

COMPUTER-BASED TEST INTERPRETATION SOFTWARE:
ITS EFFECT ON SCHOOL PSYCHOLOGIST DECISION MAKING

By

John W. Wisor

Dissertation submitted to the Faculty of the Graduate School
of the University of Maryland in partial fulfillment
of the requirements for the degree of
Doctor of Philosophy
1989

Advisory Committee:

Associate Professor William Strein, Chairman/Advisor
Associate Professor Robert Marcus
Professor Donald Pumroy
Associate Professor Arnold Spokane
Professor Henry Walbesser

Maryland
LD
3231
-M70cl
Wisor,
J. W.
Folio

ABSTRACT

Title of Dissertation: Computer-Based Test
Interpretation Software: Its
Effect on School Psychologist
Decision Making

John Wesley Wisor

Doctor of Philosophy, 1989

Dissertation directed by: William Strein
Associate Professor
Counseling and Personnel
Services

The purpose of this study was to assess the utility of interpretative software for school psychological decision making in its role as a decision aid. One hundred two professional school psychologists were provided case data and asked to make a series of diagnostic and prognostic decisions based on the case material. One subject group received case material only and each of the other two groups received one of two variations of the narrative output generated by a computer-based test interpretation software package in addition to the case material. The subjects were also asked which data sources were most influential as they made their decisions.

Diagnostic agreement among the psychologists within

each group was analyzed by Kendall's coefficient of concordance or weighted Kappa. For each decision there were no significant differences in agreement between those psychologists who had access to the decision aids and those who did not. Chi-square and Freidman analysis of variance results for similarity of diagnosis across groups were mixed with some trends suggestive of greater similarities of decisions among the subjects utilizing different variations of the computer output than among decisions made by unaided psychologists. Further the school psychologists overwhelmingly indicated that test data and behavioral observations were the most influential data sources for their decisions and that computer-based data sources were the least influential. Also there appeared to be no significant relationship between school psychologist professional experience and the perceived influence of the case data sources as well as little relationship between degree of experience in using computers to the data sources considered to be useful in the decision making process.

The results were discussed in terms of psychological decision theory. Trends in the data suggested the computer narrative was most effective in situations where it was necessary to discriminate among ambiguous decision choices rather than in more clear cut situations. It was concluded that computer-based

decision aids have the potential to debias the decision process, but that definitive changes will not come until the technology is improved and school psychologists become more familiar with the use of computers.

ACKNOWLEDGEMENTS

I would like to express my thanks to my dissertation committee members, Dr. Robert Marcus, Dr. Donald Pumroy, Dr. Arnold Spokane, and Dr. Henry Walbesser whose advice and support were invaluable to the development and completion of this project. Special thanks go to Dr. William Strein, my advisor and head of this study whose feedback, encouragement and suggestions allowed me to complete this project. Further, I would like to thank Dr. Henry Walbesser for introducing me to the potential that computers and information systems have for the world and for the professional. He was also instrumental in helping me to understand how statistics could be made meaningful. Also the ongoing encouragement and example of my friend and peer, Dr. Marie Van Melis-Wright, was critical in the completion of this project.

Thanks is due to my parents who never lost faith in their son and to my wife's parents for so often providing a quiet place for me to work when I needed it the most. However, it is my wife and best friend Jane, and our two children, Sarah and Jonathan, who have given the most. They have lived this experience with me, day by day and year by year...for far too long a time. Without Jane I would have been unable to separate my responsibilities as a father and my need to focus on this study. She was

patient, encouraging, and did more than her share for me even though her personal needs were frequently unmet and the times were rough and stressful. Sarah, in her own way, structured me into "working" after she would go to bed each night and although Jonathan is barely six months old he slept enough to allow me to do what I needed to do.

This project, although written under my name, was a project in which many individuals have fulfilled important and highly individual roles. I thank them all and hope that others may be as lucky as I to have the advisors, peers, and family that I have.

TABLE OF CONTENTS

<u>Section</u>	<u>Page</u>
LIST OF TABLES.....	vi
CHAPTER I. INTRODUCTION.....	1
Statement of the Problem.....	5
Hypotheses.....	5
Supplemental Questions.....	6
CHAPTER II. REVIEW OF THE LITERATURE.....	7
Scope and Focus.....	7
The Decision Process.....	8
Background Information.....	8
Approaches to the Decision Process.....	9
Overview.....	9
Psychological Decision Theory.....	11
Background.....	11
The Availability Heuristic.....	14
The Representativeness Heuristic.....	16
The Decision Making Process.....	19
How are Decisions Made?.....	19
Clinical/Statistical Decision Making..	23
Decision Improvement Strategies.....	27
Noncomputer-aided.....	27
Statistical/computerized decision aids.....	29
McDermott Multidimensional Assessment of Children.....	45
CHAPTER III. METHODOLOGY.....	51
Subjects.....	51
Independent Variables.....	53
Group 1: Computer Decision Aids Not Available.....	54
Groups 2 & 3: Computer Decision Aids Available.....	54
Dependent Variable.....	57
Procedures.....	58
Data Analysis.....	59
Threats to Validity.....	62
Internal/Statistical Conclusion Validity..	63
External/Construct Validity.....	65
CHAPTER IV. RESULTS.....	70
Hypothesis One.....	71

Probability Estimates for Each Handicapping/Nonhandicapping Condition.....	71
Categorical Decisions.....	74
Prognostic Probability Estimates.....	76
Hypothesis Two.....	79
Probability Estimates for Each Handicapping/Nonhandicapping Condition.....	80
Categorical Decisions.....	84
Prognostic Probability Estimates.....	87
Supplemental Question One.....	92
Group 1: Computer Decision Aids Not Available.....	95
Group 2: Computer Narrative and Summary Statistics Available.....	95
Group 3: Computer Narrative, Statistics, and Classification Available	96
Supplemental Question Two.....	97
Professional Experience as Related to Influence of Data Sources...	99
Degree of Experience Using Computers as Related to Influence of Data Sources.....	100
CHAPTER V: CONCLUSIONS.....	102
Scope and Focus.....	102
Rationale.....	102
Hypotheses.....	103
Methodology.....	104
Results and Conclusions.....	104
Limitations.....	115
Implications for Future Research.....	117
Implications for Professional Practice.....	120
APPENDIX A M-MAC User Instructions.....	122
APPENDIX B Case Data Provided Subjects.....	130
APPENDIX C M-MAC Options Utilized to Generate M-MAC Output for Groups 2 & 3.....	134
APPENDIX D M-MAC Narrative Output Provided Groups 2 & 3.....	136
APPENDIX E Cover Letter.....	141
APPENDIX F Decision/Response Sheet.....	143
APPENDIX G Response Permutations.....	147
REFERENCES.....	150

LIST OF TABLES

<u>Number</u>		<u>Page</u>
1	Extent of Computer Use by School Psychologists..	53
2	Summary Table of Kendall Coefficients of Concordance for Probability Estimates of Handicapping/Nonhandicapping Conditions for Each Subject Group.....	72
3	Mean Rankings of the Probability Estimates of Handicapping/Nonhandicapping Conditions for Each Subject Group.....	73
4	Diagnostic Agreement Exceeding Chance for Categorical Diagnoses by Type of Decision Aid Available According to Weighted Kappa.....	75
5	Percentage of Psychologists Choosing Each Categorical Handicapping/Nonhandicapping Condition According to Subject Group.....	76
6	Summary Table of Kendall Coefficients of Concordance for the Probability Estimates for Prognostic Questions 1-5 for Each Subject Group.....	77
7	Mean Rankings of the Prognostic Probability Estimates for Prognostic Questions 1-5 for Each Subject Group.....	78
8	Summary Table of Friedman Analyses of Variance for the Probability Estimates of Handicapping/Nonhandicapping Conditions for Each Subject Group.....	80
9	Pairwise Comparisons of Differences in Probability Estimates of Handicapping/Nonhandicapping Decisions for Each Subject Group.....	81
10	Percentage of Psychologists Choosing Each Probability Estimate for Handicapping/Nonhandicapping Condition According to Subject Group.....	83
11	Subject Group Comparisions for Categorical Decisions.....	85
12	Summary Table of Friedman Analyses of Variance for the Probability Estimates for Prognostic Questions 1-5 for Each Subject Group.....	87

13	Pairwise Comparisons of Differences in Probability Estimates for Prognostic Questions 1-5 for Each Subject Group.....	88
14	Percentage of Psychologists Choosing Each Probability Estimate for Prognostic Questions 1-5 by Subject Group.....	90
15	Summary Table of Outcomes of Analyses for Questions One & Two According to Type of Decision.....	91
16	Data Sources Available to Subject Groups.....	93
17	Percentage Influence of Case Data Sources by Experimental Group.....	94
18	Summary Table of Friedman Analyses of Variance of Degree of Influence of Data Source by Type of Data Source for Each of the Three Decision Making Subject Groupings.....	96
19	Coefficients of Correlation Between Years of Professional Experience as a School Psychologist & Influence of Case Data Sources.....	98
20	Significant Coefficients of Correlation Between Extent of Use of Computers in the Course of Professional School Psychological Practice and Influence of Case Data Sources.....	100

CHAPTER I: INTRODUCTION

It is impossible for the behavior of a single, isolated individual to reach any high degree of rationality. The number of alternatives he must explore is so great, the information he would need to evaluate them is so vast that even an approximation to objective rationality is hard to conceive.
(Simon, 1976 p. 23)

A major component of a school psychologist's responsibility is to administer and interpret psychological and educational tests for the purpose of assisting in the determination of eligibility for special education services in accordance with state and federal guidelines. The psychologist is bound by standards of ethical practice and/or statute to provide the information that will maximize the likelihood that each student evaluated will be provided that educational program best suited to the individual's needs. Usually this information is presented in the form of a written psychological evaluation. In practice, there tends to be a general dissatisfaction (Lidz, 1981) with the results of these evaluation reports among psychologists themselves and among their prime source of referrals, the classroom teacher. Engle (1966) suggested that school psychologists tend to feel uncertain about the treatment recommendations they have made and the general lack of a clear and direct relationship between data gleaned from the evaluation session and actual academic functioning of the student in question. Different theoretical

orientations confounded by each psychologist's idiosyncratic schema and knowledge base has significantly limited consistency of diagnostic decisions across students. Not only are some students being placed in special programs on the basis of systematic error in the decision making process, but their rights to equal protection under the law are being significantly compromised (Smith, 1984).

The choices that a psychologist makes and the parameters surrounding those choices define the quality and reliability of any diagnostic decision a psychologist makes. Good ethical practice in test interpretation and decision-making appears to rest more upon the clinical skills of the psychologist than on any definitive scientific base. Although the literature is rife with scientific evidence and scientifically-based hypotheses about the meaning of particular constellations of test scores and clinical signs, the final interpretation of a given set of data depends primarily upon the manner in which the psychologist has integrated these findings with the available data. Bazerman (1984) and Simon (1976) have hypothesized that evaluator emotional, motivational, and cognitive variables actively bias the process. Given microcomputer-based decisions aids; i.e. interpretative computer software for use by school psychologists, it is now theoretically possible to change this schema in a manner that decreases decision bias and systematic error

in the decision making process. This is because computer software typically makes decisions in a lawful, consistent manner irrespective of many of these influences. Modifications in the decision making process can be made as necessary by adjusting the algorithms on which the software is based, but the decision process can not change across clients solely on the basis of psychologist variables.

Considered together with the labor and time-saving aspects of automation and the increasing pressures to put microcomputers in every school across the nation, a move towards the active use of microcomputers as decision aids by psychologists is inevitable and is beginning to take place. As with most new forms of technology, however, it is very easy for rapidly rising expectations and positive face validity to obscure the need for sufficient study prior to implementation. Kaufman (1979), Matarazzo (1983), and others have argued that computerization is antithetical to the assessment process. Their concerns are based not so much on definitive evidence, but perhaps understandably on the implied threat to the humanity of the process and to a perceived subsequent increase in interpretive error. Somehow it is less difficult to rationalize human error than it is to rationalize machine error. Similarly, Gutkin and Ework (1985) have suggested that instead of decreasing errors in decision making; the computer has the potential to exacerbate the

uncertainties of data that are already uncertain by encouraging decision making on the basis of greater amounts of uncertain information. Others (McDermott, 1982b; Murphy, 1987) are more optimistic. The American Psychological Association's (1986) Guidelines for Computer-Based Tests and Interpretations have made a beginning attempt to address these issues by specifying that "...Psychologists offering scoring and interpretation services are able to produce appropriate evidence for the validity of the programs and procedures in arriving at interpretations" (p. 15). Therefore the use of any decision aid can only be ethically justified if its use in the decision process has been validated.

Idiosyncratic clinical interpretations and decisions are accepted practice in psychological assessment, but are difficult to evaluate. Individualized integration of data by a clinician inhibits definitive research by conceivably prescribing a separate validity study for each interpretive decision made by each clinician. The inherent unreliability and ambiguity of diagnostic classifications (Lanyon & Goodstein, 1971) further complicates any validation process. Computer-based test interpretation software, however, standardizes data integration across decisions in a manner that promotes rather than inhibits its study. Thus there is a dual opportunity to study both an emerging technology and a heretofore difficult to study decision making process.

Statement of the Problem

The purpose of this study then was to assess the utility of interpretative software for school psychological decision making in its role as a decision aid for the practicing school psychologist. More specifically, representative software was evaluated with respect to the degree to which its uses affects decisions made by psychologists in their role as diagnostician. This investigation utilized the McDermott Multidimensional Assessment of Children (McDermott & Watkins, 1985) software because it utilizes psychological test and diagnostic data from several sources that are typically used by a psychologist in the course of a psychological evaluation. It is chosen also because it represents a new generation of interpretative software because of its use of several data sources and its general availability. Also it was chosen because it was designed to overcome many of the biases inherent in psychological decision making through a utilization of decision matrices that were hoped by the authors to become a defacto standard for diagnostic decisions made by practicing school psychologists.

Hypotheses

1. Diagnostic agreement among school psychologists who use computer-based decision aids is significantly higher than diagnostic agreement of school psychologists who do not use computer-based decision aids.

2. Diagnostic decisions made by school psychologists on the basis of different applications of the same computer-based decision aid will be more similar to each other than to those diagnostic decisions made by school psychologists not utilizing computer-based decision aids.

Supplemental Questions

1. For each of the three conditions, which data sources are the most influential in the decision making process of the practicing school psychologist?

2. What is the relationship between school psychologists' professional experience level and degree of experience with using computers in clinical practice to the sources of information which they find most influential in the diagnostic decision making process?

CHAPTER II: REVIEW OF THE LITERATURE

Scope and Focus

Choices and decisions are made every day by every individual. The decision process is central to the typical activities of the school psychologist. Recent surveys (Ellis & Ray, 1983; McCullough & Olson, 1982) noted that practicing school psychologists typically spend approximately 40% to 70% of their working hours in the assessment and evaluation of referred students. Each referral typically calls for several judgments on the part of the psychologist. In most cases the psychologist is called upon to render a diagnostic judgement relevant to one or more educationally handicapping conditions as specified by state statute and federal law. Retrospectively many of these decisions will appear to have been valid or of good quality and others not as valid or good.

This review of the literature will address how decisions are made by school psychologists and how those decisions might be improved. More specifically, school psychologist decision making will be defined in terms of psychological decision theory (Kahneman & Tversky, 1972, 1973; Kozaielecki, 1975; Tversky & Kahneman, 1974, 1980); a theory of decision behavior directly relevant to the practicing school psychologist. Decisions will be examined in terms of the manner in which they are made, their validity and reliability, factors which determine

their outcome, and how the decisions might be improved through the use of decision aids such as computer-based test interpretation (CBTI) software (Fowler, Finkelstein, & Penk, 1986). Ancillary decision theories and information from other disciplines will be addressed only as they directly relate to a fuller understanding of the decisions made by school psychologists in the course of day to day practice.

The Decision Process

Background Information

The ethical, professional, and legal importance of these decisions has been the basis of a large body of research that has attempted to determine the validity and reliability of the decisions that school psychologists and other mental health or educational professionals make on a daily basis. The literature is beset by a number of basic research issues that have not yet been adequately addressed. These include generalizability of samples chosen for investigation, the problem of statistical versus practical significance of decisions made, errors in the computation or use of statistics and problems of definition and clarity of diagnostic terms (Garfield, 1978; McDermott, 1982b; Spitzer, Endicott, Cohen, & Fleiss, 1974). In spite of these difficulties the available research has concluded (Achenbach & Edelbrock, 1978; Peterson & Hart, 1978; McDermott, 1982b) there is a general lack of significant classification congruence

or consistency among the decisions offered by school psychologists and other mental health and educational professionals. There are also significant differences in the manner in which different school psychologists integrate the available data prior to making a decision (Gredler, 1986; Klett & Pumroy, 1971; Rassen, 1979). These results are, according to McDermott, a direct function of several factors including the incompatibility of available classificatory criteria, the absence of universally accepted standards for diagnostic decision making and deliberate or inadvertent noncompliance with existing standards.

Approaches to the Decision Process

Overview

There are many approaches to the decision making process and nearly as many hypotheses that attempt to explain why human judgments are often fallible and subject to inconsistency and bias. These types of conclusions have been consistently reported in the literature since Meehl's (1954) classic paper which concluded that statistical methods of decision making were superior to clinical decision making methods. Each approach to decision making is intimately tied to the philosophy and assumptions of the discipline from which it evolved. The vast majority of the research in decision processes has been discipline-specific and completed in relative isolation with only moderate

regard to research in other fields (Anderson, Deane, Hammond, & McClelland, 1981). Ambiguities and other inconsistencies in terminology across fields of study further complicate integration of decision research and as yet have precluded any comprehensive integration of the many approaches (Hammond, McClelland & Mumpower, 1980).

Hammond et al (1980), however, were able to delineate two major categories of decision theories, prescriptive decision theories and descriptive decision theories. Prescriptive theories such as generic decision theory (Keeny & Raffia, 1976) prescribe what the decision process is to be if it is rational and lawful. According to generic decision theory and its variants, choices of alternatives in a decision are a function of the probability of the occurrence of the alternative as well as the utility or usefulness of the alternative to the decision maker. The remaining theories, including psychological decision theory (Kahneman & Tversky, 1982a), are not concerned with prescriptions of what the decision process should be at all, but are wholly concerned with describing the process in terms of the effect of cognitive, environmental, and social factors that influence the decision making process.

Psychological Decision Theory

Background. Psychological decision theory (PDT) approaches to decision making attempt to explain errors made by human decision makers in terms of how probability estimates of alternatives under consideration are influenced by contextual factors. PDT further offers suggestions as to how the errors can be corrected. Persistent use is made of psychological, in contrast to statistical, concepts to explain the departure of the probability estimates of human decision makers from optimality. PDT holds that the decision task is strongly influenced by the external environment, the task itself, and by the cognitive system of the decision maker (Tversky & Kahneman, 1980). The decision making environment is considered to be multidimensional with varying and simultaneous dimensions of uncertainty and complexity which affect the ultimate decision made.

The decision maker is faced with a given amount of data that needs to be integrated or reconciled in some manner that optimizes the validity and reliability of the decision to be made. The efficacy of the decision made is partially a function of the manner in which the decision maker structures the problem and defines the data that is or is not relevant to the situation. According to Feldman's (1981) perspective to psychological decision making, people apparently do not necessarily accurately respond to the content of a

communication automatically. Instead they respond initially in terms of prepackaged stereotypes or categories whose accuracy is unchallenged until the incoming information reaches some hypothetical threshold of discrepancy with the stereotype. Similarly, Mischel's (1969) approach to personality assessment asserts that judgments of human personality are a function of the decision maker's perceptions with the human mind functioning "...like an extraordinarily effective reducing valve that creates and maintains the perception of continuity even in the face of perceptual (sic) observed changes in actual behavior..." (p. 101-102). People seem to begin with complex explanations for behavior and eliminate potential causes as the evidence warrants.

Heuristics are cognitive rules of thumb for simplifying complex problems in a manner that increases the probability that a solution will be found within a reasonable amount of time (Anderson et al 1981). According to psychological decision theory, heuristics fundamentally influence the ultimate choices that a decision maker will make (Tversky & Kahneman, 1974; 1982a). "The impossibility of having complete, reliable, predictable information about people and social interactions suggests that people adopt heuristics that enable them to make inferences and predictions from what

scanty and unreliable data are available" (Taylor, 1982 p. 191). It appears that once an uncertain situation has been perceived or interpreted in a given fashion, it is quite difficult to see it in another fashion as other interpretations of the situation are inhibited.

Kaplan (1975) explains this phenomenon in terms of the decision maker vis à vis the stimulus data on which the decision will ultimately be based. He cites four potential avenues through which individual differences across decision makers can be expressed. Judges differ in their valuation of a particular datum. Different judges differ in the conclusions that they draw on the basis of identical data and people are different in the manner in which they integrate the available information. Finally decision makers are described as demonstrating differences in pre-existing response dispositions to clients or data. In practice these differences relate to one's ability to arrive at integrated impressions from (often) inconsistent information, one's susceptibility to order effects and rigidity of first impression, and one's ability to differentiate among stimuli. These effects diminish with increased amounts of nonredundant information and increased importance of information for judgment. In the school psychology literature, Rassen (1979) analyzed the decisions made by school psychologists on student case data and noted some similar effects. She concluded that prescriptive and diagnostic

decisions were influenced by the nature of the judgement task and the salience of different diagnostic or remedial alternatives under consideration.

The availability heuristic. Research has suggested the existence of two major heuristics; the availability heuristic and the representativeness heuristic.

The availability heuristic (Nisbett, Borgida, Crandall, & Reed, 1982; Taylor, 1982; Tversky & Kahneman 1974) is a cognitive mechanism whereby individuals assess the frequency of a class or the probability of an event by the ease with which instances or occurrences come to mind. The affective strength or salience of data can increase the availability of data for consideration and therefore the subjective probability that a given decision alternative will be made (Kirsch & Ford, 1986; Kozielicki, 1975). Subjective data is more salient and more easily recalled than drier, less affectively-laden data and therefore more available to the decision maker in the decision process. For example, a suggestion that a child "may have been sexually abused" is more likely to be remembered than is the child's mid-range scores on a personality inventory.

The referral data that a psychologist receives prior to an evaluation varies in its subjectivity and affective salience. The content of this data has been shown to have a significant effect on the final decisions made

often irrespective of other pertinent case information according to conclusions drawn by Cummings et al (1986); Elders, (1977); Huebner, (1987); Huebner & Cummings (1985); Knoff, (1984); McCoy (1976); Ysseldyke & Algozzine (1981) and Zinman (1983). This occurrence is particularly apparent when the diagnostic categories are as Huebner suggests "difficult and ambiguous" and include the commonly applied handicapping classifications of "behavior disorder," "emotionally disturbed," and to a less consistent extent "learning disabled." The concept of "learning disabled," although considered to be very ambiguous in the literature and subject to inconsistent application (Dawson, 1985; Kovalski, 1985; Ysseldyke & Algozzine, 1981), is generally determined with the aid of one of several ability-achievement formulas (MacMann, 1985). It is therefore probably somewhat less susceptible to the subjective and affectively-laden referral information phenomena than the other less clearly defined diagnoses. Nonetheless Kovalski considers the learning disability decision process to involve a "confounding and confusing morass" that has not been adequately delineated and has not been empirically proven.

