0
		7	1	1	0	0
		6	1	0	0	—
		5	0	—	—	—
	C + *D* = 9	11	5	4	4	3
		10	4	3	2	2
		9	3	2	1	1
		8	2	1	1	0
		7	1	1	0	0
		6	0	0	—	—
		5	0	—	—	—

Table I. Table of Critical Values of *D* (or *C*) in the Fisher Test*,† (*Continued*)

Totals in right margin		*B* (or *A*)†	Level of significance			
			04	025	01	005
A + *B* = 12	*C* + *D* = 11	12	7	6	5	5
		11	5	5	4	3
		10	4	3	2	2
		9	3	2	2	1
		8	2	1	1	0
		7	1	1	0	0
		6	1	0	0	—
		5	0	0	—	—
	C + *D* = 10	12	6	5	5	4
		11	5	4	3	3
		10	4	3	2	2
		9	3	2	1	1
		8	2	1	0	0
		7	1	0	0	0
		6	0	0	—	—
		5	0	—	—	—
	C + *D* = 9	12	5	5	4	3
		11	4	3	3	2
		10	3	2	2	1
		9	2	2	1	0
		8	1	1	0	0
		7	1	0	0	—
		6	0	0	—	—
		5	0	—	—	—
	C + *D* = 8	12	5	4	3	3
		11	3	3	2	2
		10	2	2	1	1
		9	2	1	1	0
		8	1	1	0	0
		7	0	0	—	—
		6	0	0	—	—
	C + *D* = 7	12	4	3	3	2
		11	3	2	2	1
		10	2	1	1	0
		9	1	1	0	0
		8	1	0	0	—
		7	0	0	—	—
		6	0	—	—	—
A + *B* = 11	*C* + *D* = 8	11	4	4	3	3
		10	3	3	2	1
		9	2	2	1	1
		8	1	1	0	0
		7	1	0	0	—
		6	0	0	—	—
		5	0	—	—	—
	C + *D* = 7	1	4	3	2	2
		0	3	2	1	1
		9	2	1	1	0
		8	1	1	0	0
		7	0	0	—	—
		6	0	0	—	—
	C + *D* = 6	1	3	2	2	1
		0	2	1	1	0
		9	1	1	0	0
		8	1	0	0	—
		7	0	0	—	—
		6	0	—	—	—
	C + *D* = 5	11	2	2	1	1
		10	1	1	0	0
		9	1	0	0	0
		8	0	0		—
		7	0	—	—	—
	C + *D* = 4	11	1	1	1	0
		10	1	0	0	0
		9	0	0		
		8	0	—	—	—
	C + *D* = 3	11	1	0	0	0
		12	0	0		
		9	0		—	—
	C + *D* = 2	11	0	0		—
		11	0	—		

Table I. Table of Critical Values of *D* (or *C*) in the Fisher Test*,† (*Continued*)

Totals in right margin		B (or A)†	Level of significance			
			.05	.025	.01	.005
A + B = 12	C + D = 12	12	8	7	6	5
		11	6	5	4	4
		10	5	4	3	2
		9	4	3	2	1
		8	3	2	1	1
		7	2	1	0	0
		6	1	0	0	–
		5	0	0	–	–
		4	0	–	–	–
A + B = 12	C + D = 6	12	3	3	2	2
		11	2	2	1	1
		10	1	1	0	0
		9	1	0	0	0
		8	0	0	–	–
		7	0	0	–	–
		6	0	–	–	–
	C + D = 5	12	2	2	1	1
		11	1	1	1	0
		10	1	0	0	0
		9	0	0	0	–
		8	0	0	–	–
		7	0	–	–	–
	C + D = 4	12	2	1	1	0
		11	1	0	0	0
		10	0	0	0	–
		9	0	0	–	–
		8	0	–	–	–
	C + D = 3	12	1	0	0	0
		11	0	0	0	–
		10	0	0	–	–
		9	0	–	–	–
	C + D = 2	12	0	0	–	–
		11	0	–	–	–
A + B = 13	C + D = 13	13	9	8	7	6
		12	7	6	5	4
		11	6	5	4	3
		10	4	4	3	2
		9	3	3	2	1
		8	2	2	1	0
		7	2	1	0	0
		6	1	0	0	–
		5	0	0	–	–
		4	0	–	–	–
	C + D = 12	13	8	7	6	5
		12	6	5	5	4
		11	5	4	3	3
		10	4	3	2	2
		9	3	2	1	1
		8	2	1	1	0
		7	1	1	0	0
		6	1	0	0	–
		5	0	0	–	–

Table I. Table of Critical Values of *D* (or *C*) in the Fisher Test*,† (*Continued*)

Totals in right margin		B (or A)†	Level of significance			
			.05	.025	.01	.005
A + B = 13	C + D = 11	13	7	6	5	5
		12	6	5	4	3
		11	4	4	3	2
		10	3	3	2	1
		9	3	2	1	1
		8	2	1	0	0
		7	1	0	0	0
		6	0	0	–	–
		5	0	–	–	–
	C + D = 10	13	6	6	5	4
		12	5	4	3	3
		11	4	3	2	2
		10	3	2	1	1
		9	2	1	1	0
		8	1	1	0	0
		7	1	0	0	–
		6	0	0	–	–
		5	0	–	–	–
	C + D = 9	13	5	5	4	4
		12	4	4	3	2
		11	3	3	2	1
		10	2	2	1	1
		9	2	1	0	0
		8	1	1	0	0
		7	0	0	–	–
		6	0	0	–	–
		5	0	–	–	–
	C + D = 8	13	5	4	3	3
		12	4	3	2	2
		11	3	2	1	1
		10	2	1	1	0
		9	1	1	0	0
		8	1	0	0	–
		7	0	0	–	–
		6	0	–	–	–
	C + D = 7	13	4	3	3	2
		12	3	2	2	1
		11	2	2	1	1
		10	1	1	0	0
		9	1	0	0	0
		8	0	0	–	–
		7	0	0	–	–
		6	0	–	–	–
A + B = 13	C + D = 6	13	3	3	2	2
		12	2	2	1	1
		11	2	1	1	0
		10	1	1	0	0
		9	1	0	0	–
		8	0	0	–	–
		7	0	–	–	–
	C + D = 5	13	2	2	1	1
		12	2	1	1	0
		11	1	1	0	0
		10	1	0	0	
		9	0	0	–	–
		8	0	–	–	–
	C + D = 4	13	2	1	1	0
		12	1	1	0	0
		11	0	0	0	–
		10	0	0	–	–
		9	0	–	–	–
	C + D = 3	13	1	1	0	0
		12	0	0	0	–
		11	0	0	–	–
		10	0	–	–	–
	C + D = 2	13	0	0	0	–
		12	0	–	–	–

Table I. Table of Critical Values of D (or C) in the Fisher Test*,† (*Continued*)

Totals in right margin		B (or A)†	Level of significance			
			.05	.025	.01	.005
$A + B = 14$	$C + D = 14$	14	10	9	8	7
		13	8	7	6	5
		12	6	6	5	4
		11	5	4	3	3
		10	4	3	2	2
		9	3	2	2	1
		8	2	2	1	0
		7	1	1	0	0
		6	1	0	0	
		5	0	0	–	
		4	0	–		
$A + B = 14$	$C + D = 13$	14	9	8	7	6
		13	7	6	5	5
		12	6	5	4	3
		11	5	4	3	2
		10	4	3	2	2
		9	3	2	1	1
		8	2	1	1	0
		7	1	1	0	0
		6	1	0	–	–
		5	0	0	–	
	$C + D = 12$	14	8	7	6	6
		13	6	6	5	4
		12	5	4	4	3
		11	4	3	3	2
		10	3	3	2	1
		9	2	2	1	1
		8	2	1	0	0
		7	1	0	0	
		6	0	0		
		5	0	–		–
	$C + D = 11$	14	7	6	6	5
		13	6	5	4	4
		12	5	4	3	3
		11	4	3	2	2
		10	3	2	1	1
		9	2	1	1	0
		8	1	1	0	0
		7	1	0	0	
		6	0	0		
		5	0			–
	$C + D = 10$	14	6	6	5	4
		13	5	4	4	3
		12	4	3	3	2
		11	3	3	2	1
		10	2	2	1	1
		9	2	1	0	0
		8	1	1	0	0
		7	0	0	0	–
		6	0	0		–
		5	0	–	–	

Table I. Table of Critical Values of D (or C) in the Fisher Test*,† (*Continued*)

Totals in right margin		B (or A)†	Level of significance			
			.05	.025	.01	.005
$A + B = 14$	$C + D = 9$	14	6	5	4	4
		13	4	4	3	3
		12	3	3	2	2
		11	3	2	1	1
		10	2	1	1	0
		9	1	1	0	0
		8	1	0	0	–
		7	0	0		
		6	0	–	–	
	$C + D = 8$	14	5	4	4	3
		13	4	3	2	2
		12	3	2	2	1
		11	2	2	1	1
		10	2	1	0	0
		9	1	0	0	0
		8	0	0	0	–
		7	0	0	–	
		6	0	–	–	–
	$C + D = 7$	14	4	3	3	2
		13	3	2	2	1
		12	2	2	1	1
		11	2	1	1	0
		10	1	1	0	0
		9	1	0	0	–
		8	0	0	–	
		7	0			–
	$C + D = 6$	14	3	3	2	2
		13	2	2	1	1
		12	2	1	1	0
		11	1	1	0	0
		10	1	0	0	
		9	0	0	–	–
		8	0	0		
		7	0	–	–	
	$C + D = 5$	14	2	2	1	1
		13	2	1	1	0
		12	1	1	0	0
		11	1	0	0	0
		10	0	0		–
		9	0	0	–	
		8	0	–	–	–
$A + B = 14$	$C + D = 4$	14	2	1	1	1
		13	1	1	0	0
		12	1	0	0	0
		11	0	0	–	
		10	0	0	–	–
		9	0		–	–
	$C + D = 3$	14	1	1	0	0
		13	0	0	0	–
		12	0	0		–
		11	0			
	$C + D = 2$	14	0	0	0	
		13	0	0		
		12	0			

