DOCUMENT RESUME

ED 052 238 TM 000 643

AUTHOR Gephart, William J.
TITLE Profiling Educational Research.
INSTITUTION Phi Delta Kappa, Bloomington, Ind.
REPORT NO Occas-Pap-3
PUB DATE Jan 69
NOTE 25p.

EDRS PRICE EDRS Price MF-$0.65 HC-$3.29
DESCRIPTORS *Data Analysis, *Educational Research, Evaluation, Flow Charts, *Hypothesis Testing, Logical Thinking, Measurement Instruments, Measurement Techniques, *Profile Evaluation, *Research Methodology, Statistical Analysis

ABSTRACT

Profiling, a procedure for evaluating methodological adequacy, a prerequisite for the acceptance or rejection of the conclusions of the investigation, is discussed. Three elements of this procedure are presented: (1) structuring of a logical argument, (2) generation of data and (3) analysis of data. All three elements are employed in hypothesis testing, and the second two are employed for empirical studies. Three aspects of data generation: representativeness, treatment and measurement are presented graphically as dimensions for the research quality cube. Ordinal scales, to facilitate profiling, are given for each of the three dimensions. Problems of data analysis are noted and the development of grids for the profiling analysis procedure is explained. Flow charts for guiding the research profiling are appended. (LR)

ERIC

ED052238

U.S. DEPARTMENT OF HEALTH, EDUCATION & WELFARE
OFFICE OF EDUCATION
THIS DOCUMENT HAS BEEN REPRODUCED EXACTLY AS RECEIVED FROM THE PERSON OR ORGANIZATION ORIGINATING IT. POINTS OF VIEW OR OPINIONS STATED DO NOT NECESSARILY REPRESENT OFFICIAL OFFICE OF EDUCATION POSITION OR POLICY.

PHI DELTA KAPPA

RESEARCH SERVICE CENTER

OCCASIONAL PAPER

#3. PROFILING EDUCATIONAL RESEARCH

by

William J. Gephart
Director of Research Services
Phi Delta Kappa

January 1969

TM 000 643

ED052238

ACKNOWLEDGEMENTS

The work described herein is based on the thoughts and efforts of many predecessors. Some of them are recognized in the paper presented by my colleague, Bruce B. Bartos. Through continued interactions James Raths, Robert Ingle, Robert Remstad, and Bruce Bartos have made valuable contributions to this paper. These gentlemen and the many others who have worked at evaluation of the adequacy of educational research deserve the major credit for any advancements this paper might make to our deliberations. Its inadequacies are the responsibility of its author.

W. J. G.

PROFILING COMPLETED RESEARCH

The evaluation of the quality of completed research in education has two distinct components. The first of these components is the problem attacked in the study. The second is the methodological adequacy of the study conducted. Profiling, the procedure described in this paper, deals with the latter, the evaluation of the methodological adequacy. It avoids evaluation of the problem itself on the belief that the importance of a given problem can only be established through an historical perspective. To assert otherwise would imply the existence of a preferred value system.

Evaluation of methodological adequacy of a given piece of research is a prerequisite for the acceptance or rejection of the conclusions of that investigation. Such conclusions can be no stronger than the methods utilized in generating and analyzing the data on which the conclusions are reached. In the past we have operated on the assumption: if the methodology is sound. the conclusion can be accepted and vice versa. The faultiness of this assumption is one of the problems that have long plagued both the improvement of and use of educational research.

Research methodology is multifaceted. It involves an inherent logical argument, the selection of subjects to be studied, structuring of experiences for those subjects, measurement, and the analysis of the generated data. It is possible to have sound procedures in some of these facets and weak procedures in others; a possibility that precludes a statement that a conclusion is based either on sound or unsound methods.

The problem is further complicated. Needs for surety in varying times and professional circumstances set the quality standard for research methods. If the need for knowledge in an area is great: the methodological development crude. and the amount of risk to personal safety low, conclusions can be accepted and operated on despite weaknesses in their methodological base. In another set of circumstances this would be wholly unacceptable. Since the use to which a conclusion might be put cannot be controlled, an absolute level of quality cannot be established for each research effort.

Regardless of the knowledge needs or professional circumstances, a given conclusion ought not to be accepted, held tentatively or rejected without evaluation of the research methods underlying it. It is asserted that the profiling procedure described in this paper will facilitate the labeling of the methodology of completed research reports. When this labeling has been completed, the user of that study can make sounder decisions regarding the acceptance or rejection of its conclusions.

ELEMENTS IN PROFILING

In conducting an empirical study an investigator does numerous things. Those things are the elements on which the profiling activity focuses. They include: (1) the structuring of a logical argument; (2) the generation of data; and (3) the analysis of that data. All three items are involved in investigations which test hypotheses while only items two and three are used in studies which attempt to answer empirical questions.

THE INHERENT LOGICAL ARGUMENT is of crucial importance when a study attempts a test of a hypothesis. In effect, the investigator is trying to determine the truth or falsity of his hypothesis. He does this through a logical argument described by Polya.[1] It consists of a major premise, one or more minor premises, and a conclusion.

The major premise is typically a statement which asserts, "If the ______ ______ hypothesis is a true statement; then ______ ______ events will be observed as indicators of that truth." An example of a major premise can be seen in a study reported a few years ago by McNeil.[2] He proposes a hypothesis which asserts that teachers present different instructional treatments for the two sexes of their students. As indicators of the truth of that statement he reasoned that boys would be nominated more often than girls as recipients of certain kinds of teacher action. His major premise could be stated as,

> If the hypothesis (teachers provide different instructional treatment for boys than they do for girls) is a true statement, then systematic differences by sex will be seen when children are asked to name the students who receive specified teacher treatments.

Two kinds of minor premises have been evolved from Polya's work by Raths.[3] The first of these deals with the predicted observation. Was it or was it not seen? The premise's exact nature in a given study is determined after the data are analyzed. In the McNeil example used above, significant differences by sex were observed. The minor premise in that case would be, "There is a systematic sex differentiation in the nominations."

The second category of minor premises deals with rival hypotheses, rival or alternative explanations for the observation reported in the first minor premise. The premise is based upon the recognition that an effect in the social sciences often has multiple causes. Once an observation has been made, all its possible causes must be examined before it can be concluded that the observation supports the truth of a specific hypothesis. One of three general conditions might exist ranging from no rival hypotheses are apparent to rival hypotheses may exist to rival hypotheses are definitely involved.