Information easily retrieved from memory will also be more available to the decision maker than data not as readily retrieved (Koziielecki, 1975; Tversky & Kahneman, 1982a). It is hypothesized that a clinician, when

presented with case data, will scan his/her memory to recall cases similar to the client being evaluated. Here the subjective probability of the new client being diagnosed the same as similar clients in memory would depend upon the similarity between the cases as well as the strength of any other heuristic operating at the same time. Further, data points that have been more recently examined are more available and therefore more likely to influence a decision. Feldman (1981) states that decision making typically takes place in an informationally "noisy" environment where impressions and evaluations are formed as behavior is observed sequentially. The final judgments made are based as much, or more, on what is then available in current memory storage than upon other less available stored information. In a variation on this effect, Knoff (1980) found the order of presentation of student data influenced school psychologist and special education trainee and practitioner placement decisions.

The representativeness heuristic. People often make decisions on the basis of information that has little to do with the statistical probability of something occurring but on the basis of its subjective probability; i.e. the outcome that is most representative of or similar to the input data (Tversky & Kahneman, 1971; 1974, 1982a). The subjective probability of the event is judged by the degree to which the event is similar to its

parent population or reflects the features of the process by which it was generated. In the case of the school psychologist the child who is referred as "disruptive" is likely to be considered educationally handicapped even though the probability the child is truly handicapped is low compared to the probability that the child is normal. This ordering of events by their subjective probabilities coincides with their ordering by representativeness.

The tendency of decision makers to neglect or misinterpret base rate information through the operation of the representativeness heuristic (Bar-Hillel, 1982; Lichtenstein, Slovic, Fischhoff, Layman, & Combs, 1978; Lyon & Slovic, 1976) is enhanced by any factor that increases the perceived uniqueness of the case information. Bar-Hillel (1980) concluded people order decision-relevant information according to perceived degree of relevance. Empirical research and probability distributions associated with the particular case in question is generally perceived as less relevant because it may be perceived as merely coincidental or of low relevance compared to more specific, causal, high relevance information. The high relevance, more affectively laden information will then be considered to be more representative of the case. This systematic bias will enter into the final decision made to the general exclusion of empirical data.

When special education teachers were asked to analyze objectively defined trends in student performance (Utley, Zigmond, & Strain, 1987), those teachers who utilized observational methods made significantly different and less accurate trend analyses than those teachers using varying levels of base-rate data in addition to the observations. Had the subjects utilized the generally correct base-rate information instead of observational data that was readily representative in the situation and had they applied that information correctly to the specific cases presented; their conclusions would have been more accurate. Even when subjects were given probability statements describing the base rates for decisions to be made (Tversky & Kahneman, 1973) a significant number of ultimate decisions ignored the base rate data in favor of data biased by representativeness. Further, Nisbett, Krantz, Jepson, and Fong (1982) found that people showed no systematic differences in decisions made on the basis of "obviously" haphazardly sampled information versus more standard data samples. Apparently when specific information is available the representative heuristic frequently precludes from consideration prior base-rate defined probabilities that are necessarily the best predictors available.

The major difference between availability and representativeness is that availability evaluates the

subjective probability of an outcome by how difficult it is to retrieve or construct from memory.

Representativeness, however, evaluates subjective probability by the degree of correspondence between the available data and the population at large. The representativeness heuristic is therefore more likely to be employed when data is characterized in terms of general probabilities whereas the availability heuristic is more likely to be employed when events are "more naturally thought of in terms of specific occurrences" (Kahneman & Tversky, 1972). Because both heuristics operate simultaneously on the same data sources in real time decision making it is often difficult to ascertain which of the two heuristics is primarily operational at any given instant.

The Decision Making Process

How are Decisions Made?

Decision making in humans is best described in terms of an interplay of cognitive mechanisms of varying strength and operation that together determine the factors that are considered by the decision maker and the eventual decision outcome. Each theory of decision making and its variants are different, but there is general agreement on how decisions should optimally be made. According to Bazerman (1984), Hammond et al (1981), and Janis & Mann (1977) an optimal decision is made in terms of a standard of rational comparison where

the decision maker combines, aggregates, or organizes the probabilities and utilities associated with various decision alternatives. Decisions made by school psychologists are apparently made in a similar fashion according to a study by Scherphorn (1979). He concluded that school psychologists typically generate a large number of hypotheses and eliminate alternatives as more data is analyzed. Then they tend to test their diagnostic hypotheses with additional information and continue to collect data until a recognizable pattern occurs. He found that his subjects needed as few as four sources of data before making a first statement of what was later to become a final diagnosis, although they used eight to ten additional pieces of data as supporting evidence before finalizing the decision. A similar study by McDermott (1975) suggested similar processes.

In making a decision one evaluates information that in turn forms the basis of alternative courses of action or choices. The alternative with the highest "subjective expected utility" (Fishburn, 1972) or preference is the alternative that is normally chosen as the ultimate decision. Under optimal conditions the process occurs as just described. The actions of individuals, however, are rarely completely rational and so these optimal decision processes do not necessarily mimic actual human decision processes. Instead they are influenced in a very direct way by heuristics.

Although little is known about why heuristics survive in light of frequent error and contradictory evidence, Nisbett & Ross (1980) stated that individuals tend to use and see those covariations and patterns of learning their own specific history of learning disposes them to see. People make sense out of their world by the learning of causal relationships and the organization of events into "causal schema" (Tversky & Kahneman, 1980). Statistical probabilities relevant to day to day decisions are not learned from everyday experience because the relevant experiences are not coded appropriately nor is feedback on the validity of the decisions frequently available. Therefore, it becomes difficult to objectively discriminate among decision alternatives; necessitating the ill-advised use of particular heuristics and other potentially biasing judgment procedures. Bazerman (1984) stated that individuals have a tendency to ignore disconfirming information because it is usually easier to find supporting information for a given decision. People also tend to escalate commitment to a previously selected course of action to a point beyond that which a rational model of decision making would prescribe. It is generally felt the principles that people tend to adopt represent significant beliefs that play a substantial role in retarding the learning of correct rules.

Decision makers are hindered as well (Bazerman, 1984; Feldman, 1981) by the time and search costs necessary to get what they believe is sufficient information to make what Reitman (1984) considers to be an "ill-defined" psychological decision that does not have an explicit criterion or set of unvarying variables. For school psychologists at the mercy of heavy caseloads, daily time constraints, and the necessity to compile student information of various and or unknown validities from many sources, it is not surprising that time and search costs are often compromised in the interest of servicing the student in a timely manner. Accordingly (Simon, 1976), a decision maker who tries as best as possible to make a good optimizing decision ends up with an often unsatisfactory suboptimized solution. The solution is one that maximizes some of the utilities that were expected to be gained at the expense of losing other utilities. Simon argues that the decision maker "satisfices," rather than maximizes; i.e. searches for a course of action that is "good enough" (frequently a heuristic) that meets a minimal set of requirements. This use of "bounded" rationality allows individuals to resort to gross simplifications for dealing with complex decision problems that may or may not yield optimally valid or reliable results.

Clinical versus Statistical Decision Making

Generally speaking, the psychologist, as a trained professional, has the latitude to make diagnostic decisions in any manner and, within reason, with any data sources that appear to be situationally appropriate and within accepted ethical and legal standards. Typically the school psychologist makes decisions in an intuitive, "clinical" manner that considers student data in the context of the practitioner's training and experience. Less frequently, the psychologist will use one or more decision aids, i.e. formalized (statistical) predictor equations, computer-based test interpretive programs, or similar computer-based resources to assist in the decision making process as ethical guidelines (American Educational Research Association [AERA], American Psychological Association [APA], & National Council on Measurement in Education [NCME], 1985; APA, 1986) prohibit decision making solely on the basis of actuarial or machine processes. According to Meehl (1954), intuitive or clinical decision making is a process whereby the decision maker, on the basis of available data, formulates a hypothesis, relates a hypothesis to the subjective probabilities of events that are related to the situation and then arrives at a prediction or decision about what is expected to happen. Actuarial or statistical prediction, on the other hand, uses codifiable or quantifiable data input measures and

mechanically combines it according to a predefined (usually statistical) formula (Sawyer, 1966). Usually the decisions are described in terms of probability statements. These and other techniques reported in the literature include regression analyses (Dawes, 1982; Grebstein, 1963), configural analyses (Wiggins & Hoffman, 1968), hierarchical cluster analysis (Cleary, Humphreys, Kendrick, & Wesman, 1975), principle components varimax analysis (Achenbach & Edelbrock, 1978), multiple discriminant function analysis (Petersen & Hart, 1978; Sichel, 1982) and systems actuarial procedures (McDermott & Watkins, 1985).

On the basis of a review of several empirical studies that compared clinical decision making with decisions made on the basis of statistical or other formalized procedures, Meehl (1954) concluded that statistical procedures were approximately equal or superior to those made by a clinician with respect to comparisons made against a predefined standard or criterion. These studies generated many other studies, the majority of which tended to confirm Meehl's general conclusions (Goldberg, 1959, 1968; Grebstein, 1963; Hill et al, 1978; Lyle & Quast, 1976; Peterson & Hart, 1978; Sawyer, 1966). It was further suggested that those valid instances of clinical judgment, where they existed, could be largely duplicated by mechanical combination such as

by a simple regression equation (Dawes, 1982; Goldberg, 1970; Hill et al, 1978).

According to Dawes (1982) these types of linear models work well because decision makers are generally good at picking out the right predictor variables and coding them but they are not as good when it comes to integrating the data because of the strength of heuristics and other biasing influences. He described a series of studies in which predictions of school success were made by a variety of subjects. The actual ratings made were objective and thus could be validated based on actual data. Several statistical and random models of the judges' decisions were then developed and the same predictions were made once again. In general, all of the derived models had a higher validity than the predictions made by the human judges.

In the Peterson & Hart (1978) study, a retrospective multiple discriminant function analysis of school psychologists' clinical decisions concluded that the variables of Full Scale IQ and student race accounted for almost as much of the differences between all psychologically-defined federally designated educationally handicapping conditions (diagnoses) than did the entire set of all other available data sources. DeWitte (1976) found similar evidence of decision bias by school psychologists with respect to socioeconomic status of school children. In the case of the diagnoses of

"learning disability" and "seriously emotionally disturbed" federal regulations (United States Office of Education, 1977) require that data from many data sources (excluding race and socioeconomic status) be an implicit part of the definition. The authors concluded that the learning disabled and seriously emotionally disturbed diagnoses were completed in a "completely inferential" manner that did not take into consideration the many appropriate data sources considered to be an implicit part of the diagnoses rendered. Similarly, DSM III (American Psychiatric Association, 1980) diagnoses and similar, ICD-9 (World Health Organization, 1978) diagnostic reliabilities only approximate 60-67% (Achenbach & Edelbrock, 1978); again indicating significant inconsistency in diagnosis across clinicians.

Classification itself involves establishing orderly relationships among individuals according to similarities or specified attributes. Empirical classification methods can be used to identify subgroups of highly similar cases and by inference to define the number and nature of underlying disorders. Nonempirical methods can only approximate these relationships because of ambiguities and inconsistencies in the classification system and the difficulties inherent in attempting to classify a set of social behaviors or feelings into a discretely defined set of diagnostic parameters. Therefore, the reliability and validity of current

classification practices (McDermott, 1982b) are necessarily compromised. Meehl (1954) stated (p. 130) "...no set of percentile ranks, no graphical representations of personality components, and no paragraph of characterological descriptions can contain all of the richness of our immediate experience." It is apparent that this richness of immediate experience has a dual effect that both enhances the apparent quality and validity of clinical judgment, while at the same time introduces systematic biases that scientifically and ethically should not exist. Achenbach and Edelbrock (1978) were more pragmatic in their observations and asserted that "shot-gun approach[es]" to decision making were unacceptable for professional practice. They called instead for "sophisticated strategies of data reduction and analysis" to assure better service provision for individuals evaluated.

Decision Improvement Strategies

Noncomputer-aided. Academic credentials and experience variables do not seem to improve the quality and consistency of the clinical decisions made (Grebstein, 1963; Goldberg, 1968; Wiggins, 1973) or in the susceptibility of the decisions to outside influences unrelated to the case data (Temerlin, 1968). Instead it seems to be necessary to provide training in strategies of decision improvement (Bazerman, 1984). The first of these are so-called debiasing strategies. As a goal,

they operate to improve the permanent quality of an individual's intuition through a personal change process. Through personal change and a knowledge of potential decision biases, the trained decision maker is supposed to make less biased decisions. Essentially, decision makers are warned about the possibility of bias, the direction of the bias is described, and the decision maker is given both immediate and followup feedback regarding the accuracy of decisions made. Both Fischhoff (1977) and Lichtenstein et al (1978) have shown, however, that even when biases are explicitly described to individuals and they are asked to avoid them, the biases are still demonstrated.

Use of decision analysis techniques have also been suggested as an alternate procedure to the more comprehensive debiasing training just described. Their primary objective is to make a rational and systematic decision in each unique situation encountered as compared with more generalized behavior changes that occur as a result of effective debiasing training. This apparently is accomplished through means of ample time, ample resources, and "sophistication" in decision analytic techniques which necessitate the explicit utilization of probability and utility theory (Ginsberg, 1972). Although Bazerman (1984) and others find merit in this technique, it is generally concluded (Friedman, Howell & Jensen, 1985) there is not enough information available

on human cognitive processes to determine what a given decision maker's limitations are, how many of the basic cognitive processes actually occur in any particular decision problem, or how such processes act to affect system output. Until these processes are better understood, attempts to train individuals to make more consistent and less biased decisions will remain of limited utility.

Statistical/computerized decision aids. Bazerman (1984) and Hill et al (1978) strongly indicated that the use of actuarially-based or statistical decision models can significantly improve decision making strategies over those which are unaided. These models may circumvent many of the difficulties encountered when one attempts to change human decision making behavior through training. Shanteau (1978) categorically stated that if humans cannot be trained to avoid biases in decision making, then computers should become directly involved in the decision process. Current ethical guidelines are not generally supportive of the use of computer-based decision making procedures without concomitant review of any decision made by a trained professional. Ultimately, automated decision making methods must then be considered to be decision or judgement aids as no decision made is acceptable without human review.

Given the recent widespread availability of microcomputers, it is has become possible for the

mathematical operations and data transformations of actuarial decision making to be handled relatively painlessly in the course of the decision process. Often subjective information can be transformed into typically nominal or ordinal scaled data that can be integrated within the constraints of mathematical computations. Compilations of statistical equations utilized in this way are most accurately termed decision aids (Hammond et al, 1980) or decision support systems (Bennett, 1983) because they address themselves to the cognitive needs of the decision maker after acquisition of data.

Computer-based test interpretation software is a type of automated decision aid available for use by school psychologists. These function to assist the decision maker in the assignment of correct probabilities to various decision outcomes or alternatives. In the case of the school psychologist, they help determine probabilities related to the determination of diagnoses or intervention recommendations. They support rather than replace the decision maker and focus on improving the effectiveness of decisions made rather than just their efficiency. In decision support systems the goal is to increase the decision maker's ability to deal with complexity and uncertainty. Nonetheless, because decision rules and the models upon which they were developed vary from system to system program, users are expected to apply their own clinical judgement on how the

results of the software analysis should be applied to the decisions they are attempting to make (Brantley, 1984).

Any computerized diagnostic system is defined by its decision making logic. Technically two primary modes of decision making by computer have emerged (Miller, 1984); the logical decision tree method and the statistical method. A third mode, artificial intelligence, shows much promise, but because there are no systems currently available for use in school psychology they will not be further discussed. The overlap between the two primary modes is substantial; especially with the development of more technologically sophisticated software decision aids that utilize combinations of both modes. Statistical methods include those based primarily on multiple discriminant function analysis, Bayesian formulas, or other actuarial formulas. The decision tree methods are derived from algorithms that test input data according to responses to a set of predetermined binary yes or no type questions; each binary response branching to another predetermined question in a manner that eventually yields an overall diagnosis when the algorithm is completed. The object of the decision tree method is to choose the path between questions that will maximize the expected utility of the outcome. This method is analogous to the "cookbook" type approach used by the psychiatric community to make DSM III-R (APA, 1987) diagnoses. Advocates of this approach such as

Spitzer, Endicott, Cohen, and Fleiss (1974) and others have designed systems to render a diagnosis by having the computer program compare clinical descriptors or other input values against a predetermined standard. A diagnosis is generated when there is a sufficient match between the input data and specified criterion.

Most interpretive software systems in use today are a hybrid of the decision tree and statistical methods. The simplest systems available are test scoring programs that score test protocols administered according to test directions. They sometimes also administer the actual test. Output generally includes raw and adjusted scores and other data normally attained when a clinician scores a test manually. Usually the scores are given a descriptive interpretation on the basis of preprogrammed descriptive statements. Algorithms used in generating the descriptive statements are based both upon statistical methods and upon rules generated by any method the software author has decided upon. These types of programs offer reliability in their output, but their overall validity depends upon the validity of the algorithms programmed.

Another common type of interpretive software is the more sophisticated clinician-modelled program. An example is The Kaufman Method of WISC-R Hypothesis Generation (Ingram, 1985) for the WISC-R. The algorithms that make up these programs are written in a manner that

the output is identical to that which a renowned expert (in this case Kaufman) would generate if presented with the same input data. They typically are based upon a simple linear model derived from a regression analysis of the clinician's decisions.

So-called clinical actuarial programs provide extensive narrative descriptors and clinical hypotheses as output using clinical research findings based upon particular score patterns as well as varying amounts of clinical lore. Algorithm design is similar to the other methods and is usually a combination of statistical weighing schemes combined within a usually well-defined decision tree. There is often no attempt to integrate the data on the basis of available research data. Instead the author designs the analysis on the basis of his/her own unique approach to the assessment device.

Although most of the available interpretive software systems integrate data from one commonly used standardized test, more systems are becoming available that analyze information from more than one source. The WISC-R Factor Analysis for the Apple (Lichtenwald, 1982) is an example of the former and uses data from the WISC-R only whereas the McDermott Multidimensional Assessment of Children (M-MAC) (McDermott & Watkins, 1985) and School Psychology Information Tools (Hale, 1982) integrate a variety of data inputs from several test and nontest sources.

Other approaches to computer-based decision aids have integrated a combination of interview or questionnaire data. These, in addition to selected test results and other relevant and system specific data are used as input to generate DSM III diagnoses and in some instances treatment plans. In an ongoing study of one such system, the DIAGNO series at the Salt Lake City, Utah Veterans Administration Hospital; Johnson & Johnson (1981) and Johnson & Williams (1975; 1980) and others have developed and implemented an automated assessment system for incoming (psychiatric) patients. Hedlund, Evenson, Sletter, & Cho (1980), Miller (1984), and Fowler (1985), reviewed these approaches in terms of assessment quality, diagnostic accuracy, decision making, patient and staff acceptance, and cost efficiency. It was generally concluded that the results of the automated system were at least equal to and often superior and more internally consistent than those generated in the traditional manner. Further, completed patient analyses were received twice as fast and cost half as much as the traditional evaluations.

Space (1981) described advantages to the use of computers as decision aids because they allow the knowledge of the best clinicians to become more widely available. He stated this encourages clinical decision makers to be clearer in the articulation of their decision making rules. He also suggested that if the

decision rules were to come from expert clinicians the hit rate of correct decisions would be maximized and well above that of the average clinician. Oswald (1986) added that interpretive software systems can help to avoid overclassification of individuals through their rigid and reliable application of decision criteria.

Space also listed some arguments against the use of interpretive software as well. He indicated that because interpretive software systems generally approach an individual as a point on a continuum in comparison with the rest of the population, the analyses generated are not geared to the regularities and patterns unique to the individual. Therefore, clinically more idiographic information is lost in favor of nomothetic information. He also noted their use is eschewed because it is felt the statistical algorithms tend to result in "trivial" decisions such as placing an individual into a diagnostic category to which he/she is already known to belong. According to Nichols & Knoff (1977) and Sampson (1983), interpretive software can omit or overgeneralize information which confuses the decision maker to the point where the information is misleading or of limited utility. Gutkin & Elwork (1985) suggested that computer assistance may encourage decision making on the basis of greater and potentially more subject to error amounts of information that gives the resulting output an aura of objectivity "even when it is lacking." A less moderate

opinion (Sinnett & Albott, 1987) considered the use of computers in psychology to be "...gimmicks that allow ... the psychologist, to avoid accountability under extant ethical principles" (p. 191).

Arguments on both sides of the automation of psychological practice continue. There appears, however, to have developed a general consensus that currently available interpretive software products are not sufficient in and of themselves to make diagnostic decisions and instead should be used as a resource for the practicing psychologist

(Jacob & Brantley, 1986; Walker & Myrick, 1985).

According to Doll (1985) computer assistance should be used to support professional competence, but not extend it. Although interpretive software systems have many weakness, Murphy (1987) feels they are a valuable tool. Instead of questioning their use, the professional should begin to question the continued use of unaided clinical judgments as the sole source of professional diagnostic decisions. In a recent survey of school psychologists who were current microcomputer users (Brantley, Jacob, & Troutman, 1985) most microcomputer usage was devoted to statistical analyses, data storage, and word processing for report writing. Use of the microcomputer for test interpretation was only moderate in comparison.

A decision or judgment aid must be judged in terms of its quality to discriminate at the boundary lines

where decisions are to be made (Cronbach, 1965). When mathematical formulas are used as guides to policy-making they carry hidden value judgments that the decision maker might not be unwilling to accept if they were considered explicitly. Further, value judgments are often unavoidable when one attempts to quantify factors that have not been satisfactorily quantified by generations of psychologists and educators. Nonetheless one should keep them to a minimum to avoid inconsistencies in the overall decision system. Similarly Greenberg (1986) advises that the results of any analysis are totally dependent upon the coding systems used to obtain or structure the data used in a decision or judgment support system. Any algorithm, mathematical model, or decision tree used can only be an approximation of the real world and necessarily becomes more complex and sacrifices reality in order to gain information and computational advantages to the decision maker (Dyer & Mulvey 1983).

Irregardless of their use of data and internal programming, computers necessarily bring structure to decision making (Sidowski, Johnson, & Williams, 1980). Nisbet and Ross (1980) stated that machine-aiding can improve performance by eliminating a cognitive operation; i.e. the need to formulate cue values (the probable value of a given datum, such as a test score). They also hypothesized the degree of improvement that is obtained

through the use of a machine reflects the difficulty of the cognitive cue formulation activity the machine is performing. The harder it is for the decision maker to formulate cue values, the more helpful a machine would be that makes the cue estimation step unnecessary.

