

University of Nebraska - Lincoln

DigitalCommons@University of Nebraska - Lincoln

The Computer and the Decision-Making
Process

Buros-Nebraska Series on Measurement and
Testing

1991

4. The Validity of ComputerBased Test Interpretations of the MMPI

Lorraine D. Eyde
U.S. Office of Personnel Management

Dennis M. Kowal
U.S. Army

Francis J. Fishburne Jr.
U.S. Army, Retired

Follow this and additional works at: <https://digitalcommons.unl.edu/buroscomputerdecision>



Part of the [Educational Assessment, Evaluation, and Research Commons](#), and the [Educational Methods Commons](#)

Eyde, Lorraine D.; Kowal, Dennis M.; and Fishburne, Francis J. Jr., "4. The Validity of ComputerBased Test Interpretations of the MMPI" (1991). *The Computer and the Decision-Making Process*. 6.

<https://digitalcommons.unl.edu/buroscomputerdecision/6>

This Article is brought to you for free and open access by the Buros-Nebraska Series on Measurement and Testing at DigitalCommons@University of Nebraska - Lincoln. It has been accepted for inclusion in The Computer and the Decision-Making Process by an authorized administrator of DigitalCommons@University of Nebraska - Lincoln.

4 The Validity of Computer-Based Test Interpretations of the MMPI

Lorraine D. Eyde
U.S. Office of Personnel Management

Dennis M. Kowal
U.S. Army

Francis J. Fishburne, Jr.
U.S. Army, Retired

With advances in computer technology, computer-based test interpretations (CBTI), first developed in the early 1960s (Fowler, 1985), have proliferated (Eyde & Kowal, 1987). CBTIs have been developed and marketed for a variety of tests used in clinical, counseling, educational, and employment settings. The largest number of commercial CBTI systems are available for the Minnesota Multiphasic Personality Inventory (MMPI; Krug, 1987), the most widely used inventory of its kind in the world, which has a continuously growing literature of more than 8,000 books and articles (Holden, 1986; Lanyon, 1984).

According to Harris

CBTI refers to the automation of a set of pre-specified rules for use in analyzing, interpreting and assigning certain qualities to a response or response pattern (e.g., test score, profile pattern). The discrete rules are used to form an algorithm that guides the activity of the computer to interpret specific input data. (1987, p. 239)

Consumers of CBTIs have very little information available on the development of the algorithm or the validity of the CBTI systems. Companies selling CBTIs often do not provide a user's guide. The algorithms used in generating the computer interpretations are not available to CBTI users nor are they provided for scholarly review purposes. Notable exceptions to these business practices include Lachar's (1974) presentation of all the rules and interpretive statements for the WPS Test Report, the MMPI CBTI sold by Western Psychological Services. National Computer Systems provided the algorithms for the Minnesota Report:

Personnel Selection System, for scholarly review purposes, and gave an independent evaluation of the extent to which the interpretive statements were based on the MMPI's research literature or on the clinical judgment of the CBTI's author (Eyde, 1985).

Numerous critics have pointed out serious problems arising from the growth of CBTIs. Mitchell (1984) observed that the advent of CBTIs "presents the field of psychology with its most serious and consequential challenge of the next decade." Lanyon (1984) called attention to the exponential growth of available CBTI systems, noting that Meehl's cookbook approach to MMPI interpretation (however carefully designed) has been used to justify and market many inadequate systems. Eyde and Kowal (1987) commented that "the scientific basis for the C.B.T.I., namely the decision rules which codify the rationale and the evidence used to produce the computer interpretations, may wind up locked in a black box, inaccessible to test users" (p. 402). Also, Matarazzo (1986) decried the lack of validity evidence for CBTIs.

The problems associated with CBTIs have to do not only with the lack of validity data, but also with the problem of how to *establish* the validity of a computer interpretive report (Mitchell, 1984; Moreland, 1985, 1987; O'Dell, 1972). Mitchell (1984) notes that purists who want to do the job properly, "are faced with the task of a conducting a statement-by-statement validation involving statements generated by decision rules and decision trees of almost incomprehensible complexity."

