

## REFERENCES

and consequential structures of personality. In A. I. Rabin (Ed.), *Personality* (pp. 27-43). New York: Wiley.

vised psychological assessment. New York: Basic Books.

in a psychometric mission to clinicia. *Psychometrika*, 19, 263-270.

and beyond: Some basic issues on the prediction of behavior. 60-392.

(1988, August). *Response latency evidence for viewing personality* presented at meeting of the American Psychological Association,

*Dimensional approach to personality inventory responding*. Unpublished doctoral dissertation, University of Western Ontario, London, Canada.

(1987, August). *Reaction time and self-report psychopathological convergent and discriminant validity*. Paper presented at a meeting of the American Psychological Association, New York.

ynamics of structured personality tests. *Psychological Review*, 78, 33-52.

tial Personality Inventory types among alcoholics. In W. M. Cox (Ed.), *Assessing alcoholic personality characteristics*. *New Directions for Methodological Science* (No. 16). San Francisco: Jossey-Bass, pp. 87-100.

Vidmar, N. J. (1972). A four-dimensional interpretation of risk behavior. *Journal of Personality and Social Psychology*, 40, 483-501.

V. (1985). Construct validity and the predictability of behavior. *Journal of Social Psychology*, 49, 554-570.

(1987). Construct interpretation of Minnesota Multiphasic Personality scales: An orthogonal transformation. *Journal of Psychological Assessment*, 9, 149-160.

& Jackson, D. N. (1988). *Visual-spatial ability and psychometric validity*. Unpublished doctoral manuscript.

ests as instruments of psychological theory. *Psychological Reports*, 48, 1-10.

82). Beyond *deja vu* in searching for cross-situational consistency. 70-755.

Jackson, D. N. (1982). An application of singular value decomposition to MMPI items. *Applied Psychological Measurement*, 6, 275-283.

975). Clinical judgment of psychopathology: A model for inference. *Journal of Abnormal Psychology*, 84, 475-482.

Paunonen, S. V. (1981). Personality: Nomothetic or idiographic? A re-examination. *Psychological Review*, 88, 582-589.

tic interpretations in Rorschach testing. New York: Grune & Stratton.

ity and prediction: Principles of personality assessment. Reading, MA: Addison Wesley.

1991

In T. B. Gutkin  
& S. L. Wise (Eds.)

The computer and the  
decision-making process

# 2

## Psychodiagnostic Computing: From Interpretive Programs to Expert Systems

Marley W. Watkins

SouthWest EdPsych Services, Inc., Phoenix

Paul A. McDermott

University of Pennsylvania

As amply demonstrated by the chapters in this volume, computer applications have pervaded all aspects of psychological practice. Although thought by some to be relatively new (Nolen & Spencer, 1986), semiautomatic scoring of the Strong Vocational Interest Blank was accomplished more than 50 years ago (Campbell, 1968) and systems of computer-based test interpretation have been operational for 25 years (Fowler, 1985).

### DEVELOPMENT OF ADMINISTRATION AND INTERPRETATION PROGRAMS

Early automated programs typically focused upon the scoring or interpretation of a single psychological test. Most frequently, that test was the Minnesota Multiphasic Personality Inventory (Fowler, 1985) but the Rorschach was interpreted as well (Piotrowski, 1964). In addition to automated interpretation, there were attempts to administer existing psychological tests directly by computer. The MMPI was again the test of choice (Lushene, O'Neil, & Dunn, 1974) although the Wechsler Adult Intelligence Scale (Elwood, 1972), Slossen Intelligence Test (Hedl, O'Neil, & Hansen, 1973), Peabody Picture Vocabulary Test (Klinge & Rodziewicz, 1976), and the California Psychological Inventory (Scissions, 1976) were also administered by computer.

### Computer-administered Tests

Efforts to equate the conventional MMPI with computer-administered versions have continued unabated. White, Clements, and Fowler (1985) administered the full-length MMPI via microcomputer and standard booklet to 150 volunteer undergraduates. The two MMPI versions were generally equivalent in terms of mean scale scores, test-retest correlations, and stability of high-point codes. There was, however, a greater tendency for the computerized version to result in larger numbers of "cannot say" responses. Rozensky, Honor, Rasinski, Tovian, & Herz (1986) investigated the attitudes of psychiatric patients to computerized vs. conventional MMPI administrations. The computer group found the testing experience to be more interesting, more positive, and less anxiety-provoking than did the paper-and-pencil group. The equivalency of other conventional personality (Katz & Dalby, 1981; Lukin, Dowd, Plake, & Kraft, 1985; Skinner & Allen, 1983; Wilson, Genco, & Yager, 1986), neuropsychological (DeMita, Johnson, & Hansen, 1981), cognitive ability (Beaumont, 1981; Eller, Kaufman, & McLean, 1986), and academic (Andolina, 1982; Wise & Wise, 1987) tests to their computerized versions are also being widely explored.

The promise of parallel automated test forms has provoked investigations of the differences between computerized and conventional item presentations and their possible impact upon test reliability and validity (Hofer & Green, 1985). Jackson (1985) reviewed the evidence regarding equivalence of conventional and computerized tests and posited four methodological differences: (1) modifications in the method of presenting stimulus material; (2) differences in the task required of the examinee; (3) differences in the format for recording responses; and (4) differences in the method of interpretation. Despite these threats to equivalence, Moreland (1985) opined that "the bulk of the evidence on computer adaptions of paper-and-pencil questionnaires points to the tentative conclusion that non-equivalence is typically small enough to be of no practical consequence, if present at all" (p. 224). A more cautious note was sounded by Hofer and Green (1985). They suggested that for most computer-presented tests, "practitioners will have to use good judgment in interpreting computer-obtained scores, based on the available but inconclusive evidence" (p. 831). This conservative opinion seems well founded if automated testing is to influence the critical classification, placement, and treatment decisions made by psychologists.

### Computer-interpreted Tests

Computerized interpretation of the MMPI has remained a major line of inquiry. Honaker, Hector, and Harrell (1986) asked psychology graduate students and practicing psychologists to rate the accuracy of interpretative reports for the MMPI that were labeled as generated by either a computer or licensed psychologist. Their results demonstrated similar accuracy ratings for computer-generated

and clinician-generated reports and did not support the claim that computer-generated reports are assigned more credibility than is warranted. Butcher (1987) reviewed early MMPI systems, summarized desirable attributes of automated systems, and described the development and use of the *Minnesota Clinical Interpretive Report* (University of Minnesota Press, 1982) computerized MMPI interpretive system. Limited attention has been given to automated interpretations of other personality tests (Exner, 1987; Greene, Martin, Bennett, & Shaw, 1981; Harris, Niedner, Feldman, Fink, & Johnston, 1981; Lachar, 1984), neuropsychological measures (Adams & Heaton, 1985; Adams, Kvale, & Keegan, 1984), and ability and achievement instruments (Brantley, 1986; Hasselbring & Crossland, 1981; Johnson, Willis, & Danley, 1982; Oosterhof & Salisbury, 1985; Webb, Herman, & Cabello, 1986).

As noted by Moreland (1985), investigations of the accuracy of computer-based clinical interpretations of personality tests have been limited almost exclusively to the MMPI. A thorough review of the types of MMPI validity studies, computer interpretation systems, and outcomes are presented by Moreland (1987). He summarized these findings by concluding:

Things look pretty good for computer-based MMPI interpretations. Consumers give them high marks, and the results of properly controlled studies indicate that this high acceptance rate is not the result of generalized reports that are equally applicable to most clients. (p. 43)

In contrast, Matarazzo (1985) noted that currently available automated interpretation systems are erected upon rather tenuous empirical bases and involve varying degrees of clinical and actuarial data accumulation and interpretation which have considerable potential for harm if used in isolation. These disparate views can be reconciled by Butcher's (1987) assertion that the computerized report should be used "only in conjunction with clinical information obtained from other sources" (p. 167).

### Current Status

There has been much controversy surrounding computerized test administration and interpretation. Sampson (1983) enumerated and reviewed the potential benefits of such systems: namely, (a) better client response to the testing situation, (b) cost-effectiveness, (c) ability of the computer to do interactive testing, (d) generation of standardization data, (e) more efficient use of staff time, (f) more efficient scoring, (g) reduced error rates in scoring and administration, (h) validity of interpretation of results, and (i) potential assistance to persons with visual or auditory handicaps. Arguments against the concept of computerized assessment have been compiled by Sampson (1983) and Space (1981). Possible problems include: (a) depersonalization of the client, (b) idiographic information lost

in favor of nomothetic information, (c) poor interface between person and machine, (d) loss of efficiency with difficult clients, (e) confidentiality of client information may be at risk, (f) inability to discriminate between normal error and pathological response, and (g) introduction of bias into the testing situation. Matarazzo (1983, 1985, 1986) has been most outspoken about computerized psychological testing, arguing that automated psychological test interpretations offer considerable potential for the future, but currently fail to meet even minimal validation standards.

It is apparent from the foregoing discussion that there is no professional consensus regarding computerized administration and interpretation of psychological tests. However, comprehensive reviews of the literature and thoughtful analyses are presented by Space (1981), Fowler (1985), Hofer and Green (1985), as well as by the authors represented in this volume. Moreover, the American Psychological Association's guidelines (APA, 1986) for computer-based tests and interpretations summarize pertinent ethical, professional, and technical standards relevant to this issue.

### NOVEL ADMINISTRATION AND INTERPRETATION PROGRAMS

As observed by Hofer and Green (1985), early applications of technology in any field tend to be derivative. For example, the first automobiles were simply attempts to duplicate traditional horse-drawn carriages, pioneer television broadcasts mimicked familiar radio formats, and the first computers were used to cross-check mechanically the counts of interview cards collected by U.S. census takers. The application of computer technology to psychology is no exception. At present, computerized assessment is primarily devoted to a literal translation of existing paper-and-pencil tests or interpretive systems to the computer without modifications to take advantage of the computer's unique features. As in other technologies, psychological assessment will make revolutionary advances when novel, creative applications are computerized; not when existing applications are slavishly re-created on the computer.

#### Computer-administered Tests

*Item Types.* New types of test items can capitalize on the strengths of the computer and thereby contribute to novel and informative assessment techniques. The computer can readily capture reaction times of examinees and can present test items that involve movement, color, speech, sound, and interactive graphics. These possibilities are just beginning to be explored. For example, Jones, Dunlap, and Bilodeau (1987) utilized video games to establish dimensions of individual differences in cognitive and perceptual functioning. These comput-

erized video games contained variance not captured by conventional paper-and-pencil cognitive tests. Colby and Parkison (1985) described an innovative program which converts natural language expressions into conceptual patterns and key ideas to produce a taxonomy of neurotic patients.