Table I. Table of Critical Values of *D* (or *C*) in the Fisher Test*,† (*Continued*)

Totals in right margin		*B* (or *A*)†	Level of significance			
			.05	.025	.01	.005
A + *B* = 15	*C* + *D* = 15	15	11	10	9	8
		14	9	8	7	6
		13	7	6	5	5
		12	6	5	4	4
		11	5	4	3	3
		10	4	3	2	2
		9	3	2	1	1
		8	2	1	1	0
		7	1	1	0	0
		6	1	0	0	—
		5	0	0	—	—
		4	0			—
	C + *D* = 14	15	10	9	8	7
		14	8	7	6	6
		13	7	6	5	4
		12	6	5	4	3
		11	5	4	3	2
		10	4	3	2	1
		9	3	2	1	1
		8	2	1	1	0
		7	1	1	0	0
		6	1	0	—	—
		5	0	—	—	—
A + *B* = 15	*C* + *D* = 13	15	9	8	7	7
		14	7	7	6	5
		13	6	5	4	4
		12	5	4	3	3
		11	4	3	2	2
		10	3	2	2	1
		9	2	2	1	0
		8	2	1	0	0
		7	1	0	0	—
		6	0	0	—	—
		5	0	—	—	—
	C + *D* = 12	15	8	7	7	6
		14	[illegible]	6	5	4
		13	6	5	4	3
		12	5	4	3	2
		11	4	3	2	2
		10	3	2	1	1
		9	2	1	1	0
		8	1	1	0	0
		7	1	0	0	
		6	0	0	—	
		5	0		—	—
	C + *D* = 11	15	7	7	6	5
		14	6	5	4	4
		13	5	4	3	3
		12	4	3	2	2
		11	3	2	2	1
		10	2	2	1	1
		9	2	1	0	0
		8	1	1	0	0
		7	1	0	0	—
		6	0	0	—	
		5	0			
	C + *D* = 10	15	6	6	5	5
		14	5	5	4	3
		13	4	4	3	2
		12	3	3	2	2
		11	3	2	1	1
		10	2	1	1	0
		9	1	1	0	0
		8	1	0	0	
		7	0	0		
		6	0			—

Table I. Table of Critical Values of *D* (or *C*) in the Fisher Test*,† (*Continued*)

Totals in right margin		*B* (or *A*)†	Level of significance			
			.05	.025	.01	.005
A + *B* = 15	*C* + *D* = 9	15	6	5	4	4
		14	5	4	3	3
		13	4	3	2	2
		12	3	2	2	1
		11	2	2	1	1
		10	2	1	0	0
		9	1	1	0	0
		8	1	0	0	
		7	0	0		
		6	0	—	—	—
	C + *D* = 8	15	5	4	4	3
		14	4	3	3	2
		13	3	2	2	1
		12	2	2	1	1
		11	2	1	1	0
		10	1	1	0	0
		9	1	0	0	—
		8	0	0	—	—
		7	0	—	—	—
		6	0	—	—	—
	C + *D* = 7	15	4	4	3	3
		14	3	3	2	2
		13	2	2	1	1
		12	2	1	1	0
		11	1	1	0	0
		10	1	0	0	0
		9	0	0	—	—
		8	0	0	—	—
		7	0	—	—	—
	C + *D* = 6	15	3	3	2	2
		14	2	2	1	1
		13	2	1	1	0
		12	1	1	0	0
		11	1	0	0	0
		10	0	0	0	—
		9	0	0	—	—
		8	0	—	—	—
	C + *D* = 5	15	2	2	2	1
		14	2	1	1	1
		13	1	1	0	0
		12	1	0	0	0
		11	0	0	0	
		10	0	0	—	—
		9	0		—	—
A + *B* = 15	*C* + *D* = 4	15	2	1	1	1
		14	1	1	0	0
		13	1	0	0	0
		12	0	0	0	—
		11	0	0	—	—
		10	0		—	—
	C + *D* = 3	15	1	1	0	0
		14	0	0	0	0
		13	0	0	—	—
		12	0	[illegible]		
		11	0			—
	C + *D* = 2	15	0	0	0	
		14	0	0		
		13	0			

*Adapted from Finney, D. J. 1948. The Fisher-Yates test of significance in 2 × 2 contingency tables. *Biometrika* 35: 149-154, with the kind permission of the author and the publisher.

†When *B* is entered in the middle column, the significance levels are for *D*. When *A* is used in place of *B*, the significance levels are for *C*.

Table J. Table of Probabilities Associated with Values as Small as Observed Values of U in the Mann-Whitney Test*

$n_2 = 3$

U \ n_1	1	2	3
0	250	100	050
1	500	200	100
2	750	400	200
3		600	350
4			500
5			650

$n_2 = 4$

U \ n_1	1	2	3	4
0	200	067	028	014
1	400	133	057	029
2	600	267	114	657
3		400	200	100
4		600	314	171
5			429	243
6			571	343
7				443
8				557

$n_2 = 5$

U \ n_1	1	2	3	4	5
0	167	047	018	008	004
1	333	095	036	016	008
2	500	190	071	032	016
3	667	286	125	056	028
4		429	196	095	048
5		571	286	143	075
6			393	206	111
7			500	278	155
8			607	365	210
9				452	274
10				548	345
11					421
12					500
13					579

$n_2 = 6$

U \ n_1	1	2	3	4	5	6
0	143	036	012	005	002	001
1	286	071	024	010	004	002
2	428	143	048	019	009	004
3	571	214	083	033	015	008
4		321	131	057	026	013
5		429	190	086	041	021
6		571	274	129	063	032
7			357	176	089	047
8			452	238	123	066
9			548	305	165	090
10				381	214	120
11				457	268	155
12				545	331	197
13					396	242
14					465	294
15					535	350
16						409
17						469
18						531

U \ n_1	1	2	3	4	5	6	7
0	125	028	008	003	001	001	000
1	250	056	017	006	003	001	001
2	375	111	033	012	005	002	001
3	500	167	058	021	009	004	002
4	625	250	092	036	015	007	003
5		333	133	055	024	011	006
6		444	192	082	037	017	009
7		556	258	115	053	026	013
8			333	158	074	037	019
9			417	206	101	051	027
10			500	264	134	069	036
11			583	324	172	090	049
12				394	216	117	064
13				464	265	147	082
14				538	319	183	104
15					378	223	130
16					438	267	159
17					500	314	191
18					562	365	228
19						418	267
20						473	310
21						527	355
22							402
23							451
24							500
25							549

Table J. Table of Probabilities Associated with Values as Small as Observed Values of U in the Mann-Whitney Test* (continued)

$n_2 = 8$

U \ n_1	1	2	3	4	5	6	7	8	t	Normal
0	111	022	006	002	001	000	000	000	3 308	001
1	222	044	012	004	002	001	000	000	3 203	001
2	333	089	024	008	003	001	001	000	3 098	001
3	444	133	042	014	005	002	001	001	2 993	001
4	556	200	067	024	009	004	002	001	2 888	002
5		267	097	036	015	006	003	001	2 783	003
6		356	139	055	023	010	005	002	2 678	004
7		444	188	077	033	015	007	003	2 573	005
8		556	248	107	047	021	010	005	2 468	007
9			315	141	064	030	014	007	2 363	009
10			387	184	085	041	020	010	2 258	012
11			461	230	111	054	027	014	2 153	016
12			539	285	142	071	036	019	2 048	020
13				341	177	091	047	025	1 943	026
14				404	217	114	060	032	1 838	033
15				467	262	141	076	041	1 733	041
16				533	311	172	095	052	1 628	052
17					362	207	116	065	1 523	064
18					416	245	140	080	1 418	078
19					472	286	168	097	1 313	094
20					528	331	198	117	1 208	113
21						377	232	139	1 102	135
22						426	268	164	998	159
23						475	306	191	893	185
24						525	347	221	788	215
25							389	253	683	247
26							433	287	578	282
27							478	323	473	318
28							522	360	368	356
29								399	263	396
30								439	158	437
31								480	052	481
32								520		

* Reproduced from Mann, H. B., and Whitney, D. R. 1947. On a test of whether one of two random variables is stochastically larger than the other. *Ann. Math. Statist.* 18, 52-54, with the kind permission of the authors and the publisher.

Table K. Table of Critical Values of U in the Mann-Whitney Test*

Table K_I Critical Values of U for a One-tailed Test at α = .001 or for a Two-tailed Test at α = .002

n_1 \ n_2	9	10	11	12	13	14	15	16	17	18	19	20
1												
2												
3									0	0	0	0
4		0	0	0	1	1	1	2	2	3	3	3
5	1	1	2	2	3	3	4	5	5	6	7	7
6	2	3	4	4	5	6	7	8	9	10	11	12
7	3	5	6	7	8	9	10	11	13	14	15	16
8	5	6	8	9	11	12	14	15	17	18	20	21
9	7	8	10	12	14	15	17	19	21	23	25	26
10	8	10	12	14	17	19	21	23	25	27	29	32
11	10	12	15	17	20	22	24	27	29	32	34	37
12	12	14	17	20	23	25	28	31	34	37	40	42
13	14	17	20	23	26	29	32	35	38	42	45	48
14	15	19	22	25	29	23	36	39	43	46	50	54
15	17	21	24	28	32	36	40	43	47	51	55	59
16	19	23	27	31	35	39	43	48	52	56	60	65
17	21	25	29	34	38	43	47	52	57	61	66	70
18	23	27	32	37	42	46	51	56	61	66	71	76
19	25	29	34	40	45	50	55	60	66	71	77	82
20	26	32	37	42	48	54	59	65	70	76	82	88