Critics of prevailing practices in developing, marketing, and validating computerized applications of knowledge-based systems, may choose, as Eyde and Kowal (1985) have, to do some of the developmental work that should have been done before a computerized test product is sold.

The intent of this chapter is to describe a methodology for studying the validity of the output of CBTI systems. The research focuses on a variety of CBTI systems developed as tools for interpreting the MMPI. The MMPI is the most widely used psychodiagnostic instrument with active-duty military populations (Parkison & Fishburne, 1984). Our methodology is designed so that it may be adapted to CBTIs for other tests or self-report inventories. The study involves a comparative analysis of the accuracy, relevancy, and usefulness of the output of seven CBTI systems for patients in a military hospital which draws its patients from a wide geographical area. The research design allows us to make some inferences about the relative accuracy of CBTI systems for different profile types. A secondary objective of the research was to identify racial differences, if any, in the accuracy of the CBTIs.

This chapter will describe the study, provide basic data, and describe the results. Other chapters will cover (a) the Black/white differences in the accuracy of the CBTIs, which are minimal (Eyde, Kowal, & Fishburne, 1987); and (b) neuropsychological cases vs. nonneuropsychological cases (Fishburne, Eyde, & Kowal, 1988).

MMPI ELEMENTS FOR CBTI USE

Since a major objective of this research was to establish and apply a methodology for validating CBTIs, we will summarize some elements to aid in understanding computer interpretations of the MMPI, the inventory used in this study. Readers are referred to Anastasi (1988) and Graham and Lilly (1984) for a general introduction to MMPI use, and to Dahlstrom, Welsh, and Dahlstrom (1972), Graham (1987), Greene (1980), or Lachar (1974) for more detailed presentations.

The MMPI, a self-report inventory with 566 true–false or cannot-say (omitted) items, has an extensive history dating back to the 1930s. Its derivational groups, which consisted of both normal and clinical groups, were used in developing empirically based scoring keys, to aid in assigning psychiatric diagnostic labels to patients.

The MMPIs content includes items dealing with

Health, psychosomatic symptoms, neurological disorders, and motor disturbances; sexual, religious, political, and social attitudes; educational, occupational, family, and marital questions; and many well-known neurotic or psychotic behavior manifestations, such as obsessive and compulsive states, delusions, hallucinations, ideas of reference, phobias, and sadistic and masochistic trends. (Anastasi, 1988, p. 526)

The basic MMPI profile provides 10 “clinical” scales and 3 validity scales as described in Table 4.1. An additional validity scale, Cannot Say, which consists of the items omitted by the test taker, is usually reported. Furthermore, several hundred research scales are available.

Scale numbers are used in preference to scale names because diagnostic labels have changed since the inception of MMPI research. The scales have correlates that range far beyond those implied by the labels. Furthermore, with the increased use of the MMPI with nonhospitalized groups it is necessary to avoid the use of stigmatizing labels. From the large empirical research base and clinical lore on these scale scores it is possible to draw inferences about the test taker’s personality organization or structure, psychopathology, and other characteristics.

The validity indicators (Cannot Say, L [lie] scale, F scale [items infrequently endorsed by normal test takers] and K scale [to assess clinical defensiveness]) deal with test-taking attitudes. Greene (1980, p. 117) observes that “validity scales serve primarily to establish whether a *specific* clinical scale profile can be safely interpreted” (emphasis added). Dahlstrom et al. (1972, p. 100) differentiate the psychometric term “validity” (that is, the extent to which inferences about the test are meaningful) from its usage with the MMPI *validity indicators* in which the validity “pertains to the appropriateness or acceptability of any *one* administration of the test” (emphasis added).