Technological advances in computer hardware have made possible much more realistic graphics and sound than were exploited by Jones et al. (1987) or by Colby and Parkison (1985). Videodisk and compact digital disk developments offer interactivity with television quality visuals, digital sound, and print quality graphics (Gonsalves, 1987). With such capabilities, it might be possible to tap examinees' reactions to social situations by placing them in a simulated, but realistic, context and monitoring their character's verbalizations and movements. Vocabulary knowledge could be evaluated by providing an interactive dictionary and monitoring examinees' usage. Alternately, free responses by examinees could be compared word by word with massive tables of word frequencies. Parents and teachers could rate child behaviors by creating characters via screen animation rather than relying, as is now necessary, on written item descriptions. The advantages of using computer technology to assess human abilities, attributes, and skills in novel ways are almost unlimited and await only the development of well-researched and imaginatively implemented methods.

*Test Types.* Irrespective of types of items involved, psychological assessment must move away from the linear, fixed-item presentations necessitated by paper-and-pencil formats. With traditional tests, all examinees typically respond to the same test items. Each examinee receives items that are too easy and items that are too difficult. If test items are too difficult, an examinee might resort to random guessing or omission of responses. Easy items may dampen motivation. Conventional testing technology thereby entails a restricted range of accuracy for nonaverage examinees. Although capable of expediting the test scoring and test interpretation process, a computerized copy of conventional methods provides neither improved efficiency nor advanced psychometric properties (Weiss & Yale, 1987).

What is required is a type of test that capitalizes on the capabilities of the computer to improve test efficiency and accuracy. Such a test methodology was developed independent of computer technology, but its adaptability to computerization was immediately recognized (Weiss, 1985). Labeled *adaptive* testing, the computer presents the items to the examinee, receives and scores the item responses, chooses the next item to administer, based on the examinee's prior performance, and terminates the test when appropriate. Unlike conventional tests, adaptive test items are selected *during* rather than *before* administration. By doing so, each test item can be optimally useful for measuring each individual examinee (McKinley & Reckase, 1980).

Research on computerized adaptive testing has revealed that it is more precise and efficient than conventional testing (Weiss, 1987). As a consequence, average

test length can be reduced about 50% without compromising measurement quality (Weiss & Vale, 1987). Computerized adaptive testing has in the past been predominately restricted to academic and ability tests (Sands & Gade, 1983; Watkins & Kush, 1988). Its application to personality testing (Jackson, 1985) and to diagnostic interviews (Stein, 1987) has been described, and its utility in other areas of psychological testing has recently been speculated upon by Krug (1987). Adaptive testing, particularly when combined with novel test items, could result in dramatic improvements in the efficiency, accuracy, and relevance of psychological assessment.

### COMPUTERIZED INTERPRETATION SYSTEMS

#### Expert Systems

Computer software, like hardware, is a rapidly emerging technology. In recent years the development of artificial intelligence (AI) software has received much attention. That is, attempts to make computers exhibit, or at least simulate, different aspects of intelligent behavior. Perhaps the most popular and well-known example of AI is computerized chess. Once thought to be incapable of more than rudimentary play, chess programs have evolved to a point where they can now beat all but the best human players (Krutch, 1986).

Probably the "hottest" topic in AI is expert systems (Chadwick & Hannah, 1987). *Expert systems* are computer programs designed to reason as would most expert humans. Although still uncommon in psychology, expert system applications are relatively well established and highly publicized in medicine, economics, chemistry, geological exploration, aeronautics, and other scientific, human service, and industrial areas (Buchanan, 1985).

There is no single, universally accepted definition of an expert system. Chadwick and Hannah (1987) indicated that an expert system "is a computer program that simulates the reasoning of a human expert in a certain domain. To do this, it uses a knowledge base containing facts and heuristics, and some inference procedure for utilizing its knowledge" (p. 3). Krutch (1986) indicated that "An expert system can be described as an intelligent database that can make decisions, give advice, and come to important conclusions" (p. 3). In addition to definitions, many authors specify a number of attributes which they consider to be essential characteristics of an expert system (Buchanan, 1985).

Computerized psychological assessment systems are in their infancy and whether or not an existing application is an expert system will be widely debated (Roid, 1986). Deupree (1985) reviewed existing software and opined that WISC-R analysis programs are fundamental AI applications. It is doubtful that Waterman (1986) would agree, given that author's extensive definitional criteria and estimate of 6 person-years required to develop even a moderately difficult expert system.

#### A New Model

It seems pointless to become entangled in a definitional quagmire concerning expert psychological systems. Rather, psychologists must focus their attention on the underlying knowledge base of any computerized system. That is, after all, the area in which psychologists are expert. To this end, a two-dimensional framework is offered as a model for analysis and production of computerized psychological assessment systems. The first dimension, *scope*, refers to the scope or breadth of knowledge covered by the system. A continuous concept, scope may range from narrow to broad. The second dimension, *authority*, represents the consensus of experts regarding the verity of the underlying "knowledge" used by the program. To use a more familiar term, authority could be equated to validity and might span from low to high along its own continuum. It is possible to simplify this two-dimensional continuous model by collapsing it into four cells; that is, narrow scope with low authority, narrow scope with high authority, broad scope with low authority, and broad scope with high authority. This simplification is depicted in Fig. 2.1. Real computer systems would, of course, rarely be so well delineated or easily classified. Nevertheless, it is clear that high authority is a prerequisite to utility, irrespective of the scope of knowledge incorporated in an expert system.

*Narrow Scope and Low Authority.* For an example, consider an intelligence test interpretation program which bases its expertise on Glasser and Zimmerman's (1967) *Clinical Interpretations of the Wechsler Intelligence Scale for*

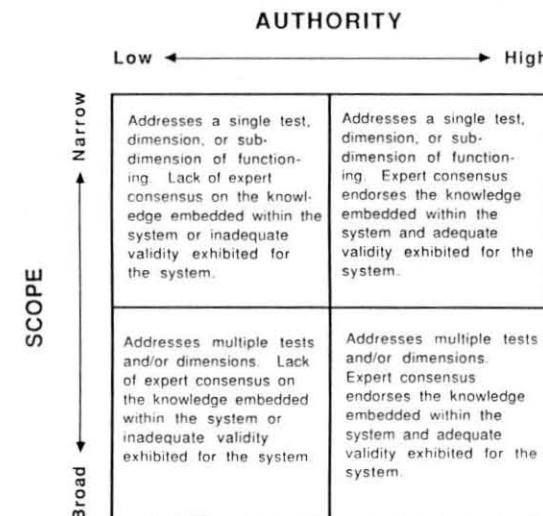


FIG. 2.1. Framework for analysis of computerized psychological assessment systems.

*Children.* Such an application necessarily would be considered of narrow scope, given its coverage of only one aspect of human functioning—intelligence. On the authority dimension, such a program's conclusions would be refuted strongly by many experts who demonstrate empirically that profile and scatter analysis of the WISC is not defensible (Kavale & Furness, 1984) and has the potential for doing more harm than good (Kramer, Henning-Stout, Ullman, & Schellenberg, 1987). Alternatively, it is quite possible for a program having very narrow scope to proceed with high authority; as, for instance, the letter capitalization program described by Watkins and Kush (1988).

A review of recent publications dealing with computerized psychological assessment (Butcher, 1987; Fowler, 1985; Jackson, 1985) reveals that most current applications are relatively narrow in scope. Even so, newer computer applications tend to rest on greater authority and should yield improved efficiency and accuracy for psychological assessment.

*Broad Scope and High Authority.* It is intuitively apparent that development of computerized psychological assessment systems with broad scope and high authority entails problems of a different nature and magnitude than those encountered during scoring or interpretation of an individual psychological test. Before attacking these problems, it would be instructive to determine if expert system developers in other disciplines have encountered similar difficulties and, if so, consider how they have dealt with them.

Perhaps medicine is the most logical field for comparison because, like professional psychology, it encompasses a vast array of human-care activities, many guided by available empirical knowledge but many more still remnants of traditional thinking and popular convention. Expert medical systems have been in use for years and efforts to develop broadly useful systems have been undertaken by several experimenters (Buchanan, 1985). It was recognized at an early stage that computer programs were more successful in narrow, constrained arenas of medicine where much hard laboratory knowledge existed, largely because expert systems which produced complicated decisions involving multiple diseases were confronted by problems of inadequate consensus concerning the underlying knowledge base (Schoolman & Bernstein, 1978). Similar problems have been noted in psychiatry, where limitations in validity of the diagnostic system itself arose as barriers to computerized expertise (Spitzer & Fleiss, 1974). This problem surfaced in many other expert system applications (Bhatnagar & Kanal, 1986) and may be described formally as *reasoning with uncertainty* or (inasmuch as empirical inquiry in the behavioral sciences never substantiates absolute truth) *reasoning with unknown certainty*.

There are striking similarities across disciplines when solutions to the uncertainty problem are reviewed. Szolovits and Pauker (1978) suggested that an expert medical system would have to use a judicious combination of categorical and probabilistic reasoning. In psychiatry, Erdman, Greist, Klein, Jefferson, and

Getto (1981) recommended a combination of statistical and clinical judgment. Bhatnagar and Kanal (1986) concluded that the management of uncertainty in automated decision making required application of numerical methods, such as probability theory, within the framework of logic.

## A PSYCHOEDUCATIONAL DIAGNOSTIC MODEL

The process of identifying, classifying, and programming for childhood developmental, social, and learning difficulties is nontrivial and realistically can be deemed broad in scope. It can be argued further and without contradiction that the existing psychoeducational diagnostic knowledge base is marked by considerable uncertainty. In fact, McDermott (1986) has characterized conventional methods of child diagnosis and classification as woefully inadequate.

On the surface, then, a computerized system for applying psychoeducational diagnostic expertise to childhood disorders seems untenable. The domain is too broad, is marked by a lack of professional consensus, and requires extensive reasoning with uncertainty. Nonetheless, the problems presented by psychoeducational diagnosis closely parallel those encountered during the development of expert systems in other disciplines and may be amenable to similar resolutions.

### Diagnostic Reliability

Meehl's (1954) seminal book on clinical and statistical classification was instrumental in sensitizing psychologists to potential reliability and validity limitations in psychodiagnostic practice. Evaluation research over the intervening years has demonstrated repeatedly that psychiatrists and psychologists are unable to render reliable psychological diagnoses (Algozzine & Ysseldyke, 1981; Cantwell, Russell, Mattison, & Will, 1979; Epps, Ysseldyke, & McGue, 1984; Freeman, 1971). Typically, agreement among child specialists has been found to be more commensurate with guesswork or unskilled decision making. For example, McDermott (1980b) observed near-chance levels of agreement among experienced psychologists' diagnoses, while Visonhaler, Weinshank, Wagner, and Polin (1983) found that single clinicians diagnosing the same cases twice achieved only 0.20 mean diagnostic agreement with themselves. The ramifications of such poor diagnostic agreement are profound because unreliable diagnoses must, by definition, be invalid (Spitzer & Fleiss, 1974).