Table K_{II} Critical Values of U for a One-tailed Test at α = .01 or for a Two-tailed Test at α = .02

n_1 \ n_2	9	10	11	12	13	14	15	16	17	18	19	20
1												
2					0	0	0	0	0	0	1	1
3	1	1	1	2	2	2	3	3	4	4	4	5
4	3	3	4	5	5	6	7	7	8	9	9	10
5	5	6	7	8	9	10	11	12	13	14	15	16
6	7	8	9	11	12	13	15	16	18	19	20	22
7	9	11	12	14	16	17	19	21	23	24	26	28
8	11	13	15	17	20	22	24	26	28	30	32	34
9	14	16	18	21	23	26	28	31	33	36	38	40
10	16	19	22	24	27	30	33	36	38	41	44	47
11	18	22	25	28	31	34	37	41	44	47	50	53
12	21	24	28	31	35	38	42	46	49	53	56	60
13	23	27	31	35	39	43	47	51	55	59	63	67
14	26	30	34	38	43	47	51	56	60	65	69	73
15	28	33	37	42	47	51	56	61	66	70	75	80
16	31	36	41	46	51	56	61	66	71	76	82	87
17	33	38	44	49	55	60	66	71	77	82	88	93
18	36	41	47	53	59	65	70	76	82	88	94	100
19	38	44	50	56	63	69	75	82	88	94	101	107
20	40	47	53	60	67	73	80	87	93	100	107	114

* Adapted and abridged from Tables 1, 3, 5, and 7 of Auble, D 1953 Extended tables for the Mann-Whitney statistic *Bulletin of the Institute of Educational Research at Indiana University*, 1, No 2, with the kind permission of the author and the publisher

Table K. Table of Critical Values of U in the Mann-Whitney Test* (*continued*)

Table K_{III} Critical Values of U for a One-tailed Test at α = .025 or for a Two-tailed Test at α = .05

n_1 \ n_2	9	10	11	12	13	14	15	16	17	18	19	20
1												
2	0	0	0	1	1	1	1	1	2	2	2	2
3	2	3	3	4	4	5	5	6	6	7	7	8
4	4	5	6	7	8	9	10	11	11	12	13	13
5	7	8	9	11	12	13	14	15	17	18	19	20
6	10	11	13	14	16	17	19	21	22	24	25	27
7	12	14	16	18	20	22	24	26	28	30	32	34
8	15	17	19	22	24	26	29	31	34	36	38	41
9	17	20	23	26	28	31	34	37	39	42	45	48
10	20	23	26	29	33	36	39	42	45	48	52	55
11	23	26	30	33	37	40	44	47	51	55	58	62
12	26	29	33	37	41	45	49	53	57	61	65	69
13	28	33	37	41	45	50	54	59	63	67	72	76
14	31	36	40	45	50	55	59	64	67	74	78	83
15	34	39	44	49	54	59	64	70	75	80	85	90
16	37	42	47	53	59	64	70	75	81	86	92	98
17	39	45	51	57	63	67	75	81	87	93	99	105
18	42	48	55	61	67	74	80	86	93	99	106	112
19	45	52	58	65	72	78	85	92	106	99	113	119
20	48	55	62	69	76	83	90	98	105	112	119	127

Table K_{IV} Critical Values of U for a One-tailed Test at α = .05 or for a Two-tailed Test at α = .10

n_1 \ n_2	9	10	11	12	13	14	15	16	17	18	19	20
1											0	0
2	1	1	1	2	2	2	3	3	3	4	4	4
3	3	4	5	5	6	7	7	8	9	9	10	11
4	6	7	8	9	10	11	12	14	15	16	17	18
5	9	11	12	13	15	16	18	19	20	22	23	25
6	12	14	16	17	19	21	23	25	26	28	30	32
7	15	17	19	21	24	26	28	30	33	35	37	39
8	18	20	23	26	28	31	33	36	39	41	44	47
9	21	24	27	30	33	36	39	42	45	48	51	54
10	24	27	31	34	37	41	44	48	51	55	58	62
11	27	31	34	38	42	46	50	54	57	61	65	69
12	30	34	38	42	47	51	55	60	64	68	72	77
13	33	37	42	47	51	56	61	65	70	75	80	84
14	36	41	46	51	56	61	66	71	77	82	87	92
15	39	44	50	55	61	66	72	77	83	88	94	100
16	42	48	54	60	65	71	77	83	89	95	101	107
17	45	51	57	64	70	77	83	89	96	102	109	115
18	48	55	61	68	75	82	88	95	102	109	116	123
19	51	58	65	72	80	87	94	101	109	116	123	130
20	54	62	69	77	84	92	100	107	115	123	130	138

* Adapted and Abridged from Tables 1, 3, 5, and 7 of Auble D 1953 Extended tables for the Mann-Whitney statistic *Bulletin of the Institute of Educational Research at Indiana University*, 1, No 2, with the kind permission of the author and the publisher

Table L. Table of Critical Values of K_D in the Kolmogorov-Smirnov Two-sample Test

(Small samples)

N	One-tailed test*		Two-tailed test†	
	α = .05	α = .01	α = .05	α = .01
3	3	—	—	—
4	4	—	4	—
5	4	5	5	5
6	5	6	5	6
7	5	6	6	6
8	5	6	6	7
9	6	7	6	7
10	6	7	7	8
11	6	8	7	8
12	6	8	7	8
13	7	8	7	9
14	7	8	8	9
15	7	9	8	9
16	7	9	8	10
17	8	9	8	10
18	8	10	9	10
19	8	10	9	10
20	8	10	9	11
21	8	10	9	11
22	9	11	9	11
23	9	11	10	11
24	9	11	10	12
25	9	11	10	12
26	9	11	10	12
27	9	12	10	12
28	10	12	11	13
29	10	12	11	13
30	10	12	11	13
35	11	13	12	
40	11	14	13	

* Abridged from Goodman, L A 1954 Kolmogorov-Smirnov tests for psychological research *Psychol Bull* 51, 167 with the kind permission of the author and the American Psychological Association

† Derived from Table 1 of Massey, F J, Jr 1951 The distribution of the maximum deviation between two sample cumulative step functions *Ann Math Statist* 22 126-127 with the kind permission of the author and the publisher

Table M. Table of Critical Values of D in the Kolmogorov-Smirnov Two-sample Test

(Large samples: two-tailed test)

Level of significance	Value of D so large as to call for rejection of H_0 at the indicated level of significance where $D = \text{maximum } \lvert S_{n_1}(X) - S_{n_2}(X) \rvert$
.10	$1.22 \sqrt{\frac{n_1 + n_2}{n_1 n_2}}$
.05	$1.36 \sqrt{\frac{n_1 + n_2}{n_1 n_2}}$
.025	$1.48 \sqrt{\frac{n_1 + n_2}{n_1 n_2}}$
.01	$1.63 \sqrt{\frac{n_1 + n_2}{n_1 n_2}}$
.005	$1.73 \sqrt{\frac{n_1 + n_2}{n_1 n_2}}$
.001	$1.95 \sqrt{\frac{n_1 + n_2}{n_1 n_2}}$

Table N. Table of Probabilities Associated with Values as Large as Observed Values of χ_r^2 in the Friedman Two-way Analysis of Variance by Ranks*

Table N_I. $k = 3$

N = 2		N = 3		N = 4		N = 5	
χ_r^2	p	χ_r^2	p	χ_r^2	p	χ_r^2	p
0	1.000	.000	1.000	.0	1.000	.0	1.000
1	.833	.667	.944	.5	.931	.4	.954
3	.500	2.000	.528	1.5	.653	1.2	.691
4	.167	2.667	.361	2.0	.431	1.6	.522
		4.667	.194	3.5	.273	2.8	.367
		6.000	.028	4.5	.125	3.6	.182
				6.0	.069	4.8	.124
				6.5	.042	5.2	.093
				8.0	.0046	6.4	.039
						7.6	.024
						8.4	.0085
						10.0	.00077

N = 6		N = 7		N = 8		N = 9	
χ_r^2	p	χ_r^2	p	χ_r^2	p	χ_r^2	p
.00	1.000	.000	1.000	.00	1.000	.000	1.000
.33	.956	.286	.964	.25	.967	.222	.971
1.00	.740	.857	.768	.75	.794	.667	.814
1.33	.570	1.143	.620	1.00	.654	.889	.865
2.33	.430	2.000	.486	1.75	.531	1.556	.569
3.00	.252	2.571	.305	2.25	.355	2.000	.398
4.00	.184	3.429	.237	3.00	.285	2.667	.328
4.33	.142	3.714	.192	3.25	.236	2.889	.278
5.33	.072	4.571	.112	4.00	.149	3.556	.187
6.33	.052	5.429	.085	4.75	.120	4.222	.154
7.00	.029	6.000	.052	5.25	.079	4.667	.107
8.33	.012	7.143	.027	6.25	.047	5.556	.069
9.00	.0081	7.714	.021	6.75	.038	6.000	.057
9.33	.0055	8.000	.016	7.00	.030	6.222	.048
10.33	.0017	8.857	.0084	7.75	.018	6.889	.031
12.00	.00013	10.286	.0036	9.00	.0099	8.000	.019
		10.571	.0027	9.25	.0080	8.222	.016
		11.143	.0012	9.75	.0048	8.667	.010
		12.286	.00032	10.75	.0024	9.556	.0060
		14.000	.000021	12.00	.0011	10.667	.0035
				12.25	.00086	10.889	.0029
				13.00	.00026	11.556	.0013
				14.25	.000061	12.667	.00066
				16.00	.0000036	13.556	.00035
						14.000	.00020
						14.222	.000097
						14.889	.000054
						16.222	.000011
						18.000	.0000006