TABLE 4.1
Sample Interpretive Inferences for Standard Minnesota Multiphasic Personality Inventory Scales

<i>Scale Name</i>	<i>Scale Abbrev.</i>	<i>Scale Number</i>	<i>Interpretation of High Scores</i>	<i>Interpretation of Low Scores</i>
--	L	--	Trying to create favorable impression by not being honest in responding to items; conventional; rigid; moralistic; lacks insight	Responded frankly to items; confident; perceptive; self-reliant; cynical
--	F	--	May indicate invalid profile; severe pathology; moody; restless; dissatisfied	Socially conforming; free of disabling psychopathology; may be "faking good"
--	K	--	May indicate invalid profile; defensive; inhibited; intolerant; lacks insight	May indicate invalid profile; exaggerates problems; self-critical; dissatisfied; conforming; lacks insight; cynical
Hypochondriasis	Hs	1	Excessive bodily concern; somatic symptoms, narcissistic; pessimistic; demanding; critical; long-standing problems	Free of somatic preoccupation; optimistic; sensitive; insightful
Depression	D	2	Depressed; pessimistic; irritable; dissatisfied; lacks self-confidence; introverted; overcontrolled	Free of psychological turmoil; optimistic; energetic; competitive; impulsive; undercontrolled; exhibitionistic
Hysteria	Hy	3	Physical symptoms of functional origin; lacks insight; self-centered; socially involved; demands attention and affection	Constricted; conventional; narrow interests; limited social participation; untrusting; hard to get to know; realistic
Psychopathic Deviate	Pd	4	Asocial or antisocial; rebellious; impulsive; poor judgment; immature; creates good first impression; superficial relationships; aggressive; free of psychological turmoil	Conventional; conforming; accepts authority; low drive level; concerned about status and security; persistent; moralistic
Masculinity/ Femininity	Mf	5	Male: aesthetic interests; insecure in masculine role; creative, good judgment; sensitive; passive; dependent; good self-control Female: rejects traditional female role; masculine interests; assertive; competitive; self-confident; logical; unemotional	<u>Male</u> : overemphasizes strength and physical prowess; adventurous; narrow interests; inflexible; contented; lacks insight <u>Female</u> : accepts traditional female role; passive; yielding to males; complaining; critical; constricted
Paranoia	Pa	6	May exhibit frankly psychotic behavior; suspicious; sensitive; resentful; projects; rationalizes; moralistic; rigid	May have frankly psychotic symptoms; evasive; defensive; guarded; secretive; withdrawn
Psychasthenia	Pt	7	Anxious; worried; difficulties in concentrating; ruminative; obsessive; compulsive; insecure; lacks self-confidence; organized; persistent; problems in decision making	Free of disabling fears and anxieties; self-confident; responsible; adaptable; values success and status
Schizophrenia	Sc	8	May have thinking disturbance; withdrawn; self-doubts; feels alienated and unaccepted; vague goals	Friendly, sensitive, trustful; avoids deep emotional involvement; conventional; unimaginative

<i>Scale Name</i>	<i>Scale Abbrev.</i>	<i>Scale Number</i>	<i>Interpretation of High Scores</i>	<i>Interpretation of Low Scores</i>
Hypomania	Ma	9	Excessive activity; impulsive; lacks direction; unrealistic self-appraisal; low frustration tolerance; friendly; manipulative; episodes of depression	Low energy level; apathetic; responsible; conventional; lacks self-confidence; overcontrolled
Social Introversion	Si	0	Socially introverted; shy; sensitive; overcontrolled; conforming; problems in decision making	Socially extroverted; friendly; active; competitive; impulsive; self-indulgent

From J. R. Graham (1978), *The Minnesota Multiphasic Personality Inventory (MMPI)*. In B. B. Wolman (Ed.), *Clinical diagnosis of mental disorders: A handbook*. New York: Plenum Press. Copyright 1978 by Plenum Press. Reproduced by permission.