### Diagnostic Error

The factors contributing to classificatory incongruity are many, complex, and incompletely understood (McDermott, 1982). Nevertheless, they may be viewed

conceptually as falling under two broad categories: inconstancy in human information processing and judgment and faults in diagnostic decision-making rules.

**Human Error.** There is often a considerable amount of disagreement among observers and judges even when they observe relatively concrete events. Thus, Koran (1975) revealed that physicians often disagreed, concerning even relatively quantifiable tasks, in one out of five instances. And so it would follow that judgments rendered under more nebulous and less-quantifiable circumstances (as so often "psychological" contexts would seem to appear) are likely to be very unreliable.

One limiting factor which may contribute to classificatory unreliability is the tendency for diagnoses to be negatively biased by client characteristics. Social class (DiNardo, 1975), gender (Broverman, Broverman, Clarkson, Rosenkrantz, & Vogel, 1970), and race (Franks, 1971) have, among other client attributes, been found to influence classification decisions. Diagnostic constancy also has been found inversely related to the information-processing load (Lueger & Petzel, 1979) and to the amount of direct probabilistic analysis required (Eddy & Clanton, 1982; Kahneman & Tversky, 1982). Sources of human error in judgment and diagnosis have been analyzed by Arkes (1981) and McDermott (1981). Judgmental impediments summarized by these authors include: (a) inconsistent theoretical orientation, (b) inability to assess covariation accurately, (c) influence of preconceived notions or expectancies, (d) minimal awareness of one's own judgment process, (e) overconfidence, (f) hindsight bias, (g) preference for unverifiable or inexclusive diagnoses, (h) inconstancy of diagnostic style, and (i) preference for a determinative diagnostic posture (i.e., the practice of responding to uncertainty by rendering rather than deferring decisions).

**Decision Rule Error.** Historically, there have been two general approaches to classification of psychoeducational disorders: clinical and actuarial. Both strategies afford important advantages as well as specific weaknesses. Quay (1986) comprehensively reviewed the foundation, development, and application of clinical diagnostic strategies. In brief, clinical methods evolved from observations by clinicians working with patients. Typically, clinicians noted the covariance of certain characteristics and determined through consensual validation that such constellations of phenomena should constitute unique diagnostic categories. Thereafter, groups of such categories were interrelated to comprise a complete clinical classification system. Examples of existing clinical systems include the American Psychiatric Association's revised *Diagnostic and Statistical Manual of Mental Disorders* (DSM-III-R; 1987) and the World Health Organization's ninth edition of the *International Classification of Diseases* (ICD-9; 1978).

Clinical decision rules are based largely on popular theory and accepted practice and are dependent on the individual psychologist for interpretation. They offer a wealth of useful constructs and recorded case experience but are

heavily reliant upon competent human judgment in weighing the elements of any specific case. Ironically, reliance on human judgment represents both the major strength and the major weakness of the clinical approach. On the positive side, humans may be more likely to identify isolated and unusual characteristics, behaviors, and patterns of behavior. However, as seemingly unique characteristics compound and become confused with the greater universe of natural human variation, dependence on clinical judgment invariably increases error.

Actuarial strategies, although often grounded in conventional theory, were derived from controlled studies of incidence and prevalence of normality and abnormality in representative general populations (McDermott, 1982). Classifications were developed by defining distinctly similar and reliable patterns of functioning, and assignment criteria were presented in the form of statistical decision rules. Individual psychologists do not interpret the decision rules because it is a straightforward matter of assigning classifications that are statistically probable and discarding those that are improbable.

Given their objective foundations and implementation, actuarial decision rules are quite reliable and control for many of the sources of human decision error that plague clinical diagnosis. Actuarial methods are limited, however, by the necessity for sound and comprehensive data concerning the characteristics of patient populations and by a general lack of the technical resources required for implementation of complex statistical decision rules.

### Minimizing Diagnostic Error

Arkes (1981) proffered three major suggestions for improvement of the accuracy and reliability of human judgment: consider alternatives, use statistical principles, and decrease reliance on memory. It is readily apparent that actuarial assessment approaches and empirical decision rules would allow the clinician to utilize statistical principles and thereby decrease diagnostic error. On the other hand, good actuarial information is frequently unavailable. Consequently, it is reasonable to regard clinical and actuarial processes as complementary, each mitigating the other's inherent weaknesses. This combination of statistical and clinical principles to improve reasoning in an uncertain domain emulates resolutions emanating from leading expert systems research (Bhatnagar & Kanal, 1986; Erdman et al., 1981; Szolovits & Pauker, 1978). Effective utilization of actuarial strategies can be facilitated by computers, which can rapidly and accurately calculate and apply a host of complicated statistical decision rules. Consideration of alternatives may be promoted by the adoption of a systematic decision process; that is, a process that capitalizes on modern decision theory (Dailey, 1971) and systems analysis (Nathan, 1967) to ensure logical sequencing and efficiency. Computerization can ensure the prompt and precise application of pertinent systems logic and guide the process so as to reduce substantially the demands made upon the clinician's memory.

## A COMPUTERIZED PSYCHOEDUCATIONAL DIAGNOSTIC SYSTEM

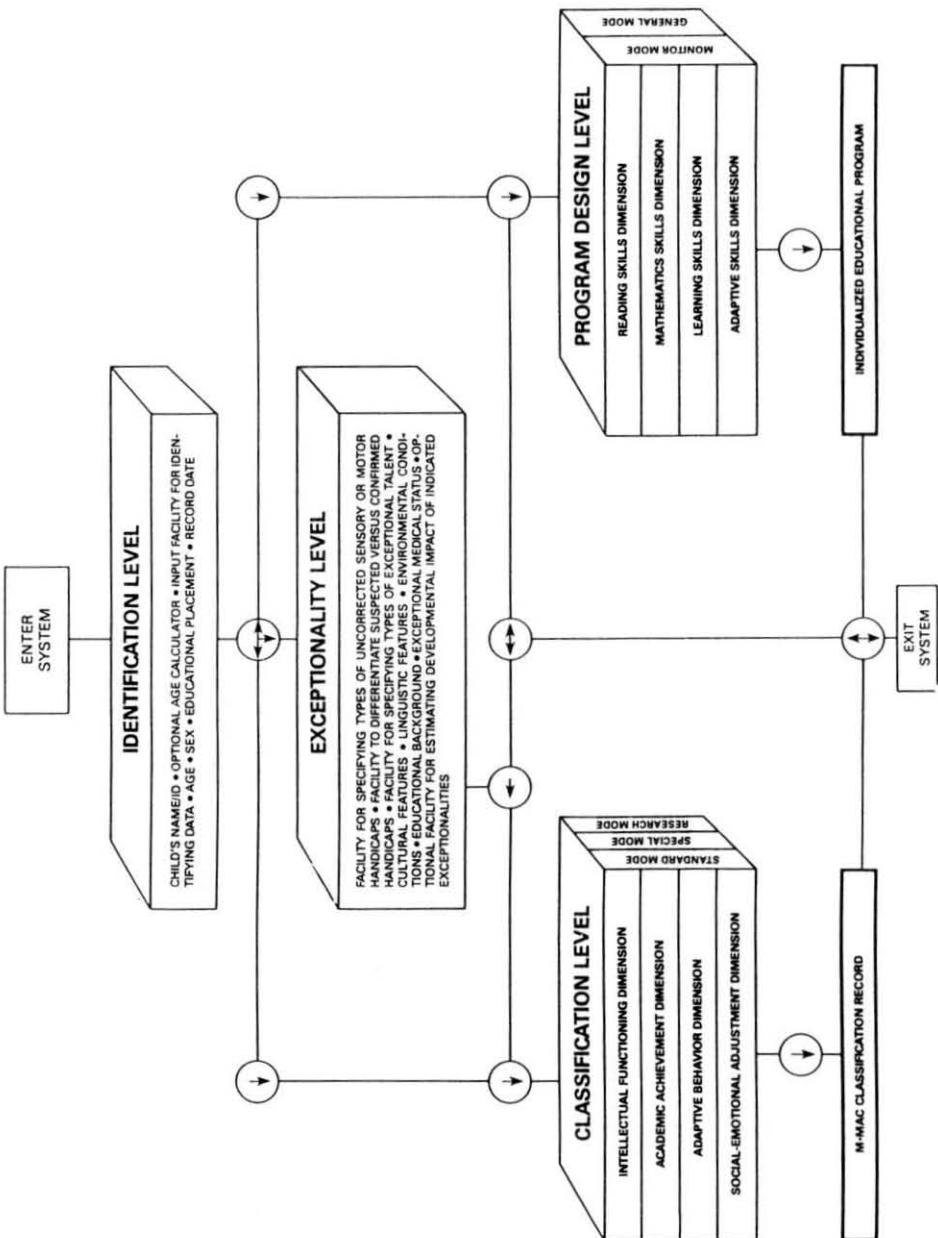
From the foregoing discussion, it is apparent that an efficient computerized diagnostic expert system should embody both clinical and actuarial methods and should implement each when optimally appropriate. Moreover, it should employ a systematic decision process to maximize consistency and reliability and thereby enhance authority. It should address multiple sources of diagnostic data (tests, demography, unusual characteristics, etc.) and dimensions of human functioning (intellectual, social, physical) to gain broad scope. The prototype of such a system was introduced by McDermott (1980a) for the diagnosis of childhood disorders. The system was described in considerable detail (McDermott, 1980c) and validated with a large group of children (Hale & McDermott, 1984; McDermott & Hale, 1982). Subsequently, its capabilities were extended and it was made operational on microcomputers (McDermott & Watkins, 1985, 1987). The remainder of this chapter is devoted to a description of that expert system.

The *McDermott Multidimensional Assessment of Children* (M.MAC) is a comprehensive microcomputer system for use by psychologists and other child specialists in assessing the psychological and educational functioning of children 2 through 18 years old. It produces objective classifications of childhood normality and exceptionality and designs instructional programs based upon actual performance in fundamental educational areas. An overview of the M.MAC system's structure and organization is presented in Fig. 2.2.

### Identification

The first component encountered in operation of the M.MAC system is the Identification Level. This preparatory stage entails collection and compilation of basic demographic information about the child, including age, grade, sex, and educational placement. This information allows the program to retrieve appropriate data (i.e., population means, standard deviations, reliability and validity coefficients, prevalence rates, etc.) from its memory for use in later levels of the system. There are almost 10,000 discrete units of statistical data stored within the M.MAC system, which are accessed by age, grade, and gender. Accurate child demographic identification is, therefore, essential for precise application of actuarial rules. Identification information also serves the traditional function of allowing the system to refer to the child by name in reports and to tailor gender references properly.

FIG. 2.2. Structure of the M.MAC system. From the microcomputer systems manual for *McDermott Multidimensional Assessment of Children*, P. A. McDermott and M. W. Watkins, 1985, 1987. New York: Psychological Corporation. Copyright (1985, 1987) by Psychological Corporation. Reproduced by permission. All rights reserved.