The final element of the logical argument is the conclusion. Its form in a given study is dependent upon the nature of the two minor premises. From the first minor premise comes information as to whether or not the truth of the hypothesis being tested is supported. If the consequents predicted are observed, support for the truth of the hypothesis is presented. If the observation is not made, support cannot be claimed. (Note: Failure to make the predicted observation does not automatically mean rejection of the hypothesis.) The second minor premise determines the strength of the conclusion. If rival hypotheses are known to be present very weak support for the truth of the hypothesis has been developed. If there is the possibility but not the probability of rival hypotheses, tentative support is generated. And finally, if no rival hypotheses are conceivable, it is credible that the hypothesis is a true statement.

THE GENERATION OF DATA, the second major facet in profiling, involves evaluation of three aspects of data generation: units studied; treatments experienced by those units; and measurement. If variation in any of these three occurs a different set of data is generated. For example, consider an investigation of the effects of test anxiety on achievement. If the study concentrates on a randomly selected group of high school seniors as subjects, one set of data will be generated. If a group of students who are divergent on a measure of test anxiety is selected as subjects, a different set of data will be generated. Given a specified group as subjects, variation in the treatment or of their experiences will cause different sets of data to be generated. Again the test anxiety problem provides an example. One set of data could be generated by a treatment in which the subjects are given information about the importance of a test and administered a test that is constructed for students at a much higher level of education than are the subjects. Still a different set of data will be generated if the students are repeatedly given a test that is very difficult. If the effects of a specific treatment on a specific group are measured by a paper and pencil test such as Sarazon's Test of Test Anxiety, one set of data would be generated. On the other hand if the seats in the classroom were wired and a galvanic skin response measure were taken, quite a different set of data would be generated.

These three aspects of data generation are displayed graphically in Figure 1.[4] The scale of unit quality or representativeness runs along the dimension OA, treatment quality OC, and measurement quality OG. A project which selected a sample perfectly representative of a population of interest would be located at Point A. on the cube. If, in that same study, a thorough programing of the content and sequence of the treatments was employed in generating data, the project would be conceptualized as being at Point B. on the quality cube. Finally, if our study employed perfectly objective, valid, and reliable measuring techniques, it would be located at Point E. on the cube.

A given study seldom reaches this level of data generation quality. Rather it falls somewhere between the extremes. To facilitate profiling ordinal scales have been developed for these three dimensions as shown below.

Dimensions for the Research Quality Cube

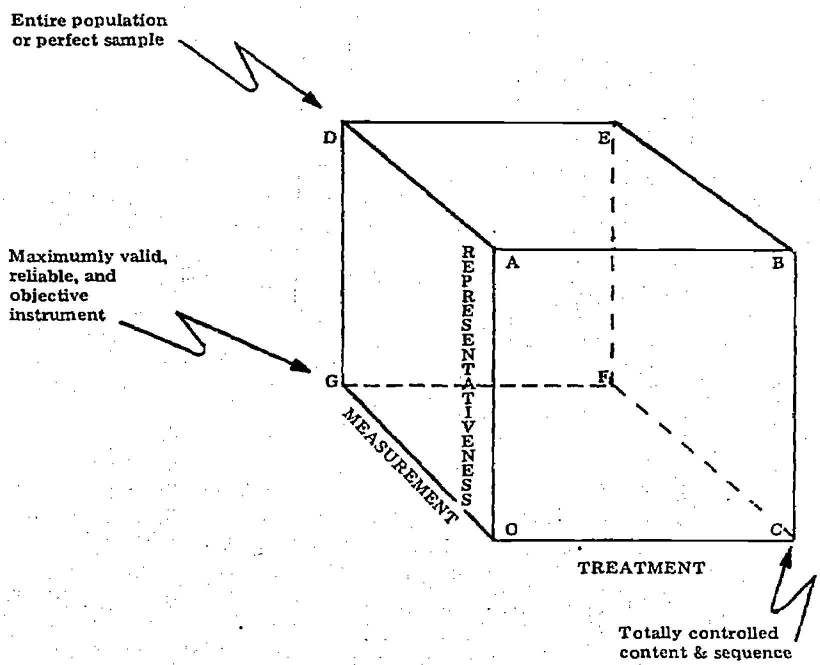
Representativeness

R_5 = The entire population was studied
R_4 = Random selection from a specified population was employed to determine which units were studied.
R_3 = Purposive sampling from a specified population established the group studied.
R_2 = Volunteers were studied.
R_1 = An unidentified group of subjects was studied.

Treatment

T_6 = A theoretically based treatment was administered and described and controls were employed for mediating variables identified in the theory AND for variables extraneous to the theory that might have an effect.
T_5 = Same as T_6 with the exception of the lack of controls for extraneous variables.
T_4 = Same as T_6 with the exception of the lack of controls for theory encompassed mediating variables and extraneous variables.
T_3 = No theory stated but the employed treatment described in detail sufficient for replication.
T_2 = Commonly known treatment administered but not

THE DATA QUALITY CUBE



described in detail.

T_1 = Something of an undescribed nature was experienced by the units studied.

Measurement

M_5= Data were generated through the use of either a commercially standardized or ad hoc instrument AND data are presented which establish high validity and reliability for its use in this measurement task.

M_4= Data generated through the use of a commercially standardized instrument and evidence presented indicating moderate validity and reliability for this application.

M_3= Data generated through a commercially standardized test but no evidence presented as to its validity and reliability for this application.

M_2= Data generated through an ad hoc instrument and evidence of moderate validity and reliability presented.

M_1= Data generated through an ad hoc instrument with either no supporting evidence as to validity and reliability or evidence indicating poor validity and reliability on either a commercially standardized or ad hoc instrument.

DATA ANALYSIS PROCEDURES are the final element in profiling. When data, typically in the form of numbers, are generated as the supporting evidence for a conclusion, understanding of the meaning of those numbers is incumbent upon the researcher and the research utilizer. That meaning is not readily apparent if there is a large quantity of numbers. Simplifying procedures have been developed; procedures which are not appropriate for all kinds of data.