The test taker's raw scores on the scales are usually transformed to linear T-scores with a mean of 50 and a standard deviation of 10 (Dahlstrom et al., 1972). In other words, the T- or standard scores are not transformed to approximate the normal distribution. There are two exceptions to this practice relevant to this chapter. Colligan, Osborne, Swenson, and Offord (1983) reported their normative data in terms of T-scores that were transformed to approximate the normal distribution. The Morris–Tomlinson Report is based on these data. Finney, whose normative data form the basis of the current Behaviordyne MMPI CBTIs, also uses normalized T-scores and, in addition, reports the Minnesota standard scores. T-scores aid in making direct comparisons among scales for test takers. Scores of 70 on the clinical scales are commonly used as cutoffs to identify potential deviancy or psychopathology.

The T-score tables are generally based on the normative data collected on Midwestern white adults before World War II (Dahlstrom et al., 1972). A major restandardization effort, using a nationwide sample, sponsored by the University of Minnesota Press, the test publisher, is under way (Holden, 1986). A modern restandardization employing Midwestern whites was reported by Colligan et al. (1983). Finney (1968) developed his norms in Kentucky. Graham and Lilly (1984, p. 238) point out that “the standardization samples used for the T-score conversions are the same normal subjects used in constructing the scales. . . . Thus, the theoretically normal or average person would have T-scores of approximately fifty on all of the scales.”

Interpretation of the MMPI generally begins with a review of test taker's scores on the validity indicators, namely, the validity profile. If the test taker appears to have responded to the inventory in a reasonably straightforward manner (e.g., has not attempted to dissimulate), then elevated scores on individual clinical scales or combinations of scales (most often the two that are most elevated) are evaluated in terms of the accumulated evidence about their meaning. Dahlstrom et al. observed that

Groups formed on the basis of the evaluation of a single scale may still be quite heterogeneous and the stable correlates may be rather different in this kind of analysis from those resulting when the groups are formed on the basis of common test *patterns* (i.e., combinations of scales). (1972, p. 178)

Since its inception, the MMPI test authors, Hathaway and McKinley, recognized the diagnostic richness of using configural analysis, for example, basing interpretations on elevations on two scales, that is, two-point codes. In general, the two-point codes are used interchangeably; that is, a 7/2 code and a 2/7 are treated the same.

With these essentials of MMPI interpretation in mind, readers may find it useful to review the general approach to MMPI interpretation used by a scientist-practitioner in his clinical work (Graham, 1977, pp. 150–151).

1. What was the test-taking attitude of the examinee, and how should this attitude be taken into account in interpreting the protocol?
2. What is the general level of adjustment of the person who produced the protocol?
3. What kinds of behaviors (symptoms, attitudes, defenses, etc.) can be inferred about or expected from the person who produced the protocol?
4. What etiology or set of psychological dynamics underlie the behaviors described?
5. What are the most appropriate diagnostic labels for the person who produced the protocol?
6. What are the implications for the treatment of the person who produced the protocol?

These six areas for which inferences may be drawn in interpreting the MMPI appear to have been used, to varying degrees, in the preparation of narrative statements for the libraries of CBTI systems for the MMPI.

CRITICISMS OF RESEARCH ON CBTI SYSTEMS

This chapter reports on a large-scale research project on the validity of the output of CBTI systems for the MMPI, based on a modification of Moreland's 1980 research plan (W. G. Dahlstrom, personal communication, November 20, 1985; Moreland, 1985, 1987). Our research plan took into consideration Moreland's criticism of research on the validity of CBTI systems, his recommendations for future research, and advice to consumers evaluating CBTI systems.

Moreland's (1985, 1987) criteria for evaluating CBTI research served as a model for developing our design. The dependent variable was ratings, by experi-

enced clinical psychologists, of the accuracy of all individual narrative statements or sentences from each CBTI system. Existing (file drawer) case histories or self-report questionnaires (for subclinical normal cases) were used as the criteria against which raters made their evaluations. The independent variables were seven CBTI systems, the nature of the MMPI profiles evaluated (e.g., profile types), and the race of the subjects.