### Exceptionality

As denoted by the flow chart in Fig. 2.2, the Exceptionality Level is an optional component of a case study. Its purpose is to allow the classification process to consider unusual personal features of the child or the child's environment that might affect diagnosis. The psychologist informs the M.MAC system about sensory and physical handicaps, special language and cultural features, health problems, environmental stress, and educational disadvantage. The examiner also characterizes, based upon medical records and best clinical judgment, each factor as either confirmed or suspected.

Confirmed or suspected exceptional conditions can produce a variety of consequences in later M.MAC analyses. Each exceptional factor is regarded as a possible threat to the validity of formal assessment and each is systematically analyzed for its potential impact. In cases where exceptionailities are determined to be indirect threats to validity, the M.MAC system produces cautionary notices and may append a "provisional" label to a diagnosis which could be secondary to identified exceptional factors. An exceptionality which constitutes direct interference with a child's performance results in alteration of decision rules in subsequent classificatory analyses. As a simple example, confirmed vision impairment evokes alterations in use of the WISC-R performance IQ score. Furthermore, the exceptionality level permits the psychologist to identify talents and evaluate the extent to which a child has coped with exceptional circumstances.

### Classification

Classification is based upon four principal dimensions of child functioning: intellectual functioning, academic achievement, adaptive behavior and social-emotional adjustment. When proceeding through the successive classification dimensions, the psychologist may select from 24 separate assessment instruments and methods. These are listed in Table 2.1. Scores obtained from these devices are entered into the M.MAC system and processed in relation to normative statistics and child population characteristics (major *actuarial* components of the system's knowledge base).

As detailed in Fig. 2.3, a wide variety of analyses are performed within and across dimensions. There are commonalities among all data entry formats and analyses across classification dimensions that contribute to ease of use and functionality. Standardized instruments used for data collection in each dimension supply a bewildering array of scores. Many instruments naturally provide standard scores based upon a mean of 100 and standard deviation of 15, but some scores are based upon a mean of 100 and standard deviation of 16. Other instruments use standard scores with a mean of 50 and standard deviation of 10, whereas many scales provide only raw scores. To reduce confusion, M.MAC automatically calculates standard scores for instruments that report only raw scores and then applies the mixed categorical-dimensional approach to classifica-

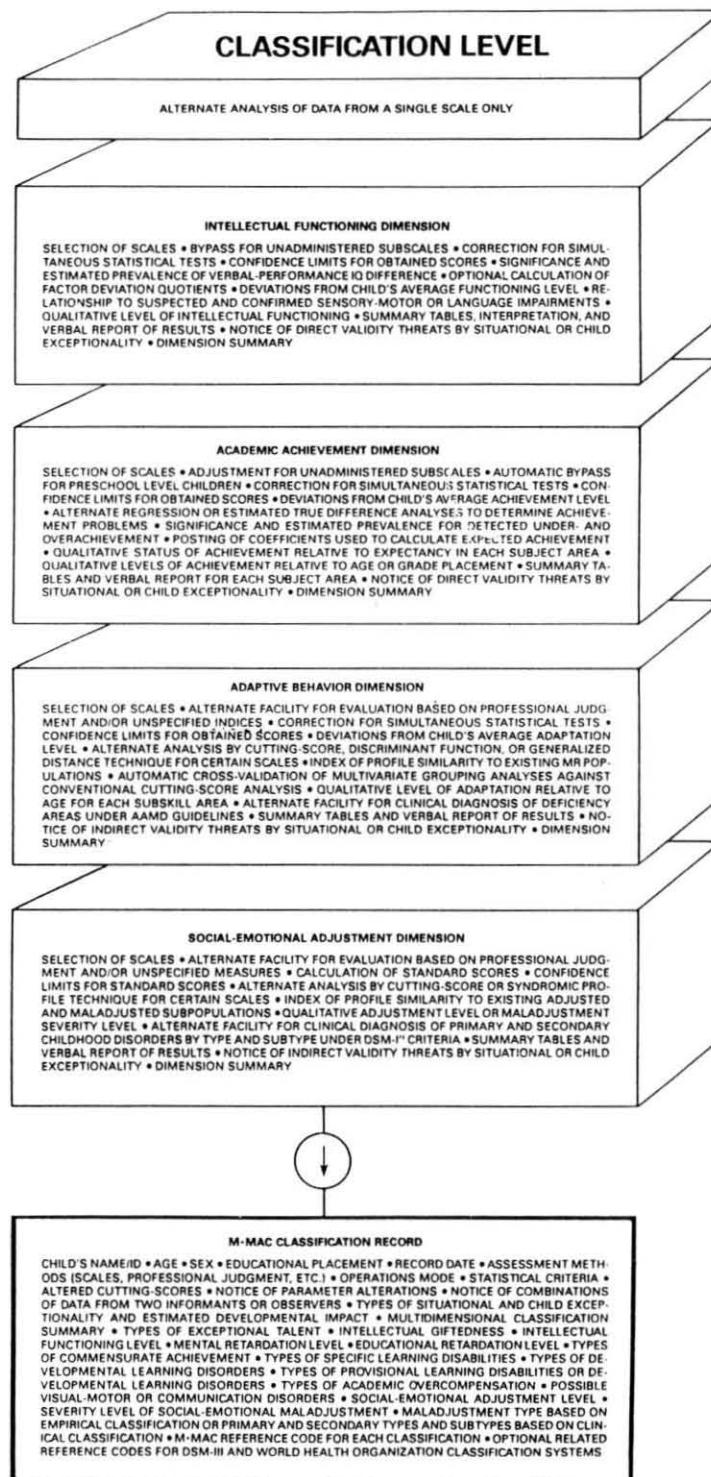
TABLE 2.1  
Assessment Scales and Methods Supported by the Four M.MAC Classification Dimensions

<b>INTELLECTUAL FUNCTIONING</b>
Wechsler Intelligence Scale for Children-Revised
Wechsler Preschool and Primary Scale of Intelligence
Wechsler Adult Intelligence Scale-Revised
Stanford-Binet Intelligence Scale
McCarthy Scales of Children's Abilities
Peabody Picture Vocabulary Test-Revised
<b>ACADEMIC ACHIEVEMENT</b>
Basic Achievement Skills Individual Screener
Peabody Individual Achievement Test
Woodcock-Johnson Tests of Achievement
Woodcock Reading Mastery Test
KeyMath Diagnostic Arithmetic Test
Wide Range Achievement Test-Revised
<b>ADAPTIVE BEHAVIOR</b>
Adaptive Behavior Inventory for Children
AAMD Adaptive Behavior Scale-School Edition
Vineland Adaptive Behavior Scales
Vineland Social Maturity Scale-Revised
Professional judgment/Other indices (AAMD guidelines)
<b>SOCIAL-EMOTIONAL ADJUSTMENT</b>
Bristol Social Adjustment Guides
Connors Teacher Rating Scale
Kohn Problem Checklist and Social Competence Scale
Louisville Behavior Checklist
Revised Behavior Problem Checklist
Professional judgment/Other indices (DSM-III criteria)

tion advocated by Cromwell, Blashfield, and Strauss (1975), whereby underlying standard score ranges are associated with terminology that describes comparable levels of functioning.

Another common classification feature is application of only those test scales and subscales for which construct validity has been demonstrated through factor- or cluster-analytical research. The only exception to this general rule is within the academic achievement dimension, where reliance on factorial constructs not recognized by school and society would create unnecessary confusion. The Peabody Individual Achievement Test (PIAT) provides a good example of this exception to the general rule. The PIAT measures and reports scores for five widely accepted academic areas (Mathematics, Reading Recognition, Reading Comprehension, Spelling, and General Information) but has been found by Wikoff (1978) to contain only two factors. Utilizing empirically derived factor scores in such a case would not foster clear communication with teachers, parents, or students.

Derived standard scores are reported across all dimensions, along with upper and lower score limits based upon confidence in reliability. Within an area of functioning, the deviation of each subarea from a child's own average level is analyzed (Davis, 1959) and the increased risk of error associated with multiple statistical comparisons is automatically controlled through Bonferroni correc-



tions (Silverstein, 1982). Additionally, reports of statistical significance are supplemented, whenever possible, by actuarial knowledge of prevalence; that is, the percentage of children in the general population showing deviations as serious as currently being manifested (Silverstein, 1981a, 1981b).

Beyond these commonalities, the classification level can be operated in one of three separate modes: Standard, Special, or Research. The mode chosen is dependent on the flexibility required by the psychologist. Each mode enables the examiner to select appropriate actuarial information, adjust classificatory criteria for special circumstances, or alter data bases of actuarial information. Functions and features of these operational modes are summarized in Fig. 2.4.

The Standard mode is automatically chosen by the M-MAC system unless the examiner specifies otherwise. This mode applies general population norms, conventional cutting-scores, standard prevalence levels, and conventional probability test levels. Operation under the Standard mode is recommended by the authors (McDermott & Watkins, 1985, 1987), unless exceptional circumstances intervene, because it guarantees a *reference standard* for assessment, thereby lending comparability to decisions across psychologists, agencies, and regions. The Special mode is intended for special needs arising in regular practice while the Research mode is reserved for applied research and needs not arising in everyday practice. Further detailed descriptions and applications of M-MAC's operational modes are provided by Glutting (1986a), McDermott (1990), and McDermott and Watkins (1985, 1987).

The M-MAC system produces 113 empirical and 35 clinical classifications. For a given child, at least one or as many as four classifications are rendered for each dimension. Each classification may be accompanied by values specifying qualitative level of functioning (e.g., mild, adequate, etc.) and by specific subtype designations (e.g., attention deficit disorder with hyperactivity, without hyperactivity, etc.). In addition, psychologists may elect to have DSM-III and ICD-9 codes accompany each M-MAC classification.

Although a complete discussion of all M-MAC classification features and logic is beyond the scope of this chapter, several examples are provided to demonstrate the multidimensional nature of diagnoses and complex interplay of clinical and actuarial methods. Fig. 2.5 illustrates the basic logic for differential classification of cognitive functioning. Review of this figure reveals that the M-MAC system first examines the child's intellectual functioning in relation to the prespecified *mild mental retardation cutting-score* value. In Standard Mode, this value is set at two standard deviations below the mean, in congruence with accepted diagnostic standards (Grossman, 1977). Based upon this rule, an obtained IQ equal to or greater than the cutting-score value precludes the classifica-

FIG. 2.3. Classification-level system. From the microcomputer systems manual for McDermott Multidimensional Assessment of Children, P. A. McDermott and M. W. Watkins, 1985, 1987. New York: Psychological Corporation. Copyright (1985, 1987) by Psychological Corporation. Reproduced by permission. All rights reserved.