These estimation steps, according to Simon and Reed (1976) reduce the decision maker's so-called problem space, i.e. the relative number of alternatives and data sources the decision maker has to choose from. In the case of school psychologists who were presented with a typical assessment problem and were provided with no available decision aids (Gredler, 1986) approximately 25% of the professionally trained school psychologists chose data sources that yielded a decision making "competency" score equivalent to that which would be attained if data sources would have been randomly chosen for consideration. Although it could be stated that 75% of the school psychologists made choices that improved upon random selection, the fact that the decision making competency of one in four psychologists did not improve upon random choice is ethically irresponsible.

Friedman, Howell, and Jensen (1985), designed an automated simulation that looked at how the quality of a human judgment is affected by preprocessing applied to raw data when such processing is carried out by a human with or without machine aid. The authors obtained direct measures of estimation (decision making) performance on

varied decision tasks. Levels of mechanical aid that ranged from unaided decisions, as is commonly utilized in clinical decision making, to a mechanical computation of cues important to the final decision were varied across type of decision task. Decision making performance improved systematically as the level of mechanical aid increased in the absence of any requirement for the subject to estimate cue values prior to making a decision. The act of estimating cue values alone, when factored out separately added significantly to the accuracy of the decisions made. Insofar as the results of this study were concerned, merely requiring a preprocessing estimation step either by use of a structured decision step or through the use of computer-integrated data markedly enhanced the accuracy of the final diagnostic judgment made. Similarly, Snizek & Reeves (1986) investigated the impact of what was termed a feature cue; i.e. summarized data reflecting prior performance on a cue probability learning task. Results indicated the presence of the feature cue greatly improved prediction achievement and accuracy. Similar results were obtained in a study by Utley et al (1987) where accuracy of teachers' analysis of student performance was compared with and without summarized data available. Although the diagnostic judgments made in these studies are not as complex as those required in psychological diagnosis, nor was the mechanical aiding as

complex as that normally used in software-assisted decision making, the results of both are strongly suggestive of the effect that preprocessed data can have on decisions rendered.

There is no currently available literature that has experimentally manipulated the diagnostic decision process of the school psychologist by varying the influence of interpretive software decision aids. Green, however, (1982) designed a study where a group of experienced psychiatrists, psychologists, and social workers evaluated three different interpretive software reports. Two of the reports were based on the Minnesota Multiphasic Personality Inventory (MMPI) and one was based on the Millon Clinical Multiaxial Inventory (MCMI). One hundred different cases were judged on the basis of the manner in which the interpretive software decision affected their understanding of their clients. The interpretive software reports were rated on a modified Likert scale across categories related to adequacy of information, confirmation of knowledge, addition of relevant information, exclusion of important case information and degree of inclusion of trivial and misleading information. The raters found significant differences among the three interpretive software systems in terms of information accuracy; especially in the extent to which the report "extended" or "enriched" their knowledge. A trend in the data seemed to indicate the

programs that did not appear to exclude as much data were rated as adding more to the knowledge of the user than the programs with less information.

In early interpretive programs the intent was to simulate, as nearly as possible, the analyzing and synthesizing functions of the human test interpreter and to produce a report similar in style and content to a report written by a clinician (Fowler, 1969; Klett & Pumroy, 1971). Interpretive software developed by Kleinmuntz (1963) for use with the MMPI (Hathaway & McKinley, 1942) was the first computer-based decision aid to be developed for the interpretation of a psychological test. Software for the MMPI continues to be the most widely used and available type of interpretive software product (Butcher, Keller, & Bacon, 1985). Since that time, a number of other commercial and noncommercial interpretive systems have become available, for example The Minnesota Report (Butcher, 1979), The Caldwell Report (Caldwell, 1971), and the WPS Test Report (Lachar, 1974).

Other than the MMPI, interpretive software decision aids are probably most widely available for the Wechsler Intelligence Scale for Children-Revised (Wechsler, 1974) and the other Wechsler series intelligence tests. They are most widely used to gain an estimate of the individual's cognitive abilities and to assist in the prediction of an individual's level of achievement in school. There are a wide variety of interpretive schema

available for the Wechsler series (e.g. Bannatyne, 1974; Kaufman, 1979; Sattler, 1982). Interpretation of the Wechsler series is rather complex given these interpretive schemas, factors related to chronological age, considerations of standard errors of measurement and the calculations and interpretations involved in determining statistically significant differences between subtest scaled scores and or IQ scores (Fisher & Jenkins, 1986). Further, the clinician may, and often does, choose to adopt one primary method of interpreting Wechsler profiles and excludes others because of the effects of training, habit, time constraints, and the complexities of some of the methods. Because of these factors Wechsler interpretive programs have been considered to be of some assistance to the clinician by reducing the opportunity for clerical and mathematical errors. They also make available to the clinician models for interpretation that can generate hypotheses for further consideration in the decision making process.

In a review of the WISC-R 90 program (Trifletti, Trifletti, & Trifletti, 1983) Quinlan & Quinlan (1984) described it as producing "voluminous" material that was not considered to be pertinent and that might imply serious psychopathology when psychopathology clearly was not indicated. Watkins (1986) reviewed the same program and was concerned that it did not make available evidence supportive of the validity of interpretations. He

concluded the program had serious technical and professional limitations and failed to adequately control for the possibility of misunderstanding or misuse of output. Fisher & Jenkins (1986) compared three different commonly available interpretive software programs, the WISC-R Computer Report (Nicholson, 1982), the WISC-R Monitor (Mercadal, 1982) and The WISC-R Analysis (Lichtenwald, 1982) by entering actual case data from one case into each of the three programs. In their nonstatistical comparative analysis they found many differences across the programs; both in generated content and interpretive statements, as well as in the respective approaches to the integration of the data. Some of these approaches appeared to integrate data in a manner inconsistent with commonly accepted practice. Although Fisher & Jenkins made no attempt to address the validity of the program output or to determine its affect on the decision making process of the user, they termed the programs "useful" in both clinical and training areas. Wiese and Grossman (1986), in a review of The Explorer (Vance & Paparella, 1983), and Hoffman (1985) in a review and comparison of The Explorer and nine other software programs for the Wechsler series found similar inconsistencies. They variously described some of the programs as having some "useless" features, "erratic" operation, "lacking in professional sophistication and polish," and "impressive to the non-professional

but...mundane to the practicing school psychologist." At the same time, the value to the profession varied from "weak" to "useful" to "pleasing" with "great [but unspecified] professional value."

Replogle (1982) developed a WISC-R interpretive program which was validated in a study that compared case analyses completed by practicing school psychologists with analyses generated by the interpretive software. The judges, licensed psychological examiners and psychologists with varying degrees of advanced education and experience, rated the analyses on the basis of the content and accuracy of the analyses, as well as on its strengths and weaknesses. A MANOVA analysis of the results indicated the interpretive software reports were rated significantly higher than the reports generated by human examiners.

The validity of the interpretive software systems in general has been attacked vigorously (Fowler, 1980; Matarazzo, 1983) because, for the most part, the validity of the resulting interpretations have yet to be established. For example, in a review of attempts to validate MMPI interpretive programs, Moreland (1985) was unable to reach any definitive conclusions about the their overall validity. Eyde, Kowal, & Fishburne (1986), in a comparison of nine different systems, were only able to conclude that interpretive systems are "good to use as long as good clinical judgment is used." Nonetheless

these decision aids have generally, yet variably, been determined to be well-received by clinicians. (Space, 1981).

It is difficult to determine the number of interpretive systems and similar aids that have implications for assessing human behavior. In 1984 Krug listed 190 of these products. As new products are continually being developed, this estimate should be considered to be very conservative.

McDermott Multidimensional Assessment of Children.

McDermott & Watkins (1985) refined classification analysis into what they termed "systems actuarial multidimensional classification analysis." It was presented as an alternative to traditional clinical and actuarial classification currently being utilized by school psychologists. Their intent was to develop a comprehensive and standardized system for the classification of childrens' functioning that was objective, reliable, valid, logically consistent, and useful to both practitioners and researchers. It was made available as a microcomputer-based software program, the McDermott Multidimensional Assessment of Children (M-MAC) (McDermott & Watkins, 1985). Prior versions of the software were developed as early as 1977 with a program entitled MAC (McDermott, 1977; 1980) developed for use on a mainframe computer.

The systems actuarial approach used in M-MAC systematically, and in decision tree style, considers the status of the characteristics of a child or the relationship between two such characteristics and then moves on in a predetermined sequence of such considerations. At each step questions are asked regarding the statistical probability of relationships between variables such as general intelligence, school achievement, measures of adaptive behavior, and indicators of social-emotional adjustment. Decision rules are actuarial in nature and reportedly modelled after national standards for differential diagnosis, such as American Association on Mental Deficiency (AAMD) criteria for a diagnosis of mental retardation, learning disability according to federal definition, and academic over and underachievement according to regression formulae.

McDermott and Watkins noted several advantages over more conventional diagnostic systems. First, clinical diagnoses are based primarily on experiential probability where the probability of relationships between child characteristics and consequent decisions is typically drawn from a clinician's past experiences and from their education and training. In contrast, M-MAC diagnoses are based on the statistical probability of such relationships and are derived through a normative perspective of child development. Second, the systems

actuarial approach does not focus exclusively on one or another dimension of classification as do many clinical methods. Rather, the software is programmed to assume that all children share common dimensions of personal qualities and differ only in the intensity of the relationships among these qualities. Hence, every child can be conceivably characterized on the dimensions of intellectual functioning, academic achievement, adaptive behavior, and social-emotional adjustment. Finally, the M-MAC systems actuarial approach provides classifications that are assumed to be maximally consistent across psychologists and congruent with ideal standards and accepted definitions.

M-MAC requires that the school psychologist provide certain data as system input; i.e. demographic information, standard scores for general intelligence and academic achievement, and standard scores or clinical judgments for adaptive behavior and social-emotional adjustment. The software program, in addition to the algorithms described, stores reliability and validity coefficients, means, standard deviations and other statistical information for 30 commonly and uncommonly used tests and rating surveys used by school psychologists. Therefore the clinician has a wide range of data input sources to choose among. Provision is also made so consideration can be made for special factors that may influence the validity of the actual

test scores utilized. These include, but are not limited to the existence of handicapping conditions, special linguistic considerations such as being raised in a bilingual household, cultural factors, and environmental factors.

Output is variable depending upon the choices of the clinician. The narrative output reflects the individual's performance on the assessment measures plus cautionary statements, when necessary, to indicate that the conclusions generated may lack validity. Diagnoses based on the ICD-9 and the DSM-III classification system are optionally available as are certain diagnoses not included in those systems. See Appendices A and D for details and examples of narrative output.

There is probably more research data on the M-MAC system than any other interpretive system in common use other than the MMPI systems. Nevertheless, few studies have addressed the validity or utility of M-MAC. In fact most of the studies noted were completed on M-MAC's earlier version of MAC. McDermott & Hale, 1982 and McDermott (1982a) used both highly experienced and newly trained child psychologists to evaluate and classify 73 children referred to outpatient clinics. The same test scores were used as input to MAC. The result was that conjoint agreement for each diagnostic category between psychologists and MAC was 86% in excess of chance. Reliability of diagnostic classification was assessed as

well. MAC's 100% reliability was compared to that of expert psychologists who agreed with each other at 76.5% above chance and newly trained psychologists agreeing at 4.6% above chance. According to the program manual, the current version of M-MAC was successfully field tested in five states by so-called "child specialists" (McDermott et al, 1984). In a subjective review by Thomas (1986), M-MAC was considered "cumbersome" to use in the context of day to day school psychological practice. Bracken (1986), in another review, suggested its value to the profession was minimal and could "...stimulate a feeling of technological pseudosophistication." Other "research" per se is unavailable. Although McDermott and his colleagues appear to be satisfied with the validity, reliability, and utility of M-MAC (McDermott, 1986) there is not as yet sufficient evidence the system does any more than provide what could eventually turn out to be an accepted diagnostic standard in child classification.

Rubin (1986) has concluded that computer-generated decision aids should play a role as a general screen and guide to be interpreted by the clinician but not used as an absolute standard. This is similar to the use of x-rays or laboratory test results in traditional medicine. By utilizing the computer as assisting the psychologist to interpret the data (Loesch, 1986; Walker & Myrick, 1985) fears of school psychologists being replaced by "intelligent" computers (Altemose & Williamson, 1981) can

be reduced. Instead, the computer can be considered to be a valuable assistant. In 1971 Klett & Pumroy described the state of the art regarding psychologists' use of computers. They were unable to determine what indeed the value of the computer would be to the practicing psychologist. Their question has remained largely unanswered. Therefore, the major question now is not one of how valid are the diagnostic classifications generated by M-MAC, but how does the use of computer-based interpretive software such as M-MAC affect the diagnostic decisions made by school psychologists?

CHAPTER III: METHODOLOGY

The major questions of this study are designed for analysis as an experimental posttest-only control group design and the supplemental questions as a pre-experimental design (Campbell & Stanley, 1963) that has been modified with random assignment of subjects. After a discussion of the subjects, variables, and procedures of the study; a description of the data analysis is followed by threats to its validity.

Subjects

Two hundred one active members of the National Association of School Psychologists (NASP) were selected at random from its current membership roster as subjects for this study. They were selected because Association membership is generally representative of the school psychologists practicing in the United States. Upon receipt of the subjects' names from the membership roster, the names were randomly assigned to one of three experimental groups by use of a random number table. Sixty-seven subjects were allotted to each of the three groups for an initial total sample of 201 school psychologist subjects. Out of the 201 questionnaires mailed to the subjects, 102 were actually returned for a total return rate of 50.7%. There were no significant differences in rate of return across the treatment groups $\chi^2(2, N = 201) = .3582, p = .84$. Of the total questionnaires returned three were discarded as

unusable; one because the subject had analyzed the data with microcomputer software which would confound any treatment effects and the others because the individuals were not practicing school psychologists. Further, some of the responses from thirteen other returned questionnaires were partially unusable because of missing data or because some of the responses (from nine of the subjects) made to the stimulus data were illogical and inconsistent with reality. See Appendix G for the responses omitted. These types of unusable responses were spread relatively evenly across the three subject groups.

Experience of subjects as school psychologists was not significantly different across the three experimental groups, $\chi^2(6, N = 96) = 3.48, p = .18$. Most of the respondees (92%) had more than one year of experience as a practicing school psychologist with the remaining 8% having had less than one year of experience. Nearly 37% of the respondees had over 10 years of experience.

Each of the subjects was asked to state the degree of experience they had had in using any type of computer within their day to day practice of school psychology. Again, when all possible uses of computers were considered together, the three subject groups did not differ significantly in their experience $\chi^2(6, 96) = .6948, p = .71$. As a total group, on a scale of one indicating no use of computers in current

Table 1

Extent of Computer Use by Practicing School Psychologists

<u>Type of Use</u>	<u>Median</u>	<u>Interquartile Range</u>	
Word Processing	3.00	4	
Report Writing	1.00		3
Test Scoring	1.00	2	
Test Interpretation	1.00		1
Case Management	1.00	1	
Counseling/Therapy	1.00		0
Other Uses	1.00	0	
Composite of All Uses	2.00		1

Note. Values range from 1 (No Use of Computers) to 5 (Frequent Use of Computers).

professional practice, to a maximum possible of five indicating frequent use of computers in professional practice; the overall median computer use was two with an interquartile range of one. See Table 1. Each of seven individual types of computer use were analyzed separately to determine if computer use was different across groups. In every instance there was no significant difference in computer use across subject groups. Word processing and report writing uses were most frequently reported with test scoring and test interpretation uses next most frequently reported.

Independent Variables

Availability of a microcomputer-based interpretive software decision aid was the major independent variable utilized in this study and was varied across the three subject groups. Each of the two experimental subject

groups were provided microcomputer-generated narratives in addition to case data typically available to a school psychologist in the course of a psychological evaluation. A third control group received the case data, but not the narrative. See Table 15 located in Chapter IV.

Group 1: Computer Decision Aids Not Available

The subjects in Group 1 (the control traditional decision makers) were asked to make two diagnostic decisions and a series of prognostic decisions on the basis of case data provided by the experimenter. No constraints were placed on the manner in which they made their decisions. In this way the decision making process generally paralleled the decision processes involved in actual field-based school psychological decision making. See Appendix B for the case data provided the Group 1 subjects and Appendix F for the decisions that the subjects were instructed to make.

Groups 2 & 3: Computer Decision Aids Available

The remaining two groups of subjects were asked to make the same diagnostic and prognostic decisions on the basis of the same case data and test scores which were available for the nonsoftware-aided decision makers. Each of the subjects was also provided with a copy of one form of output from an interpretive software program, the McDermott Multidimensional Assessment of Children (M-MAC) (McDermott & Watkins, 1985). The output of many software decision aids, including M-MAC, can be varied according

to different parameters defined by the user. Two out of the many possible output alternatives generated by M-MAC were utilized; a different alternative to each of the two experimental groups.

Group 2 subjects each received an M-MAC analysis that integrated case data separately across each of two domains of assessment (intellectual abilities and academic achievement). Subjects were provided printed microcomputer output which included statistically analyzed case data and an associated descriptive narrative. Group 3 school psychologist subjects received the same narrative output as described above except that it also included a multidimensional diagnostic classification. For a detailed explanation of how these alternatives were generated see Appendix A. Refer to Appendix D for the actual microcomputer narrative output.

Both computer generated output narratives included obtained WISC-R IQs, standard confidence limits, factor deviation quotients and their confidence limits, and a paragraph describing interpretations of differences among the deviation quotients. Also included was the overall qualitative level of intellectual functioning. Wide Range Achievement Test-Revised scores were analyzed by M-MAC in a similar manner. Included in the analysis were the obtained subscale scores and their respective upper and lower 95% confidence intervals. Also included were the deviation of the subscale scores from the student's

average performance level and the results of a regression analysis highlighting any discrepancy between observed and expected academic achievement. Finally a statement indicating the relative frequency of discrepancies between expected and obtained achievement in these situations in the general population was provided. A summary table concluded the output for the achievement domain and presented narrative that recorded the child's achievement status relative to both age peers and to intellectual functioning across reading, spelling and arithmetic. The Group 3 narrative also included a diagnostic classification printed at the end of the output.

Examples of Groups 2 and 3 stimulus materials are located in Appendices B and D. These stimulus materials are adapted from actual M-MAC program output, but the original output was reformatted to conserve space in mailing. A few minor changes were also made in the actual narrative in order to enhance control of external variables.

In order to better approximate the kinds of information available to the typical psychologist making a decision, a narrative (Appendix B) including reason for referral, behavioral observations, test scores, and pertinent background information was included for both the experimental groups and the control group. All subjects received the same narrative and a listing of

relevant standardized test scores. Because of the intrinsic subjective nature of projective personality instrument interpretation, objective descriptors (response themes, picture descriptions, etc.) were used in lieu of any scoring or statements "describing" the child's personality. Test scores used in the M-MAC analysis were limited to those available indicated within the case data sheet provided all subjects.

Dependent Variable

In addition to the case materials and decision aids described, each subject was provided with an instruction/response sheet (Appendix F). The subjects were asked to consider the data in the packet and place an "x" beside the diagnostic and prognostic decisions which best described the case material. Listed for consideration, in a random order determined through the use of a random number table, were the four nonmedical handicapping conditions currently defined by state and federal regulations as educational handicapping. A fifth possible diagnosis, "no educationally handicapping condition" was also available for choice and was included in the list. The subjects were also asked to make diagnoses in the form of probability estimates for each of the handicapping conditions. Here the subject psychologists entered probability estimates for each category that ranged from 0% probability to 100% probability that the child described in the case data

could be considered to be handicapped under each of the five labels previously noted. Also they were asked to respond, with similar probability estimates, to five prognostic statements relating to the child described in the case material.

The case data presented for consideration by all subjects was data obtained by a professional school psychologist from an actual test protocol which had all identifying information removed and replaced by fictitious demographic data in a manner that maintained the confidentiality of the original case material. The narrative paragraph was adapted from, but did not duplicate, actual case material to further disguise the true identity of the individual actually evaluated. The content of the case material met the definitional requirements of learning disability and at the same time generally precluded a diagnosis of seriously emotionally disturbed or multihandicapped. Descriptors used in the background information and teacher comments were constructed in a manner that typified that received in the course of a school psychologist's day to day practice. See Appendix B for the actual case material.

Procedures

Each subject was provided a cover letter (Appendix E), an instruction/response sheet (Appendix F), and the appropriate case material based upon the randomly assigned experimental condition (Appendices B & D). They

were instructed to make their diagnostic decisions and probability estimations as previously described. After the decisions had been made; they were instructed to open a sealed piece of paper and enter requested information. More specifically they were asked to enter demographic information and answer questions related to their use of computers in professional practice, whether or not they used an interpretive software or other decision aid in the decision process, and what sources of data they found to be influential as they made their decision. They were then instructed to place the completed forms in an enclosed preaddressed and stamped envelope. All subjects were requested to separately return a stamped post card with their name and address to assist the experimenter in determining which subjects needed to be sent follow-up notices. The first follow-up reminder notices were sent three weeks after the date of the original materials mailing. Another follow-up request took place two weeks later. Data analysis began after approximately 16 weeks had elapsed from the original mailing. No further responses were received after that time.

Data Analysis

The two major hypotheses of this study were designed to examine how the use of computer-based interpretive software by school psychologists affects clinical decisions made by school psychologists. Hypothesis One focused upon the degree of agreement among decision

makers and hypothesized that agreement among school psychologists who use computer-based decision aids is higher than diagnostic agreement among school psychologists who do not use computer-based decision aids. Each source of data, i.e. diagnostic decision, categorical decision, or prognostic decision was analyzed by statistical analyses appropriate for their scaler properties as well as for meeting standards for an indicator of diagnostic agreement. The probability estimates of handicapping/nonhandicapping conditions as well as the five prognostic probability estimates for each subject group were analyzed using Kendall's coefficient of concordance (W), which describes the association between the characteristics of k related samples. In this case, the relationship or association of the probability estimates for each of the handicapping conditions or each of the prognostic statements were analyzed. According to McDermott's (in press) discussion of the differences between statistical measures of agreement per se versus statistical measures which indicate relationships only among ratings, Kendall's coefficient of concordance is the appropriate statistical measure of agreement for this kind of ordinal scale data. The categorical decision, however, was analyzed by weighted kappa (K_m) (Fleiss, 1971; Light, 1971), a measure of internal consistency that allows level of agreement for nominal scale data to be measured

among more than two observers. Weighted kappa provides an outcome measure that estimates the joint agreement of m observers that exceeds chance. Significance tests and multiple comparisons appropriate for use with Kendall's coefficient of concordance (Minium, 1978) and for weighted kappa (McDermott, in press) were completed as required by the comparisons stated in Hypothesis One.