The determination of the correct procedure in a given study is not an exact science. In developing a procedural flow chart for the profiling of educational research, sixteen schemes were identified which were supposed to assist in the selection of the correct statistic for given sets of data. Some of these were incomplete schemes in that they purported to deal only with limited kinds of statistical analysis.[5] Some imply a comprehensiveness but fail to be definitive as they list a number of statistics appropriate for a given set of conditions.[6]

Since a single comprehensive grid or table for selecting the correct analytic procedure could not be found a second task was undertaken. Existing statistical procedures were catalogued and the assumptions underlying them were listed. An effort to build a comprehensive selection procedure by analyzing these items has to this point been unsuccessful. (A colleague at Indiana University[7] has just recently attacked this problem using Guttman's Facet design and Analysis Technique[8] with initially promising results.)

Because of these problems three grids have been generated for profiling the data analysis procedures. The first of these deals with analytic procedures for sample description. It includes measures of central tendency and dispersion and classifies the procedures by levels of measurement, i.e., nominal, ordinal, and interval-ratio. The second grid is used when an associational analysis is desired. It has identical labels for its rows and columns which refer to the nature of the measurement on the two variables to be correlated. The categories in this case are:

1. Continuous variables (age, height, I.Q., achievement, etc.)
2. Forced dichotomy (number of persons over and under 100 I.Q., number of persons weighing over and under 150 pounds, etc.)
3. True dichotomy (student-nonstudent, male-female, etc.)

Given the nature of the two variables on which an associational analysis is desired the grid can be used to select the appropriate statistic. Four special cases exist and are shown with the grid. Three of these are instances in which more than two variables are involved. The final case covers correlation among ordinal variables.

The third grid deals with inferential statistics, instances in which a generalization about the relationship between the numbers generated by observation of some sample are indicative of observations that could be made on the entire population. The categorizing elements on this grid are the number of dependent and independent variables, the level of measurement, and the number of groups. Again the determination of the appropriate level on each category for a given set of data leads to the recommended statistic.

The use of these grids leads to a specific statistic (in the inference grid there is the possibility of alternatives). Through the article the analytic procedure actually used can be identified. Two quality categories follow from a comparison of the statistic used and the statistic appropriate for the data and purpose of the study: first, the statistic used is identical with the statistic identified and appropriate; second, they are different. In the former the research is profiled as appropriately analyzed; in the latter, as inappropriately analyzed.

PROFILING SUMMARIZED: When a study has been analyzed and profiled, it has been described on the following basis:

A. Is it (I) a test of a hypothesis, or (2) an answer to an empirical question?

A_1. If it is a test of a hypothesis, is the strength of conclusion: I The hypothesis is very little more credible; II more credible; or III very much more credible?

B. What is the quality of the data generation procedure ($r_i t_i m_i$)?

C. Is the data: (a) appropriately analyzed; or (b) inappropriately analyzed.

It should be noted that a single project may consist of several substudies, each of which may be profiled separately. A decisional flow chart has been developed for arriving at the profile for a given study. It is appended. Your reactions regarding its adequacy are welcomed.

It is believed that through profiling completed research their adequacies and inadequacies can be made apparent and can more readily be considered as the conclusions of the research are weighed in decision situations. One further benefit is seen. Studies of such profiles should pinpoint problems that could keep research methodologists busy for years to come.

ERIC

FOOTNOTES

1. George Polya, PATTERNS OF PLAUSIBLE INFERENCE: MATHEMATICS AND PLAUSIBLE REASONING Volume II, Princeton, New Jersey: Princeton University Press, 1954.

2. John McNeil, "Programed Instruction Versus Usual Classroom Procedures in Teaching Boys to Read." AMERICAN EDUCATIONAL RESEARCH JOURNAL 2:113-19; March, 1964.

3. James Raths, "Plausible Logic in Educational Research" Paper delivered at AERA Convention, Feb. 1964. Reprinted in EDUCATIONAL RESEARCH: SELECTED READINGS. W. J. Gephart & R. B. Ingle Eds., Columbus, Ohio: Charles E. Merrill, Publisher, 1969.

4. William J. Gephart and Robert B. Ingle, EDUCATIONAL RESEARCH: SELECTED READINGS. Columbus, Ohio: Charles E. Merrill, Publisher, (In press 1969).

5. See the grid on the inside cover of Siegel's NONPARAMETRIC STATISTICS, New York: McGraw-Hill Book Co., Inc. 1956. An unpublished paper presenting a grid for selecting the appropriate correlation coefficient developed by W. J. Gephart, the flow chart for analysis of variance in K. D. Hopkins & R. A. Chadbourn, "A Scheme for Proper Utilization of Multiple Comparisons in Research and a Case Study." AMERICAN EDUCATIONAL RESEARCH JOURNAL, 4:407-12, Nov. 1967.

6. See the grid presented by D. V. Tiedeman and M. Tatsuoka in HANDBOOK OF RESEARCH ON TEACHING. N. L. Gage Ed., Chicago: Rand McNally Publishers, 1963.

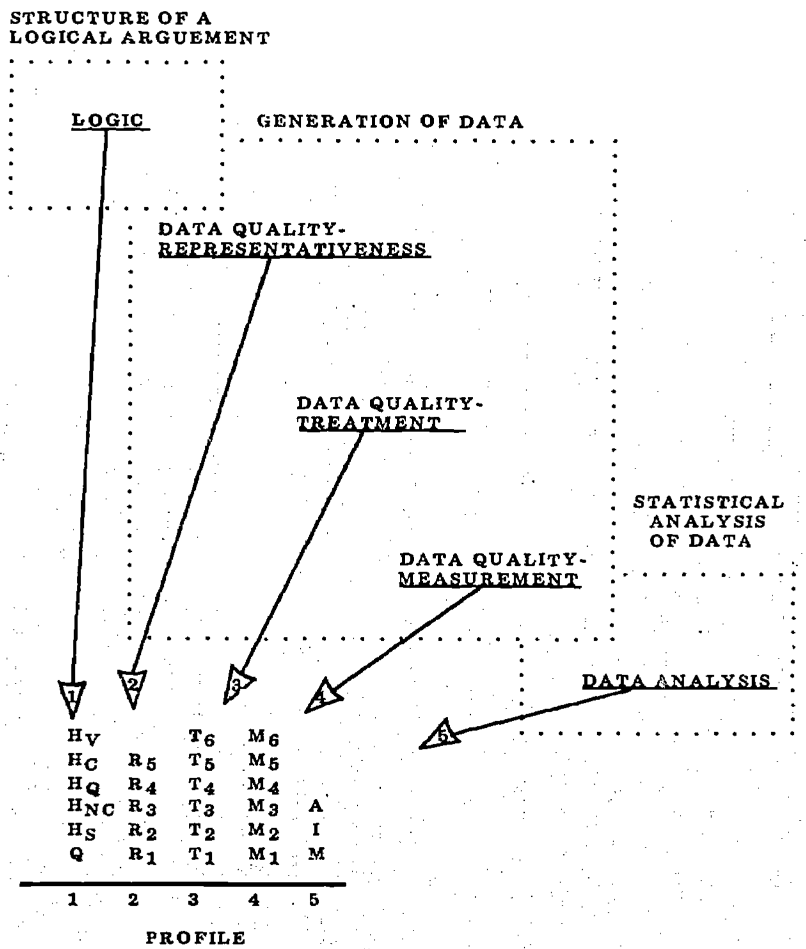
7. Donald Ary, Unpublished mimeo. Indiana University, Bloomington, Indiana, 1968.