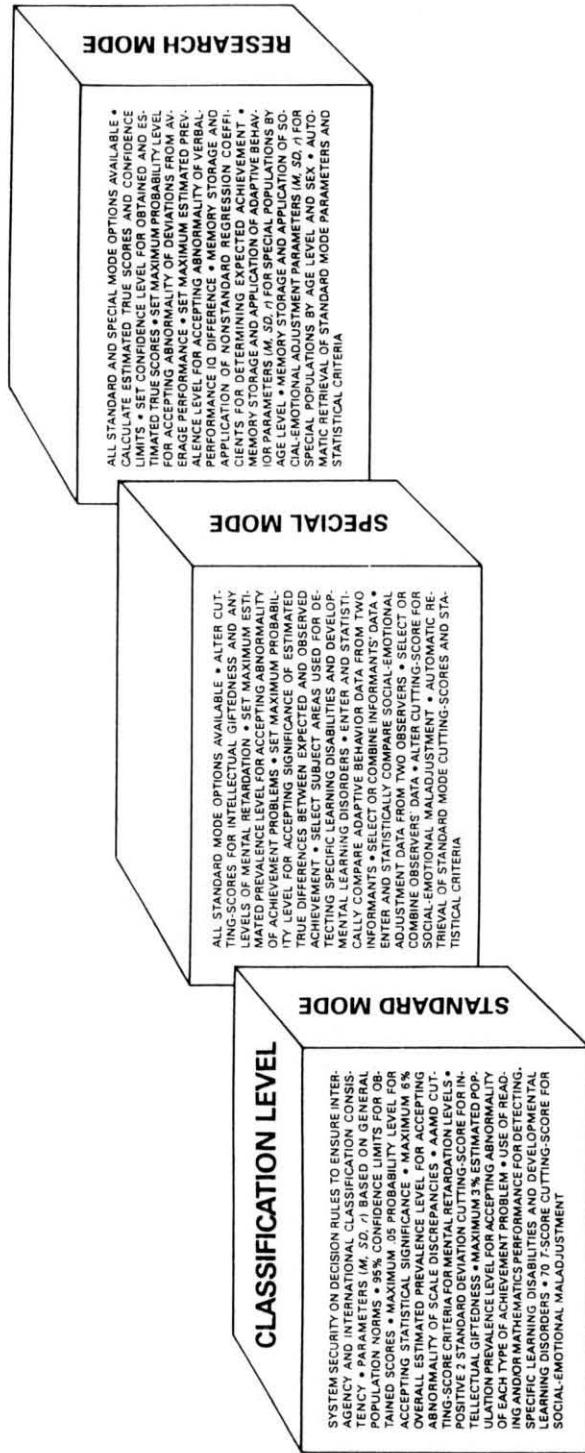


FIG. 2.4. Operation modes of the classification level. From the microcomputer systems manual for McDermott Multidimensional Assessment of Children, P. A. McDermott and M. W. Watkins, 1985, 1987. New York: Psychological Corporation. Copyright (1985, 1987) by Psychological Corporation. Reproduced by permission. All rights reserved.

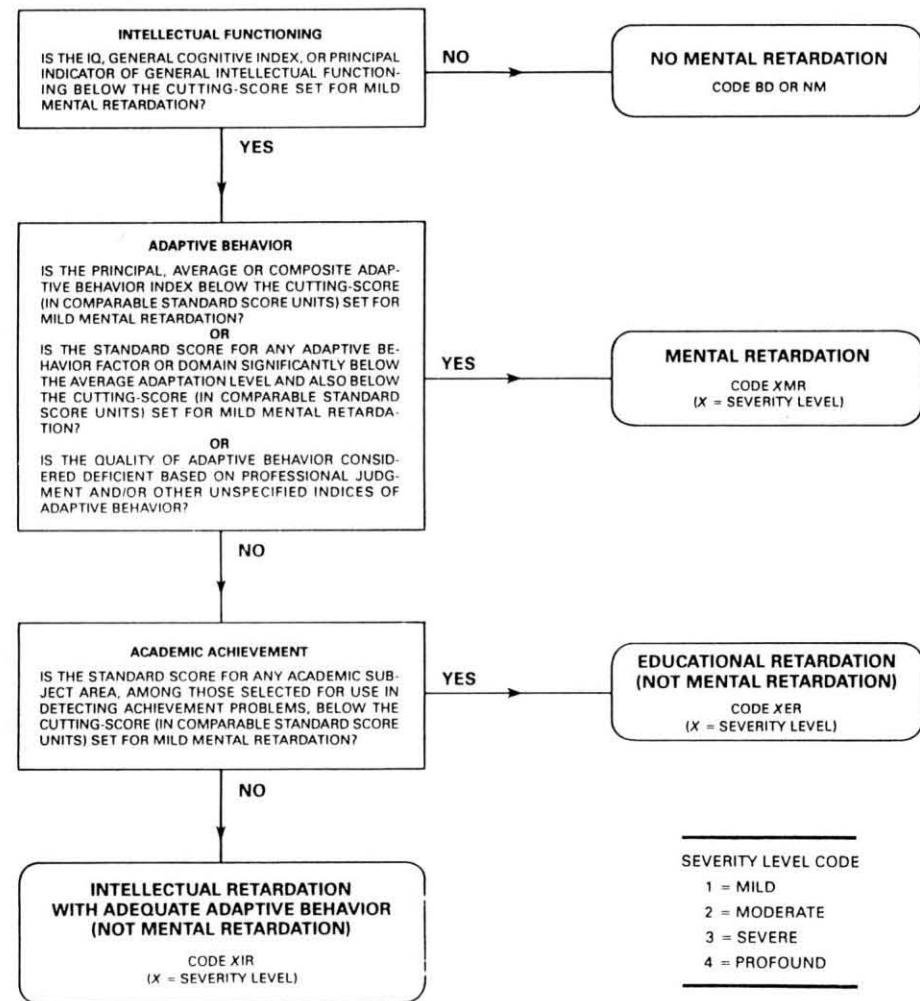
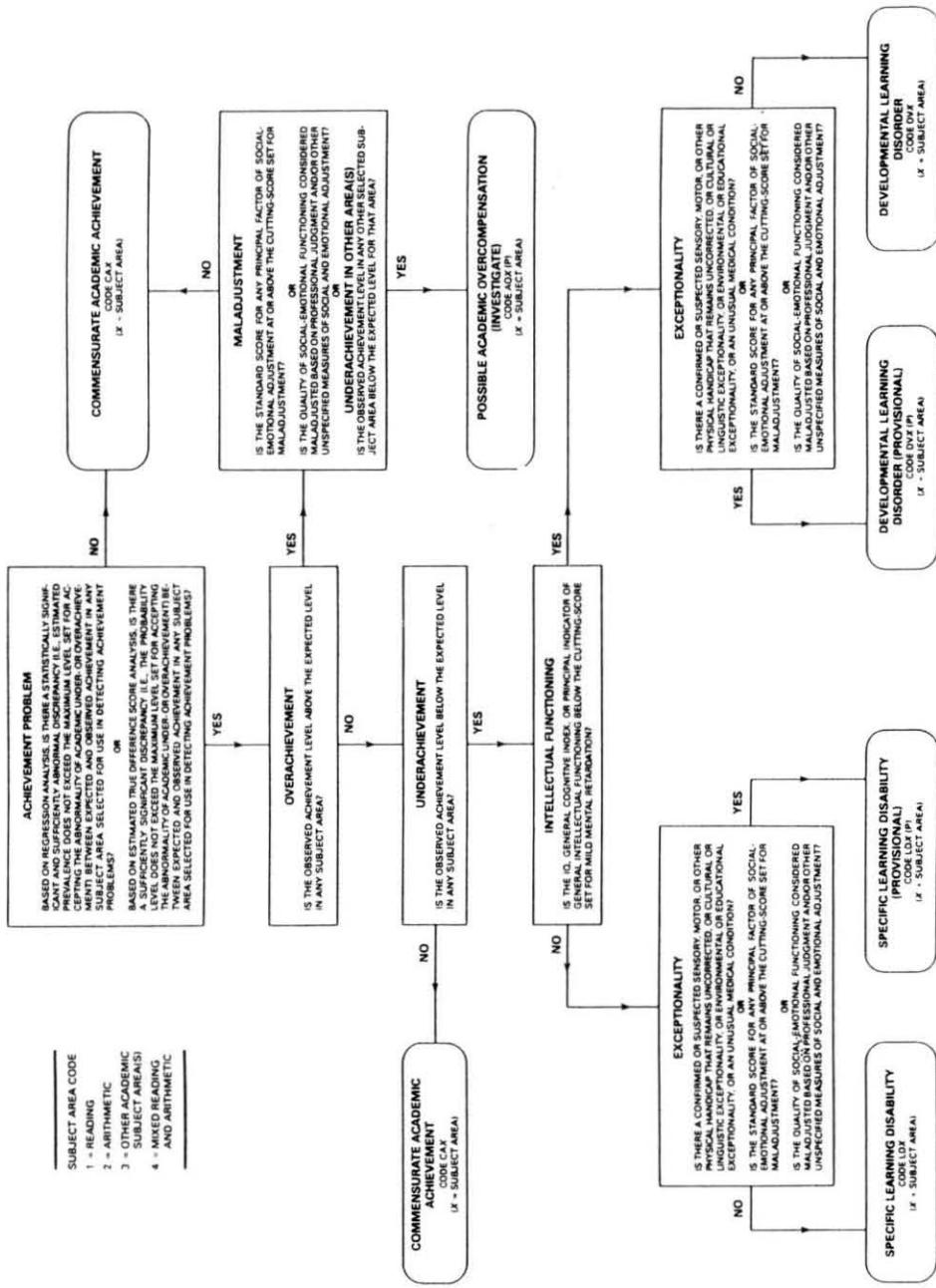


FIG. 2.5. M.MAC systems-actuarial logic for classification of intellectual proficiency and retardation. From the microcomputer systems manual for McDermott Multidimensional Assessment of Children, P. A. McDermott and M. W. Watkins, 1985, 1987. New York: Psychological Corporation. Copyright (1985, 1987) by Psychological Corporation. Reproduced by permission. All rights reserved.

tion of mental retardation. An obtained IQ lower than the cutting-score value invites consideration, sequentially, of adaptive behavior and academic achievement. Adaptive behavior may be determined empirically or clinically, but must be considered deficient by one of these two methods to result in a mental retardation diagnosis (Grossman, 1983; APA, 1987).

Differential classification of academic functioning is modeled in Fig. 2.6. For



each subject area considered, achievement is approached from three perspectives: qualitatively compared with other children of like age or grade, deviation of subareas from the child's average level of academic performance, and discrepancy between levels of expected and observed academic performance. The first two perspectives allow the psychologist to understand better the child's academic performance in relation to other children's skills and in relation to the child's own skills. That is, nomothetic and idiographic analysis, respectively.