Moreland's (1985, 1987) literature review brings out factors that should be considered in efforts to determine the accuracy of CBTI interpretations. In particular, the design should require raters to evaluate specific interpretive statements; limiting them to global accuracy ratings will limit the usefulness of the ratings for improving the CBTI system. He stressed the need for maximizing the number and variety of cases and the need for developing procedures for selecting an unbiased sample and he noted the importance of assessing rater reliability. He points out the merits of basing ratings on external criteria such as records or special research instruments, which provide raters with a standard criterion, in preference to studies in which clinicians evaluate the accuracy of CBTI's by using their own unsystematic observations on patients.

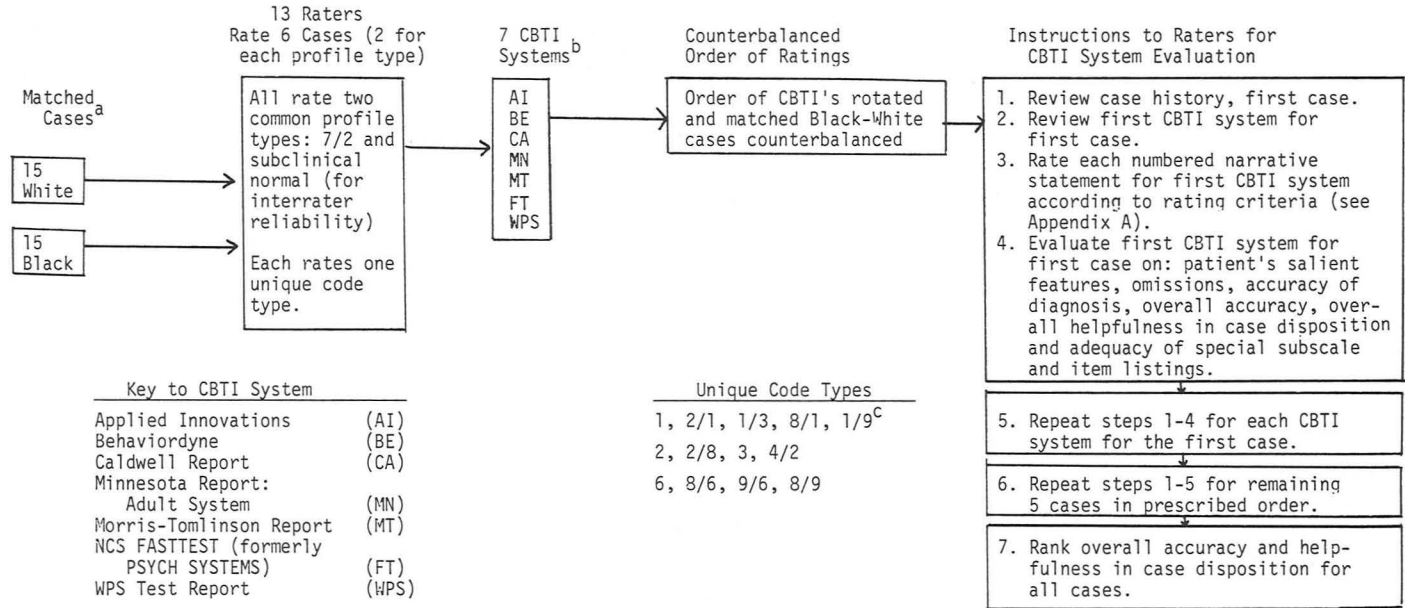
Moreland's (1987) review indicates that few existing commercial CBTI systems have been so evaluated and they are often evaluated for only a limited number of types of profiles. Moreland reviewed comparative studies of clinicians' ratings of the global accuracy of five CBTI systems. Only two of these (the Minnesota Report and the WPS Test Report) are currently marketed. He also examined four studies which evaluated five CBTI systems against external criteria; three of these (Behaviordyne, the Caldwell Report, and the WPS Test Report) continue to be commercially available. He found that the number of cases and the profile types evaluated tended to be limited in number in these latter four studies. Raters were sometimes students, such as psychiatric residents, rather than fully qualified clinicians and the evaluation of interrater reliability was infrequent.