Hypothesis Two proposed that the decisions of school psychologists who have available different variations of the same computer decision aid will not be significantly different from each other. More specifically, the ultimate diagnostic/prognostic decisions made will not be significantly different even though the computer output of the decision aid is qualitatively different. The probability estimates of each handicapping/nonhandicapping condition as well as (separately) the probability estimates of the five prognostic questions were analyzed in a stepwise manner utilizing Friedman's analysis of variance for dependent variables (Gibbons, 1985). First, for each subject group, the probability estimates for each handicapping/nonhandicapping condition were compared according to a Friedman analysis to ascertain variation of the estimates across the handicapping/nonhandicapping conditions. Post hoc pairwise multiple comparisons of the handicapping/nonhandicapping conditions were completed to more precisely delineate the patterns of

subject responses; i.e. agreement across diagnostic categories. Finally, the patterns of agreement within each subject group were compared across the three subject groups to answer the overall hypothesis. For the categorical decisions chi-square analyses were utilized to test the relationships between the diagnostic decisions and subject groups for the comparisons called for in the hypothesis.

Supplemental Questions One and Two were designed to more closely examine the subjective effects of different data sources upon the decisions made by the school psychologist subjects.

Supplemental Question One was approached as a Kruskal-Wallis analysis of variance for ordinal data. Appropriate post hoc multiple comparisons were calculated to ascertain differences in influence among the respective data sources available to each individual subject group.

Supplemental Question Two was analyzed by correlating degree of professional school psychological experience with ratings on influence of data sources. A second correlation was similarly performed between degree of experience in using computers and subject ratings on the influence of data sources.

Threats to Validity

So the data and conclusions generated by this study can be put into proper perspective, threats to the

validity of the results of this study will be described based primarily upon Cook & Campbell's (1979) perspective.

The primary differences between the major hypotheses and the supplemental questions involve the availability of appropriate comparisons across subject groups. The designs for the supplemental questions do not necessarily include simultaneous comparisons with control groups. Therefore conclusions were generated based on the data from (often) one subject group which necessarily precluded controlling for many sources of variation. Consequently, conclusions generated for the supplemental questions should be considered to be preliminary analyses of relationships rather than the more definitive experimental conclusions of the major questions. Unless otherwise noted, sources of validity and invalidity to be described below will be considered to apply to both the major hypotheses and supplemental questions. However, the degree to which they apply will necessarily vary according to the question asked.

Internal/Statistical Conclusion Validity

According to Cook and Campbell, the analysis should be sufficiently sensitive so that reasonable conclusions about the covariation of the variables and the casual relationships, if any, between the variables can be made. An analysis of the various components of internal/statistical conclusion validity finds the design

of the study was generally moderately sensitive to these issues, although it is clear that adjustments in methodology would be indicated in similar studies in the future.

The dilemma of statistical power, i.e. Type II error, is a serious consideration when sample sizes are small and alpha values are set relatively low as they were in this study. Although the possibility of Type I error was minimized, there remains uncertainty as to whether actual treatment effects were overlooked because of the sample size of approximately 30 subjects per condition and because of the variation in statistical power of the nonparametric tests used. Although the Type II error rate per analysis perhaps was increased by these factors, Type I error was mitigated by maintaining an alpha level of .05 for the major comparisons. It was also minimized by making adjustments in alpha levels for the many multiple comparisons called for in the analyses. Similarly, the alpha level was made more stringent for the correlations in Supplemental Question Two in order to reduce the number of random, yet significant correlation coefficients that could be expected to be generated in the course of the 120 correlation calculations therein.

Other threats to validity may have been a function of basing the results of the analysis on one particular case or collection of stimulus data points instead of many. The use of one case to reflect the population of

cases is statistically weak and is limited in its ability to show desired phenomena. This is because there are not enough occasions for accessing the systematic presence or absence of effect of the major variables. The use of several cases would have decreased the possibility that any one aspect of the case material may have inadvertently contributed to uncontrolled variability in the outcome measures. This effect was probably most pronounced in the obtained weighted Kappa values for the categorical data of Hypothesis One.

Random assignment of the subjects to treatment or nontreatment control groups, however, effectively minimized most of the other potential sources of internal invalidity. Further those threats which randomization does not account for, according to Cook & Campbell, (imitation of treatments; resentful demoralization of respondents receiving less desirable treatments, etc.) are not pertinent to this study and have a low probability of affecting its internal validity.

External/Construct Validity

External validity is considered to be the degree to which the results of the study can be generalized to situations of interest in the environment or, in the case of construct validity, the ability to generalize to other related abstract constructs. Here the pertinent comparisons relate to the extent to which the data reflect the day to day decision making practice of

professional school psychologists. Threats to the validity of the study would tend to decrease the applicability of the data generated herein and decrease the intrinsic value of the study.

Considered as a whole construct validity was fairly adequate considering this study is one of the first studies to explore many of the constructs associated with computer-aided decision making with professional school psychologists. Although the data cannot be directly applied to all of the many constructs related to computer-aided decision making among school psychologists, it does provide a firm base that future research can build upon.

The term "agreement" as specified in Hypothesis One is one such example. According to McDermott (in press) many different statistical measures have been used to define the construct of agreement. The analyses utilized in Hypothesis One meet specific standards for the concept of agreement as described in the Data Analysis subsection in this chapter and can be considered to enhance the construct validity of this study. One drawback, however, is that the literature has not completely embraced McDermott's stringent criteria for agreement. As such many older studies and some recent studies do not define the term "agreement as stringently as does McDermott. Therefore the construct of agreement as defined herein

may be different than that predicated in some of the related literature.

To decrease reliance on one source of data for an outcome measure (mono-operation bias) and to provide stimulus materials that approximate the various types and formats of decisions made in day to day practice of school psychologists, three related, yet different types of decisions were made by the subjects on the basis of the case data. Mono-method bias; i.e. the use of one method only to gather data, was not mitigated, but is nonetheless is considered to have little effect on the validity of the study. This is because it was the intent of this study to maximize external validity per se by closely approximating decision making in the real environment of a school psychologist. Therefore, different presentation methods were not considered to be as important as the need to maximize the parallel between the method and what actually occurs in the field.

Other threats to construct validity such as evaluation apprehension, experimenter expectancies, and hypothesis guessing within experimental conditions can be considered to have had little negative impact on the interpretation of this study. This is primarily a design issue as the stimulus materials were fashioned and presented in a manner that approximated decisions made day to day more than they did that of an artificially contrived experiment. Further the experimenter himself

had no direct contact with the subjects except via the letter requesting participation in the study.

Threats to external validity per se were generally minimal. As the subjects were randomly selected members of the largest professional organization in the United States representing the interests of school psychologists there is a reasonable probability the individuals chosen reflect the population of school psychologists at large. Also because they were randomly selected the probabilities are lessened that there was an interaction of selection of subject and treatment effect. Any definitive statement would be premature because approximately 50% of the subjects chosen did not choose to participate in the study. An analysis of their years of professional experience indicated the vast majority of the subjects had more than one year of experience which increases the probability that the results reflect the decision making patterns of experienced professionals rather than those individuals with little experience in the field.

With respect to the case data itself and the manner in which the subjects were asked to make their decisions, the probability of external invalidity was significantly moderated. This was accomplished by designing the stimulus materials in a manner that closely approximated the type and volume of data sources available to a school psychologist as he/she generates diagnostic decisions.

Overall, the threats to validity of this study are considered to be moderate and in general similar to the validity of the majority of similar studies cited in the literature. Decisions and conclusions related to this study should therefore be tempered with the considerations just described.

CHAPTER IV: RESULTS

The results of the major hypotheses will be presented followed by the results of the supplemental questions.

The two major hypotheses were investigated in terms of subject responses to each of three different data points considered to reflect the kinds of diagnostic and prognostic decisions typically made by practicing school psychologists. The subjects generated probability estimates for each of four handicapping and one nonhandicapping conditions as well as for five prognostic questions; all of which were based on case material provided each subject. Also each subject was asked to choose the primary handicapping condition (if any) represented by the case material.

The decisions recorded by the subjects had different scaler properties. Therefore different statistical tests and data transformations were required across the respective analyses. The data was further adjusted by omitting consideration of one of the handicapping conditions, mental retardation. This action was taken because it was not chosen by any of the subjects in any of the experimental groups as either a categorical diagnostic decision nor was it considered to be anything but 0% probable where probability estimates were called for. This category was dropped from the analysis because, according to McDermott (in press), categories

never chosen cannot be considered to be viable categories for an analysis of agreement. At most any effect of the diagnostic category of mental retardation was evenly distributed across the three subject groups and should therefore have had no differential effect on any outcome measures of agreement.

Hypothesis One

Hypothesis One related directly to the concept of diagnostic agreement. It posited that diagnostic agreement among school psychologists who used computer-based decision aides would be significantly higher than diagnostic agreement among school psychologists who did not use the decision aids. Results for each of the three types of decisions are presented separately as well as summarized in Table 15.

Probability Estimates for Each

Handicapping/Nonhandicapping Condition

The probability estimates made for each of three handicapping and one nonhandicapping condition were considered to reflect ordinal rankings ranging from less probably handicapped (or nonhandicapped, as appropriate) to more probably handicapped/nonhandicapped. An analysis of the probability estimates made by the subject decision makers indicated that agreement (concordance) according to the criteria of McDermott (in press) within each decision making group ranged from Kendall's W ($3, N = 29$) = .496 for the decision makers who did not have

Table 2

Summary Table of Kendall Coefficients of Concordance for
Probability Estimates of Handicapping/Nonhandicapping
Conditions for Each Subject Group

<u>W</u>	<u>Chi-Square</u>	<u>df</u>	<u>Significance</u>
Group 1: Computer-Based Decision Aids Not Available			
.496	43.17	4	* $p < .001$
Group 2: Computer Narrative & Statistics			
.351	29.47	4	* $p < .001$
Group 3: Computer Narrative, Statistics, Classification			
.503	45.30	4	* $p < .001$

Note: The number of subject responses in each group were similar; i.e. 29 subjects in Group 1, 28 subjects in Group 2, and 30 subjects in Group 3.

access to computer-based decision aids to W (3, $N = 28$) = .351 for the decision makers who had access to the M-MAC narrative and statistics to W (3, $N = 30$) = .503 for the decision makers who also had access to the M-MAC classification. In each instance the coefficients of concordance were significantly different from zero. See Table 2. As the magnitude of Kendall's W ranges from zero to a maximum of one these results were indicative of at least moderate agreement among rankings for each of the three subject groups. Fisher's r to Z score

Table 3

Mean Rankings of the Probability Estimates of
Handicapping/Nonhandicapping Conditions for Each
Subject Group

	LD	NEH	SED	LANG
Group 1: Computer-Based Decision Aids Not Available				
	3.52	1.48	2.21	2.79
Group 2: Computer Narrative & Statistics				
	3.46	2.04	2.02	2.48
Group 3: Computer Narrative, Statistics, Classification				
	3.73	1.85	2.12	2.30

Note. LD indicates "Learning Disabled."
 NEH indicates "Not Educationally Handicapped."
 SED indicates "Seriously Emotionally Disturbed."
 LANG indicates "Language Disabled."

The ranks range from values of 1 through 5 with higher values indicating a higher probability of the indicated handicapping/nonhandicapping condition.

transformation (Minium, 1978) was used to test hypotheses of differences in magnitude of these coefficients of concordance from subject group to subject group at a 95% confidence level. In no instance were any of the coefficients of concordance obtained within any of the subject groups significantly different from their counterparts in any other subject group. Further the mean rankings of the conditions as reported in Table 3

also tend to indicate general consistence in mean rankings of handicapping/nonhandicapping condition. Therefore, the hypothesis that diagnostic agreement for the probability estimates of handicapping/nonhandicapping conditions among school psychologists who use computer-based decisions aids is significantly higher than diagnostic agreement among school psychologists who do not use the aids is not supported by the data.

Categorical Decisions

When asked to choose a primary handicapping condition, the subjects made their choice on the basis of discrete categories, i.e. learning disabled, language disabled, seriously emotionally disturbed, mentally retarded, and not educationally handicapped.

A weighted Kappa analysis for each of the three experimental groups found weighted Kappa to be not significant at a 95 percent confidence level. Although not significant, the weighted Kappa coefficients were very close to a value of zero and suggestive of neither real agreement nor real disagreement exceeding chance among raters within each group. See Table 4 for specific details. Obtained weighted Kappa coefficients from each of the two computer-assisted subject groups were compared for significant differences and no significant differences in agreement were obtained. Neither were there significant differences between the weighted Kappa coefficients for Groups 1 and 3 and for Group 1 as

Table 4

Diagnostic Agreement Exceeding Chance for Categorical
Diagnoses by Type of Decision Aid Available According to
Weighted Kappa

	χ_m	<u>Variance</u>	<u>p</u>
Group 1: Computer-Based Decision Aids Not Available			
	-.032	.036	$p > .05$
Group 2: Computer Narrative and Statistics Available			
	-.033	.179	$p > .05$
Group 3: Computer Narrative, Statistics, Classification			
	-.036	.054	$p > .05$

compared to Group 2. When the actual percentage of psychologists choosing each category were compared across the three groups (Table 5) the psychologist subjects did not consistently choose one category as the primary handicapping condition. With the exception of apparently somewhat moderate agreement for the category of learning disability there was little agreement within groups for the other handicaps. Therefore the hypothesis that diagnostic agreement among school psychologists who use computer-based decision aids is significantly higher than diagnostic agreement among school psychologists who do not use computer-based decision aids cannot be supported for the categorical decisions herein.

Table 5

Percentage of Psychologists Choosing Each Categorical
Handicapping/Nonhandicapping Condition According to
Subject Group

	LD	NEH	SED	LANG
Group 1: Computer-Based Decision Aids Not Available				
	64.7%	17.6%	8.8%	2.9%
Group 2: Computer Narrative and Statistics Available				
	71.0%	9.7%	3.2%	12.9%
Group 3: Computer Narrative, Statistics, & Classification				
	84.4%	3.1%	3.1%	6.3%

Note. LD indicates "Learning Disabled."
 NEH indicates "Not Educationally Handicapped."
 SED indicates "Seriously Emotionally Disturbed."
 LANG indicates "Language Disabled."
 The percentages do not add up to 100% across the rows
 because of missing data from a few cases.

Prognostic Probability Estimates

The responses to the five prognostic questions (See Appendix F) were analyzed in the same manner as the prognostic probability estimates for the handicapping conditions i.e., for agreement according to Kendall's coefficient of concordance (W). Here again (except for prognostic question four) the numerical probability estimates were considered to be ordinal rankings in terms of lesser to greater probabilities of future difficulties

Table 6

Summary Table of Kendall Coefficients of Concordance for
the Probability Estimates for Prognostic Questions 1-5
for Each Subject Group

<u>W</u>	<u>Chi-Square</u>	<u>df</u>	<u>Significance</u>
Group 1: Computer-Based Decision Aids Not Available			
.571	75.37	4	*p < .001
Group 2: Computer Narrative & Statistics			
.507	54.79	4	*p < .001
Group 3: Computer Narrative, Statistics, & Classification			
.587	65.74	4	*p < .001

Note: The number of subject responses in each group were similar; i.e. 33 subjects in Group 1, 27 subjects in Group 2, and 28 subjects in Group 3.

in the educational and vocational environment for the student described in the case data. For prognostic question four, rankings of a greater magnitude were indicative of fewer difficulties in the future. In each instance, agreement within the three decision making conditions were analyzed separately. An analysis of the probability estimates made by the subject decision makers indicated that agreement (concordance) within each decision making group ranged from $\bar{W} (4, N = 33) = .571$ for the noncomputer-aided psychologists

Table 7

Mean Rankings^a of the Prognostic Probability Estimates
for Prognostic Questions 1-5 for Each Subject Group

	Q1	Q2	Q3	Q4	Q5
Group 1: Computer-Based Decision Aids Not Available					
	4.29	3.26	3.67	1.77	2.02
Group 2: Computer Narrative & Statistics					
	4.13	2.74	4.04	1.98	2.11
Group 3: Computer Narrative, Statistics, & Classification					
	4.39	3.13	3.75	1.71	2.02

Note. See decision/response sheet in Appendix F for the actual prognostic questions.

^aThe ranks range from values of 1 through 5 with higher values indicating a higher probability of future adjustment difficulties of the individual described by the case data except for question four where the higher values indicate a lesser probability of future adjustment difficulties.

to \underline{W} (4, \underline{N} = 27) = .507 for those psychologists who had access to the M-MAC narrative and statistics to \underline{W} (4, \underline{N} = 28) = .587 for the decision makers who also had access to the M-MAC classification. In each instance each Kendall coefficient was significant and indicated that the agreement within each group was significantly different than zero. Please refer to Table 6. Fisher's \underline{x} to \underline{z} score transformation was again used to test hypotheses of differences in magnitude of the coefficients of concordance from subject group to subject group at a 95%

confidence level. In no instance were any of the coefficients of concordance obtained within any of the subject groups significantly different from their counterparts in any other subject group. Further, the mean rankings of the conditions as reported in Table 7 also tend to indicate general consistence in mean rankings of prognostic probability estimates. Therefore the hypothesis that agreement of the probability estimates for the prognostic questions among the psychologists who used computer-based decisions aids is significantly higher than agreement among school psychologists who did not use the aids is not supported by the data.

Hypothesis Two

Hypothesis Two compared decisions made with and without the use of computer-based decision aids in terms of the differences and similarities in the actual decisions made. It was hypothesized that decisions made in connection with two variations of the same interpretive software program would be more similar to each other than to those decisions made without the use of decision aids. The same data points, probability estimates for handicapping/nonhandicapping conditions, categorical handicapping conditions, and prognostic probability estimates for the five questions were used to answer the hypothesis. The three types of decision data points will be discussed separately.

Table 8

Summary Table of Friedman Analyses of Variance for the
Probability Estimates of Handicapping/Nonhandicapping
Conditions for Each Subject Group

<u>N</u>	<u>Chi-Square</u>	<u>df</u>	<u>Significance</u>
Group 1: Computer-Based Decision Aids Not Available			
29	39.00	3	* $p < .001$
Group 2: Computer Narrative & Statistics			
28	23.15	3	* $p < .001$
Group 3: Computer Narrative, Statistics, & Classification			
30	38.35	3	* $p < .001$

Probability Estimates for Each
Handicapping/Nonhandicapping Condition

A series of Friedman analyses of variance for dependent variables; one for each of the three subject groups, were completed. Each was significant and indicated the rankings of each of the four probability estimates by the subjects within each group were significantly different from each other. See Table 8. Pairwise multiple comparisons were completed for all possible combinations of handicapping/nonhandicapping conditions (Table 9) in order to compare differences in rankings among the diagnostic decisions within each

Table 9

Pairwise Comparisons of Differences in Probability
Estimates of Handicapping/Nonhandicapping Decisions for
Each Subject Group

		LD	NEH	SED	LANG
LD	Group 1		*p < .05	*p < .05	NS
	Group 2		*p < .05	*p < .05	*p < .05
	Group 3		*p < .05	*p < .05	*p < .05
NEH	Group 1			NS	*p < .05
	Group 2			NS	NS
	Group 3			NS	NS
SED	Group 1				NS
	Group 2				NS
	Group 3				NS

Note. LD indicates "Learning Disabled."
 NEH indicates "Not Educationally Handicapped."
 SED indicates "Seriously Emotionally Disturbed."
 LANG indicates "Language Disabled."
 Group 1 indicates "Computer Decision Aids not Available."
 Group 2 indicates "Computer Narrative & Statistics."
 Group 3 indicates "Computer Narrative, Statistics, & Classification."

subject group. As indicated by Table 9 the two computer-assisted groups agreed with each other in terms of the significance or nonsignificance for 100% of the available paired comparisons of the rankings of the diagnostic categories. The pairwise comparisons of the rankings of the subjects who did not have computer decision aids available agreed to a lesser extent with the two computer-aided conditions. Only 66% of the comparisons

were in agreement with those comparisons of the computer-aided conditions.

Although both groups of computer-assisted psychologists ranked the probability of learning disability to be significantly higher than that of a language disability ($p < .05$), the rankings of these two handicaps did not differ significantly in the noncomputer-aided condition. Similarly, the computer-assisted subjects showed no significant difference in the rankings of the probability estimates of language disability and of not educationally handicapped, but the noncomputer-aided psychologists ranked the probability of language disabled to be significantly higher than that of not educationally handicapped ($p < .05$). Descriptive comparisons of the raw data (Table 10) suggest little variation across the three subject groups in the psychologists' choices of probability estimates for the handicapping/nonhandicapping conditions. This is especially true for the conditions of learning disability and not educationally handicapped. However, the pattern of responses for the category of seriously emotionally disturbed presented somewhat of a trend for differences between the estimates made by those subjects who had access to the computer aid and those who did not. The computer-assisted decision makers tended to rate seriously emotionally disturbed as less probable than the

Table 10

Percentage of Psychologists Choosing Each Probability
Estimate for Handicapping/Nonhandicapping Condition
According to Subject Group

Probability Estimate of:		0%	25%	50%	75%	100%
LD	Group 1	2.9%	8.8%	14.7%	52.9%	5.9%
	Group 2	3.3%	10.0%	30.0%	43.3%	6.7%
	Group 3	3.1%	3.1%	25.0%	50.0%	12.5%
NEH	Group 1	58.8%	23.5%	5.9%	0.0%	0.0%
	Group 2	43.3%	36.7%	10.0%	3.3%	0.0%
	Group 3	46.9%	34.4%	9.4%	3.1%	6.3%
SED	Group 1	17.6%	32.4%	38.2%	0.0%	0.0%
	Group 2	43.3%	33.3%	10.0%	3.3%	3.3%
	Group 3	21.9%	50.0%	18.8%	3.1%	0.0%
LANG	Group 1	8.8%	26.5%	23.5%	26.5%	2.9%
	Group 2	26.7%	36.7%	13.3%	16.7%	0.0%
	Group 3	18.8%	46.9%	18.8%	9.4%	0.0%

Note: The percentages do not add up to 100% across the rows because of missing data from a few cases.

nonassisted decision makers who tended to waver more at the undecided 50% probability level. Similarly the percentage response by group in Table 10 further suggests a trend on the part of those with the decision aids towards ratings of lesser probabilities of a language handicap than the noncomputer-assisted psychologists.

Together, these data indicate that Hypothesis Two is supported by the data for the probability estimates of

handicapping conditions because the diagnostic decisions made by school psychologists on the basis of different applications of the same computer decision aid were indeed more similar to each other than to those diagnostic decisions made by school psychologist subjects not utilizing computer decision aids.