Discrepancy between expected and observed academic performance forms the foundation for classification of academic functioning. Expected achievement is the level of academic performance that would be manifested if essential elements in a child's life were to remain relatively constant and if no extraordinary assistance or interference with the child's learning were to occur. When observed achievement is markedly discrepant from expectancy, it suggests that something unusual may be influencing, either positively or negatively, academic performance.

Discrepancies between expected and observed achievement have been operationalized through a variety of methods, most of which have been demonstrated to be fatally flawed (Reynolds, 1985). Consistent with accepted theory, the M.MAC system utilizes level of general intellectual functioning to estimate academic expectancy (Kirk & Bateman, 1962). Discrepancy is calculated through regression analysis, employing the standard error of discrepancy from prediction (Thorndike, 1963) or, when certain actuarial data are unavailable, through estimated true difference analysis, using the standard error of measurement of estimated true difference (Stanley, 1971). These methods have been determined to be statistically and professionally sound (Glutting, McDermott, & Stanley, 1987; Reynolds, 1985).

Achievement in any given subject area may be found to be higher, lower, or reasonably consistent with expected levels. Underachievement is, of course, indicative of a learning problem and the M.MAC system logic displayed in Fig. 2.6 outlines the reasoning process which would result in diagnosis of a learning disability or developmental learning disorder. Overachievement suggests that learning has been inordinately induced, rather than inhibited. Such induction may be correlated with maladaptive social-emotional functioning. McDermott (1990) has noted that educators rarely assess for overachievement or consider the possibility of attendant social-emotional maladaptation. M.MAC systematizes the analysis of achievement to assess *both* possibilities and thereby ensure that all possible diagnostic alternatives are considered.

FIG. 2.6. M.MAC systems-actuarial logic for classification of academic functioning. From the microcomputer systems manual for McDermott Multidimensional Assessment of Children, P. A. McDermott and M. W. Watkins, 1985, 1987. New York: Psychological Corporation. Copyright (1985, 1987) by Psychological Corporation. Reproduced by permission. All rights reserved.

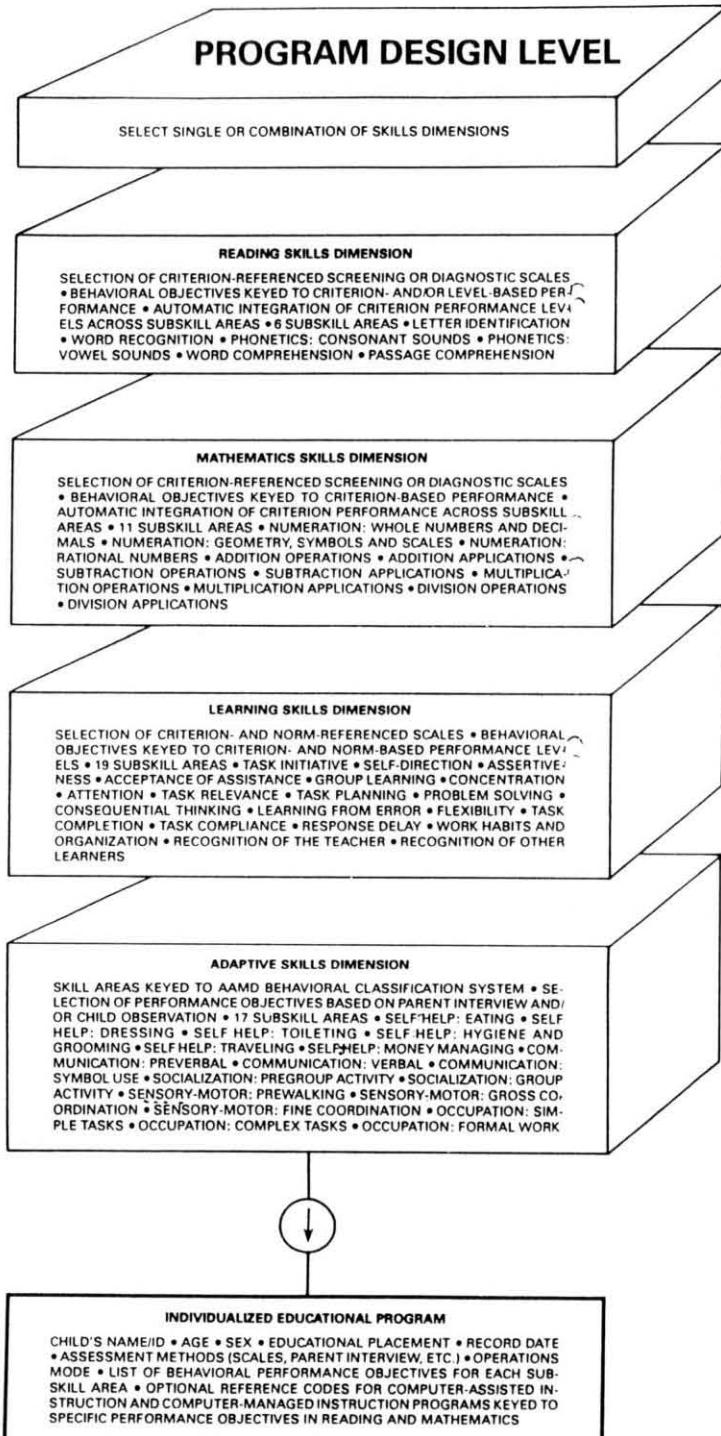


TABLE 2.2  
 ASSESSMENT SCALES AND METHODS SUPPORTED BY THE MMAC  
 PROGRAM DESIGN DIMENSION

**READING SKILLS**

Basic Achievement Skills Individual Screener-Reading  
 Woodcock Reading Mastery Test  
 Stanford Diagnostic Reading Test-Red Level  
 Stanford Diagnostic Reading Test-Green Level  
 Stanford Diagnostic Reading Test-Brown Level

**MATHEMATICS SKILLS**

Basic Achievement Skills Individual Screener Math  
 KeyMatic Diagnostic Arithmetic Test  
 Stanford Diagnostic Mathematics Test-Red Level  
 Stanford Diagnostic Mathematics Test-Green Level  
 Stanford Diagnostic Mathematics Test-Brown Level

**LEARNING SKILLS**

Study of Children's Learning Styles  
 Guide to the Child's Learning Style

**ADAPTIVE SKILLS**

Parent Interview / Observation of Child

### Program Design

The classification of childhood normality and exceptionality is only one facet of the M.MAC system. Once exceptionality is evident, it is vital to focus upon what a child knows, through more specific second-stage assessments, and to determine what steps may be necessary to promote learning and development. The Program Design level serves this function.

As seen in Fig. 2.7, there are four major dimensions of educational assessment and programming: reading, mathematics, learning, and adaptive skills. Although educational treatment plans for a child are unlikely to involve all four dimensions, the psychologist may elect to utilize as many as deemed necessary. For each selected dimension, the data collection method is specified (i.e., tests, teacher observations, clinical observations, or parent interview) and obtained data are entered into the system for analysis and design of remedial programs. Available instruments and methods are displayed in Table 2.2.

As in classification, there are several overarching concepts which apply to all program design dimensions. Namely, the system embodies a basic skills orientation, is objective, utilizes performance-based objectives, sequences objectives hierarchically, designs individualized programs, and is versatile. It is impossible within the limitations of this chapter to describe all aspects of the program design dimension. However, detailed descriptions and applications are provided by Glutting (1986b), McDermott (1990), and McDermott and Watkins (1985, 1987).

FIG. 2.7. Program design-level organization and features. From the microcomputer systems manual for McDermott Multidimensional Assessment of Children, P. A. McDermott and M. W. Watkins, 1985, 1987. New York: Psychological Corporation. Copyright (1985, 1987) by Psychological Corporation. Reproduced by permission. All rights reserved.

***Basic Skills Orientation.*** Preference for a basic skills orientation reflects the logic that proficiency in certain basic skills, irrespective of exceptionality, is a fundamental prerequisite to successful school and social adjustment. Primary skills covered by the M.MAC system include: reading and using written language; understanding and applying mathematics concepts; using effective learning strategies; and being reasonably self-sufficient in such adaptive behaviors as personal care, communication, socialization, sensory-motor, and vocational functions.

*Objectivity.* Educational programs covering vital basic skills must be objectively developed and based upon well-validated instruments intended for diagnostic educational programming. They must dispense with subjective opinions and unspecified criteria which have, unfortunately, been the norm (McDermott, 1990). The M.MAC system analyzes item responses, observed mastery levels, and other criterion-referenced performances of children and converts those observed performances into content-congruent basic skills objectives.

*Performance-based Objectives.* Assessment should lead to objectives which are stated in behavioral or verifiable terms. This does not imply a "behavioral" theoretical orientation, but simply reflects the reality that behavioral objectives are universally understood, provide criteria for judging attainment, and are easy to explain to parents and students. Specialists will, of course, apply the system's behavioral objectives in accordance with their theoretical orientation and within the context of each child's unique needs.

*Hierarchical Sequence of Objectives.* A comprehensive compilation of behavioral objectives which encompasses each primary basic skill area would be voluminous. Unstructured educational application of objectives is likely to be inefficient, if not ineffectual. When structured and aligned along educationally and psychologically meaningful dimensions, they can contribute to an orderly and effective educational program.

The M.MAC system contains 1,111 objectives distributed across 4 basic skill areas and 53 subskill areas. Within each subskill area, objectives are ordered hierarchically so that foundation skills precede other skills which are dependent or more difficult. Fig. 2.8 illustrates a representative selection by the M.MAC system from a hierarchical sequence of objectives within subskill areas in the mathematics domain. In areas where subskills are interdependent (e.g., paragraph comprehension skills rest upon word comprehension skills which, in turn, require certain letter identification and phonics skills, etc.), M.MAC objectives are integrated so that performance objectives selected in one subskill hierarchy do not outpace those in other hierarchies. This approach is compatible with conventional curricula and is particularly useful for building skills through step-by-step approximations.

FIG. 2.8. Sample mathematics educational program generated by M.MAC program design level. From the microcomputer systems manual for McDermott Multidimensional Assessment of Children, P. A. McDermott and M. W. Watkins, 1985, 1987. New York: Psychological Corporation. Copyright (1985, 1987) by Psychological Corporation. Reproduced by permission. All rights reserved.

*Individualization.* Individualized education programs are far too often oriented to the resources and needs of the school, teacher, or placement rather than to the needs of the child. As noted by McDermott (1990), this is not necessarily the fault of educators, but simply reflects the lack of resources necessary for production of truly individualized programs. M.MAC helps resolve this problem by applying systems-actuarial logic to educational program design; that is, by objectively analyzing a child's actual academic performance to guide a systematic selection of comprehensive skills hierarchies and thereby identify performance objectives directly related to the child's demonstrated educational needs.