Moreland (1985) recommended that raters focus on identifying irrelevant (e.g., redundant) statements and separate these statements from those whose accuracy should be rated. He also recommended identifying significant omissions in the CBTI's content (1987). The present study endeavors to incorporate these recommendations in its design.

WRAMC RESEARCH DESIGN AND SETTING

Overview of Research Methodology

The general methodology for this study is outlined in Fig. 4.1. Hospital patients, whose records met specific test and demographic criteria, were selected to form a research sample, stratified by profile type, and within these constraints selected



^aSee selection criteria in Table 4.2 for matching according to MMPI profile types.

^bSee Table 6 for selection of CBTI companies.

^cThis code type was rated but was lost in the mail.

FIG. 4.1. Methodology for the WRAMC study of the validity of CBTI systems for the MMPI.

randomly. Experienced Army clinical psychologists rated the accuracy of MMPI CBTIs generated by several CBTI systems, using case history materials as the criterion. They all rated two pairs of Black/white cases matched on the basis of profile type, making it possible to obtain data on interrater reliability. Each rater also rated one unique code type for a matched pair of Black/white cases. Thus each clinician rated six subjects: (a) a pair of 7/2 code type cases, (b) a pair of cases without significant elevations on MMPI scales, and (c) a pair representing some different (i.e., not 7/2) code type. Each rater rated each numbered sentence of a CBTI for a subject within the context of each paragraph.

Nature of Hospital Population

Our clinical subjects were drawn from inpatient and out-patient files at Walter Reed Army Medical Center (WRAMC), covering a period of 3 years (1983–1985) during which at least 1,500 MMPIs were administered and interpreted. WRAMC draws its inpatient population from active-duty military personnel or their dependents from the east coast of the United States, Europe, and from a large group of retired military families in metropolitan Washington, D.C., area. Patients referred to WRAMC are evaluated for complex diagnostic or treatment problems (including neurological cases) or for determinations of fitness to continue to serve on active duty.

Inpatients at the hospital may be admitted from the local area or from one of the feeder hospitals within the military system. Requests for psychological evaluation may occur at any point in the course of the patient's stay in the hospital; the majority of the requests for psychological evaluations are made within the first 2 weeks of the patient's admission. At the point of the patient's discharge from the hospital, a narrative summary of the patient's hospitalization is prepared by the treating physician. This summary will include all of the pertinent information gathered on the patient over the course of his stay in the hospital and provides the most comprehensive overview of the patient's status at the time of discharge. Although the time between admission and discharge may vary, depending on the nature of the patient's case, it is not unusual for a 6-month period to exist between admission of the patient and dictation of the narrative summary. Thus, the psychological evaluation may have occurred some months prior to the final narrative summary with intervening events accounting for changes in the patient's status.

Psychological reports are also provided for out-patients who are generally referred to the hospital from nearby military installations. Reports include an evaluation of the patient's salient personality and a diagnostic evaluation of possible psychopathology. Treatment recommendations often are not made.

Neuropsychological evaluations are provided by psychologists, largely for inpatients who have experienced a neurological event. The patient's brain-based functioning is evaluated, salient personality characteristics described, and treatment recommendations made.

Selection of Subjects

The criteria for selecting the Black and white subjects, matched by profile type, for our study of the accuracy of CBTIs, are given in Table 4.2. The subjects were restricted to active-duty males, from 20 to 29 years old, who were inpatients or out-patients in WRAMC from 1983 through 1985 or soldiers who were in the Army's normative study of the MMPI (Fishburne & Parkison, 1984).

Each clinical case selected had (a) a case history, (b) an MMPI answer sheet, and (c) met the raw score criteria, set for the basic validity scales in consultation with W. Grant Dahlstrom. These were: Cannot say ≤ 49 , Lie ≤ 10 , F ≤ 21 . The case history may be an inpatient report, an outpatient report, or a neuropsychological evaluation. The subclinical subjects met criteria (b) and (c) and had completed an anonymous self-report questionnaire covering, for example, military disciplinary actions and treatment for emotional problems.