Table N_{II}. $k = 4$

N = 2		N = 3		N = 4			
χ_r^2	p	χ_r^2	p	χ_r^2	p	χ_r^2	p
.0	1.000	.2	1.000	.0	1.000	5.7	.141
.6	.958	.6	.958	.3	.992	6.0	.105
1.2	.834	1.0	.910	.6	.928	6.3	.094
1.8	.792	1.8	.727	.9	.900	6.6	.077
2.4	.625	2.2	.608	1.2	.800	6.9	.068
3.0	.542	2.6	.524	1.5	.754	7.2	.054
3.6	.458	3.4	.446	1.8	.677	7.5	.052
4.2	.375	3.8	.342	2.1	.649	7.8	.036
4.8	.208	4.2	.300	2.4	.524	8.1	.033
5.4	.167	5.0	.207	2.7	.508	8.4	.019
6.0	.042	5.4	.175	3.0	.432	8.7	.014
		5.8	.148	3.3	.389	9.3	.012
		6.6	.075	3.6	.355	9.6	.0069
		7.0	.054	3.9	.324	9.9	.0062
		7.4	.033	4.5	.242	10.2	.0027
		8.2	.017	4.8	.200	10.8	.0016
		9.0	.0017	5.1	.190	11.1	.00094
				5.4	.158	12.0	.000072

Table O. Table of Probabilities Associated with Values as Large as Obs…ed Values of H in the Kruskal-Wallas One-way Analysis of Variance by Ranks*

n_1	n_2	n_3	H	
2	1	1	2 7000	500
2	2	1	3 6000	200
2	2	2	4 5714	067
			3 7143	200
3	1	1	3 2000	300
3	2	1	4 2857	100
			3 8571	133
3	2	2	5 3572	029
			4 7143	048
			4 5000	067
			4 4643	105
3	[illegible]	1	5 1429	043
			4 5714	100
			[illegible]	129
3	3	2	[illegible]	011
			5 3611	032
			5 1389	061
			4 5556	100
			4 2500	121
3	3	3	7 2000	004
			6 4889	011
			5 6889	029
			5 6000	050
			5 0667	086
			4 6222	100
4	1	1	3 5714	200
4	2	1	4 8214	057
			4 5000	076
			4 0179	114
4	2	2	6 0000	014
			5 3333	033
			5 1250	052
			4 4583	100
			4 1667	105
4	3	1	5 8333	021
			5 2083	050
			[illegible]	057
			4 0556	093
			3 8889	129
4	3	2	6 4444	008
			6 3000	
			5 4444	
			5 4000	[illegible]
			4 5111	098
			4 4444	102
4	3	3	6 7455	010
			6 7091	013
			5 7909	046
			5 7273	050
			4 7091	092
			4 7000	101
4	4	1	6 6667	010
			6 1667	022
			4 9667	048
			4 8667	054
			4 1667	082
			4 0667	102
4	4	2	7 0364	[illegible]
			6 8727	011
			5 4545	046
			5 2364	052
			4 5545	098
			4 4455	103
4	4	3	7 1439	010
			7 1364	011
			5 5985	049
			5 5758	051
			4 5455	099
			4 4773	102
4	4	4	7 6538	008
			7 5385	011
			5 6923	049
			5 6538	054
			4 6539	097
			4 5001	104
5	1	1	3 8571	143
5	2	1	5 2500	036
			5 0000	048
			4 4500	071
			4 2000	095
			4 0500	119

Table O. Table of Probabilities Associated with Values as Large as Observed Values of H in the Kruskal-Wallis One-way Analysis of Variance by Ranks* (*continued*)

n_1	n_2	n_3	H	p
5	2	2	6 5333	008
			6 1333	013
			5 1600	034
			5 0400	056
			4 3733	090
			4 2933	122
5	3	1	6 4000	012
			4 9600	048
			4 9711	052
			4 0178	095
			3 8400	123
5	3	2	6 9091	009
			6 8218	010
			5 2509	049
			5 1055	052
			4 6509	091
			4 4945	101
[illegible]	3	3	7 0788	009
			6 9818	011
			5 6485	049
			5 5152	051
			4 5333	097
			4 4121	[illegible]
5	4	1	6 9545	008
			6 8400	011
			4 9855	044
			4 8600	056
			3 9873	098
			3 9600	102
	4	2	7 2045	009
			7 1182	010
			5 2727	049
			5 2682	050
			4 5409	098
			4 5182	101
5	4	[illegible]	7 4449	010
			7 3949	011
			5 6564	049
			5 6308	050
			4 5487	099
			4 5231	103
5	4	4	7 7604	009
			7 7440	011
			5 6571	049
			5 6176	050
			4 6187	100
			4 5527	102
	5	1	7 3091	009
			6 8364	011
			5 1273	046
			4 9091	053
			4 1091	086
			4 0364	105
5	5	2	7 3385	010
			7 2692	010
			5 3385	047
			5 2462	051
			4 6231	097
			4 5077	100
5	5	3	7 5780	010
			7 5429	010
			5 7055	046
			5 6264	051
			4 5451	100
			4 5363	102
5	5	4	7 8229	010
			7914	010
			5 6657	049
			5 6429	050
			4 5229	099
			4 5200	101
5	5	5	8 0000	009
			7 9800	010
			5 7800	049
			5 6600	051
			4 5600	100
			4 5000	102

* Adapted and abridged from Kruskal W H and Walli[illegible] A 19[illegible] Use of ranks in one-criterion variance analysis *J Amer Statist Ass* 47 614–617 with the kind permission of the authors and the publisher (The corrections to this table given by the authors in Errata [illegible] *Statist Ass* 48 910, have been in[illegible]orporated.)

Table P. Table of Critical Values of r_s, the Spearman Rank Correlation Coefficient*

N	Significance level (one tailed test)	
	05	01
4	1 000	
5	900	1 000
6	829	943
7	714	893
8	643	[illegible]
9	600	783
10	564	746
12	506	[illegible]
14	456	645
16	425	601
18	399	564
20	377	534
22	359	508
24	343	485
26	329	465
28	317	448
30	[illegible]	432

*Adapted from Olds, E G 1938 Distributions of sums of squares of rank differences for small numbers of individuals *Ann Math Statist* 9 133-148 and from Olds E G 1949 The 5% significance levels for sums of squares of rank differences and a correction *Ann Math Statist* 20 117-118 with the kind permission of the author and the publisher

Table Q. Table of Probabilities Associated with Values as Large as Observed Values of *S* in the Kendall Rank Correlation Coefficient

S	Values of N				S	Values of N		
	4	5	8	9		6	7	10
[illegible]	625	592	548	540	1	500	500	500
	375	408	452	460	[illegible]	360	386	431
4	167	242	360	381	5	235	281	364
6	042	117	274	306	7	136	191	300
8		042	199	238	9	068	119	242
10		0083	138	179	11	028	068	190
12			089	130	13	0083	035	146
14			054	090	15	0014	015	108
16			031	060	17		0054	078
18			016	038	19		0014	054
20			0071	022	21		00020	036
22			0028	012	23			023
24			00087	0063	25			014
26			00019	0029	27			0083
28			000025	0012	29			0046
30				00043	31			0023
32				00012	33			0011
34				000025	35			00047
36				0000028	37			00018
					39			000058
					41			000015
					43			0000028
					45			00000028

*Adapted by permission from Kendall, M G, *Rank correlation methods* Charles Griffin & Company, Ltd, London, 1948, Appendix Table 1, p 141

Table R. Table of Critical Values of s in the Kendall Coefficient of Concordance*

k	N					Additional values for N = 3	
	3†	4	5	6	7	k	s
Values at the 05 level of significance							
3			64 4	103 9	157 3	9	54 0
4		49 5	88 4	143 3	217 0	12	71 9
5		62 6	112 3	182 4	276 2	14	83 8
6		75 7	136 1	221 4	335 2	16	95 8
8	[illegible]	101 7	183 7	299 0	453 1	18	107 7
10	60 0	127 8	231 2	376 7	571 0		
15	89 8	192 9	349 8	570 5	863 9		
20	119 7	258 0	468 5	764 4	1 158 7		
Values at the 01 level of significance							
3			75 6	122 8	185 6	9	75 9
4		61 4	[illegible]	176 2	265 0	12	103 5
5		80 5	[illegible]	229 4	343 8	14	121 9
6		99 5	176 1	282 4	422 6	16	140 2
8	[illegible]	137 4	242 7	388 3	579 9	18	158 6
10	85 1	175 3	309 1	494 0	737 0		
15	131 0	269 8	475 2	758 2	1 129 5		
20	177 0	364 2	641 2	1,022 2	1,521 9		

*Adapted from Friedman, M 1940 A comparison of alternative tests of significance for the problem of *m* rankings *Ann Math Statist* 11 86-92 with the kind permission of the author and the publisher

†Notice that additional critical values of s for N = 3 are given in the right hand column of this table

Table S. Table of Factorials

N	N!
0	1
1	1
2	2
3	6
4	24
5	120
6	720
7	5040
8	40320
9	362880
10	3628800
11	39916800
12	479001600
13	6227020800
14	87178291200
15	1307674368000
16	20922789888000
17	355687428096000
18	6402373705728000
19	121645100408832000
20	2432902008176640000

Table T. Table of Binomial Coefficients

N	$\binom{N}{0}$	$\binom{N}{1}$	$\binom{N}{2}$	$\binom{N}{3}$	$\binom{N}{4}$	$\binom{N}{5}$	$\binom{N}{6}$	$\binom{N}{7}$	$\binom{N}{8}$	$\binom{N}{9}$	$\binom{N}{10}$
0	1										
1	1	1									
2	1	2	1								
3	1	3	3	1							
4	1	4	6	4	1						
5	1	5	10	10	5	1					
6	1	6	15	20	15	6	1				
7	1	7	21	35	35	21	7	1			
8	1	8	28	56	70	56	28	8	1		
9	1	9	36	84	126	126	84	36	9	1	
10	1	10	45	120	210	252	210	120	45	10	1
11	1	11	55	165	330	462	462	330	165	55	11
12	1	12	66	220	495	792	924	792	495	220	66
13	1	13	78	286	715	1287	1716	1716	1287	715	286
14	1	14	91	364	1001	2002	3003	3432	3003	2002	1001
15	1	15	105	455	1365	3003	5005	6435	6435	5005	3003
16	1	16	120	560	1820	4368	8008	11440	12870	11440	8008
17	1	17	136	680	2380	6188	12376	19448	24310	24310	19448
18	1	18	153	816	3060	8568	18564	31824	43758	48620	43758
19	1	19	171	969	3876	11628	27132	50388	75582	92378	92378
20	1	20	190	1140	4845	15504	38760	77520	125970	167960	184756