Categorical Decisions

Chi-square analyses were utilized to test the relationships between the diagnostic decisions and subject groups. Two separate analyses were performed. One compared the diagnostic decisions made without the assistance of computer decision aids with those made with interpretive software assistance. The second analysis compared the decisions made on the basis of the two variations of M-MAC output; i.e. the categorical decisions of the Group 2 subjects compared with those of the Group 3 subjects. See Table 11. In order to minimize the number of cells in the analysis with an expected frequency of less than five it was necessary to collapse the ratings. The three subject groups were collapsed into two groups. The noncomputer-aided decision group remained intact, but the two groups of subjects who had a microcomputer decision aid available were combined together. Also, the actual diagnostic decisions were combined together in two separate ways. Because of the preponderance of "learning disability" choices among the raters across all three subject groups

Table 11

Subject Group Comparisons for Categorical Decisions

<u>Comparison</u>	<u>Chi-Square</u>	<u>df</u>	<u>Significance</u>
Group 1 vs Groups 2 & 3 Combined			
Learning Disabled/ Nonlearning Disabled	.98	1	p > .05
Handicapped/ Nonhandicapped	2.11	1	p > .05
Group 2 vs Group 3			
Learning Disabled/ Nonlearning Disabled	1.06	1	p > .05
Handicapped/ Nonhandicapped	.30*	1	p > .05

Note: All Chi square values are after Yates correction.
 Group 1 indicates "Computer Decision Aids not Available."
 Group 2 indicates "Computer Narrative & Statistics."
 Group 3 indicates "Computer Narrative, Statistics, & Classification."

*Analysis with at least one cell with an expected frequency of less than five which could make the data somewhat misleading.

and because of the considerable uncertainty regarding the term "learning disability" in the literature (Dawson, 1985), the diagnostic categories were collapsed into two categories, "learning disabled" and "not learning disabled." A second recombination was derived based upon the concept of educational handicap with the original

data being transformed into the categories of "not handicapped" and "handicapped." Nevertheless, the analysis that compared the two computer-aided conditions together on the dimension of "handicapped" and "nonhandicapped" had at least one cell with an expected frequency of less than five which could make the results of that particular analysis somewhat misleading. This analysis is noted with an asterisk in Table 11. All other expected frequencies were five or greater.

In each instance chi-square was not significant and the categorical decisions therein cannot be considered to be different for the different groups. As expected the decisions made by the two groups who had access to the microcomputer decisions aids were not significantly different from each other. Therefore the hypothesis that categorical diagnostic decisions made on the basis of different applications of the same computer-assisted decision aid will be more similar to each other than to those decisions not utilizing decision aids cannot be totally supported. At best the data indicate there are no significant differences between the categorical decisions made among the subject groups across the dimensions of learning disabled and nonlearning disabled as well as the dimensions of handicapped and nonhandicapped.

Table 12

Summary Table of Friedman Analyses of Variance for the
Probability Estimates for Prognostic Questions 1-5 for
Each Subject Group

Decisions	<u>Chi-square</u>	<u>df</u>	Significance
Group 1: Computer-Based Decision Aids Not Available			
33	61.32	4	* $p < .001$
Group 2: Computer Narrative & Statistics			
27	45.85	4	* $p < .001$
Group 3: Computer Narrative, Statistics, & Classification			
28	57.52	4	* $p < .001$

Prognostic Probability Estimates

The five prognostic statements were analyzed for agreement on the basis of a second series of Friedman analyses of variance (Table 12); again one for each of the three subject groups. As with the probability estimates of handicapping/nonhandicapping conditions just described, all three analyses were significant and indicated that the rankings of the five prognostic probability estimates by the subjects within each group were significantly different from each other. Pairwise multiple comparisons were completed for all possible combinations of prognostic probability estimates one through five in order to compare differences in rankings

Table 13

Pairwise Comparisons of Differences in Probability
Estimates for Prognostic Questions 1-5 for Each Subject
Group

		Q1	Q2	Q3	Q4	Q5
Q1	Group 1		NS	NS	* $p < .05$	* $p < .05$
	Group 2		NS	NS	* $p < .05$	* $p < .05$
	Group 3		NS	NS	* $p < .05$	* $p < .05$
Q2	Group 1			NS	* $p < .05$	NS
	Group 2			NS	NS	NS
	Group 3			NS	NS	NS
Q3	Group 1				* $p < .05$	* $p < .05$
	Group 2				* $p < .05$	* $p < .05$
	Group 3				* $p < .05$	NS
Q4	Group 1					NS
	Group 2					NS
	Group 3					NS

among the prognostic decisions within each subject group. The two computer-assisted groups agreed with each other in terms of significant or nonsignificant differences between pairs of prognostic questions for 90% of the available paired comparisons. See Table 13. The pairwise comparisons of the rankings of the subjects who did not have computer decision aids available agreed to a somewhat lesser extent with the two computer-assisted conditions. Ninety percent of the comparisons in Group 1 were in agreement with the analogous Group 2 comparisons and 80% of the comparisons of Group 1 were consistent

with those in Group 3. The exceptions to agreement among the pairwise comparisons were not consistent from group to group. For the comparison of Question 3 ("The child will be referred again as an educationally handicapped child.") and Question 5 ("The child will have difficulty competing in the vocational marketplace as an adult") the difference in rankings was not statistically significant for the subjects who had the computer classification available, but there was a significant difference in the mean rankings for the subjects of the two other groups. The comparison of Question 2 ("The child will demonstrate future social adjustment problems.") with Question 4 ("The child will successfully graduate from high school."), however, found the rankings of the questions to be significantly different for Group 2 psychologists only. Descriptive comparisons of the raw data (Table 14) across the three subject groups indicate little, to, in one instance, at most moderate, (Question 2 only) variation in percentage of psychologists' choices of probability estimates for the prognostic questions. The data therefore cannot support Hypothesis Two and instead suggests that there is no real difference across the three subject groups in terms of the prognostic probability estimates made by the respective subjects.

Table 14

Percentage of Psychologists Choosing Each Probability
Estimate for Prognostic Questions 1-5 by Subject Group

Probability Estimate of:		0%	25%	50%	75%	100%
Q1 ^a	Group 1	0.0%	0.0%	0.0%	32.4%	64.7%
	Group 2	0.0%	3.3%	6.7%	50.0%	40.0%
	Group 3	0.0%	0.0%	3.1%	40.6%	56.3%
Q2	Group 1	0.0%	2.9%	26.5%	35.3%	35.3%
	Group 2	0.0%	16.7%	30.0%	43.3%	6.7%
	Group 3	0.0%	31.3%	40.6%	18.8%	3.1%
Q3	Group 1	0.0%	0.0%	8.8%	50.0%	38.2%
	Group 2	0.0%	0.0%	10.0%	56.7%	30.0%
	Group 3	0.0%	6.3%	12.5%	46.9%	31.3%
Q4	Group 1	8.8%	26.5%	44.1%	14.7%	2.9%
	Group 2	0.0%	36.7%	46.7%	13.3%	0.0%
	Group 3	0.0%	46.9%	37.5%	9.4%	0.0%
Q5	Group 1	2.9%	23.5%	44.1%	17.6%	8.8%
	Group 2	0.0%	26.7%	50.0%	16.7%	3.3%
	Group 3	3.1%	25.0%	46.9%	25.0%	0.0%

Note: The percentages do not add up to 100% across the rows because of missing data from a few cases.

^aPlease refer to Appendix F for the actual question content.

Table 15

Summary Table of Outcomes of Analyses for Questions One
and Two According to Type of Decision

	<u>Hypothesis One</u>	<u>Hypothesis Two</u>
Probability Estimates of Handicapping/ Nonhandicapping Conditions	Not Supported	Supported
Categorical Handicapping/ Nonhandicapping Conditions	Not Supported	Not Supported
Prognostic Probability Estimates of Questions 1-5	Not Supported	Not Supported

Supplemental Questions One and Two were designed to more closely examine the subjective effects of different data sources upon the decisions made by the school psychologist subjects.

Supplemental Question One

The case materials provided each of the subjects formed the basis of Supplemental Question One. It asked, for each of the three experimental conditions, which data sources were considered by the subjects to be the most influential in the their decision-making.

Each of the three subject groups had different data sources available during the decision making process and by necessity they were analyzed separately by subject group. At least four of the data sources were available to all of the subjects; six of them to those subjects in experimental Group 2 and a total of seven to experimental Group 3. Table 16 lists the data sources that were available to each subject group. Table 17 reflects the percentage reported influence of the case data sources as well as their median influence and associated interquartile ranges. The ratings were made on an ascending ordinal scale where a rating of one indicated "no influence" and five was indicative of "much influence."

A series of Friedman analyses of variance for ordinal rankings were used to analyze the differences in subject rankings for each particular subject group.

Table 16

Data Sources Available to Subject Groups

	Group 1	Group 2	Group 3
Reason for Referral	*	*	*
Background Information	*	*	*
Behavioral Observations	*	*	*
Test Data	*	*	*
M-MAC Statistics		*	*
M-MAC Narrative		*	*
M-MAC Classification			*

Note: A "*" indicates that the data source was available to the subjects within the group.

Table 17

Percent Influence of Data Sources by Experimental Group

		Group 1	Group 2	Group 3
Reason for Referral	N% No Influence	26.5%	26.6%	12.5%
	N% Moderate Influence	38.2%	33.3%	43.8%
	N% Much Influence	31.4%	33.4%	43.8%
	Median	3.0	3.0	3.0
	Interquartile Range	2.0	2.0	1.0
Backgrd Informa- tion	N% No Influence	2.9%	13.3%	9.4%
	N% Moderate Influence	29.4%	26.7%	37.5%
	N% Much Influence	64.7%	56.7%	53.2%
	Median	4.0	4.0	4.0
	Interquartile Range	1.5	1.0	1.0
Behavior Observa- tions	N% No Influence	0.0%	3.3%	0.0%
	N% Moderate Influence	17.6%	20.0%	31.3%
	N% Much Influence	79.5%	73.4%	68.7%
	Median	4.0	4.0	4.0
	Interquartile Range	1.0	1.5	1.0
Test Data	N% No Influence	0.0%	3.3%	6.5%
	N% Moderate Influence	5.9%	13.3%	12.9%
	N% Much Influence	91.2%	80.0%	78.2%
	Median	4.0	4.0	4.0
	Interquartile Range	1.0	1.0	1.0
M-MAC Stats	N% No Influence		56.7%	54.8%
	N% Moderate Influence		23.3%	32.3%
	N% Much Influence		16.3%	12.9%
	Median		2.0	2.0
	Interquartile Range		1.0	1.0
M-MAC Nar- rative	N% No Influence		56.7%	64.5%
	N% Moderate Influence		26.7%	25.8%
	N% Much Influence		14.4%	9.7%
	Median		2.0	2.0
	Interquartile Range		1.0	1.0
M-MAC Classi- fica- tion	N% No Influence			80.2%
	N% Moderate Influence			19.4%
	N% Much Influence			0.0%
	Median			2.0
	Interquartile Range			1.0

Note: The percentages do not sum to 100% because of missing data.

Where there was a significant main effect of differences in data influences, post hoc pairwise multiple comparisons were applied to determine the relative magnitudes of these differences. Each subject group will be discussed separately.

Group 1: Computer Decision Aids Not Available

The main effect of data influences was significant here $\chi^2 (3, N = 33) = 17.88, p < .001$ and indicated that the rankings of the data influences by the subjects were significantly different from each other. See Table 18. Pairwise comparisons between types of data influences indicated that Behavioral Observations and Test Data were considered by the psychologists to be significantly more influential than Reason for Referral. As differences between all other available comparisons were nonsignificant, Behavioral Observations and Test Data are apparently the most influential data sources for the psychologists who did not have computer decision aids available to them.

Group 2: Computer Narrative and Summary Statistics Available

Here too the main effect of data sources was significant in a Friedman analysis of variance $\chi^2 (5, N = 28) = 42.30, p < .001$. See Table 18. An analysis of the fifteen available pairwise comparisons found Behavioral Observations and Test Data to be significantly more influential ($p < .05$) than the

Table 18

Summary Table of Friedman Analyses of Variance of Degree of Influence of Data Source by Type of Data Source for Each of the Three Decision Making Subject Groupings

<u>N</u>	<u>Chi-Square</u>	<u>df</u>	<u>Significance</u>
Group 1: Computer-Based Decision Aids Not Available			
33	17.88	3	*p < .001
Group 2: Computer Narrative & Statistics			
28	42.30	5	*p < .001
Group 3: Computer Narrative, Statistics, & Classification			
30	100.35	6	*p < .001

computer-related data sources, i.e. the M-MAC Narrative and the M-MAC Statistics. As all other comparisons were nonsignificant, these two data sources appear to have been considered by the school psychologist decision makers to be subjectively more influential than any of the other data sources.

Group 3: Computer Narrative, Statistics, and Classification Available

As with the other subject groups, the main effect of data sources was significant in a Friedman analysis of variance $\chi^2 (6, N = 30) = 100.35$ $p < .001$. See Table 18. Differences between the individual data sources generally followed the pattern indicated for Group 2

UNIVERSITY OF MARYLAND

subjects with no significant differences between pairs of data sources that were not computer-generated; i.e. for example Reason for Referral, Background Information, etc. The same was true for the three computer-generated data sources. However, except for Reason for Referral, all of the other noncomputer-generated data sources were considered to be significantly higher in influence than the M-MAC Classification ($p < .05$). Like the two other groups, the data reflects that the subjects of Group 3 indicated that with the exception of Reason for Referral, the noncomputer generated data sources were considered to be more influential than the computer-generated data sources.

The data presented suggests that apparently the subjects seemed to be rather more consistently influenced by the test data or the test data in association with other traditional data sources than they were of the data generated by M-MAC. The M-MAC diagnostic classification itself seemed to have been considered as the least influential data source by those individuals to whom it was presented as an option.

Supplemental Question Two

Supplemental Question Two attempted to ascertain what the relationship was between school psychologists' professional experience level and degree of experience with using computers in clinical practice to the sources of information which they found to be most influential in

Table 19

Coefficients of Correlation Between Years of Professional
Experience as a School Psychologist and Influence of Case
Data Sources

	Group 1	Group 2	Group 3
Reason Referral	.08	.10	-.15
Background Information	.12	.10	-.18
Behavioral Observations	.03	.05	-.12
Test Data	-.31	.14	-.19
M-MAC Statistics		-.13	-.24
M-MAC Narrative		-.08	-.10
M-MAC Classification			.05

Note. $p > .05$ in each instance.

the diagnostic decision-making process. The data was analyzed by correlating degree of experience as a school psychologist with each of the individual ratings on influence of data sources. A second correlation was executed between degree the diagnostic decision-making process. The data was analyzed by correlating degree of experience as a school psychologist with each of the

individual ratings on influence of data sources. A second correlation was executed between degree of experience in using computers with ratings on the influence of individual data sources. Each correlation was completed for each experimental group individually. Years of experience was treated as ordinal data with four ascending categories of experience ranging from less than one year to over ten years of experience. Influences of data sources and extent to which the subjects had used computers within their clinical practice were also considered to be ordinal data on an ascending scale from one to five. A rating of one indicated either no influence or experience and five indicated much influence or frequent use, respectively. See Appendix F, the decision/response sheet. Kendall's tau was utilized as the measure of association for each analysis.

Professional Experience as Related to Influence of Data Sources

The data indicate that there was no significant relationship between years of professional experience as a school psychologist and influence of any of the data sources; traditional or computer-generated; for subjects individually by treatment groups. See Table 19. A preliminary analysis had also indicated that there were no significant relationships when all of the subjects were considered together as one group.

Table 20

Significant-Coefficients of Correlation Between Extent of
Use of Computers in the Course of Professional School
Psychological Practice and Influence of Case Data Sources

	Group 1	Group 2	Group 3
<u>Test Scoring</u>			
Reason for Referral	.50*		
<u>Test Interpretation</u>			
M-MAC Statistics			.55*

Note: $P < .01$ in each instance. All of the other 100 plus correlation coefficients generated cannot be considered to be significantly different from zero.

Degree of Experience Using Computers as Related to
Influence of Data Sources

Considering that the number of correlations computed exceeded 100 for this analysis, it is estimated that chance alone would generate a few significant correlations irrespective of any real relationships. Therefore an alpha level of .01 was utilized to test the significance of correlation coefficients obtained and to minimize Type I error. With the exception of two significant correlations (Table 20); one each for the noncomputer-aided subjects and the subjects who had the

M-MAC classification available, no statistically significant relationships were observed between use of computers in professional practice and influence of data sources. For the noncomputer-aided group reason for referral was positively related to use of a computer for test scoring. For the subjects who had access to the M-MAC classification, the use of a computer for test interpretation was positively related to the influence of M-MAC statistics. Otherwise there appeared to be no significant relationships between degree of professional experience in the use of computers and the influence of the data sources available to the subjects for any of the three subject groups.

CHAPTER V: CONCLUSIONS

Scope and Focus

This chapter study will first review the rationale for the study in the context of the current literature followed by the results, conclusions and limitations of the study. Implications for future research and professional practice follow.

Rationale

Ethical, professional and legal standards mandate the decisions made by school psychologists in the course of day to day practice be valid and free from intentional bias. The decisions to be made, however, are decisions that have historically been considered to be biased and of limited validity and reliability. The final interpretation of a given set of data seems to depend primarily upon the manner in which the psychologist has idiosyncratically integrated the available data.

The decision making literature reflects the general opinion that decisions can be debiased through the use of decision aids. Attempts to debias the decision process and thereby make the decisions more consistent across practitioners have been variably successful at best. The meager school psychological and educational literature that is available lends support to the premise the availability and use of preprocessed and summarized data

from a mechanical/nonhuman source improves the quality and consistency of decisions ultimately made. Computer-based interpretative software is one such data source and therefore has the potential to significantly decrease bias and systematic error in the decision making process. Ethical practice, however, currently prohibits the independent use of interpretative software without practitioner review. Instead, computer-based interpretive software must be considered to be a decision aid or simply another data source to be utilized by the psychologist in the course of day to day decision making. The purpose of this study, then, was examine the utility of computer-based decision aids in their role as decision aids for the professional school psychologist.

Hypotheses

It was specifically hypothesized that diagnostic agreement among school psychologists who had computer-based decision aids available would be significantly higher than diagnostic agreement of school psychologists who did not have the decision aids available. A second hypothesis proposed that actual decisions made by school psychologists who used different variations of the same computer-based decision aid would be more similar to each other than to those diagnostic decisions made by psychologists who did not have access to computer-based decision aids. Supplemental Question One investigated which data sources were considered by school

psychologists to be most influential in their decision making. Relationships between professional experience level and degree of experience in using computers with sources of information considered to be most influential were analyzed in Supplemental Question Two.

Methodology

In order to test these hypotheses, randomly selected members of the National Association of School Psychologists were randomly assigned to either one of two experimental groups or to a control group. Each subject was provided information designed to qualitatively and quantitatively reflect case data materials that a school psychologist typically has prior to making a diagnostic or prognostic decision on a referred child. The two experimental groups of subjects also received one of two variations of the output from a currently available computer-based test interpretation software package. Each psychologist was asked to make diagnostic and prognostic decisions that were designed to reflect the types of decisions made by school psychologists. They were also asked which data sources were most influential in the decisions made, as well as relevant demographic information.

Results and Conclusions

Hypothesis One proposed that diagnostic agreement among school psychologists who used computer-based decision aides would be significantly higher than

diagnostic agreement among school psychologists who did not use the decision aids. The hypothesis was not supported with the data for any of the three types of decisions made by the subjects.

According to psychological decision theory (Tversky & Kahneman, 1980), data that is readily available in memory and/or that is representative or brings to mind similar situations have been found to directly influence the ultimate decisions made. In theory, the computer generated output should have become readily available in short term memory; i.e. a discrete cue for the decision maker indicating that learning disability was the diagnosis of choice. Diagnostic agreement should then have been higher than when no decision aid was available because each psychologist subject would have been exposed to and affected by the same major cues in the computer output. The cues available to the nonaided subjects, however, varied according to how the individual psychologist personally structured the decision problem and the cues which were idiosyncratically most salient. Agreement among the nonaided decision makers should have been lower as a consequence. This did not happen and is probably most explainable in terms of the data from Supplemental Question One where the subjects were asked to indicate which data sources were most influential. In each instance, and for each subject group, the computer-based

data influences were rated lower than any of the traditional data sources common to subjects in all three groups. Apparently the computer output did not have affective salience sufficient to outweigh the heuristics operationalized through a reading of the case data. It is also possible that the salience was not sufficient to overcome the influence of cumulative years of decision making based upon time worn frames of reference. This was apparently the case in Nisbett et al (1982), where decisions made with the standard, objective data sources were no different than those made on the basis of "obviously haphazardly sampled information." Further, there continues to exist a substantial degree of uncertainty regarding the validity of the content of the computer-generated reports and perhaps of computers in general. This is especially probable given that among practicing school psychologists; experience with computers other than for word processing is so minimal (Table 1 and Brantley et al, 1985). The uncertainty may also be a function of the considerable debate in the literature regarding unresolved ethical and practical problems relating to the use of computers in a profession that has traditionally considered decision making as an art rather than as a hard science.

Hypothesis Two proposed there would be no significant differences in decisions made in connection with two variations of the same interpretive software

program. Also the ultimate decisions made with the use of a decision aid would be significantly different from those decisions made without the use of a decision aid. Here, the less stringent concept of similarity of decisions was tested. The data indicated that for at least one kind of judgement, probability estimates of handicapping/nonhandicapping conditions, there was a differential effect of availability of decision aid. Also, as hypothesized, there appeared to be a trend in the data to suggest that different variations of M-MAC did not have a differential effect on the decisions made. This is not surprising given the degree to which the subjects indicated that the computer-based data sources influenced them.

Suggestions are that for these specific judgments, the computer output was sufficiently salient to frame the case data in a manner that was representative of the concept of learning disability and excluded serious consideration of the concept of language disability. The nonaided psychologists, however, seemed to have made their decisions on the basis of a schema wherein the differences between the concepts of learning disability and language disability were not clear cut enough to indicate one as substantially different than the other.

The major differences in judgments appeared to be a function of the more probable categories of "learning disability" and "not educationally handicapped" vis à vis

a less probable category, "language disability." Whereas the computer-assisted psychologists consistently reported learning disability to be more probable than language disability, the noncomputer-aided psychologists did not make the distinction. Similarly the computer-aided psychologists found little difference in the probabilities of not educationally handicapped and language disabled whereas the other subjects did. A closer examination of the actual percentages of responses for each handicapping/nonhandicapping condition (Table 10) seems to reinforce this trend. Here again the trends were not apparent for the rather straightforward conditions of learning disability and not educationally handicapped, but were more powerful for the conditions that were less likely or less clear cut; i.e. seriously emotionally disturbed and language disabled. The subjects in the noncomputer-aided conditions tended to make ratings which were more uncertain (50% probability) than the subjects in the computer-aided conditions who gave these handicapping conditions generally lower and more definitive probability estimates.