Subjects were not screened on the basis of their K scores, a measure of clinical defensiveness, because it is not appropriate to use this score for rejecting a total profile. Furthermore, only the scales which deal with clinical syndromes were used to select code types; hence, scores on scales 5 (Masculinity-femininity) and 0 (Social Introversion) were not considered.

The matched pairs of subjects were chosen to maximize the number and nature of MMPI code types in the study. Inpatients and outpatients were included as were psychiatric, medical, and neuropsychological cases. Thirty-three spike and two-point code types involving elevations of $T \geq 70$ were sought, representing a range of frequently occurring code types (see Table 4.3). Black-white cases were matched for four spike profiles and 9 two-point code types.

We began searching for the code types as listed in Table 4.3 by searching the 1983 WRAMC files for the first white case for the first code type, a spike 1 profile. All other code types were ignored until the 1' profile was found. Then we searched until we found the next code type, a 1/2 case. If the code type we were seeking could not be found by going through the 1983 files, we followed

TABLE 4.2
Case Selection Criteria for Black/White Pairs Matched on MMPI Code Type in WRAMC Validity Study

Inpatient or outpatient, Walter Reed Army Medical Center, 1983-1985
Active duty uniformed personnel
Male
Age 20-29
Documented case history/background information
Race: White (Caucasian) or Black non-Hispanic
MMPI Answer Sheet Available
Validity Profile Scores
Cannot say ≤ 49 raw score
Lie ≤ 10 raw score
F ≤ 21 raw score
Among 33 spike profile and two-point code types ($T \leq 70$), using stratified sample or in Army normative study and net above criteria with $T \leq 70$ on clinical scales and had no record of disciplinary actions or inpatient or outpatient treatment for emotional problems.

TABLE 4.3
Spike Profile and Two-Point MMPI Code Types of Matched Black/White Cases
Sought for WRAMC CBTI System Validity Study

1 ^a	12/21 ^a	13/31 ^a	14/41	18/81 ^a	19/91 ^b	
2 ^a	23/32	24/4 ^a	26/62	27/72 ^a	28/8 ^a	29/92
3 ^a	34/43	36/63	38/83	39/93		
4	46/64	47/74	48/84	49/94		
6 ^a	67/76	68/8 ^a	69/96 ^a			
7	78/87	79/97				
8	89/98 ^a					
9						

Note. Systematic search was made for code types listed in Graham (1977), Greene (1980), Lachar (1974), or were present in 1% or greater of two-point code types in Appendix M, Tables 9, 11, 13, and 15 in Dahlstrom, Welsh, and Dahlstrom (1972). Also, two subclinical normal profiles were included in the study.
^a Spike and two-point code types included in the study.

^b Ratings were completed, but the data were lost in the mail.

the same procedures for the 1984 and 1985 files. That cycle continued for the remaining white code types, in the order given in Table 4.3. The same steps were taken in the search for the Black cases, beginning with the 1983 files. Cases that could not be matched by race were discarded.

We followed the same procedures for selecting white and Black subclinical normal cases from the U. S. Army's normative study of the MMPI (Fishburne & Parkison, 1984; Parkison & Fishburne, 1984), which covered active-duty males from age 18 to 33, who were stationed throughout the United States and Europe. All the subjects in the normative study, on the average, had 12 years of education, and were, on the average, 25. The two subclinical normal cases were drawn from the sample of 1,032 subjects who met the MMPI validity criteria, scored ≤ 5 on the Carelessness Scale (Greene, 1980) and had IQ scores of at least 75 on the Shipley Institute of Living Scale. The subjects met these MMPI validity criteria: Cannot say ≤ 29 ; and $F \leq 24$. The 1,930 subjects in the normative study were also screened on a 43-item background information questionnaire, which may be obtained from the authors. Soldiers who reported any of the following background factors were excluded: felony convictions, court-martials, a psychiatric hospitalization, a suicide attempt, psychiatric treatment, or treatment for a drug or alcohol problem. A total of 898 subjects from the normative study were excluded on the basis of test scores or legal, behavioral, or treatment criteria. Most of these subjects were excluded because of invalid test scores.