Glossary

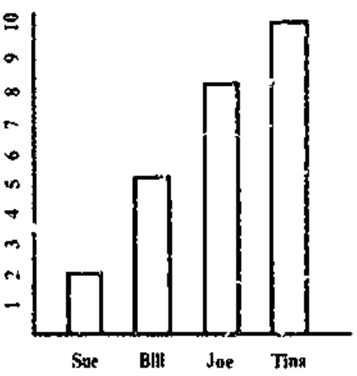
ABILITY TESTS
Tests that purport to measure an individual's over-all facility in doing given things. Often a distinction is attempted between that facility which results from heredity and that which results from learning. In such cases, *ability* tests are usually applied to the "native" aspect and *achievement* tests to the learned aspect.
EX—IQ tests, Dominance test

ACHIEVEMENT TESTS
Tests that purport to measure an individual performance or competence relative to a given subject, usually a subject taught in the schools. Achievement tests are concerned with learned outcomes (generally knowledge and/or understanding) rather than "native" capacity or ability to learn the subject.
EX—Metropolitan Achievement Tests, MacGinitie

AGE EQUIVALENTS
A method of expressing scores on standardized tests. The raw score typical of pupils of different ages is determined and then any pupil's raw score may be converted to the age to which it pertains. Usually given in years and months.
EX—Mental age = 12 6, reading age = 10 4

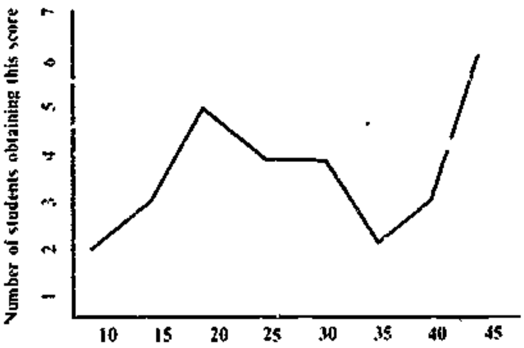
APPLIED RESEARCH
Aims to solve an immediate practical problem. It is research performed in relation to actual problems and under conditions in which they are found in practice.
EX—Is *oral presentation or written presentation* more effective in *improving students' performance on tests*? Is *reading comprehension improved* by using the *individualized approach* as opposed to the *traditional reading group approach*?

BAR GRAPH
Any graphic presentation that uses bars of various length to symbolize differences in quantity, size, amount, etc.

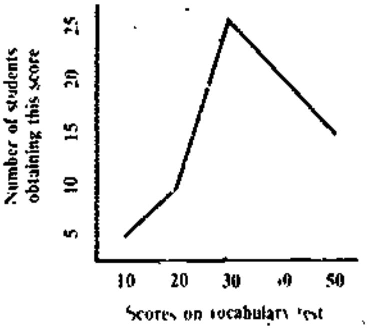


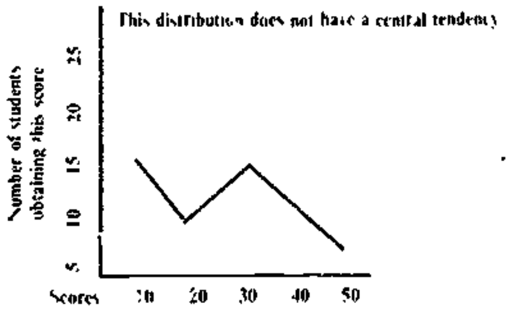
BASIC RESEARCH
Has as its aim obtaining data that can be used to formulate, expand, or evaluate theory. Its essential aim is to expand the frontiers of knowledge without regard to practical application, though the results may be used to solve practical problems.
EX—Is there a relationship between disruptive behavior and reading achievement? Is there a relationship between oral language skills and written language skills?

BIMODAL
A distribution of measures, particularly test scores, with two foci of central tendency rather than one. A superficial indication of bimodality is the presence of two modes separated by scores or score intervals whose frequency is appreciably less than that of the modes. Bimodality in a distribution can be suggestive of several attributes of the group or of the test or other measuring procedure in use. It often indicates that the group which is bimodal involves two subgroups having important mean differences as to age, mentality, reading ability, nationality, etc.
Ex—



CENTRAL TENDENCY
In a distribution of scores or other measures, the point or interval at which a plurality or majority of scores tends to cluster. Unless there is such a clustering, the distribution has no central tendency.
EX—All distributions have a central tendency





CHECK LISTS

A device used in observation to direct attention to factors to be observed and sometimes to provide space for recording ratings or comments relative to them

EX—

1 Can read accurately at a rate of 15 words/minute
2 Can discriminate between two similarly spelled words
3. Can follow orally given direction

C.A. (CHRONOLOGICAL AGE)

A child's age expressed in years and months Used in reckoning the intelligence quotient and any other index involving a comparison between skill or knowledge and age.

EX—Billy is five years and three months old His C A = 5 3

CLASSIFICATION

One of four basic forms of measurement (types of measurement symbols). Involves the establishment of categories (classification), the designation of symbols for the categories, and then the assignment of the symbols to phenomena according to the category to which they belong. This is sometimes referred to as the Nominal Level of measurement

EX—Blood typing, draft classifications A B, C, D F as course marks

COEFFICIENT OF CORRELATION (r)

A measure of the degree with which the variation of one variable is associated with variation of another variable

EX—

	Intelligence	Grades	Conclusion
Joe	118 (2)	3 5 gpa	
Sue	103 (4)	2 5 gpa	If you have high in-
Fred	110 (3)	2 0 gpa	telligence you'll
Linda	130 (1)	4 0 gpa	have high grades

CONCEPT

An abstraction from observed events. It is a word that represents the similarities or common aspects of objects or events that are otherwise quite different from one another The purpose of a concept is to simplify thinking by including a number of events under one general heading.

EX—Words such as chair, dog, tree, liquor and thousands of others in our language represent common aspects of otherwise diverse things

CONSTRUCTS

Higher level abstractions that cannot be easily illustrated by pointing to specific objects or events

EX—Problem-solving ability, motivation, justice or intelligence

CORRELATION

The statistical technique used for measuring the degree of relationship between two variables is called *correlation* Correlation shows us the extent to which values in one variable are linked or related to values in another variable An important use of such measurement is in prediction When correlational analysis indicates some degree of relationship between two variables, we can use the information about one of them to make predictions about the other

EX—Having found that intelligence and achievement are correlated, one can make predictions about the future achievement of school children from the results of a test of intelligence given at the beginning of the school year The accuracy of such prediction is a function of the degree of relationship, that is, the extent of the correlation The higher the correlation, the more accurate the predictions.

CRITERION

Anything with which a measuring procedure is compared in determining its validity Specifically a measuring procedure for a given phenomenon for which exemplary validity is claimed or assumed and with which other similar procedures are asked to have high positive correlations

EX—To show your reading program's effectiveness, you decide all children must advance one year in ability to read The improvement goal of one year is your criterion

CUMULATIVE FREQUENCY

A column in a conventional tabulation of scores or other measures that shows the frequency of scores up to and including any given interval.

EX—

No of students receiving score	Score	Cumulative frequency
1	98	9
2	95	8
3	90	6
2	78	3
1	75	1

DERVIED SCORE

A test score that has been converted to an index of rank, scale position, or classification, as distinct from a raw score, which is the number of correct responses or the immediate numerical weight given the test Percentile rank, standard scores, mental age

EX—A child gets 9 spelling words right out of 10, he got a 90% He did better than all the rest of the class His derived score is A

DEPENDENT VARIABLES

Variables that are a consequence of or dependent upon antecedent variables In research studies, the dependent variable is the phenomenon that is the object of study and investigation It is the one that must always be assessed

EX—This is sometimes called assigned variable

DESCRIPTION

An informal type of measurement expression used to indicate the status of phenomena in which ordinary language is used The information is not quantified This is also called the Nominal level of measurement.

EX—Scale rank and classification symbols associated with appraisal of citizenship, study habits, social adjustment

DESCRIPTIVE RESEARCH

Describes and interprets *what* is. It is concerned with conditions or relationships that exist, practices that prevail, beliefs, points of view, or attitudes that are held, processes that are going on, effects that are being felt, or trends that are developing

EX—There are several subcategories of descriptive research

a Case studies
b Surveys
c Developmental studies
d Follow-up studies
e Documentary analysis
f Trend studies
g Correlational studies

DEVIATION

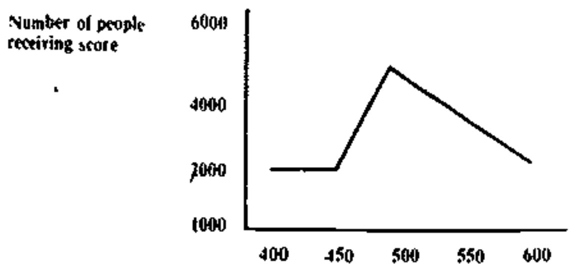
Departure from a given condition In particular, the numerical difference between a test score or other measure of an individual and given point of reference, usually the mean of a group of test scores or other measures

EX—The class average on a test = 85, Jill received a 35 This is a large deviation

DISTRIBUTION

A table or graph showing the scores or other measures found for a group, so arranged that the number who have a given score or who fall within a given range of scores is apparent

EX—



EQUIVALENT FORM

Either of two forms of a measuring instrument, particularly a standardized test which is parallel in content, difficulty, and norms, but different as to terms
EX—Stanford-Binet Forms L and M

EVALUATION

The process of assigning symbols to phenomena. These symbols signify the worth of the phenomena relative to some scheme of value
EX—Grading student's paragraphs A, B, C, D or F

EXPERIMENTAL DESIGN

The conceptual framework within which the experiment is conducted. It serves two functions. It provides opportunity for the comparisons required by the hypotheses of the experiment and it enables the experimenter through his statistical analysis of the data to make a meaningful interpretation of the results of the study
EX—
One group → pretest → treatment → post-test
Exp. group → pretest → treatment I → post-test
Control group → treatment II →

EXPERIMENTAL RESEARCH

A scientific investigation in which an investigator manipulates and controls one or more independent variables and observes the dependent variable or variables for variation concomitant to the manipulation of the independent variables. Its major purpose is to determine "what may be."
EX—Will subjects receiving individualized instruction achieve more than the students receiving traditional reading group instruction?