*Versatility.* As previously noted, current expert systems must utilize both actuarial and clinical methodologies to enhance their authority. The program design component also embodies such a felicitous combinatorial approach. Even automated program development may, however, benefit from the interactive guidance of specialists with expertise and personal knowledge of a child's functioning. This added versatility is provided by two operational modes: Monitor and General.

The Monitor mode permits educational programs to be previewed and modified. It allows programs based upon measured criterion-referenced performance to be subsequently refined through professional judgment so as to best meet the unique needs of each child. Under the General mode, assessment moves directly from data input to data analysis to production of an educational program without preview or alteration of system-selected objectives.

Another aspect of versatility is represented in Fig. 2.8 under the "CAI/CMI CODES" heading. This column refers to computer-assisted instruction (CAI) and computer-managed instruction (CMI) resources which might assist children in achieving mastery of selected performance objectives (Kulik & Kulik, 1987; Kulik, Kulik, & Bangert-Drowns, 1985). CAI/CMI Codes are cross-referenced in the M.MAC manual to identify the title and publisher of specific software packages referenced by M.MAC. Thus, the computer can be used by the psychologist as an assessment tool and by the child as an instructional aid.

## SUMMARY

Computerized psychological systems must be viewed in light of their scope and authority; that is, the breadth and verity of their underlying knowledge base. Most current psychological applications are relatively narrow in scope and derivative in application. Even so, some do promise improved efficiency, economy, and reliability. Automated psychological systems of broad scope continue to be rare. The M.MAC system is an exception. It applies a judicious combination of the salient aspects of actuarial and clinical reasoning, decision theory, and sys-

tems analysis to the psychoeducational assessment of children. The system contains almost 10,000 discrete units of actuarial data and its reasoning is guided by thousands of decision rules. Its authority is established through adherence to standards formulated by appropriate national professional organizations, and through reliance upon some 250 empirical investigations. The M.MAC is a comprehensive, objective, reliable, and versatile system which enhances the validity of psychoeducational diagnosis. As such, it may serve as a model for future developments in computerized psychological expert systems.

## REFERENCES

- Adams, K. M., & Heaton, R. K. (1985). Automated interpretation of neuropsychological test data. *Journal of Consulting and Clinical Psychology*, 53, 790-802.
- Adams, K. M., Kvale, V. I., & Keegan, J. F. (1984). Performance of three automated systems for neuropsychological interpretation based on two representational tasks. *Journal of Clinical Neuropsychology*, 6, 413-431.
- Algozzine, B., & Yssel'ye, J. E. (1981). Special education services for normal children: Better safe than sorry? *Exceptional Children*, 48, 238-243.
- American Psychological Association. (1986). *Guidelines for computer-based tests and interpretations*. Washington, DC: Author.
- American Psychiatric Association. (1987). *Diagnostic and statistical manual of mental disorders* (3rd Ed. revised). Washington, DC: Author.
- Andolina, M. (1982). Reading tests: Traditional vs. computerized. *Classroom Computer News*, 2, 39-40.
- Arkes, H. R. (1981). Impediments to accurate clinical judgment and possible ways to minimize their impact. *Journal of Consulting and Clinical Psychology*, 3, 323-330.
- Beaumont, J. G. (1981). Microcomputer-aided assessment using standard psychometric procedures. *Behavior Research Methods & Instrumentation*, 13, 430-433.
- Bhatnagar, R. K., & Kanal, L. N. (1986, October). *Mechanizing decision making with uncertain information*. Paper presented at the Expert Systems in Government Symposium, McLean, VA.
- Brantley, J. C. (1986, April). *Decision rule comparisons among eight computer programs for WISC-R profile analysis*. Paper presented at a meeting of the National Association of School Psychologists, Hollywood, FL.
- Broverman, I. K., Broverman, D. M., Clarkson, F. E., Rosenkrantz, P. S., & Vogel, S. R. (1970). Sex-role stereotypes and clinical judgments of mental health. *Journal of Consulting and Clinical Psychology*, 34, 1-7.
- Buchanan, B. G. (1985). *Expert systems: Working systems and the research literature* (Report No. KSL-85-37). Stanford, CA: Department of Computer Science, Stanford University.
- Butcher, J. N. (1987). Computerized clinical and personality assessment using the MMPI. In J. N. Butcher (Ed.), *Computerized psychological assessment* (pp. 161-197). New York: Basic Books.
- Campbell, D. P. (1968). The Strong Vocational Interest Blank: 1927-1967. In P. McReynolds (Ed.), *Advances in psychological assessment*. Palo Alto, CA: Science and Behavior Books.
- Cantwell, D. P., Russell, A. T., Mattison, R., & Will, L. (1979). A comparison of DSM-II and DSM-III in the diagnosis of childhood psychiatric disorders. *Archives of General Psychiatry*, 36, 1208-1213.
- Chadwick, M., & Hannah, J. (1987). *Expert systems for microcomputers*. Blue Ridge Summit, PA: Tab Books.

Colby, K. M., & Parkison, R. (1985). Linguistic conceptual-patterns and key-idea profiles as a new kind of property for a taxonomy of neurotic patients. *Computers in Human Behavior*, 1, 181-194.

Cromwell, R. L., Blashfield, R. K., & Strauss, J. S. (1975). Criteria for classification systems. In N. Hobbs (Ed.), *Issues in the classification of children* (Vol. 1, pp. 4-25). San Francisco: Jossey-Bass.

Dailey, C. (1971). *Assessment of lives*. San Francisco: Jossey-Bass.

Davis, F. B. (1959). Interpretation of differences among average and individual test scores. *Journal of Educational Psychology*, 50, 162-170.

DeMita, M. A., Johnson, J. H., & Hansen, K. E. (1981). The validity of a computerized visual searching task as an indicator of brain damage. *Behavior Research Methods & Instrumentation*, 13, 592-594.

Deupree, C. (1985, Summer). Artificial intelligence and the expert system. *CTASP Newsletter*, pp. 3-5.

DiNardo, P. A. (1975). Social class and diagnostic suggestion as variables in clinical judgment. *Journal of Consulting and Clinical Psychology*, 43, 363-368.

Eddy, D. M., & Clanton, C. H. (1982). The art of diagnosis: Solving the clinicopathological exercise. *New England Journal of Medicine*, 306, 1263-1267.

Eller, B. F., Kaufman, A. S., & McLean, J. E. (1986). Computer-based assessment of cognitive abilities: Current status/future directions. *Journal of Educational Technology Systems*, 15, 137-147.

Elwood, D. J. (1972). Validity of an automated measure of intelligence in borderline retarded subjects. *American Journal of Mental Deficiency*, 7, 90-94.

Epps, S., Ysseldyke, J. E., & McGue, M. (1984). "I know one when I see one"—Differentiating LD and non-LD students. *Learning Disability Quarterly*, 7, 89-101.

Erdman, H. P., Greist, J. H., Klein, M. H., Jefferson, J. W., & Getto, C. (1981). The computer psychiatrist: How far have we come? Where are we heading? How far dare we go? *Behavior Research Methods & Instrumentation*, 13, 393-398.

Exner, J. E. (1987). Computer assistance in Rorschach interpretation. In J. N. Butcher (Ed.), *Computerized psychological assessment* (pp. 218-235). New York: Basic Books.

Fowler, R. D. (1985). Landmarks in computer-assisted psychological assessment. *Journal of Consulting and Clinical Psychology*, 53, 748-759.

Franks, D. J. (1971). Ethnic and social status characteristics of children in E.M.R. and L.D. classes. *Exceptional Children*, 34, 327-338.

Freeman, M. (1971). A reliability study of psychiatric diagnosis in childhood and adolescence. *Journal of Child Psychology and Psychiatry*, 12, 43-54.

Glasser, A. J., & Zimmerman, I. (1967). *Clinical interpretation of the Wechsler Intelligence Scale for Children*. New York: Grune & Stratton.

Glutting, J. J. (1986a). The McDermott Multidimensional Assessment of Children: Applications to the classification of childhood exceptionality. *Journal of Learning Disabilities*, 19, 331-335.

Glutting, J. J. (1986b). The McDermott Multidimensional Assessment of Children: Contribution to the development of individualized education programs. *Journal of Special Education*, 20, 431-445.

Glutting, J. J., McDermott, P. A., & Stanley, J. C. (1987). Resolving differences among methods of establishing confidence limits for test scores. *Educational and Psychological Measurement*, 47, 607-614.

Gonsalves, P. (1987). Interview: Trip Hawkins, President, Electronic Arts. *CD-I News*, 1, 6-8.

Greene, R. L., Martin, P. W., Bennett, S. R., & Shaw, J. A. (1981). A computerized scoring system for the Personality Inventory for Children. *Educational and Psychological Measurement*, 41, 233-235.

Grossman, H. J. (1977). *Manual on terminology and classification in mental retardation* (rev. ed.). Washington, DC: American Association on Mental Deficiency.

Grossman, H. J. (1983). *Classification in mental retardation*. Washington, DC: American Association on Mental Deficiency.

Hale, R. L., & McDermott, P. A. (1984). Pattern analysis of an actuarial strategy for computerized diagnosis of childhood exceptionality. *Journal of Learning Disabilities*, 17, 30-37.

Harris, W. G., Niedner, D., Feldman, C., Fink, A., & Johnston, J. (1981). An on-line interpretive Rorschach approach: Using Exner's comprehensive system. *Behavior Research Methods and Instrumentation*, 13, 588-591.

Hasselbring, T. S., & Crossland, C. L. (1981). Using microcomputers for diagnosing spelling problems in learning-handicapped children. *Educational Technology*, 21, 37-39.

Hedl, J. J., O'Neil, H. F., & Hansen, D. N. (1973). Affective reactions toward computer-based intelligence testing. *Journal of Consulting and Clinical Psychology*, 40, 217-222.

Hofer, P. J., & Green, B. F. (1985). The challenge of competence and creativity in computerized psychological testing. *Journal of Counseling and Clinical Psychology*, 53, 826-838.

Honaker, L. M., Hector, V. S., & Harrell, T. H. (1986). Perceived validity of computer- versus clinician-generated MMPI reports. *Computers in Human Behavior*, 2, 77-83.

Jackson, D. N. (1985). Computer-based personality testing. *Computers in Human Behavior*, 1, 255-264.

Johnson, D. L., Willis, J., & Danley, W. (1982). A comparison of traditional and computer-based methods of teaching students to administer individual intelligence tests. *AEDS Journal*, 16, 56-64.

Jones, M. B., Dunlop, W. P., & Bilodeau, I. M. (1987). Comparison of video game and conventional test performance. *Simulation & Games*, 17, 435-446.

Kahneman, D., & Tversky, A. (1982). The psychology of preferences. *Scientific American*, 246, 160-173.

Katz, L., & Dalby, S. (1981). Computer and manual administration of the Eysenck Personality Inventory. *Journal of Clinical Psychology*, 37, 586-588.