8. Described by Philip Runkle, "Some Recent Developments in Research Methodology." 1965 ERIC Document ED 010 221.

APPENDIX

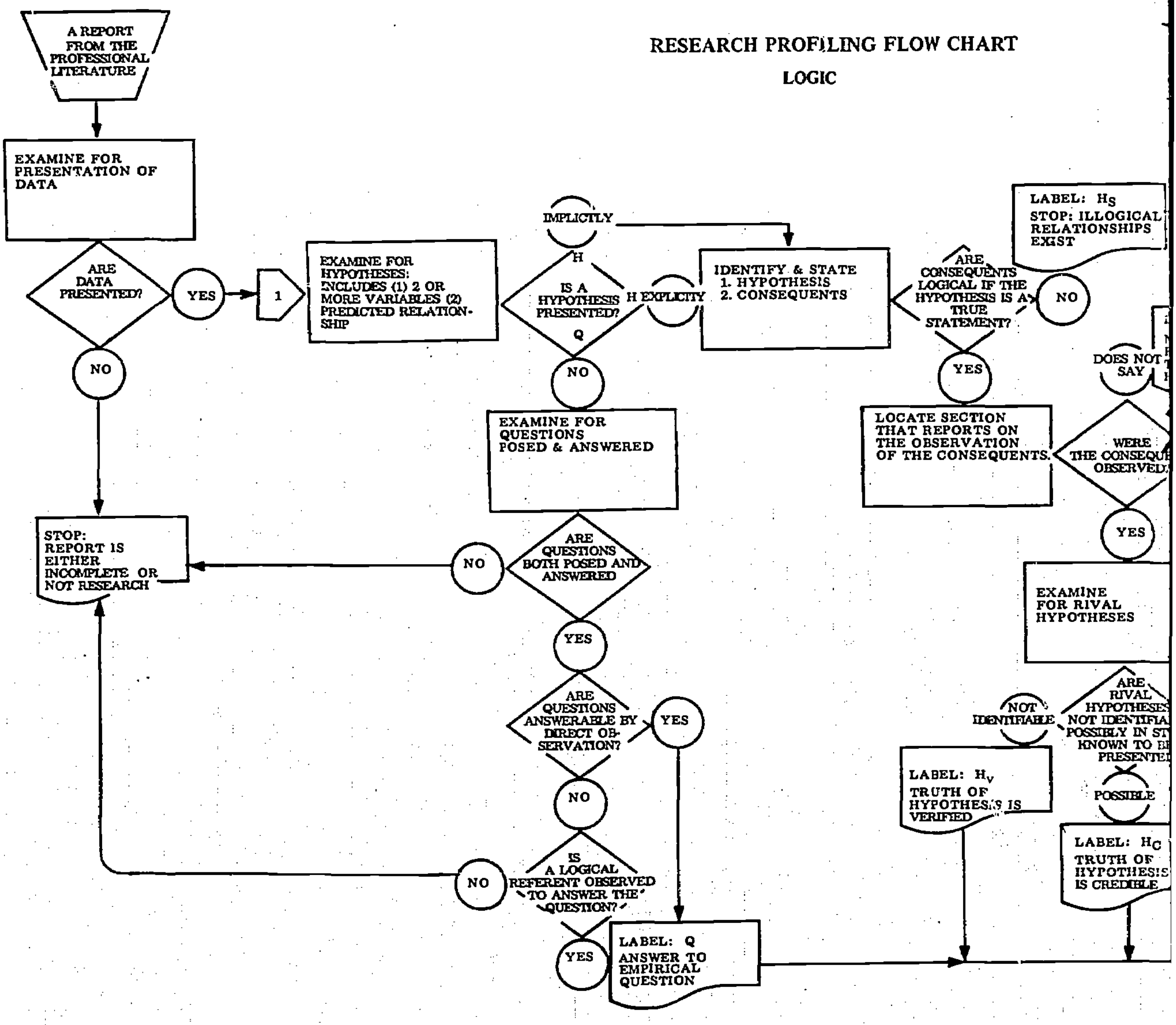
FLOW CHART FOR GUIDING THE RESEARCH PROFILING