EX POST FACTO RESEARCH

Similar to experimental research except investigator cannot directly manipulate independent variables
EX—Did my students achieve less than a comparable class because they didn't have a regular teacher?

EXTERNAL CRITERION

One needs an external criterion that is known to be a measure of the variable involved and can be used to compare one's predictions. Success in college, as reflected by grade point average is a clearly defined external criterion for validating those tests that are constructed for the selection of college applicants
EX—Number of library books read outside of class assignments is an *external criterion* of reading enjoyment

FREQUENCY

Refers in statistics to the number of times a score is repeated or to the number of scores appearing in a given interval
EX—Joe and Sally got 85% on the spelling test. Fred, Donna, Shirley and Bob got 80%. Frequency for 85 = 2. Frequency for 80 = 4

FREQUENCY DISTRIBUTION

A systematic arrangement of individual measures from highest to lowest. The use of this technique merely involves making a list of the individual measures in a column, with the highest measure at the top, the next highest, second from the top, continuing down until the lowest measure is recorded at the bottom of the column
EX—

# of words correct	# of people rec'd score frequency
10	2
7	6
5	5
3	2
1	3

GRADE EQUIVALENT

The grade for which the ability is typical
EX—Kathy is achieving at the 4th grade level, 3rd month on the Metropolitan Achievement tests. Her grade equivalent is 4.3

HISTORICAL RESEARCH

A procedure supplementary to observation. A process by which the historian seeks to test the truthfulness of the reports of observations made by others. Its major purpose is to tell "what was."
EX—Tracing the evaluation of the open classroom

HYPOTHESIS

A tentative proposition suggested as a solution to a problem or as an explanation of some phenomenon. It presents in simple form a statement of the researcher's expectations relative to a relationship between variables with the problem. It is then tested in a research study
EX—Students who attend a remedial reading clinic five hours a week will improve their scores on the Metropolitan Primary Achievement Tests significantly more than students who attend the clinic for only three hours a week

INDEPENDENT VARIABLES

Variables that are antecedent to the dependent variable are called independent variables. This is the factor that is measurably separate and distinct from the dependent variable but may relate to the dependent variable. Many factors that may function as independent variables are discriminate aspects of the environment, such as, social class, home environment, and classroom conditions. In addition, characteristics of the individual himself such as age, sex, intelligence and motivation—may be independent variables that can be related to the dependent variable
EX—A child's height (dependent variable) would be dependent to a certain extent upon his age (independent variable). These terms are often used even in the absence of empirical or theoretical reasons for considering one to be the antecedent and the other to be the consequence. They are used to indicate the direction of prediction—from individuals' positions on the independent variable to their positions on the dependent variable. This is sometimes called the active variable. Examples of Dependent and Independent variables:

1. *Reading achievement* (D.V.) is affected by *Self-concept* (I.V.)
2. *Word knowledge* (D.V.) is dependent on *Social economic status* (I.V.)
3. *Reading achievement* (D.V.) is dependent on *Reading enjoyment* (I.V.)

INFERENTIAL STATISTICS

The process of going from the part to the whole. A population comprises all the possible cases (persons, objects or events) that constitute a known whole. A sample is a portion of a population
EX—A representative sample of 1000 six year old children obtain a mean raw score of 48 on the WISC. It is then inferred that the "average" 6 year old will obtain a score of 48 of the WISC

INFERRED DIMENSION
A property or quality of a phenomenon not itself observable but imput or *inferred* to a phenomenon.
EX—A child's knowledge is measured by an I Q test

INTERCORRELATION
A term applied to each of the correlations among a group of tests Usually displayed in tables showing the correlation of each test with each of the other test. They are then used to show the extent of interrelationships among a certain group of tests
EX—If a child scores high on the reading comprehension tests in Gates-MacGinite test, then he will probably score high on the vocabulary test.

INTERVAL SCALE
Not only indicates the relative position of individuals but also provides additional information about these positions because this type of scale uses predetermined equal intervals Such scales do not necessarily have a true zero point. Arbitrary zero points may be used, but such points are by no means absolute. Consider intelligence tests, for example. In these tests there are zero points and it is conceivable that one's score could be zero, but zero scores in these tests do not mean zero intelligence. For this reason it is not possible to compare an intelligence test score of 75 with a score of 150 and say the latter score is twice as high as the former
EX—Number of correct spelling words on an exam Score on the Stanford-Binet test.

LEVEL OF SIGNIFICANCE
A statistical term used to indicate the amount of confidence in whether or not the difference between two means, two percentages or other comparable measures is statistically significant (not due to chance) Also referred to as significance of difference and statistical difference.
EX—If Suzie got a 93% on a spelling test and Bill got a 90%, is Suzie a significantly better speller, or is her better score simply due to chance?

MEASUREMENT
The assignment of a symbol, often a number, so as to characterize the status of a phenomenon relative to some dimension, usually by indicating its scale position, its rank, or its classification per this dimension.
EX—Joe got 15 out of 20 spelling words correct, or 75% of them correct This is a measurement of correct replies If the score he received is the fourth highest test score, it is a measurement of his rank compared to others

MEAN
The most widely used measure of central tendency is the mean, which is popularly known as the average or *arithmetic average* It is the sum of all the values in a distribution divided by the number of cases. In terms of a formula it is: $\bar{X} = EX/N$
where, $\bar{X}$ = the mean
E = the sum of
X = each of the values in the distribution
N = number of cases

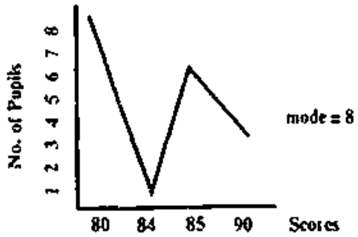
EX—The average or mean Reading test score -

Jo -	8
Sally -	3
Mike -	6
Tom -	7

24/4 = 6 = ave

MEDIAN
The score or point that divides a distribution of scores into two equal groups with half of the scores falling above and half below Used as a representative score or a measure of central tendency
EX—Scores 4 12 18 21 26 18 is the center score, it is the median

MODE
The score or measure that occurs most frequently in a distribution
EX— 3 students got 90 on their exam
6 students got 85 on their exam
1 student got 84 on his/her exam
8 students got 80 on their exam



NOMINAL SCALE
The simplest type of scale and provides the lowest level of quantification of the objects to be measured A nominal scale simply sorts objects, or classes of objects, into mutually exclusive categories. Our data will only tell us how many of the subjects belong to each group, or how many students of a class are boys and how many are girls. Dividing individuals into such categories as smokers and nonsmokers, Democrats, Republicans, and independents, elementary, junior high, and secondary, tall and short, and so on, are all examples of nominal scales.
EX—When we label the experimental units in a study as groups A, B ., and D, or when we divide the students of a class into boys and girls, we are using a nominal scale in each of these examples

NORMAL CURVE
A symmetrical distribution of measures with the same number of cases at specified distances below the mean as above the mean Its mean is the point below which exactly 50% of the cases are located. The median and the mode in such a distribution are identical values and coincide with the mean. In a normal curve, most of the cases concentrate near the mean
EX—

NORMS
Statistics based upon a standardization group or a group that is purported to be representative of a much larger population These norms are thus assumed to be representative of large groups
EX—All fifth-grade children or all twelve-year-olds, grade, age percentile, and standard score norms are the most common forms

OBSERVATION
The most widely used and usually most crude method of behavioral measurements Involves direct perception of the dimensions of the phenomenon being measured With appropriate attentional, perceptual, and recording aids, observation can be a highly reliable procedure
EX—Frequency count—Phillip got out of his seat without permission six times in the fifty minute reading period Interval count—Mary was not attending to the lesson for 40% of the thirty second intervals recorded

OPERATIONAL DEFINITION
Ascribes meaning to a concept or construct by specifying the operations that must be performed in order to measure the concept. This type of definition is essential in research, since data must be collected in terms of observable events. When one defines a concept or construct operationally, he chooses discriminable events as indicators of the abstract concept and devises operations to obtain data relevant to the concepts. An operational definition thus refers to the operations by which an investigator may measure a concept. These are essential to research because they permit investigators to measure abstract concepts and constructs and permit a scientist to move from the level of constructs and theory to the level of observation, upon which science is based.
EX—Operationalized definition of achievement—scores obtained on the Stanford Diagnostic Achievement Test. Operationalized definition of reading enjoyment—number of books read outside of class, not for assignments.

ORDINAL SCALE
The use of the ordinal scale permits the sorting of objects or classes of objects on the basis of their standing relative to each other. This scale not only categorizes but also ranks the objects on the basis of some criterion. A teacher who ranks his students on the basis of their intelligence, achievement, class participation, discipline, creativity, or any other characteristic is making use of an ordinal scale.
EX—Rank in class, percentile rank, percentiles

POPULATION
Used in an abstract sense in measurement and statistics to indicate any given group of things, the total group in question not just part of it.
EX—All the pupils in the sixth grade in your school district is the population from which your sample (the children in your sixth grade class) is taken.

PRACTICE EFFECT
It is known that a performance of any task affects a reperformance of that task, usually in the direction of improvement. *Practice effect* is the term for the significance of such reperformance when the same test is administered to the same individual more than once.
EX—When pupils do better on a quiz the second time it is given in a week, is this because they know the material better or because they have had practice with the question.