Care should be taken not to overinterpret these results because of their descriptive nature. Nonetheless, subjectively, it is possible to describe the outcome in terms of the operation of heuristics. Most of the psychologist subjects found the case data to be representative of learning disability irrespective of

decision aid available. However, when the psychologists were forced to make a judgment regarding the probability of the less well-defined diagnoses of seriously emotionally disturbed or language disabled, the most readily available data, i.e. the computer narrative, was inconsistent with these diagnoses. The narrative probably functioned to increase the probability that these two categories would be considered to have low probabilities. However, where the computer decision aid was not available, the raters responded with a higher percentage of uncertain (50% probability) ratings. This would suggest that perhaps the availability of the computer output indicating learning disability may have been influential in the determination of probabilities for the more ambiguous and difficult to diagnose (Ysseldyke & Algozzine, 1981) conditions of seriously emotionally disturbed and language disabled. The decisions of the nonaided psychologists, however, seemed to vary with changing decision choices and perhaps different cues within the case data as Rassen (1979) indicated in her analysis of school psychologist decisions.

When comparisons were made across the three groups for the categorical decisions, there were no significant differences in diagnostic decisions. These findings, however, need to be tempered by the necessity of collapsing the categorical decisions into two sets of

dichotomous judgments, "learning disabled" versus "nonlearning disabled" and "handicapped versus nonhandicapped" so that adequate expected frequencies within the chi-square analyses were available. This may have obscured any real treatment effects; especially considering the effects noted so far have involved the less clear cut diagnoses of seriously emotionally disturbed and language disabled.

The probability estimates for the prognostic questions did not conform to the hypothesis of more similarity between the computer-assisted groups than with the noncomputer-assisted groups of subjects. In fact, the data are more suggestive of a general consistency across the three groups in the prognostic decisions made. Unlike the probability estimates of the handicapping/nonhandicapping conditions where there was relative consistency across the computer-assisted subject groups, the pattern with the prognostic decisions was less clear cut. The pattern of responses to the questions across groups is perhaps best explained as a function of a combination of chance variation in decisions associated with the influence of heuristics common to the decision makers as a whole. The questions centered around relationships between disparate, yet related questions about expectations of future events. The probability of these events may or may not be directly related to current data and will be affected to

an undetermined extent by unknown and unpredictable factors. The data sources most available or representative at the time the prognostic decisions were made were probably not evoked by the computer-generated output. Instead they probably were related to an undetermined data source that had a generally consistent effect among all of the psychologist subjects. This is reasonable as the computer output directly addresses diagnosis; not the more inferential prognosis.

For the supplemental questions the school psychologist subjects, irrespective of treatment group, indicated test data and behavioral observations were the most influential data sources as they made their decisions. Further, the psychologists who had access to the computer aids found them to be significantly less influential than the noncomputer data sources. A trend in the data indicated the M-MAC diagnostic classification was considered to be the least influential of all of the data sources. Degree of professional experience as a psychologist was not significantly related to the influence of the data sources and should probably be ruled out as a contributing factor to this outcome. Degree of experience in using a computer within professional practice does not appear to be a significant factor in this outcome either, because with two exceptions, a significant relationship was not established between computer experience and the perceived

influence of data sources. The two exceptions were a significant correlation between experience in the use of computers for test scoring indicated by the nonaided subjects and a significant relationship between use of computers for test interpretation and M-MAC statistics by the subjects who had the M-MAC classification available. These might best be considered artifacts of the correlation matrix; especially the positive relationship between reason for referral and experience with test scoring. The positive relationship between experience with the use of computers for test interpretation and the influence of the M-MAC statistics for the psychologists who had the M-MAC classification is more logical in light of the data. This could be because individuals who use decision aids would be expected to find the M-MAC classification influential. However, correlations that would be reasonably related to this one such as those involving the M-MAC narrative and the M-MAC classification were not significant for any of the subject groups. As such the probability is increased that this significant correlation should also be considered to be a random occurrence with little apparent meaning within the context of the overall study.

From these data it is apparent the computer-generated output was considered by the school psychologists to be of relatively minor influence in the decisions they made generally irrespective of their

professional experience and their experience with computers. This factor may be related to the findings reported in Bar-Hillel (1982), Tversky & Kahneman, (1973), Nisbett et al (1980) and others in the literature of psychological decision making. These studies have generally concluded that empirical data has been perceived as generally less relevant or even coincidental to the question at hand as compared to more specific, personal, and therefore more affectively laden material. Apparently, the affectively laden material is perceived as more representative of similar instances and decisions were made on that basis instead of on the more objective data sources. In the school psychology and education literature this factor has been demonstrated too, as described in DeWitte (1976), Hersh, (1971), and the series of studies completed in the early 1980's by Ysseldyke and his associates (Ysseldyke & Allongize, 1981; Ysseldyke & Pianta, 1983, etc.) who generally concluded that decisions regarding educationally handicapping conditions are apparently often made on the basis of factors other than objective data. Thomas (1983) summarized the situation when he concluded that decision making in school psychology is often based on factors unrelated to formal assessment and instead is predicated on the context of the decision.

In general, then, the computer-based test interpretive software program utilized in this study

should not be considered to be of significant utility in its role as a decision aid to the practicing school psychologist. Trends in the data, however, suggest that the utility of computer-based decision aids should not yet be ruled out and that more study is required before any definitive judgments can be made. The failure of the computer-generated output to overcome the effects of the salience of other sources of data is actually not too surprising given the status of computer-generated data in today's society. Although computers have been available for decision making in society for over four decades, it has not been until relatively recently they have become readily available for day to day use in the workplace and/or homes of more than a minority of the populace. In the practice of school psychology, they are still used by a minority of practicing psychologists and therefore subject to misperceptions and an incomplete understanding of how they can function in the workplace. They still apparently present as a somewhat impersonal machine whose operations are antithetical to the highly interpersonal nature of school psychological practice. Therefore, any output generated is affectively paled by comparison to the "real" test scores and behavioral observations they have been trained to utilize in their decision making.

The literature surrounding the role of computers in psychological practice and in particular diagnosis and

assessment has been particularly partisan and reflective of strong opinions on either side of the issue. A recent article by Kramer (1988) essentially concluded there are no firm conclusions regarding the use of computers in psychology and but instead "enormous" challenges need to be overcome prior to acceptance of their use in psychology's mainstream. With these considerations it is not surprising the affective salience of computer-generated output as a decision aid was not of a sufficient magnitude to overcome the availability of other more traditional data sources.

Limitations

Each study has its limitations. The most salient of these limitations will be described in the following paragraphs.

Perhaps the single most salient limitation in the design of this study was the inclusion of only one referred case for diagnostic and prognostic consideration among the subjects. The case did indeed reflect the kind of case that is typically received as a referral for school psychological decision making. However, by using one case, the entire analysis hinged on the interaction of the specific demand characteristics of the child described therein to the exclusion of other cases with other demand characteristics. Therefore, the results obtained should be tempered with the fact that there were probably not enough occasions for assessing the

systematic presence or absence of the major treatment effects. A design that utilized several cases randomly assigned to the subjects would have been preferable. Along similar lines, it would have been preferable to utilize more than one computer-based decision aid rather than to use two different applications of M-MAC. The major difference between the narrative output of the M-MAC classification (Group 3) and the M-MAC narrative and statistics (Group 2) was one paragraph at the end of the narrative. It is possible that some of the subjects overlooked or otherwise might not have considered the entire output provided. The use of more than one software package would also have enhanced the construct and external validity of the study.

Given the trends of the descriptive data, it is entirely possible that treatment effects in the major hypotheses were obscured by factors related to inadequate sample size and perhaps the power of the statistical analyses utilized. It would be well-advised to substantially increase the subject sample in any future analysis of this type so the possibility of Type II error could be diminished more than it apparently was in this study.

By the same token, limitations inherent in analyzing data of an ordinal nature; data which does not, by definition, have a single underlying distribution on which to base assumptions, may have contributed to an

undetermined loss of power. Although it was not necessary to meet as many of the underlying assumptions that would have been necessary if the data would have been of interval or ratio scale; the analyses utilized were necessarily less powerful than generally would have been attainable had the stricter assumptions and scaler qualities been met for the use of parametric statistical methods.

Another limitation of the analysis was in the chi-square analyses of the categorical data of Hypothesis Two where it was necessary to significantly collapse the diagnostic categories into two separately analyzed dichotomous categories of "learning disabled;" "not learning disabled" and "handicapped;" "not handicapped." Although the derived categories themselves are relevant to the actual decisions made in school psychology, it is possible the transformed categories masked any differential treatment effect and again contributed to possible Type II error.

Finally, a field study would have been preferable as the methodology only approximated the conditions and data sources available to the school psychologist decision maker.

Implications for Future Research

It is clear from these results that attempts at debiasing psychologist decisions with computer assistance is not a simple phenomenon. Future research should

address itself to both minimizing the limitations of the current research and building upon the data and methods herein in a manner that will more clearly delineate the strengths and weaknesses of computer-based decision aids.

Future research should include a much larger sample of cases for the decision makers to analyze. This would effectively allow a much wider and much closer to reality sampling of variation across cases reviewed. It would also allow for more confidence to be placed in the outcome statistics.

Attention should be given to the more global problem of relating the diagnostic category to the intervention. There is a need to more clearly investigate the efficacy of the decisions made instead of concentration upon measures of consistency alone. Although consistency is one aspect of validity; computer decisions aids and attendant research should concentrate upon development of systems whereby decisions can be empirically linked to effective intervention. Irrespective of how consistent decisions are; if the decisions have no treatment validity they are worthless in terms of instructional intervention.

Computer-based decisions aids available for today's school psychologists are only rudimentary approximations of what the technology is in other disciplines and what it will be in the future. Current and emerging technologies should therefore be actively utilized in

research and development. The effect of the decision aid upon the ultimate decision made by the decision maker should be integrated into the software development process rather than investigated piece meal post hoc. In this way, the software can be tailored to function as a decision aid rather than as a maker of decisions whose decisions must be reviewed.

In a similar vein, it is critical that subjective, available, affectively laden data points be integrated into the decision process in a manner wherein they can contribute to the treatment validity of the decision to be made rather than probably mitigate its effect. Computer-based expert systems do have this potential and should be developed for the use of psychologists as part of any ongoing research effort. By effectively integrating the subjective nuances inherent in any psychologist decision there are increased probabilities that computer decision aids might be more fully embraced by practicing school psychologists.

Further some trends in the data indicated that there was perhaps a differential effect of the computer-based decision aid. Any follow-up research on existing decision aids should more closely investigate those areas of the present research where the trends were most evident; i.e. where the decision choices were more ambiguous or of relatively low incidence as in the handicapping conditions of seriously emotionally

disturbed and language handicapped. This is because confining research to high incidence handicaps such as learning disability cannot truly provide an accurate predictor of the decision aid's utility due to the necessity to overcome the powerful effects of chance.

Implications for Professional Practice

Until computers and computer decision aids become more widely available and accepted by practicing school psychologists it will continue to be much easier for decisions to be made solely on the basis of clinical judgment. The profession tends to tolerate a moderate amount of inconsistency in judgment which decreases a need to utilize decision aids. Further, the software-based decision aids currently on the market are not sufficiently utilitarian and powerful to overcome their negative press. It still remains just as easy to base a decision upon clinical judgment as it is to base it upon the output of a software package whose validity is unknown. Nonetheless, the development of computer-based decision aids has made it possible for some individuals in the profession to realize there is indeed an alternative to pure clinical judgment. Even if the current software packages do not meet the needs, the mere fact they do exist and have a demonstrated potential to make consistent and potentially valid decisions makes it more likely that the profession will gradually move more towards their use over the coming decades. Also the

currently available software packages can play an important role in training for psychological decision making. Their use with clinicians in training will increase the probability that a new generation of clinicians will be more familiar with their use and will not consider them to be the threat that many currently practicing clinicians do. The results of this study will not change how the profession feels about computer-based decision aids, but it should provide an incentive to improve both the software packages themselves, the research surrounding them, and, most importantly, the quality of decisions that are made.

APPENDIX A

M-MAC User Instructions

Many first generation interpretive software packages rigorously prescribe the data that the user enters into the system as well as the manner in which the data are combined by the system. M-MAC, however, allows the user rather wide latitude in determining what data is included for consideration by the software as well as the specific algorithms and decision options available. The system is designed in a manner which requires the user to make a series of choices which determine how M-MAC evaluates the data as well as what is presented in the program output. Because flexibility is built into the program the user is able to tailor the classification process according to specific situational, administrative, or client-specific needs.

M-MAC diagnostic classification must be completed under one of three so-called operations modes; Standard Mode, Special Mode, and Research Mode. The operation mode controls the quantity and complexity of options available to the user. For the purposes of this study all experimental conditions utilizing M-MAC operated under the Standard Mode. Here all basic decision rules are predetermined by the system's algorithms. Nonetheless the user has the ability to exercise several options which will be detailed shortly. Program output is conducted under standards detailed in the M-MAC manual and are assumed by the program authors to generate systematic identical classifications across all users

irrespective of the Standard Mode user options employed. McDermott & Watkins (1985) stated that use of Standard Mode output in classification decisions would assure that each decision meets an "international and interagency reference standard" instead of localized and inconsistent classifications commonly reported in the literature (See Chapter 2). Special Mode and Research Mode allow the user to alter many of the statistical parameters and decision rules according to special needs. As Special and Research Modes are not utilized in this study, the reader is directed to the M-MAC manual for detailed information.

User defined options within the Standard Mode do not alter the statistical parameters used in the algorithms or other decision rules. Instead the user is allowed a series of options which define what information is actually utilized by the system, which classifications are generated, and the content of the actual output narrative. It is these parameters which were altered as levels two and three of the independent variable. They are most easily explained and understood within the context of a step-by-step explanation of the operation of M-MAC which follows.

After a series of title and introductory screens, the user is instructed to provide demographic and identifying information on the student to be classified. The information requested is student name or identifier,

sex, evaluation date, date of birth, and education level (grade/type of program currently enrolled within).

Variations in much of this information will influence system output primarily in the areas of the application of appropriate normative and other reference tables to specific child demographic variables; areas not under the direct control of the user in Standard Mode.

The user is then asked to make the first major decision which will have a significant impact on the eventual narrative output of the program. "Enter exceptionality level?" Here the system asks the user to report certain idiographic personal and situational factors relating to the cultural, environmental, linguistic, and medical/physical background of the child. Further the user is asked to estimate the extent of the child's general adaptation to these circumstances. For example screen number 2 asks: "Please indicate any uncorrected physical handicaps that you know or suspect the child to have." If the user enters yes for any of the listed handicaps, then the system will automatically branch to another series of questions in an attempt to determine the extent to which the handicaps are influencing the child's daily functioning or may have influenced the child's test performance.

Other options available are related to the user's determination of which data sources are to be included in the resulting analysis. Within the Standard Mode the

user may analyze data on the basis of one chosen dimension (i.e. intellectual, achievement, adaptive behavior, or social-emotional adjustment) or on the basis of a multidimensional classification which includes and integrates findings from all four dimensions. Permutations of two to three dimensions are precluded. Choice of a single dimension analyses data from one of the scales only. If the user was interested only in the intellectual domain, then data from a selected intelligence test would be analyzed only. On the other hand, if a multidimensional classification was chosen, the system would analyze results from each of the scales separately. Then, at the end of the printed narrative, it would integrate the results and print a multidimensional diagnostic classification which classifies each child according to the dimensions of intellectual functioning, academic achievement, adaptive behavior, and social-emotional adjustment. Optionally the system generates DSM-III and ICD-9 diagnostic descriptors.

Within each dimension, the user may select from one of several supported tests or other assessment methods. The intellectual functioning dimension supports a total of six individually administered intelligence tests including the entire Wechsler series. The academic achievement dimension includes also six achievement measures ranging from the Wide Range Achievement Test-Revised to the Woodcock-Johnson Tests of Achievement.

The adaptive behavior dimension supports a total of five scales including the Vineland Social Maturity Scale-Revised and the social-emotional adjustment dimension includes six different rating scales or behavior checklists including the Bristol Social Adjustment Guides.

Further, the user may choose to assess dimensional behavior on the basis of professional judgment instead of upon standardized instruments. Alternately, the user may choose to assess some dimensional behavior on the basis of professional judgment instead of upon standardized instruments. This option is available only on the dimensions of adaptive behavior and social-emotional adjustment. Under the adaptive behavior dimension, the user is asked to indicate whether the quality of the child's adaptive functioning is good, adequate, or deficient and to specify the methods used in reaching that conclusion. Possible methods include the professional judgment method, other, adaptive behavior indices, or both methods. If the quality of the behavior is considered to be deficient, the system will display the eight major areas of adaptive behavior recognized by the American Association on Mental Deficiency (Grossman, 1977; 1983) and asks that the areas deemed deficient be indicated. These will be listed in the output narrative under the adaptive behavior dimension. If none of the eight major areas was indicated by the user then the

narrative will report only that adaptive behavior is deficient.

Similarly, under the social-emotional adjustment dimension, the user may choose to utilize "other" data input as an assessment of adjustment in this dimension. The user is asked to specify whether the quality of the child's current social-emotional adjustment is considered to be well adjusted, adequately adjusted, or maladjusted. The user must also specify the basis upon which the judgment was made, i.e. professional judgment, other measures of social-emotional adjustment, or both methods. It is the intent of the program authors to refer to the other measures of social-emotional adjustment to apply to nonsystem-supported objective and projective personality instruments. When the social-emotional adjustment is considered to be maladjusted, the system asks the user to indicate a primary type of maladjustment in terms of DSM-III diagnostic criteria. This however is optional and instead the user may choose the term "other" instead of one of the diagnostic descriptors. The final narrative will include the DSM-III descriptor if entered and if not the narrative will read "nonspecific type" maladjustment. Next M-MAC presents the user with a selection menu for a secondary type maladjustment. The same options are available as for primary maladjustment and if no secondary maladjustments are evident, then the user may enter "none" and the narrative will not report a

secondary type of maladjustment.

When all of the requested information is entered into the system, the system will, within the span of less than five minutes, analyze the data according to both system defined and user defined parameters and generate several pages of narrative output. The narrative output, when studied and considered as part of the decision process functions as an interpretive software decision aid.

See Appendix D for copies of output used in this study and Appendix C for specific information regarding options actually utilized to generate this output.

APPENDIX B**Case Data Provided to All Decision Maker Subjects**

Decision makers in each level of the study received the data reproduced on the following pages. Those decision makers who received the interpretive software narrative also received this same case data in order to control for differences in data sources across the decision maker subjects.

The case material was adapted from an actual case referred for psychological evaluation by a school psychologist. The author attempted to attain a fine balance between the need to maintain confidentiality of all child data and a concomitant need to maximize the generalizability of the study by using actual data sources that a school psychologist would typically have available prior to a diagnostic decision. Therefore all identifying information was changed as was some of the actual data in order to protect the confidentiality of the original child referred. However, much of the raw test and observational data was retained in a manner that maintains confidentiality and simultaneously provides a highly typical case on which to make a diagnostic decision.

NAME: Child A
 GRADE 2, attending a public elementary school
 D.O.B.: 11/26/78
 AGE: 8 years 4 months
 SEX: Male
 DATE OF EVALUATION: 4/1/87
 PRIMARY LANGUAGE OR MODE OF COMMUNICATION: English

REASON FOR REFERRAL AND BACKGROUND INFORMATION: The child has been experiencing academic difficulties and demonstrating inappropriate classroom behaviors since at least his first grade year. According to his mother he was the product of a normal pregnancy. Developmental milestones reportedly were within normal limits and he has experienced no major illnesses or medical emergencies. A history of poor school attendance and tardiness appears to have exacerbated the difficulties he has been experiencing although some improvements have been noted in this area of concern recently. An evaluation from a child development center when he was five years old indicated that he was experiencing "serious perceptual impairments" that were suggestive of potential learning problems. Resource help on a nonhandicapped basis for fine motor weakness was provided in kindergarten as were speech and language services. He failed to master first grade objectives and repeated the first grade.

Stimulant medication has been prescribed for use on a daily basis. His current teachers report that his classroom behaviors tend to vary according to whether or not he has received his medication. His regular classroom teacher reports that when the medication is effective he tends to do work presented and listen attentively. However, when he has not taken it or it has not taken effect he tends to "...come on like gangbusters..." refusing to stay seated or do his classwork and "totally" disrupts the class. He is also described as being rude to adults, refuses to follow directions and sometimes appears to be "sulky and angry." Another teacher described him as an "emotionally disturbed, extremely sensitive individual who is oversensitive, rigid, and stubborn."

He was further described by his classroom teacher as "bright" and "knowledgeable." Academically reading and written language skills are especially difficult for him. Arithmetic performance is generally adequate whenever he is able to attend to instruction. He is a classroom leader and although he experiences some interpersonal difficulties with his peers he is generally liked by them.

BEHAVIORAL OBSERVATIONS:

The child was observed in his classroom during a time when the teacher was working with a small reading group in one corner of the room and the aide was available to assist the other children who were completing seat assignments. He was frequently out of his seat (preferential seating near the teacher's desk), exploring most areas of the room and seemed to lose himself within the activity of the busy classroom. He had work to do but generally refused to do it on his own except for one two minute interval during the forty-five minute observation when he worked quietly on task. He tended to ignore verbal requests to stay on task by both the

interval during the forty-five minute observation when he worked quietly on task. He tended to ignore verbal requests to stay on task by both the teacher and the aide. On two instances staff sat down with him to assist him with his work and on both occasions he worked cooperatively and appeared to be trying to do his best.

In the one to one evaluation session he presented as a generally cooperative individual variably motivated to do his best in the testing. Affect was positive across the session and interpersonal skills were adequate for his age although he frequently burped for attentional purposes. He generally followed directions but tended to test the limits of the examiner and passively refused some tasks. Out of seat behaviors were common. Firm examiner structure was generally sufficient to bring him back on task each time. Little overt anxiety was evident. He tended to "talk" his way through many of the tasks. Some of the visual perceptual tasks appeared to be somewhat confusing for him. Although he turned one of the Bender designs upside down because he could not "...see it like that..." his actual copy of the design was appropriately aligned. He reversed the colors in two out of his three block design errors and did not realize that the colors were reversed until mentioned by the examiner.

TEST SCORES:

WISC-R: VIQ 85	PIQ 100	FSIQ 91		
Information	7		Picture Completion	15
Similarities	10		Picture Arrangement	9
Arithmetic	9		Block Design	7
Vocabulary	5		Object Assembly	10
Comprehension	7		Coding	9
Digit Span	6			

BENDER GESTALT:

3 Koppitz developmental errors: (angulation; disproportion; integration)
 4 Koppitz emotional indicators: (confused order; large size; reinforced lines; second attempt)

WRAT-R:

Standard Scores: Reading 64 Spelling 64 Arithmetic 67

PROJECTIVE PERSONALITY TESTS: (TAT, Incomplete Sentences, Human Figure Drawings, Kinetic Family Drawing)

TAT themes: major theme: sadness (in 8 out of 10 cards) with no overt personal action that turned into happiness (5 cards) and more sadness or death (3 cards); other themes: surprise, aggression, uncertainty

Incomplete sentences: short abrupt action laden statements; strong emotional bond with family members; otherwise benign

Drawings: KFD: each individual drawn separately...does not include self; HFD: normal size; includes all developmentally expected features plus teeth in both figures; female face shaded; story/description: male 8 years old; female 4 years...both actively playing indoors at home.