FACETS OF THE RESEARCH PROCESS



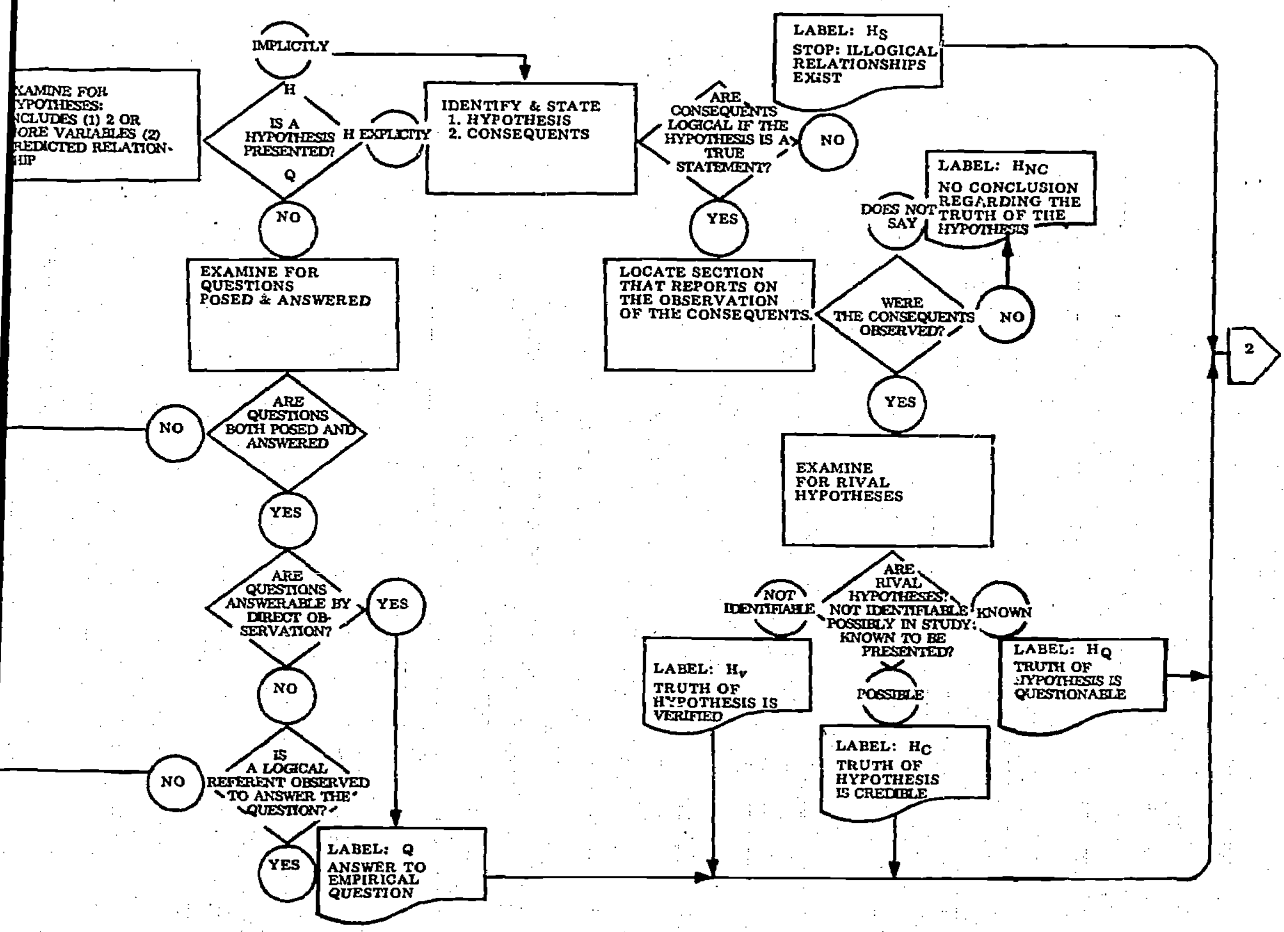
RESEARCH PROFILING FLOW CHART

LOGIC



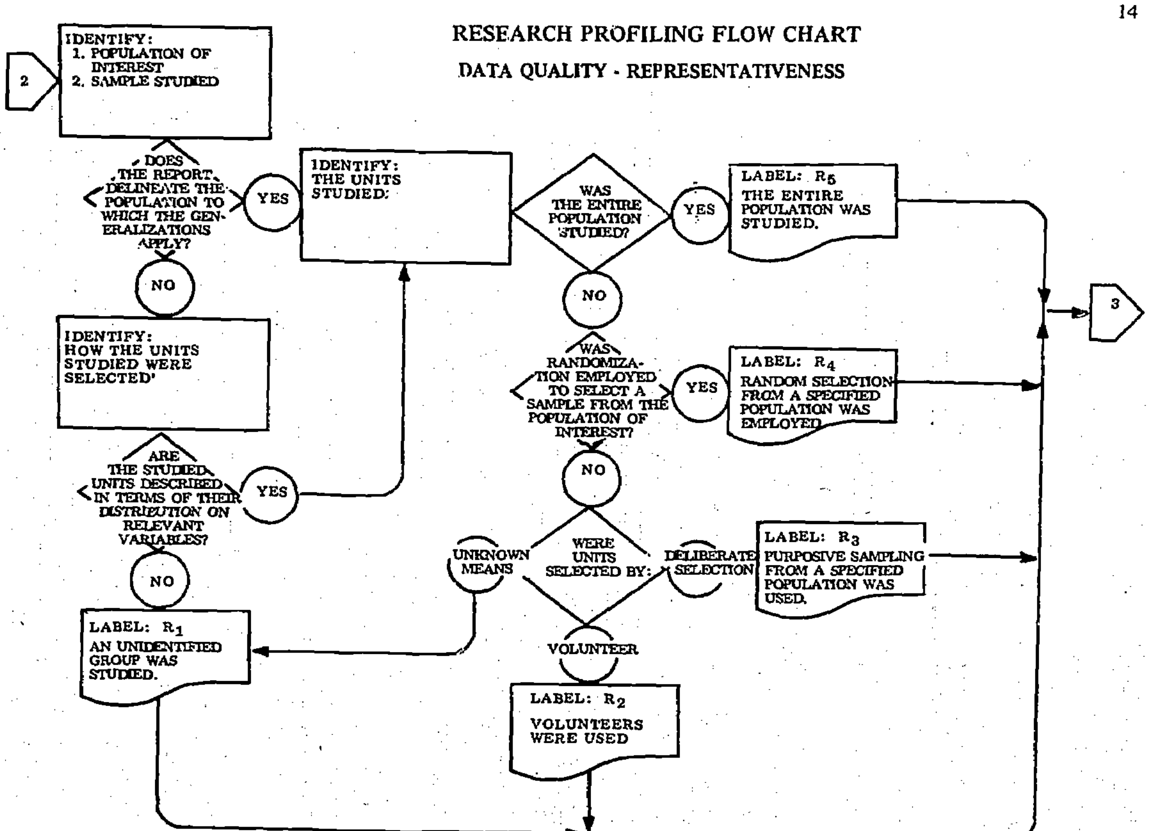
RESEARCH PROFILING FLOW CHART

LOGIC



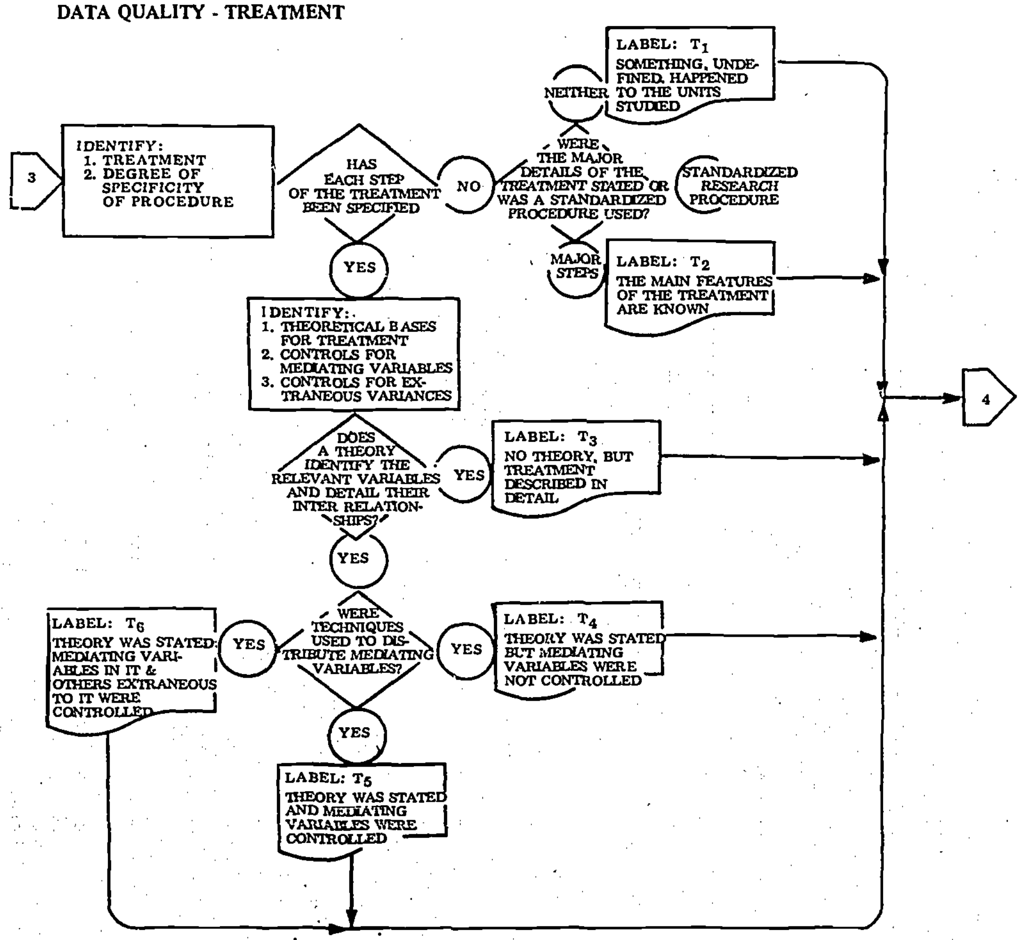
RESEARCH PROFILING FLOW CHART

DATA QUALITY - REPRESENTATIVENESS



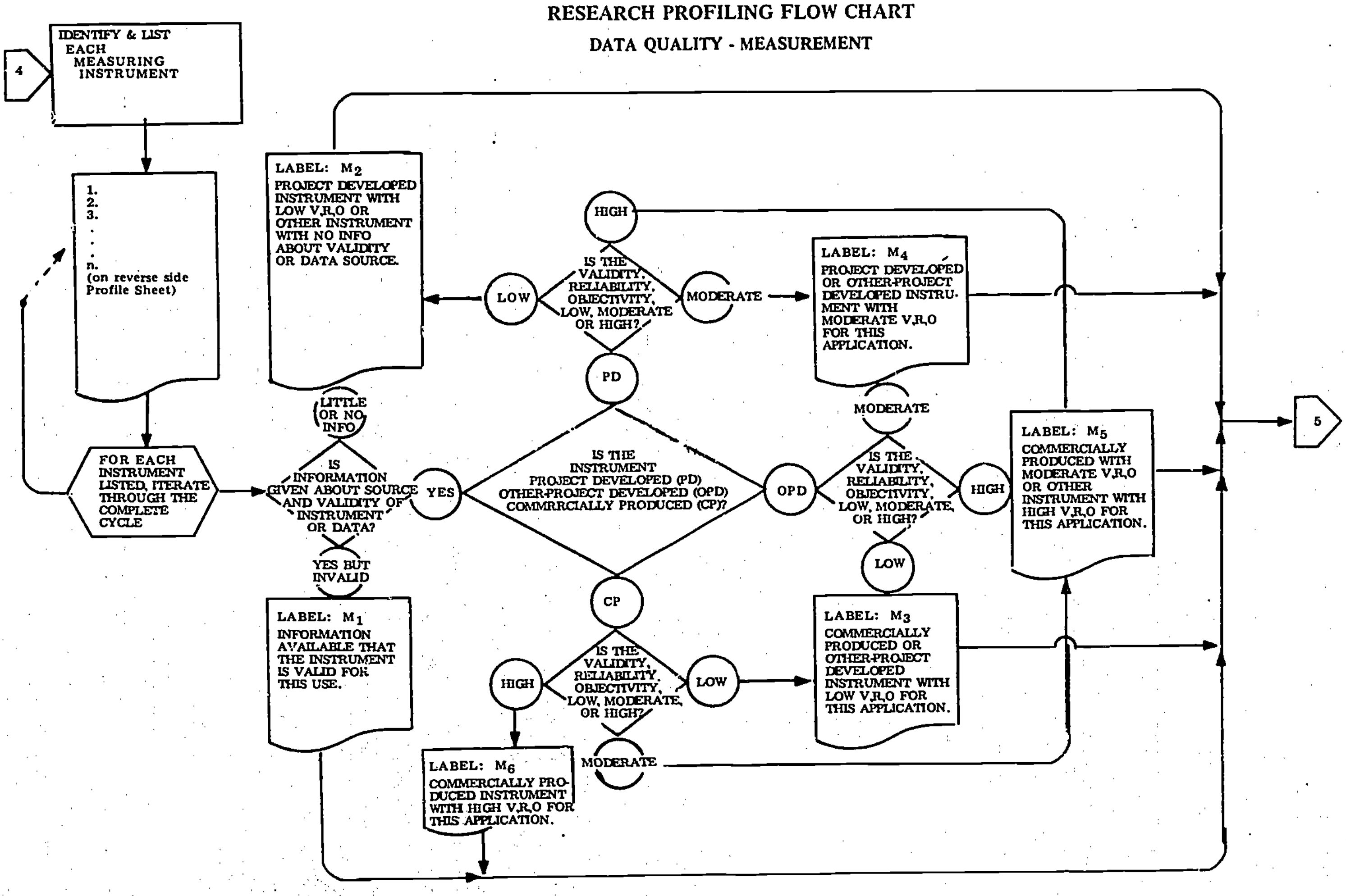
RESEARCH PROFILING FLOW CHART

DATA QUALITY - TREATMENT



RESEARCH PROFILING FLOW CHART

DATA QUALITY - MEASUREMENT