Kavale, K. A., & Furness, S. R. (1984). A meta-analysis of the validity of Wechsler scale profiles and recategorizations: Patterns or parodies? *Learning Disability Quarterly*, 2, 136-156.

Kirk, S. A., & Bateman, B. (1962). Diagnosis and remediation of learning disabilities. *Exceptional Children*, 29, 73-76.

Klinge, V., & Rodziewicz, T. (1976). Automated and manual intelligence testing of the Peabody Picture Vocabulary Test on a psychiatric adolescent population. *International Journal of Man-Machine Studies*, 8, 243-246.

Koran, L. M. (1975). The reliability of clinical methods, data and judgments. *New England Journal of Medicine*, 293, 695-701.

Kramer, J. J., Henning-Stout, M., Ullman, D. P., & Schellenberg, R. P. (1987). The viability of scatter analysis on the WISC-R and the SBIS: Examining a vestige. *Journal of Psychoeducational Assessment*, 5, 37-47.

Krug, S. E. (1987). Microtrends: An orientation to computerized assessment. In J. N. Butcher (Ed.), *Computerized psychological assessment* (pp. 15-25). New York: Basic Books.

Krutch, J. (1986). *Experiments in artificial intelligence for microcomputers*. Indianapolis: Howard W. Sams.

Kulik, J. A., & Kulik, C. C. (1987). Review of recent literature on computer-based instruction. *Contemporary Educational Psychology*, 12, 222-230.

Kulik, J. A., Kulik, C. C., & Bangert-Drowns, R. L. (1985). Effectiveness of computer-based education in elementary schools. *Computers in Human Behavior*, 1, 59-74.

Lachar, D. (1984). *Multidimensional description of child personality: A manual for the Personality Inventory for Children*. Los Angeles: Western Psychological Services.

Lueger, R. J., & Petzel, T. P. (1979). Illusory correlation in clinical judgment: Effects of amount of information to be processed. *Journal of Consulting and Clinical Psychology*, 47, 1120-1121.

Lukin, M. E., Dowd, E. T., Plake, B. S., & Kraft, R. G. (1985). Comparing computerized versus traditional psychological assessment. *Computers in Human Behavior*, 1, 49-58.

Lushene, R. E., O'Neil, H. F., & Dunn, T. (1974). Equivalent validity of a completely computerized MMPI. *Journal of Personality Assessment, 38*, 353-361.

Matarazzo, J. D. (1983). Computerized psychological testing. *Science, 221*, 323.

Matarazzo, J. D. (1985). Clinical psychological test interpretations by computer: Hardware outpaces software. *Computers in Human Behavior, 1*, 235-253.

Matarazzo, J. D. (1986). Computerized clinical psychological test interpretations: Unvalidated plus all mean and no sigma. *American Psychologist, 41*, 14-24.

McDermott, P. A. (1980a). A computerized system for the classification of developmental, learning, and adjustment disorders in school children. *Educational and Psychological Measurement, 40*, 761-768.

McDermott, P. A. (1980b). Congruence and typology of diagnoses in school psychology: An empirical study. *Psychology in the Schools, 17*, 12-24.

McDermott, P. A. (1980c). A systems-actuarial method for the differential diagnosis of handicapped children. *Journal of Special Education, 14*, 7-22.

McDermott, P. A. (1981). Sources of error in the psychoeducational diagnosis of children. *Journal of School Psychology, 19*, 31-44.

McDermott, P. A. (1982). Actuarial assessment systems for the grouping and classification of schoolchildren. In C. R. Reynolds & T. B. Gutkin (Eds.), *The handbook of school psychology* (pp. 243-272). New York: Wiley.

McDermott, P. A. (1986). The observation and classification of exceptional child behavior. In R. T. Brown & C. R. Reynolds (Eds.), *Psychological perspectives on childhood exceptionality: A handbook* (pp. 136-180). New York: Wiley-Interscience.

McDermott, P. A. (1990). Applied systems-actuarial assessment. In T. B. Gutkin & C. R. Reynolds (Eds.), *The handbook of school psychology* (2nd ed.). New York: Wiley.

McDermott, P. A., & Hale, R. L. (1982). Validation of a systems-actuarial computer process for multidimensional classification of child psychopathology. *Journal of Clinical Psychology, 38*, 477-486.

McDermott, P. A., & Watkins, M. W. (1985). *McDermott Multidimensional Assessment of Children: 1985 Apple II version* [Computer program]. New York: Psychological Corporation.

McDermott, P. A., & Watkins, M. W. (1987). *McDermott Multidimensional Assessment of Children: 1987 IBM version* [Computer program]. New York: Psychological Corporation.

McKinley, R. L., & Reckase, M. D. (1980). Computer applications to ability testing. *AEDS Journal, 13*, 193-203.

Meehl, P. E. (1954). *Clinical versus statistical prediction*. Minneapolis: University of Minnesota Press.

Moreland, K. L. (1985). Computer-assisted psychological assessment in 1986: A practical guide. *Computers in Human Behavior, 1*, 221-233.

Moreland, K. L. (1987). Computerized psychological assessment: What's available. In J. N. Butcher (Ed.), *Computerized psychological assessment* (pp. 26-49). New York: Basic Books.

Nathan, P. E. (1967). *Cues, decisions, and diagnoses: A systems-analytic approach to the diagnosis of psychopathology*. New York: Academic Press.

Nolen, P., & Spencer, M. (1986). School psychology and computerized data handling systems: A confrontation. *Computers in the Schools, 2*, 137-143.

Oosterhof, A. C., & Salisbury, D. F. (1985). Some measurement and instruction related considerations regarding computer-assisted testing. *Educational Measurement: Issues and Practice, 4*, 19-23.

Piotrowski, Z. A. (1964). Digital-computer interpretation of ink-blot test data. *Psychiatric Quarterly, 38*, 1-26.

Quay, H. C. (1986). Classification. In H. C. Quay & J. S. Werry (Eds.), *Psychopathological disorders of childhood* (3rd ed., pp. 1-34). New York: Wiley.

Reynolds, C. R. (1985). Critical measurement issues in learning disabilities. *Journal of Special Education, 18*, 451-475.

Roid, G. H. (1986). Computer-based test interpretation: The potential of quantitative methods of test interpretation. *Computers in Human Behavior, 1*, 207-219.

Rozensky, R. H., Honor, L. F., Rasinski, K., Tovian, S. M., & Herz, G. I. (1986). Paper-and-pencil versus computer-administered MMPIs: A comparison of patients' attitudes. *Computers in Human Behavior, 2*, 111-116.

Sampson, J. P. (1983). Computer-assisted testing and assessment: Current status and implications for the future. *Management and Evaluation in Guidance, 15*, 293-298.

Sands, W. A., & Gade, P. A. (1983). An application of computerized adaptive testing in U. S. Army recruiting. *Journal of Computer-Based Instruction, 10*, 87-89.

Schoolman, H. M., & Bernstein, L. M. (1978). Computer use in diagnosis, prognosis, and therapy. *Science, 200*, 926-931.

Scissoms, E. H. (1976). Computer administration of the California Psychological Inventory. *Measurement and Evaluation in Guidance, 9*, 22-25.

Silverstein, A. B. (1981a). Reliability and abnormality of test score differences. *Journal of Clinical Psychology, 37*, 392-394.

Silverstein, A. B. (1981b). Verbal-Performance IQ discrepancies on the WISC-R: One more time. *Journal of Consulting and Clinical Psychology, 49*, 465-466.

Silverstein, A. B. (1982). Pattern analysis as simultaneous statistical inference. *Journal of Consulting and Clinical Psychology, 50*, 234-240.

Skinner, H. A., & Allen, B. A. (1983). Does the computer make a difference? Computerized versus face-to-face versus self-report assessment of alcohol, drug, and tobacco use. *Journal of Consulting and Clinical Psychology, 51*, 267-275.

Space, L. G. (1981). The computer as psychometrist. *Behavior Research Methods & Instrumentation, 13*, 595-606.

Spitzer, R. L., Endicott, J., Cohen, J., & Fleiss, J. L. (1974). Constraints on the validity of computer diagnosis. *Archives of General Psychiatry, 31*, 197-203.

Spitzer, R. L., & Fleiss, J. L. (1974). A re-analysis of the reliability of psychiatric diagnosis. *British Journal of Psychiatry, 125*, 341-347.

Stanley, J. C. (1971). Reliability. In R. L. Thorndike (Ed.), *Educational measurement* (2nd ed., pp. 356-442). Washington, DC: American Council on Education.

Stein, S. J. (1987). Computer-assisted diagnosis for children and adolescents. In J. N. Butcher (Ed.), *Computerized psychological assessment* (pp. 145-160). New York: Basic Books.

Szolovits, P., & Pauker, S. G. (1978). Categorical and probabilistic reasoning in medical diagnosis. *Artificial Intelligence, 11*, 115-144.

Thorndike, R. L. (1963). *The concepts of over- and under-achievement*. New York: Teachers College Press.

University of Minnesota Press. (1982). *User's guide for the Minnesota Report*. Minneapolis: National Computer Systems.

Visonhaler, J. F., Weinshank, A. B., Wagner, C. C., & Polin, R. M. (1983). Diagnosing children with educational problems: Characteristics of reading and learning disabilities specialists, and classroom teachers. *Reading Research Quarterly, 18*, 134-164.

Waterman, D. A. (1986). *A guide to expert systems*. Reading, MA: Addison-Wesley.

Watkins, M. W., & Kush, J. (1988). Assessment of academic skills of learning disabled students with classroom microcomputers. *School Psychology Review, 17*, 79-86.

Webb, N. M., Herman, J. L., & Cabello, B. (1986). Diagnosing students' errors from their response selections in language arts. *Journal of Educational Measurement, 23*, 163-170.

Weiss, D. J. (1985). Adaptive testing by computer. *Journal of Consulting and Clinical Psychology, 53*, 774-789.

Weiss, D. J., & Vale, C. D. (1987). Computerized adaptive testing for measuring abilities and other psychological variables. In J. N. Butcher (Ed.), *Computerized psychological assessment* (pp. 325-343). New York: Basic Books.

White, D. M., Clements, C. B., & Fowler, R. D. (1985). A comparison of computer administration with standard administration of the MMPI. *Computers in Human Behavior*, 1, 153-162.

Wikoff, R. L. (1978). Correlational and factor analysis of the Peabody Individual Achievement Test and WISC-R. *Journal of Consulting and Clinical Psychology*, 46, 322-325.

Wilson, F. R., Genco, K. T., & Yager, G. G. (1986). Assessing the equivalence of paper-and-pencil vs. computerized tests: Demonstration of a promising methodology. *Computers in Human Behavior*, 1, 265-275.

Wise, S. L., & Wise, L. A. (1987). Comparison of computer-administered and paper-administered achievement tests with elementary school children. *Computers in Human Behavior*, 3, 15-20.

World Health Organization. (1978). *Ninth revision of the International Classification of Diseases*. Geneva: Author.