APPENDIX C

M-MAC Options Utilized to Generate M-MAC Output for Groups 2 & 3

The following user options were employed to generate the M-MAC narrative stimulus materials for psychologist subjects in Groups 2 and 3.

MODE: Standard

EXCEPTIONALITY LEVELS Not entered

DATA SOURCES:

INTELLECTUAL	WISC-R with all subtest scores entered
ACHIEVEMENT	WRAT-R with Reading, Spelling, & Arithmetic
ADAPTIVE	Omitted
SOCIAL-EMOTIONAL	Omitted
PRINT DSM-III & ICD-9	No for Group 2
DIAGNOSTIC	Yes for Group 3
CLASSIFICATIONS?	

Note that the only difference between the data provided the Group 2 and Group 3 psychologists was that Group 3 psychologists received an M-MAC diagnostic classification and Group 2 subjects did not. See also Table 16.

APPENDIX D**M-MAC Narrative Output Provided Groups 2 & 3**

The psychologist subjects in Groups 2 and 3 received one of two variations of M-MAC narrative output. The narrative provided Group 2 subjects included everything on the next three pages with the exception of the "M-MAC CLASSIFICATION SUMMARY" at the end of the final page. Group 3 subjects received the next three pages as is; including the classification summary.

The following information is adapted from the
McDermott Multidimensional Assessment of Children
Copyright 1985
The Psychological Corporation

RECORD DATE: 04/01/87

METHODS:

WECHSLER INTELLIGENCE SCALE FOR CHILDREN - REVISED (WISC-R)
WIDE RANGE ACHIEVEMENT TESTS - REVISED (WRAT-R)
PROFESSIONAL JUDGMENT REGARDING SOCIAL-EMOTIONAL ADJUSTMENT
UNSPECIFIED MEASURES OF SOCIAL-EMOTIONAL ADJUSTMENT

I N T E L L E C T U A L F U N C T I O N I N G D I M E N S I O N
WECHSLER INTELLIGENCE SCALE FOR CHILDREN - REVISED (WISC-R)

OBTAINED INTELLIGENCE QUOTIENTS

95% CONFIDENCE LIMIT	VERBAL	PERFOR- MANCE	FULL SCALE
	85	100	91
UPPER	93	109	97
LOWER	77	91	85

OBTAINED FACTOR DEVIATION QUOTIENTS

95% CONFIDENCE LIMIT	VERBAL COMPREHEN.	PERCEPTUAL ORGANIZAT.	FREEDOM FR DISTRACT.
	84	102	87
UPPER	92	111	98
LOWER	76	92	76

ANALYSIS OF INTELLIGENCE QUOTIENTS:

MAXIMUM OVERALL ESTIMATED PREVALENCE LEVEL FOR ACCEPTING
ABNORMALITY OF VERBAL-PERFORMANCE IQ DIFFERENCE = 6%
THE 15-POINT DIFFERENCE BETWEEN VERBAL AND PERFORMANCE SCALE QUOTIENTS IS
NOT SUFFICIENTLY UNCOMMON TO BE CONSIDERED TO BE ABNORMAL, SINCE
DIFFERENCES THIS LARGE MAY OCCUR WITH AS MANY AS 25% OF CHILDREN WITHIN
THE CHILD'S AGE RANGE. SUCH A DIFFERENCE WOULD PROBABLY REPRESENT A
NATURAL VARIATION IN GENERAL INTELLECTUAL FUNCTIONING. HOWEVER, THE
DIFFERENCE IS SIGNIFICANT STATISTICALLY ($P < .05$) AND, THEREFORE, THE
VERBAL IQ IS USED AS THE PRINCIPAL INDICATOR OF THE CHILD'S CURRENT LEVEL
OF INTELLECTUAL FUNCTIONING.

ANALYSIS OF FACTOR SCORES:

MAXIMUM PROBABILITY LEVEL FOR ACCEPTING STATISTICAL SIGNIFICANCE
OF DIFFERENCES AMONG FACTOR DEVIATION QUOTIENTS = .15
THE DEVIATION QUOTIENT FOR THE PERCEPTUAL ORGANIZATION FACTOR IS
SIGNIFICANTLY GREATER THAN THAT FOR THE VERBAL COMPREHENSION FACTOR
($P < .05$). THIS RESULT IS CONSISTENT WITH THE OBTAINED VERBAL-PERFORMANCE
DISCREPANCY FAVORING THE PERFORMANCE SCALE IQ.

THE FREEDOM FROM DISTRACTIBILITY FACTOR QUOTIENT IS DEPRESSED SIGNIFICANTLY BELOW THAT FOR THE PERCEPTUAL ORGANIZATION FACTOR ($P < .15$). THIS MAY INDICATE THAT THE CHILD HAS COMPARATIVELY LESS ABILITY SUSTAINING ATTENTION IN CERTAIN PROBLEM SOLVING SITUATIONS, OR THAT THE CHILD IS SOMEWHAT LESS PROFICIENT IN RETAINING AND COGNITIVELY MANIPULATING NUMERICAL DATA. THESE POSSIBILITIES SHOULD BE STUDIED.

LEVEL OF FUNCTIONING:

CURRENTLY THE CHILD IS FUNCTIONING THE LOW AVERAGE LEVEL OF GENERAL INTELLIGENCE (IQ = 85). THIS ESTIMATE IS SUBJECT TO ERROR VARIATION. HOWEVER, WITH 95% CONFIDENCE, IT SHOULD NOT BE EXPECTED TO EXCEED THE UPPER LIMIT OF AN IQ OF 93 OR LOWER LIMIT OF AN IQ OF 77.

ACADEMIC ACHIEVEMENT DIMENSION
WIDE RANGE ACHIEVEMENT TESTS - REVISED (WRAT-R)

OBTAINED SUBSCALE SCORES

95% CONFIDENCE LIMIT	READING	SPELLING	ARITHMETIC
	64	64	67
UPPER	71	72	80
LOWER	57	56	54

DEVIATION OF SUBSCALE SCORES FROM CHILD'S AVERAGE PERFORMANCE LEVEL
AVERAGE = 65

LEVEL OF PERFORM.	READING	SPELLING	ARITHMETIC
----------------------	---------	----------	------------

>CHILD'S AVERAGE

CHILD'S SCORE	64	64	67
------------------	----	----	----

<CHILD'S AVERAGE

MAXIMUM PROBABILITY FOR ACCEPTING STATISTICAL SIGNIFICANCE
OF DEVIATIONS FROM AVERAGE PERFORMANCE = .05

DISCREPANCY BETWEEN EXPECTED & OBSERVED ACHIEVEMENT
(REGRESSION ANALYSIS)

LEVEL OF ACHIEVEMENT.	READING	ARITHMETIC
EXPECTED	90	90
OBSERVED	64	67
DISCREPANCY	-26	-23
STAT. SIGN.	.0001	.005
EST. % PREVALENCE	1	2
COEFFICIENT	.68	.69

MAXIMUM OVERALL ESTIMATED PREVALENCE LEVEL FOR ACCEPTING
ABNORMALITY OF DISCREPANCY FROM EXPECTED ACHIEVEMENT = 6%

READING:

COMPARED TO THE AVERAGE PERFORMANCE OF CHILDREN IN THE SAME AGE RANGE, THE CHILD'S PRESENT ACHIEVEMENT IN READING IS AT THE MILDLY DEFICIENT LEVEL. THIS LEVEL OF PERFORMANCE IS MARKEDLY DISCREPANT ($P < .0001$) FROM EXPECTANCY BASED ON THE CHILD'S ESTIMATED INTELLECTUAL FUNCTIONING, THUS INDICATING APPRECIABLE UNDERACHIEVEMENT IN READING. THIS DEGREE OF UNDERACHIEVEMENT IS UNLIKELY TO OCCUR IN MORE THAN 1% OF THE CHILD'S AGE PEERS.

SPELLING:

COMPARED TO THE AVERAGE PERFORMANCE OF CHILDREN IN THE SAME AGE RANGE, THE CHILD'S PRESENT ACHIEVEMENT IN SPELLING IS AT THE MILDLY DEFICIENT LEVEL.

ARITHMETIC:

COMPARED TO THE AVERAGE PERFORMANCE OF CHILDREN IN THE SAME AGE RANGE, THE CHILD'S PRESENT ACHIEVEMENT IN ARITHMETIC IS AT THE MILDLY DEFICIENT LEVEL. THIS LEVEL OF PERFORMANCE IS MARKEDLY DISCREPANT ($P < .005$) FROM EXPECTANCY BASED ON THE CHILD'S ESTIMATED INTELLECTUAL FUNCTIONING, THUS INDICATING APPRECIABLE UNDERACHIEVEMENT IN ARITHMETIC. THIS DEGREE OF UNDERACHIEVEMENT IS UNLIKELY TO OCCUR IN MORE THAN 2% OF THE CHILD'S AGE PEERS.

DIMENSION SUMMARY:

ACADEMIC AREA	STATUS RELATIVE TO AGE RANGE	STATUS RELATIVE TO INTELLECTUAL FUNCT.
READING	MILDLY DEFICIENT	UNDERACHIEVEMENT
SPELLING	MILDLY DEFICIENT	
ARITHMETIC	MILDLY DEFICIENT	UNDERACHIEVEMENT

M-M A C C L A S S I F I C A T I O N S U M M A R Y

MULTIDIMENSIONAL CLASSIFICATION

DESCRIPTION	RELATED DSM-III
LOW AVERAGE INTELLECTUAL FUNCTIONING	
MIXED LEARNING DISABILITY (PROVISIONAL)	315.50

APPENDIX E

Cover Letter



THE UNIVERSITY OF MARYLAND

COLLEGE OF EDUCATION
Division of Human and Community Resources
College Park, Maryland 20742

Counseling and Personnel Services
JCI, 454-2020

December 1, 1987

Dear

I am writing to you to request approximately twenty minutes of your assistance and expertise. I am a doctoral candidate in school psychology interested in the school psychologist's decision-making process.

I have enclosed data relevant to a school-based case referred for psychological evaluation in order to assist in the determination of the child's eligibility for special educational services.

Please review the case data and decide whether the child is educationally handicapped. If handicapped, please indicate the primary handicapping condition.

Enter your decision on the enclosed (blue) sheet of paper. After you have made your decision, open the blue sheet and answer the five short questions that request primarily demographic information. I have enclosed a self-addressed stamped envelope for your reply as well as a stamped postal card that I request that you mail as well. All responses will be anonymous. I anticipate completing the data analysis no later than April, 1988 and will be happy to send copies of my findings to those interested.

Thank you in advance for your cooperation. It is my hope that the data collected herein will contribute to an increased understanding of the school psychologist's decision process and that this project will be of interest to you.

Sincerely,

John W. Wisor
John W. Wisor, M.A.
School Psychologist

William Strein
William Strein, D.Ed.
Associate Professor

APPENDIX F**Decision/Response Sheet**

There were three separate versions of the response form on the next page; one for each subject group. The variations occur only on the question beginning "To what extent were the following sources of information...." Group 1 subjects were asked to indicate degree of influence from among reason for referral, background information, observations, and test data only. Group 2 subjects were given these data sources, plus the M-MAC statistics and M-MAC narrative to respond to. Group 3 subjects were additionally provided the data source of M-MAC classification. See also Table 16.

PLEASE ENTER YOUR DECISIONS BELOW.
DO NOT UNSTAPLE THIS SHEET UNTIL YOU HAVE ENTERED YOUR DECISIONS.

What is the probability that the child is:
(Circle the percentage that best represents your estimation for each category below.)

Language Disabled	0%—25%—50%—75%—100%
Learning Disabled	0%—25%—50%—75%—100%
Mentally Retarded	0%—25%—50%—75%—100%
Seriously Emotionally Disturbed	0%—25%—50%—75%—100%
Not Educationally Handicapped	0%—25%—50%—75%—100%

Now consider that you are a member of a special education admission, review, and dismissal team and must cast a vote to determine which handicapping condition, if any, is the child's primary handicapping condition. Mark an "X" by the descriptor below that best represents your decision.

☐ Language Disabled
☐ Learning Disabled
☐ Mentally Retarded
☐ Seriously Emotionally Disturbed
☐ Not Educationally Handicapped

Now suppose that no additional educational interventions were provided to this child. How probable is each of the following prognostic statements? (Circle one probability estimate for each statement below.)

The child will experience significant difficulties in the future in at least one of the major academic areas.

0%—25%—50%—75%—100%

The child will demonstrate future social adjustment problems.

0%—25%—50%—75%—100%

The child will be referred again for consideration as an educationally handicapped child.

0%—25%—50%—75%—100%

The child will successfully graduate from high school.

0%—25%—50%—75%—100%

The child will have difficulty competing in the vocational marketplace as an adult.

0%—25%—50%—75%—100%

NOW PLEASE REMOVE THE STAPLE, TURN THIS SHEET OVER, AND ANSWER THE FIVE SHORT QUESTIONS THEREON. AFTER YOU HAVE COMPLETED THE SHORT QUESTIONNAIRE ON THE OTHER SIDE OF THIS SHEET YOU MAY USE THE SPACE BELOW TO ENTER ANY COMMENTS YOU MIGHT HAVE.

Now that you have checked off your diagnostic decision, please answer the following questions.

Did you use any resource or reference materials as you were making the diagnostic decision you just made? YES _____ NO _____

If you answered yes to the previous question, please check what kind of resource that you used.

TEXTBOOK _____ OTHER WRITTEN MATERIAL _____
COLLEAGUE _____ COMPUTER SOFTWARE _____
OTHER _____ SOFTWARE TITLE _____

To what extent were the following sources of information influential in the diagnostic decision you just made? Circle the number which reflects the influence of each of the sources.

	NO INFLUENCE	MODERATE INFLUENCE	MUCH INFLUENCE
REASON FOR REFERRAL	1-----2-----3-----4-----5		
BACKGROUND INFORMATION	1-----2-----3-----4-----5		
OBSERVATIONS	1-----2-----3-----4-----5		
TEST DATA	1-----2-----3-----4-----5		

How long have you been a practicing (school) psychologist?

LESS THAN 1 YEAR _____ 2-5 YEARS _____
6-10 YEARS _____ OVER 10 YEARS _____

To what extent have you used computers of any type in your current professional (not recreational) practice? Circle the applicable number.

	NO USE	SOME USE	FREQUENT USE
WORD PROCESSING	1-----2-----3-----4-----5		
REPORT WRITING	1-----2-----3-----4-----5		
TEST SCORING	1-----2-----3-----4-----5		
TEST INTERPRETATION	1-----2-----3-----4-----5		
CASE MANAGEMENT	1-----2-----3-----4-----5		
COUNSELING/THERAPY	1-----2-----3-----4-----5		
OTHER	1-----2-----3-----4-----5		

Thank you for taking the time to participate in this study. There is space for any comments you might have on the other side of this sheet. When you have finished please place this sheet into the enclosed self-addressed and stamped envelope and mail it. The self-addressed and stamped postcard should also be mailed separately. If you desire a copy of the results of this study, please mark the appropriate space on the postcard as well.

Thank you again for your valuable assistance and patience.

APPENDIX G

Response Permutations Provided by Subjects
 to the Series of Questions:
 "What is the Probability that the Child is...?"

Each subject was asked to make a probability estimate for each of the handicapping/nonhandicapping conditions. The entries that follow is the entire set of permutations of the probability judgments made by the subjects. The handicapping condition of "Mentally Retarded" was considered to have a probability of 0% in each permutation and was therefore omitted from this list.

LD ¹	NEH ²	SED ³	LANG ⁴	
75	25	25	0	
75	25	25	75	
50	25	25	25	
75	0	0	0	
75	25	0	25	
75	25	25	25	
50	75	25	25	
75	25	0	12	
0	75	0	0	
50	0	50	0	
25	0	0	25	
100	0	0	25	
75	0	50	25	
100	0	25	50	
75	0	50	50	
75	50	12	50	omit ⁵
100	75	0	0	omit
75	0	0	50	
100	0	25	25	
75	0	25	25	
100	0	50	75	
25	0	50	25	
50	0	25	25	
75	25	25	50	
75	25	50	25	
75	0	25	0	
50	100	25	50	omit
100	25	0	25	
50	50	25	50	
75	0	0	25	
75	0	0	75	
50	25	25	50	
50	25	25	75	
25	75	75	25	omit

75	75	25	25	omit
25	25	50	25	
25	50	25	25	
25	75	25	25	
75	0	50	75	
0	75	75	0	omit
75	0	25	50	
100	75	0	75	omit
100	0	25	75	
100	100	0	0	omit
25	25	50	0	
75	25	50	50	
75	0	0	100	
25	0	50	75	
75	25	12	50	
50	50	0	25	
87	0	25	50	
50	0	25	75	
75	0	50	0	
75	100	0	75	omit
25	25	0	50	
50	0	25	50	
0	50	50	0	
50	0	75	25	

¹LD indicates "Learning Disabled."

²NEH indicates "Not educationally handicapped."

³SED indicates "Seriously emotionally disturbed."

⁴LANG indicates "Language disabled."

⁵Omit indicates that this is one of the permutations whose meaning was uninterpretable and therefore omitted from the data analysis.

REFERENCES

- Abelson, R. P. (1976). Script processing in attitude formation and decision making. In J. S. Carroll & J. W. Payne (Eds.), Cognition and social behavior. New York: Erlbaum.
- Achenbach, T. M., & Edelbrock, C. S. (1978). The classification of child psychopathology: Review and analysis. Psychological Bulletin, 85, 1275-1301.
- Altemose, J. R., & Williamson, K. B. (1981). Clinical judgment vs. the computer: Can the school psychologist be replaced by a machine? Psychology in the Schools, 18, 356-363.
- American Educational Research Association, American Psychological Association, & National Council on Measurement in Education. (1985). Standards for educational and psychological testing. Washington DC: American Psychological Association.
- American Psychiatric Association. (1980). Diagnostic and statistical manual of mental disorders (Third Edition) Washington, DC: Author.
- American Psychiatric Association. (1987). Diagnostic and statistical manual of mental disorders (Third Edition) Revised. Washington, DC: Author.
- American Psychological Association. (1986). Guidelines for computer-based tests and interpretations. Washington, DC: Author.

- Anderson, B. F., Deane, D. H., Hammond, K. R., & McClelland, G., (1981). Concepts in judgment and decision research. New York: Praeger.
- Anderson, N. H. (1981). Foundations of information integrating theory. New York: Academic Press.
- Argote, L., Seabright, M. A., & Dyer, L. (1986). Individual versus group use of base-rate and individuating data. Organizational Behavior and Human Decision Processes, 38, 65-75.
- Bannatyne, A. (1974). Diagnosis: A note on recategorization of the WISC scaled scores. Journal of Learning Disabilities, 7, 272-273.
- Bar-Hillel, M. (1982). Studies of representativeness. In D. Kahneman, P. Slovic, & A. Tversky (Eds.), Judgement under uncertainty: Heuristics and biases (pp. 69-83). New York: Cambridge University Press.
- Bazerman, M. (1984). Judgment in managerial decision making. New York: John Wiley & Sons.
- Bennett, J. L. (Ed.). (1983). Building decision support systems. Reading, MA: Addison-Wesley Publishing Co.
- Bracken, B. (1986). [Review of McDermott multidimensional assessment of children.] Computers in Human Behavior, 2, 309-313.
- Brantley, J. (1984). Differences among software decision rules for WISC-R analysis [Summary]. Convention '84 Proceedings National Association of School Psychologists, 376.

- Brantley, J., Jacob, S., & Troutman, L. (Summer, 1985).
The CTASP survey of computer use, attitudes and
ethical concerns among psychologists in the schools.
CTASP Newsletter, (Available from NASP Washington,
DC), 6-10.
- Butcher, J. (1979). The Minnesota Report [Computer
program]. Minneapolis, MN: National Computer Systems.
- Butcher, J. Keller, L., & Bacon, S. (1985). Current
developments and future directions in computerized
personality assessment. Journal of Consulting and
Clinical Psychology 53, 803-815.
- Caldwell, C. (1971). The Caldwell report [Computer
program]. Los Angeles: Caldwell Report, Inc.
- Campbell, D. & Stanley, J. (1963). Experimental and
quasi-experimental designs for research. Chicago:
Rand McNally College Publishing Company.
- Cleary, T. A., Humphreys, L. G., & Kendrick, S. A. &
Wesman, A. (1975). Educational uses of tests with
disadvantaged students. American Psychologist, 30,
15-41.
- Cronbach, L. (1956). Assessment of individual
differences. Annual review of psychology. 7, 173-196.
- Cummings, J. A., Huebner, E. S., & McLeskey, J. (1986).
Psychoeducational decision-making: Reasons for
referral vs test data. Professional School
Psychology, 1, 253-260.

- Dawes, R. (1982). The robust beauty of improper linear models in decision making. In D. Kahneman, P. Slovic, & A. Tversky (Eds.), Judgement under uncertainty: Heuristics and biases (pp. 391-407). New York: Cambridge University Press.
- Dawson, M. (1985). LD: Professional dilemma or diagnostic scandal [Summary]? Convention '85 Proceedings National Association of School Psychologists, 133-135.
- DeWitte, L. (1976). The effects of client race and social class on the attitudes and decisions of school psychologists. Dissertation Abstracts International, 37, 3603-3604B.
- Doll, B. (Winter, 1985). The ethical computer. CTASP Newsletter, (Available from NASP Washington, DC), 7-8.
- Dyer, J. S. & Mulvey, J. M. (1983). Integrating optimization models with information systems for decision support. In J.L. Bennett (Ed.), Building decision support systems (pp. 89-110), Reading, MA: Addison-Wesley Publishing Co.
- Edwards, W. (1961). Behavioral decision theory. Annual Review of Psychology, 12, 473-498.
- Einhorn, H. J. (1982). Learning from experience and suboptimal rules in decision making. In D. Kahneman, P. Slovic, & A. Tversky (Eds.), Judgement under uncertainty: Heuristics and biases (pp. 268-283). New York: Cambridge University Press.

- Elders, D. (1977). Biased teacher referrals and their effects upon school psychologists' diagnostic judgments and expectancies. Dissertation Abstracts International, 38, 1297A.
- Ellis, C. R., & Ray, S. (1983). Use of the computer to assist WISC-R administration. [Summary]. Convention '83 Proceedings National Association of School Psychologists, pp.163-164.
- Engle, M. (1966). Time and the reluctance to diagnose. Journal of School Psychology, 4, 1-8.
- Eyde, L., Kowall, D., & Fishburne, F. (1986, August). The validity of computer-based test interpretations of the MMPI. Paper presented at the 94th Annual Convention of the American Psychological Association, Washington, DC.
- Exner, J. (1974). The Rorschach: a comprehensive system (Vol. 1). New York: Wiley.
- Feldman, J. (1981). Beyond attribution theory: Cognitive processes in performance appraisal. Journal of Applied Psychology, 66, 127-148.
- Fischhoff, B. (1977). Perceived informativeness of facts. Journal of Experimental Psychology: Human Perception and Performance, 3, 349-358.
- Fishburn, P. (1972). Personalistic decision theory: Exposition and critique. In H. S. Brinkers (Ed.). Decision-making. Columbus, OH (pp. 4-41).