RESEARCH PROFILING FLOW CHART

ANALYSIS

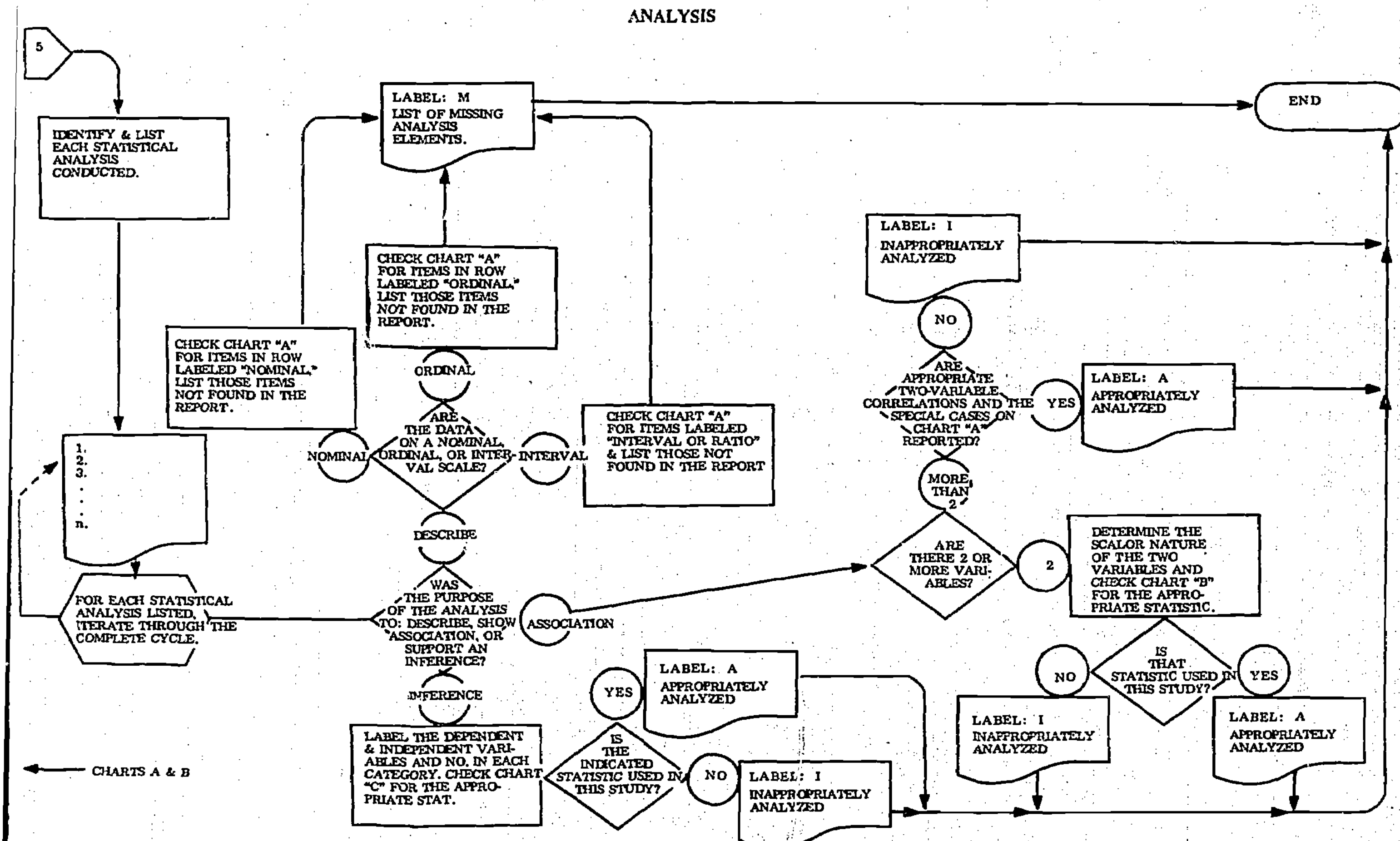


CHART A

POPULATION DESCRIPTORS

	DISTRIBUTION	CENTRAL TENDENCY	DISPERSION
NOMINAL	FREQUENCY IN EACH CATEGORY	MODE	
ORDINAL	FREQUENCY IN EACH SCALOR POSITION	MEDIAN	SEMI-INTERQUARTILE RANGE
INTERVAL/ RATIO	FREQUENCY IN EACH INTERVAL	MEAN	STANDARD DEVIATION

CHART B - MEASURES OF ASSOCIATION

VARIABLE 1*

	CONTINUOUS	FORCED DICHOTOMY	DICHOTOMY
CONTINUOUS	PEARSON r	BISERIAL r	POINT BISERIAL r
FORCED DICHOTOMY	BISERIAL r	TETRACHORIC r	(NONE AVAILABLE USE CHI SQUARE)
DICHOTOMY	POINT BISERIAL r	NONE AVAILABLE	FOUR FOLD r OR PHI COEFFICIENT

* INTERVAL DATA

SPECIAL CASE
2 VARIABLES

RANK DATA SPEARMAN'S RHO

SPECIAL CASE
MORE THAN 2 VARIABLES

INTERVAL DATA MULTIPLE R
ORDINAL DATA KENDALL'S W
NOMINAL DATA CONTINGENCY COEFF. C

ERIC Full Text Provided by ERIC

CHART C

INDEPENDENT VARIABLE

DEPENDENT VARIABLE(S)