PRETEST
Any measuring instrument (usually an achievement test) administered prior to a period of instruction, an experiment, or other circumstance of interest. As a rule pretests are used to establish the initial status of pupils so that the amount of their learning may be judged from the results of a later retest.
EX—Students are given the Metropolitan Primary Achievement test in September and again in May. The tests in September would be a pretest.

PROBABILITY
As applied to behavioral measurement, the concept that any measure or statistic is somewhat subject to chance variation. Hence it deviates from some theoretically "true" measure. Such deviation is commonly called error and its probable extent can be determined and stated mathematically. *See Level of Significance*
EX—There is .05 or 5% chance that these scores were obtained by chance.

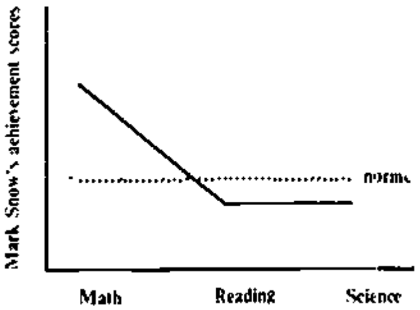
PRODUCT ANALYSIS
A basic procedure of educational evaluation in which the things that pupils produce in the course of instruction are appraised in appropriate ways and given scores or ratings.
EX—Compositions, outlines

PRODUCT MOMENT FORMULA
A widely used formula for the correlation coefficient. Let Zx be the standard score for variable x. If the pairs of Zx's and Zy's for each individual are multiplied, then added for all individuals and divided by the number of cases, the result is the product moment formula for the correlation coefficient. The correlation coefficient is the mean of the set of products of standard scores for the two variables.

EX— $r = \Sigma \frac{(Zx\,Zy)}{n}$

Zx = Z - scores for all x
Zy = Z - scores for all y
n = number of subjects

PROFILE
An analytic graphic presentation of a pupil's scores on a test battery, scores on parts of a given test, marks in several school subjects, ratings on several personality variables, etc.
EX—



RANDOM SAMPLING
The basic characteristic of random sampling is that all members of the population have an equal and independent chance of being included in the sample. That is, for every pair of elements x and y, x's chance of being selected equals y's chance, and the selection of x in no way affects y's probability of selection.
EX—Mary, Joe and Sally are in Ms. Brown's class. She decides to choose 2 of them to do an experiment. She puts all of their names in a hat. They all have an equal chance to be chosen and if one is chosen this does not affect the chances of the others being chosen.

RANGE
The difference between the highest and lowest scores in a given distribution of scores.
EX—If the highest score in a distribution is 74 and the lowest is 30, the range would be R = 74 - 30 = 44

RANKING
The process of ordering the constituents of a group in terms of some dimension. Rank numbers indicate the relative position of the constituents.
EX—Scores on a reading achievement test

S	63	61	60	60	60	55	51
R	1	2	4	4	4	6	7.5

RATING
A direct appraisal of a dimension in terms of some descriptive scale or verbal classification scheme.
EX—Children are rated by their teacher for their disruptive behavior 1 - very disruptive, 2 - average, 3 = quiet

RATIO SCALE
The highest level of measurement is provided by a ratio scale. In addition to having equal intervals, a ratio scale measures from a

meaningful zero Most physical measures have a meaningful zero The scale used in education measurements are seldum of this level of measurement.

EX—Using a ratio scale we can say that John is 48 inches tall, Ralph is 45 inches tall, and Paul is 44 inches tall, but using an interval scale we are only able to say that John is 3 inches taller than Ralph, who is one inch taller than Paul Not only can we say that the difference between 60 and 90 pounds is the same as the difference between 90 pounds and 120 pounds, but we can say that 120 pounds is twice as heavy as 60 pounds We can do this because zero weight is an actual possibility

RAW SCORE

The first quantitative untreated result obtained in sconng a test

EX—Bill got 98% on vocabulary test
Jill got 83% on vocabulary test
Tony got 78% on vocabulary test

READING GRADE

A type of "norm" score derived from standardized tests that states a pupil's ability to read in terms of grade equivalents Reading grade means the school grade whose average performance is must like that of the pupil in question. By interpolation, the reading grade be fractional As with reading age, reading grade refers only to a given standardized test

EX—Ann received a 6.3 on the Gates-MacGinities Test She is performing at sixth grade, three month level of achievement.

RELIABILITY

The extent to which a measuring device is constant in measuring whatever it measures.

EX—Will Greg receive approximately the same score on the Reading Achievement test if he takes it a second time two weeks after he took it the first time?

RESEARCH PROBLEM

A question concerning the relationships existing between sets of events (variables) in education. Research is conducted in order to find answers to these questions. One of the most fruitful sources for the beginning researcher is his experience as an educational practitioner. Decisions must be made daily about the probable effects of educational experience on pupil behavior

EX—What is the effectiveness of using verbal instructions compared to written ones?

RETEST (also called a Post-test)

A test readministered at the end of a period of instruction or other activity, the result of which is to be compared with an earlier administration of a test.

EX—A list of spelling words are given at the beginning of the week to determine whi ' words a child needs to learn After working on the words all week, a test is given Friday to find if the pupils learn the words The test on Friday would be considered a retest or post-test

RHO (Q)

The rank-difference measure of correlation Individuals are assigned ranks with respect to each of two variables, and for each individual the difference (d) in rank is determined. These differences are squared and summed for all cases and substitution is made in the following formula

EX— $Q(rho) = \frac{1 - 6 \Sigma d^2}{N(N^2 - 1)}$

SAMPLE

A sample is a portion of a population

EX—The children in Ms Smith's class is a sample of the population of sixth grade students in that district Reading Group A is Ms Smith's class is a sample of her sixth grade class.

SCALING

Measurement in terms of defined and precise units that represent given amounts or degrees of some dimension. Scale numbers indicate the number of units and hence the amount or degree of the dimension Scale numbers refer to a fixed point of reference, usually a zero.

EX—Rate your agreement with this statement on a scale of 1 to 5
1 = strongly agree
2 = agree
3 = undecided
4 = disagree
5 = strongly disagree

SCORING

A process of assigning a score (usually a number or letter symbol) to a test or pupil product. For a test, this is often done by comparing a paper with the key, marking the questions answered correctly and adding up the total.

EX—Bill got 9 words right Score = 9
Judy got 15 words right Score = 15
Ruth got 12 words right Score = 12

SELF-EVALUATION

Any of many concepts and procedures concerned with an individual observing and judging his own performance, achievement, or adjustment.

EX—Coopersmith Self-Esteem Inventory

STANDARD DEVIATION

An index of variation in a group of mesures It represents the square root of the mean of the squared deviations of the individual measures.

EX— $SD = \sqrt{\frac{\Sigma d^2}{N-1}}$

d^2 = difference between score and mean
n = number of subjects

STANDARD SCORE (z score)

A general term referring to any of a number of scores that indicate how many standard deviations a measurement is above or below the mean. It is found by determining the difference between the raw score (X) and the mean ($\bar{X}$) and dividing by the standard deviation (S)

EX— $Z = \frac{X - \bar{X}}{S}$

X = Bills score
$\bar{X}$ = Class average
X = Standard deviation (equation needed)

STANDARDIZED TESTS

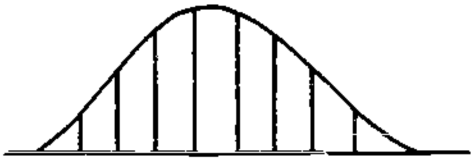
Tests, usually published, which have been preadministered to a population of known characteristics and yield scores in terms of this population This population is selected so as to be a representative sample of the total population for which the test is designed

EX—Stanford-Binet Intelligence Test, Metropolitan Primary Achievement Test, Ginn Reading Achievement Test

STANINE

Any one of nine intervals on a scale of standard scores The "stanine" (abbreviation for standard-nine) scale spans the normal curve in nine intervals of size equal to one half of a standard score The stanine intervals have values from 1 to 9 and the middle interval, 5, extends from standard score − ¼ to + ¼

EX—



STATISTICAL PROCEDURES

Basic methods of handling quantitative information in such a way as to make that information meaningful. These procedures have two principal advantages for the researcher First, they enable him/her to describe and summarize his observations Such techniques are called *descriptive statistics*. Second, they help him/her determine how reliably (s)he can infer that phenomena observed in a limited group, *a sample*, will also occur in the unobserved larger population of concern, from which the sample was drawn. In other words, how well (s)he can employ inductive reasoning to infer that what (s)he observes in the part will be observed in the whole. For problems of this nature (s)he will need to employ *inferential statistics*
EX—Finding the mean score—*descriptive statistics* Finding if there is significant differences-*inferential statistics*

STATISTIC(S)

Any derived quantity obtained from a set of raw scores or measures
EX—N, mean, standard deviation, median, mode, quartile deviation, correlation coefficient

TESTS

Any of a great number of procedures in which individuals respond to a common stimulation in comparable ways and which yield a measure of the individuals with respect to one or more dimensions
EX—Achievement tests, Personality tests, Spelling tests, Performance tests, Ability tests

VALIDATION

The process of establishing on the basis of empirical data the validity of a test, usually a standardized one, by comparing its results with one or more criteria Typically involves, as a minimum item analysis, correlation of results with other test scores, analysis of distributions of scores, and determination of reliability
EX—See test manual or Mental Measurement Yearbook by Buros to find out about a test validation

VALIDITY

The extent to which an instrument measures what it is supposed to measure
EX—A test measures a pupil's reading comprehension not size of vocabulary or general knowledge

VARIABLE

A concept that can take on different values
EX—It can vary within an individual from one time to another, between individuals at the same time, between the averages for groups, and so on. Social class, sex, motivation, intelligence quotient, and spelling test scores are other examples of variables Educational researchers are interested in determining how such variables are related to each other