- Fisher, G. L. & Jenkins, S. J. (1986). Computer software: An aid in interpreting the Wechsler Intelligence Scale for Children-Revised. Measurement and Evaluation in Counseling and Development, 19, 53-59.
- Fleiss, J. L. (1971). Measuring nominal scale agreement among many raters. Psychological Bulletin, 76, 378-382.
- Fowler, R. (1969). The current status of computer interpretation of psychological tests. American Journal of Psychiatry (Supplement) 125, 21-27.
- Fowler, R. (1980). The automated MMPI. In J. B. Sidowski, J. H. Johnson, & T. A. Williams (Eds.). Technology in mental health care delivery systems. Norwood, NJ: Ablex Publishing Corporation. pp. 69-84.
- Fowler, R. (1985). Landmarks in computer-assisted psychological assessment. Journal of Consulting and Clinical Psychology, 53, 748-759.
- Fowler, D., Finkelstein, A., & Penk, W. (1986, August), Measuring treatment response by computer interview. Paper presented at the 94th Annual Convention of the American Psychological Association, Washington, DC.
- Friedman, L., Howell, W. C., & Jensen, C. R. (1985). Diagnostic judgment as a function of preprocessing of evidence. Human Factors, 25, 665-673.
- Garfield, S. (1978). Research problems in clinical diagnosis. Journal of Consulting and Clinical Psychology, 46, 596-607.

- Gibbons, J. (1985). Nonparametric methods for quantitative analysis. Columbus, OH: American Sciences Press, Inc.
- Ginsberg, A. S. (1972). The diagnostic process viewed as a decision problem. In J. Jacquez (Ed.) Computer diagnosis and diagnostic methods (pp. 203-240). Springfield, IL: Charles C. Thomas.
- Goldberg, L. R. (1959). The effectiveness of clinicians' judgments: The diagnosis of organic brain damage from the Bender-Gestalt Test. Journal of Consulting Psychology, 23, 25-33.
- Goldberg, L. R. (1968). Simple models or simple processes American Psychologist, 23, 25-33.
- Goldberg, L. R. (1970). A rationale for a method of improving on clinical inference. Psychological Bulletin, 73, 422-432b.
- Grebstein, L. (1963). The relative accuracy of actuarial prediction, experienced clinicians, and graduate students in a clinical judgment task. Journal of Consulting Psychology, 37, 127-132.
- Gredler, G. R. (1986, August). Assessment of problem-solving judgment of school psychologists: Implications for staff development. Paper presented at the 94th Annual Convention of the American Psychological Association, Washington, DC.
- Green, C. (1982). The diagnostic accuracy and utility of the MMPI & MCMI computer interpretive reports. Journal

- of Personality Assessment, 46, 359-369.
- Greenberg, A. (1986). Change process research. Journal of Clinical Psychology, 54, 4-9.
- Grossman, H. J. (Ed.). (1977). Manual on terminology and classification in mental retardation (Rev. ed.).
- Grossman, H. J. (Ed.). (1983). Classification in mental retardation. Washington, DC: American Association on Mental Deficiency.
- Gutkin, T. & Elwork, A. (1985). The behavioral sciences in the computer age. Computers in Human Behavior. 1, 19-34.
- Hale, R. (1982). School psychology information tools [Computer program]. State College, PA: author.
- Hammond, K. R. (1966). The psychology of Egon Brunswik. New York: Holt, Rinehart, & Winston.
- Hammond, K. R., McClelland, G. H., & Mumpower, J. (1980). Human judgment and decision making. New York: Praeger Publishers.
- Hathaway, S. R. & McKinley, J. C. (1942). A multiphasic personality schedule. Journal of Psychology. 14, 73-84.
- Hedlund, J., Evenson, R., Sletten, I., & Cho, D. (1980). The computer and clinical prediction. In J. B. Sidowski, J. H. Johnson, & T. A. Williams (Eds.). Technology in mental health care delivery systems (pp. 201-235). Norwood, NJ: Ablex Publishing Corporation.
- Heider, F. (1958). The psychology of interpersonal

relations. New York: Wiley.

Hersh, J. B. (1971). Effects of referral information on testers. Journal of Consulting and Clinical Psychology, 37, 116-122.

Hill, P. H., Bedau, H. A., Chechile, R. A., Crochtiere, W., Kellerman, B. L., Ounjian, D., Pauker, S. G., & Rubin, J. Z. (1978). Making decisions. Reading, MA: Addison-Wesley Publishing Company.

Hoffman, R. (Fall, 1984). [Review of the Rorschach diagnostic and interpretive program and the Bender report interpretive software programs]. CTASP Newsletter, (Available from NASP Washington, DC), 6-7.

Hoffman, R. (Fall, 1985). [Software review] CTASP Newsletter, (Available from NASP Washington, DC), 1-10.

Huebner, R. (1987). The effects of type of referral information and test data on psychoeducational decisions. School Psychology Review, 16, 382-390.

Huebner, F. S., & Cummings, J. A. (1985). The impact of sociocultural background and assessment data upon school psychologists' decisions. Journal of School Psychology, 23, 157-166.

Ingram, R. (1985). The Kaufman method of WISC-R hypothesis generation [Computer program]. New York: John Wiley & Sons.

Jacob, S. & Brantley, J. (Winter, 1986). Computer applications in school psychology: Proposed guidelines

for good practice. CTASP Newsletter, (Available from NASP Washington, DC), 8-12.

Janis, I. L., & Mann, L. (1977). Decision making. New York: The Free Press.

Johnson, J. H. & Johnson, J. N. (1981). Psychological considerations related to the development of computerized testing stations. Behavior Research Methods and Instrumentation, 13, 421-424.

Johnson, J. & Williams, T. (1975). The use of on-line computer technology in a mental health admitting system. American Psychologist, 30, 388-390.

Johnson, J. & Williams, T. (1980). Using on-line computer technology to improve service response and decision-making effectiveness in a mental health admitting system. In J. B. Sidowski, J. H. Johnson, & T. A. Williams (Eds.). Technology in mental health care delivery systems. Norwood, NJ: Ablex Publishing Corporation. pp. 237-249.

Kahneman, D. & Tversky, A. (1972). Subjective probability: A judgment of representativeness. Cognitive Psychology, 3, 430-454.

Kahneman, D. & Tversky, A. (1982a). Intuitive prediction: Biases and corrective measures. In D. Kahneman, P. Slovic, & A. Tversky (Eds.), Judgement under uncertainty: Heuristic and biases (pp. 414-421). New York: Cambridge University Press.

- Kaplan, M. (1975). Information integration in social judgment: Interaction of judge and informational components. In M. Kaplan & S. Schwartz (Eds.). Human judgment and decision processes (pp. 139-171). New York: Academic Press.
- Kaufman, A. (1979). Intelligent testing with the WISC-R. New York: Wiley Interscience.
- Keeney, R. & Raiffa, H. (1976). Decisions with multiple objectives: Preferences and value trade-offs. New York: Wiley.
- Kirsch, M., & Ford, J. K. (1986, August). The effect of type of data on information integration strategies. Paper presented at the 94th Annual Convention of the American Psychological Association, Washington, DC.
- Kleinmuntz, B. (1963). Personality test interpretation by digital computer. Science, 139, 416-418.
- Klett, J. C., & Pumroy, D. K. (1971). Automated procedures in psychological assessment. In P. McReynolds (Ed.), Advances in psychological assessment (Vol. 2). Palo Alto, CA: Science and Behavior Books, Inc.
- Knoff, H. (1980). A comparison of school psychology and special education trainees and practitioners: The effect of presentation order of intelligence and adaptive behavior information on educational placement decisions, and attitudes toward group process variables in child study teams. Dissertation Abstracts

International, 41, 2518A.

- Knoff, H. (1984). Placement decisions revisited: IQ doesn't always tip the scale. Exceptional Children, 51, 122-128.
- Kovaleski, J. (1985). Consultation-based verification: A position paper on the assessment of learning disabilities [Summary]. Convention '85 Proceedings National Association of School Psychologists, pp. 15-137.
- Kozielecki, J. (1975). Psychological decision theory. Boston: D. Reidel Publishing Co.
- Kramer, J. (1988). Computer-based test interpretation in psychoeducational assessment: An initial appraisal. Journal of School Psychology, 26, 143-153.
- Krug, S. (1984). Psychware. Kansas City, MO.: Test Corporation of America.
- Lachar, D. (1974). WPS test report [Computer program]. Los Angeles: Western Psychological Services.
- Lanyon, R. I. & Goodstein, L. D. (1982). Personality assessment, 2nd Edition. New York: Wiley.
- Lichtenstein, S. & Slovic, P. (1971). Reversals of preference between bids and choices in gambling decisions. Journal of Experimental Psychology, 89, 46-55.
- Lichtenstein, S. Slovic, P., Fischhoff, B., Layman, M., & Combs, B. (1978). Judged frequency of lethal events. Journal of Experimental Psychology: Human Learning

- and Memory, 4, 551-578.
- Lichtenwald, T. (1982). WISC-R factor analysis program [Computer program]. Oshkosh, WI: Happ Electronics.
- Lidz, C. S. (1981). Improving assessment of school children. San Francisco, Josey-Bass, Inc.
- Light, R. L. (1971). Measures of response agreement for qualitative data: Some generalizations and alternatives. Psychological Bulletin, 76, 365-377.
- Loesch, L. (1986). Computer-assisted assessment: A reaction to Meir and Geiger. Measurement and Evaluation in Counseling and Development, 19, 35-37.
- Lyle, O. & Quast, W. (1976). The Bender-Gestalt: Use of clinical judgment versus recall scores in prediction of Huntington's disease. Journal of Consulting & Clinical Psychology, 44, 229-232.
- Lyon, D. & Slovic, P. (1976). Dominance of accuracy information and neglect of base rates in probability estimation. Acta Psychologica, 40, 287-298.
- Machover, K. (1949). Personality projection in the drawing of the human figure. Springfield, IL: Charles C. Thomas.
- MacMann, G. (1985). The actuarial classification of underachievement: A simulation of technical adequacy [Summary]. Convention '85 Proceedings National Association of School Psychologists, pp. 46-47.
- Matazarro, J. (1983). Computerized psychological testing. Science, 221, 323.

- McCoy, S. A. (1976). Clinical judgments of normal childhood behavior. Journal of Consulting and Clinical Psychology, 44, 710-714.
- McCullough, S. (Fall, 1984). 1984 NASP CTASP membership survey. CTASP Newsletter, (Available from NASP Washington, DC), 1-5.
- McCullough, S. & Olson, K. (1982). Computer technology applications in school psychology practice and training. [Summary]. Convention '82 Proceedings National Association of School Psychologists, pp. 186-187.
- McDermott, P. (1975). Characteristics of diagnostic decision-making behaviors among groups of pre- and post-practicum trainee and experienced school psychologists. Dissertation Abstracts International, 36, 3520A.
- McDermott, P. (1977). Program MAC (Multidimensional Actuarial Classification): A FORTRAN program for differential diagnosis of mental and learning deficiency and other disorders in children (Computer program). Philadelphia: University of Pennsylvania.
- McDermott, P. (1980). A systems actuarial method for the differential diagnosis of handicapped children. Journal of Special Education, 14, 7-22.
- McDermott, P. (1982a). Validation of a systems-actuarial computer process for multidimensional classification of child psychopathology. Journal of School

- Psychology, 38, 477-485.
- McDermott, P. (1982b). Actuarial assessment systems for the grouping and classification of school children. In C. R. Reynolds & T. B. Gutkin (Eds.), The handbook of school psychology (pp. 243-272). New York: Wiley.
- McDermott, P. (1986, August). Computer-based remedial instruction linked directly to assessment. Paper presented at the 94th Annual Convention of the American Psychological Association, Washington, DC.
- McDermott, P. (in press). Agreement among diagnosticians or observers: Its importance and determination. Journal of School Psychology.
- McDermott, P. & Hale, R. (1982). Validation of a systems actuarial computer process for multidimensional classification of child psychopathology. Journal of Clinical Psychology, 38, 477-486.
- McDermott, P. & Watkins, M. (1985). McDermott multidimensional assessment of children [Computer program] (M-MAC). San Antonio, TX: The Psychological Corporation.
- McDermott, P., Fridy, J., Lambert-Newton, S., Nester, A., Pies, D., Samuels-Singer, M., Sutton-Smith, K. & Watkins, M. (1984, April). Field implementation of M-MAC computerized IEP's in school psychology. Symposium conducted at the Annual Convention of the National Association of School Psychologists, Philadelphia, PA.

- McReynolds, P. (Ed.). (1971). Advances in psychological assessment (Vol. 2). Palo Alto, CA: Science and Behavior Books, Inc.
- Meehl, P. E. (1954). Clinical versus statistical prediction. Minneapolis, MN: University of Minnesota Press.
- Mercadal, D. (1982). WISC-R monitor [Computer program]. Burlington, NC: Southern Microsystems.
- Miller, M. (1984). Computerized models of psychiatric diagnosis. In M. Schwartz (Ed.), Using computers in clinical practice (pp. 225-240). New York: The Haworth Press, Inc.
- Minium, E. (1978). Statistical reasoning in psychology and education. New York: John Wiley & Sons.
- Mischel, W. (1969). Continuity and change in personality. American Psychologist, 24, 1012-1018.
- Moreland, K. (1985). Validation of computer-based test interpretations: Problems and prospects. Journal of Consulting and Clinical Psychology, 53, 816-825.
- Murphy, K. (1987). The accuracy of clinical versus computerized test interpretations. American Psychologist, 42, 192-193.
- Nichols, M., & Knoff, I. (1977). Refining computerized test interpretations: An indepth approach. Journal of Personality Assessment, 41, 157-159.
- Nicholson, C. (1982). WISC-R monitor [Computer program]. Casper, WY: Clinician's Digest.

- Nisbett, R. Borgida, E. Crandall, R. & Reed, H. (1982).
In D. Kahneman, P. Slovic, & A. Tversky (Eds.),
Judgement under uncertainty: Heuristics and biases
(pp. 445-459). New York: Cambridge University Press.
- Oswald, J. (1986, August). Publishers' perspective on the
use of computerized assessment. Paper presented at
the 94th Annual Convention of the American
Psychological Association, Washington, DC.
- Passini, F. & Norman, W. (1966). A universal conception
of personality structure? Journal of Personality and
Social Psychology, 4, 44-49.
- Peterson, C. & Hart, D. (1978). Use of multiple
discriminant function analysis in evaluation of a
state-wide system for identification of educationally
handicapped children. Psychological Reports, 43, 743-
755.
- Piersma, H. (1986, August). Computer-generated diagnoses:
How do they compare to clinician judgment? Paper
presented at the 94th Annual Convention of the
American Psychological Association, Washington, DC.
- Quinlan, D. & Quinlan, S. (1984). WISC-R 90 program:
WISC-R 90 Edudisk (Software review). Computers in
Psychiatry and Psychology, 6, 23-24.
- Rassen, R. (1979). An investigation of the judgment
policies of school psychologists in diagnosing and
prescribing for students. Dissertation Abstracts
International, 40, 1969-1970A.

- Reitman, W. (1984). Heuristic decision procedures, open constraints, and the structure of ill-defined problems. In M. W. Shelley & G. L. Bryan (Eds.), Human judgments and optimality. New York: Wiley. pp. 282-315.
- Replogle, W. H. (1982). A controlled comparison of psychological examiners and computer generated interpretations of the WISC-R. Dissertation Abstracts International, 43, 1900A.
- Reschly, D., Genshaft, J., & Binder, M. (1986). The 1986 NASP survey: Comparison of practitioners, NASP leadership, and university faculty on key issues. Washington, DC: National Association of School Psychologists.
- Ross, M. & Sicoly, F. (1982). Egocentric biases in availability and attribution. In D. Kahneman, P. Slovic, & A. Tversky (Eds.), Judgment under uncertainty: Heuristics and biases (pp. 163-178). New York: Cambridge University Press.
- Rubin, S. (1986, August). Using computerized testing in the psychiatric hospital. Paper presented at the 94th Annual Convention of the American Psychological Association, Washington, DC.
- Sampson, J. (1983). Computer-assisted testing and assessment: Current status and implications for the future. Measurement and Evaluation in Guidance, 15, 293-299.

- Sattler, J. (1982). Assessment of children's intelligence and special abilities (2nd ed.) Boston: Allyn & Bacon.
- Sawyer, J. (1966). Measurement and prediction, clinical and statistical. Psychological Bulletin, 66, 178-200.
- Scherphorn, T. (1979). Informational use by school psychologists for special education placement decisions. Dissertation Abstracts International, 41, 172A.
- Schneider, D. (1973). Implicit personality theory: A review. Psychological Bulletin, 79, 294-309.
- Schwartz, M. (Ed.). (1984). Using computers in clinical practice. New York: The Haworth Press, Inc.
- Shanteau, J. (1978). When does a response error become a judgmental bias? Commentary on "Judged frequency of lethal events." Journal of Experimental Psychology: Human Learning and Memory, 4, 579-581.
- Sichel, A. (1982). Evaluation of three models for classification of exceptional children [Summary]. Convention '83 Proceedings National Association of School Psychologists, pp. 58-59.
- Sidowski, J., Johnson, J., Williams, T. (Eds.). (1980). Technology in mental health care delivery systems. Norwood, NJ: Ablex Publishing Corporation.
- Simon, H. (1976). Administrative behavior: A study of decision-making process in administrative organization (3rd ed.). New York: The Free Press.

- Simon, H. & Reed, S. (1976). Modeling strategy shifts in a problem solving task. Cognitive Psychology, 8, 86-97.
- Sinnett, E. & Albott, W. (1987). Computerized interpretations: Additional issues concerning ethics, standards, and regulation of practice. American Psychologist, 42, (2), 191-192.
- Smith, M. L. (1984). Distorted indicators in special education: A response to Ginsberg. Evaluation and Program Planning, 7, 13-14.
- Snizek, J. A. & Reeve, A. P. (1986). Feature cues in probability learning: Data base information and judgment. Organizational Behavior and Human Decision Processes, 37, 297-315.
- Space, L. (1981). The computer as psychometrician. Behavior Research Methods & Instrumentation, 13, 595-606.
- Spitzer, R., Endicott, J., Cohen, J., & Fleiss, J. (1974). Constraints on the validity of computer diagnosis. Archives of General Psychiatry, 31, 197-203.
- Taylor, S. (1982). The availability bias in social perceptions and interaction. In D. Kahneman, P. Slovic, & A. Tversky (Eds.), Judgment under uncertainty: Heuristics and Biases (pp. 190-200). New York: Cambridge University Press.

- Temerlin, M. (1968). Suggestion effects in psychiatric diagnoses. Journal of Nervous and Mental Disorders, 147, 349-353.
- Thomas, A. (Summer, 1984). [Review of the WISC-R Monitor interpretative software program]. CTASP Newsletter, (Available from NASP Washington, DC), 1-5.
- Thomas, A. (Fall, 1986). Computer questions that come my way. CTASP Newsletter, (Available from NASP Washington, D.C), 1, 10.
- Trifletti, D. T., Trifletti, R. M., & Trifletti, R. J. (1983). WISC-R-90 [Computer program]. Jacksonville, FL: Precision People.
- Trope, Y. (1982). Inferences of personal characteristics on the basis of information retrieved from one's memory. In D. Kahneman, P. Slovic, & A. Tversky (Eds.), Judgment under uncertainty: Heuristics and Biases (pp. 378-387). New York: Cambridge University Press.
- Tversky, A. & Kahneman, D. (1971). Belief in the law of small numbers. Psychological Bulletin, 76, 2, 105-110.
- Tversky, A. & Kahneman, D. (1973). Availability: A heuristic for judging frequency and probability. Cognitive Psychology, 5, 207-232.
- Tversky, A. & Kahneman, D. (1974). Judgment under uncertainty: Heuristics and biases. Science, 185, 1124-1131.

- Tversky, A. & Kahneman, D. (1980). Causal schemas in judgments under uncertainty. In M. Fishbein (Ed.). Progress in social psychology. Hillsdale, NJ: Erlbaum.
- Tversky, A. & Kahneman, D. (1982a). Availability: A heuristic for judging frequency and probability. In D. Kahneman, P. Slovic, & A. Tversky (Eds.), Judgment under uncertainty: Heuristics and biases (pp. 163-178). New York: Cambridge University Press.
- Tversky, A. & Kahneman, D. (1982b). Evidential impact of base rates. In D. Kahneman, P. Slovic, & A. Tversky (Eds.), Judgment under uncertainty: Heuristics and biases (pp. 190-200). New York: Cambridge University Press.
- United States Office of Education. (1977). Rules and regulations implementing Public Law 94-142 (The Education for All Handicapped Children Act of 1975). Federal Register, 42.
- Utley, B., Zigmond, N., & Strain, P., (1987). How various forms of data affect teacher analysis of student performance. Exceptional Children, 53, 411-422.
- Vance, B. & Paparella, C. (1983). The explorer [Computer program]. Novato, CA: Academic Therapy.
- Walker, N. W., & Myrick, C. C. (1985). Ethical considerations in the use of computers in psychological testing and assessment. Journal of School Psychology, 23, 51-57.

- Watkins, M. (1986). Review of WISC-R 90 clinical report. In J. Mitchell (Ed.). The ninth mental measurements yearbook (Vol. 1). Lincoln, NB: University of Nebraska Press. (p. 1354).
- Wechsler, D. (1974). Manual of the Wechsler intelligence scale for children-revised. New York: The Psychological Corporation.
- Wiese, M., & Grossman, F. (1986). The Explorer (Software reviews). Computers in human behavior, 2, 89- 91.
- Wiggins, J. (1973). Personality and prediction: Principles of personality assessment, Reading, MA: Addison-Wesley.
- Wiggins, N. & Hoffman, P. (1968). Three models of clinical judgment. Journal of Abnormal Psychology, 73, 70-77.
- World Health Organization. (1978). Mental Disorders: Glossary and guide to their classification in accordance with the ninth revision of the International Classification of Diseases. Geneva: Author.
- Ysseldyke, J., & Algozzine, B. (1981). Diagnostic classification decisions as a function of referral information. The Journal of Special Education, 15, 429-435.

- Ysseldyke, J. & Pianta, B. (1983). Psychoeducational decision making: Generalizations and implications for training and practice [Summary]. Convention '83 Proceedings National Association of School Psychologists, 253-254.
- Zinman, A. (1983). Decision making in school psychological assessment. Dissertation Abstracts International, 44, 3335A.