	NOMINAL 1	NOMINAL>1	ORDINAL 1	ORDINAL>1
NOMINAL 1	FISHER'S EXACT PROB. FOR 2x2 TABLE MC NEMAR'S TEST FOR SIGNIFICANCE OF CHANGES COCHRAN'S Q TEST FOR SEVERAL RELATED PROPORTIONS CHI-SQUARE TEST FOR INDEPENDENCE METHODS FOR MAXIMIZING PROBABILITY OF CORRECT CLASSIFICATION		SIGN TEST MEDIAN TEST MANN-WHITNEY U TEST	
NOMINAL>1			FRIEDMAN'S 2-WAY ANOVA	
ORDINAL 1	SIGN TEST MEDIAN TEST MANN-WHITNEY U TEST KRUSKAL-WALLIS 1-WAY ANOVA KOLMOGOROV-SMIRNOV 1-SAMPLE TEST	FRIEDMAN'S 2-WAY ANOVA	INDEX OF ORDER ASSOCIATION	ANALYSIS OF V WITH TREND A
ORDINAL>1				
INTERVAL 1	ANALYSIS OF VARIANCE	INDEPENDENT t ANALYSIS OF VARIANCE ANALYSIS OF COVARIANCE	ANALYSIS OF VARIANCE	
INTERVAL>1	CORELATION FACTOR ANALYSIS MULTIPLE DISCRIMINANT ANALYSIS MULTIPLE REGRESSION ANALYSIS	MULTIPLE DISCRIMINANT FUNCTION		

Most of the measures shown are located with reference to the lowest order of data that should be used with them. One should always be able to transform the observed data downward i.e. interval can be considered ordinal or nominal, ordinal can be considered nominal.

CHART C

INDEPENDENT VARIABLE(S)

	NOMINAL>1	ORDINAL 1	ORDINAL>1	INTERVAL 1	INTERVAL>1
...ILE ...NCE OF CHANGES RELATED PROPORTIONS ...NCE ...BILITY OF CORRECT CLASSIFICATION		SIGN TEST MEDIAN TEST MANN-WHITNEY U TEST		ANALYSIS OF VARIANCE	HOTELLING'S T MAHALANOBIS' D^2 FISHER'S DISCRIMINANT FUNCTION
		FRIEDMAN'S 2-WAY ANOVA		ANALYSIS OF VARIANCE CORRELATION FACTOR ANALYSIS MULTIPLE DISCRIMINANT ANALYSIS	RAO'S V_k MULTIPLE DISCRIMINANT FUNCTION
TEST	FRIEDMAN'S 2-WAY ANOVA	INDEX OF ORDER ASSOCIATION	ANALYSIS OF VARIANCE WITH TREND ANALYSIS		
	INDEPENDENT t ANALYSIS OF VARIANCE ANALYSIS OF COVARIANCE	ANALYSIS OF VARIANCE		REGRESSION ANALYSIS	MULTIPLE REGRESSION ANALYSIS
SIS S	MULTIPLE DISCRIMINANT FUNCTION			MULTIPLE REGRESSION ANALYSIS	CANONICAL CORRELATION

...e located with reference to the
...be used with them. One should
...bserved data downward i.e. in-
...r nominal, ordinal can be

This table is an adaptation of Tatsuoka and Tiedeman's Table 1 in "Statistics As An Aspect of Scientific Method in Research on Teaching" pgs. 154-155, in HANDBOOK OF RESEARCH ON TEACHING, N. L. Gage (Ed); Rand McNally, Chicago, 1963, 1218pp.

ERIC
Full Text Provided by ERIC

H_V		T_6	M_6	
H_C	R_5	T_5	M_5	
H_Q	R_4	T_4	M_4	
H_{NC}	R_3	T_3	M_3	A
H_S	R_2	T_2	M_2	I
Q	R_1	T_1	M_1	M
1	2	3	4	5

RESEARCH PROFILE SHEET

REPORT TITLE: ________________________________

____

AUTHOR ____________________ SOURCE: ______________

____ STOP The report is either not research or it is an incomplete part of the research process.

1 LOGIC

____ Q Answer to an Empirical Question

____ H_s Stop, Illogical relationship in the test of the hypothesis.

____ H_{nc} No conclusion can be reached from this test of the hypothesis.

____ H_q Hypothesis is questionable.
(Rival hypotheses must be considered a cause of the consequents)

____ H_c Hypothesis is credible.
(Rival hypotheses may be considered a cause of the consequents)

____ H_v Hypothesis is verified.
(Rival hypotheses cannot be considered as a cause of the consequents)

2 DATA QUALITY - REPRESENTATIVENESS

____ R_1 An unidentified group of subjects was studied.

____ R_2 Volunteers were studied.

____ R_3 Purposive sampling from a specified population established the group studied.

____ R_4 Random selection from a specified population established the group studied.

____ R_5 The entire population was studied.

3 DATA QUALITY - TREATMENT

____ T_1 No theory; something undefined happened to the units studied.

____ T_2 No theory; treatment description incomplete, or detailed elsewhere.

____ T_3 No theory; treatment described in detail in the report.

____ T_4 Theory stated but no controls on variables.

____ T_5 Theory stated and mediating variables controlled.

____ T_6 Theory stated, mediating variables controlled, and techniques used to distribute possible extraneous variances.

4 DATA QUALITY - MEASUREMENT

____ M_1 Available information indicates instrument is invalid for this use.

____ M_2 Project Developed instrument with low validity (V), reliability (R), objectivity (O), or other instrument with no info about validity or data source.

____ M_3 Used Commercially-Produced-or-Other-Project-Developed instrument with low V,R,O for this application.

____ M_4 Used Project Developed instrument or Other-Project Developed instrument with moderate V, R, O for this application.

____ M_5 Used instrument which was Project Developed with high V, R, O or Other-Project developed with high V,R,O or Commercially Produced with moderate V, R, O for this application.

____ M_6 Used Commercially Produced instrument with high V, R, O for this application.

5 STATISTICAL ANALYSIS

____ A Appropriately analyzed

____ I Inappropriately analyzed

____ M Missing items - incomplete analysis