VARIANCE

The mean of the squared deviation scores

EX— $S^2 = \Sigma \frac{(X-\bar{X})^2}{N}$

$X - \bar{X}$ = difference between score and mean
N = number of subjects

References

Andrews, Frank M., Laura Klem, Terrence Davidson, Patrick M O'Malley, Willard L. Rodgers. *A Guide for Selecting Statistical Techniques for Analyzing Social Science Data*, University of Michigan Ann Arbor, Michigan, 1976.
This guide attempts to record the sequential decisions a social scientist might make in selecting a particular statistic or statistical technique appropriate for a given analysis from the vast array of available techniques Systematic but highly condensed overview of many of the statistics in current use and the purposes for which each is intended
Instructions and Comments on the Use of this Guide, p 1
The Decision Tree, p. 3

Barnes, F P., *Research for the Practitioner in Education*, Department of Elementary School Principals, NEA Washington, 1964
Easy to read. Basic concepts, examples and charts
The Research Process, p. 12
Researchable Problems, p 27
Development of Hypotheses, p 29
Population and Samples, p 33
Independent and Dependent Variables, p 43
Statistical Tests of Probability, p 46
Designs and Decisions, p. 52
Nonparametric and Parametric Methods, p 75
Measurement Scales, p 78
Stating and Evaluating the Null Hypothesis, p 79
Reporting Results, p 108
The Research Literature, p 112
Staff Training and Follow-up, p 129

Buros, O K., *Tests in Print—A Comprehensive Bibliography of Tests For Use in Education, Psychology and Industry*, Gryphon Press Highland Park, N J 1961
Intermediate Requires understanding of reliability and validity List of tests in following areas Character and personality, Intelligence, Miscellaneous Vocations, Mathematics, English Reading, Social Studies, Science, Foreign Languages, Sensory-motor, Business Education, Achievement Batteries, Fine Art, Multi-Aptitude Batteries, Publishers Directory and Index, Title Index, Name Index

Buros, Oscar Krisen, *Tests in Print II— An Index to Tests, Test Reviews and the Literature on Specific Tests*, Gryphon Press, Highland Park, N.J., 1974
Assist educators, personnel workers, and psychologists in the selection and use of tests Comprehensive bibliography of all known tests

Dixon, Wilfred J., *Introduction to Statistics* 2nd edition, McGraw-Hill Book Co., Inc., New York, 1957
Beginning level of difficulty Glossary of terms used in each chapter. Discussion questions, class exercises and problems
Application of Statistics, p 2
Population or Universe, p 31
Sample, p 32
Design of Experiments, p 35
Various Measures of Central Value and Dispersion p 70
Statistical Hypothesis, p 88
Level of Significance, p 89
Second Type of Error, Beta, p 91
Test of Statistical Hypothesis, p 93
Alternatives and Two Types of Error, p 244
Nonparametric vs Parametric Statistics, p 299

George A. Ferguson, *Statistical Analysis in Psychology and Education* McGraw-Hill Book Co., Inc New York, 1959
Object of this book is to introduce students and research workers in psychology and education to concepts and applications of statistics. Emphasis is placed on the analysis and interpretation of data resulting from the conduct of experiments The book may be used as a text in a statistics course
. Exercises at the end of each chapter
Basic Ideas in Statistics, p 1
Essential Ideas of Sampling, p 112
Tests of Significance, p 131
Chi-square, p 157
Rank Correlation Methods, p. 179
Selecting Nonparametric Tests, p 264
Errors of Measurement, p 275

Gallo, A.J. and Miller, E., *Interpreting Education Research*, Wm C Brown Co., Publishers, Dubuque, Iowa, 1965
-Advanced intermediate level of difficulty Theoretical background.
Basic Nature of Education Research, p 3
Methods of Data Collection, p 25
Concepts that Serve as Basis for Statistical Data Analysis, p 103
Statistical Design of Experiments in Education, p 153
Typical Applications of Research, p 299

Hays, William L., *Statistics for the Social Sciences*, 2nd edition, Holt Rinehart and Winston, Inc New York, 1973
Intermediate advanced In a relatively nonmathematical form, but in somewhat more detail than is customary in such texts More emphasis on the theoretical than the applied and computational aspects of the methods
Exercises at the end of the chapters with solutions to the odd-numbered problems
Pearson's product moment r, p 623
Contingency Coefficient, p 745
Chi-square, p 721
Central Tendency and Variability, p 215
Levels of Measurement Scales, p 81

Quinn, McNemar, *Psychological Statistics*, 4th edition, John Wiley and Sons, Inc New York, 1969
Intermediate Emphasis on interpretation and assumptions rather than computation Exercise and questions for each chapter at the end of the book
Probability and Hypothesis Testing, p 39
Mode, p 14
Median, p 14
Range, p 19
Mean, p 16

Runyon, R P and A Haber, *Fundamentals of Behavioral Statistics*, 3rd edition, Addison-Wesley Publishing Co, 1976
Easy reading, basic concepts, examples, charts, and exercises
Definition of Statistical Analysis, p 3
Basic Mathematical Concepts, p 15
Frequency Distributions and Graphs Technique, p 37
Percentile, p 63
Standard Deviation and the Standard Normal Distribution, p 109
Correlation, p 125
Probability, p 177
Introduction to Statistical Inference p 207
Power and Power Efficiency of Statistical Test, p 307
Statistical Inference with Ordinally Scaled Variables, p 339
Glossary of Symbols (Parametric), p 368

Siegel, Sidney, *Nonparametric Statistics for the Behavioral Sciences* McGraw-Hill Book Co., Inc. New York, 1956
Intermediate level of difficulty. Nonparametric statistical tests, function and rationale, method for small and large samples, an example of a problem using the test, summary of the procedure, power-efficiency and references for each test
Null Hypothesis, p 7
Choice of Statistical Test, p 7
Level of Significance and Sample Size, p 8
Levels of Measurement, p 20
Power Efficiency, p. 21
Parametric vs Nonparametric, p. 30
Appendix with Statistical Tables, p 245

Slakter, Malcolm J., *Statistical Inference for Educational Researchers*, Addison-Wesley Publishing Co. Reading, Massachusetts, 1972
Moderate. It is assumed student has mathematical background equivalent to a student completing elementary algebra in high school; more advanced mathematical symbols of concepts explained in the text. Text for a one semester course for students planning to be involved with educational research. Attempt to blend theory and application, study assumptions of statistical models, effects of violation of assumptions on inference process. Many examples and problems
Statistical Hypothesis, p 253
Types of Decisions and Errors, p 255
General Overview of Statistical Hypothesis Testing, p 257
Choice of Region of Rejection, p 261
Power Function, p 265
Answers to Problems, p 382

Wallis, W A., and H V Roberts, *Statistics a New Approach*. The Free Press New York, 1956
Intermediate or advanced level of difficulty. It gives the theoretical background and examples of basic concepts such as
Some Uses of Statistics in Social Sciences, p 23
Samples and Populations, p 100
Randomness in Sampling, p 115
Recording Data, p 142
The Art of Organizing Data, p 167
Descriptive Statistics, p 211
Mode Median Mean, p 213–219
Association, p 268
How to Read a Table, p 270
Randomness and Probability, p 309
Statistical Tests and Decision Procedures, p 384
Design of Investigations, p 475

Staff

Project Director—Dr. Frederick A Rodgers

Frederick A. Rodgers is a Professor of Early Childhood and Elementary Education at the University of Illinois—Champaign/Urbana. His primary fields of research are curriculum development, design and planning, program evaluation, social studies programs, and educational policy. His recent work has concentrated on the development of reading materials for young children outlining theoretical aspects of curriculum development.

Research Assistant—Ms. Terry Peters

Terry Peters is a graduate student in Special Education at the University of Illinois—Champaign/Urbana. Her primary fields of study deals with helping young children with special learning problems and developing materials for selected student populations. She has placed some emphasis on measurement and research problems associated with young children who have difficulty learning.

Research Assistant—Ms Marty Markward

Marty Markward is a graduate student in Social Work at the University of Illinois—Champaign/Urbana Her primary field of study deals with social and emotional problems that affect the school performance of school age children

Research Assistant—Ms Susan Herzog

Susan Herzog is a graduate student in Early Childhood Education at the University of Illinois—Champaign/Urbana

Program Officer

Shirley A. Jackson
Dept. of Education
Basic Skills Program
Washington, D.C. 20202

Steering Committee

Mrs. Effie Games
Title I Building Reading Resource Teacher
J O. Wilson School
6 and K St N E.
Washington, D C
202–724-4707

Dr Grace Green
English Language Arts Supervisor
South Colonie Schools
Lisha Hill Jr High School
Waterman Avenue
Albany, New York 12205

Dr. John Guthrie
Director, Research
International Reading Association
800 Barksdale Road
Newark, Delaware 19711
302–731-1600

Roselmina Indrisano
Boston University
School of Education
765 Commonwealth Avenue
Boston, Massachusetts 02215
617–353-3267

Ms Shirley Jackson
U S. Office of Education
Right to Read, DHEW
Donohoe Bldg
400 Maryland Ave., S W
Washington, D.C 20202

Dr Alvin D Loving, Sr.
National Center on Black Aged
1730 M Street, N W
Washington, D.C 20036
202–637-8400

Dr. Ora B McConner
Director, Bureau of Pupil Personnel Services
228 North LaSalle Street Room 912
Chicago, Illinois 60601
312–641-3990

Dr. Marion McGuire
Director, Graduate Reading Center
University of Rhode Island
Kingston, Rhode Island 02881
401–792-5835

Dr. Richard Petre
State Director Right to Read
Department of Education
International Tower Building
P O Box 717, BWI Airport
Baltimore, Maryland 21240
301–796-8300

Dr Frederick A. Rodgers (Contractor)
315 Education Building
College of Education
University of Illinois
Urbana, Illinois 61801
217–333-1844

U S GOVERNMENT PRI ''NG OFFICE 1980 O—302-238