Background of the Subjects

The background of the 28 subjects reported here is given in Tables 4.4 and 4.5.

The common cases, assigned to all raters, consisted of a pair of Black and white cases from WRAMC matched for the 7/2 code type and a pair of Black and white soldiers from the Army normative study with subclinical, that is, all clinical scales < 70 T MMPI profiles. The common cases had the equivalent of

TABLE 4.4
Background Characteristics of Common Cases Evaluated by All Raters

<i>Code Type</i>	<i>Race</i>	<i>Age</i>	<i>Marital Status</i>	<i>Years of Education</i>	<i>Nature of Subject</i>
72/27	White	20	Single	12	Outpatient psychiatric ^a
72/27	Black	21	Married	12	Inpatient psychiatric
Subclinical normal	White	24	Married	12	Normative study
Subclinical normal	Black	22	Married	12 (GED)	Normative study

^a Involved neuropsychological evaluation.

TABLE 4.5
Background Characteristics of Unique Cases Evaluated by Only One Rater

<i>Rater</i>	<i>Code Type</i>	<i>Background Characteristics</i>					<i>Nature of Subject</i>
		<i>Age</i>	<i>Race</i>	<i>Marital Status</i>	<i>Years of Education</i>		
1	1	29	White	--	--	inpatient	medical
1	1	24	Black	single	12	inpatient	medical
2	2/1	22	White	married	13	outpatient	medical ^a
2	2/1	20	Black	--	--	outpatient	medical
3	1/3	23	White	married	12	inpatient	psychiatric
3	1/3	21	Black	single	13	inpatient	psychiatric
4	8/1	24	White	single	16	inpatient	medical ^a
4	1/8	27	Black	single	12	inpatient	medical ^a
6	2	26	White	married	12	inpatient	psychiatric
6	2	25	Black	single	12	inpatient	medical ^a
7	2/8	23	White	divorced	14	inpatient	psychiatric
7	2/8	27	Black	--	--	inpatient	psychiatric
8	3	26	White	married	--	inpatient	medical
8	3	24	Black	married	12	inpatient	medical ^a
9	4/2	29	White	divorced	--	inpatient	psychiatric
9	4/2	20	Black	single	--	inpatient	psychiatric
10	6	22	White	married	16	inpatient	medical ^a
10	6	27	Black	married	--	inpatient	psychiatric
11	8/6	20	White	married	12	inpatient	psychiatric
11	8/6	21	Black	single	12	inpatient	psychiatric
12	9/6	21	White	--	13	inpatient	medical ^a
12	9/6	24	Black	single	12(GED)	inpatient	psychiatric
13 ^b	8/9	20	White	single	11	inpatient	psychiatric
13	8/9	27	Black	single	--	inpatient	psychiatric

^a Involved neuropsychological evaluation.

^b Materials provided to the rater for this code type included a psychological report written in response to a referral on the patient, which was not congruent with an 8/9 code type.

12 years of education (including G.E.D.), three out of four were married, and they were between 20 and 24. The 7/2 white case was a psychiatric out-patient involved in a neuropsychological evaluation, and the 7/2 Black case was a psychiatric inpatient.

The unique cases consisted of 12 pairs matched for race and MMPI spike or two-point code type. Twelve different MMPI spike or two-point codes were included. Each pair was evaluated by only one rater. The age of the subjects ranged from 20 to 29 years. Ten were single, eight married, two divorced, and data were not available for four subjects. Nine of them had 12 years or its equivalent of education, six had completed 13 to 16 years, one had 11 years education, and data were not available for eight subjects. Of the six subjects with more than 12 years of education, four involved neuropsychological examinations. Seven of the 24 cases involved a neuropsychological evaluation. Twenty-two subjects were inpatients, two out-patients. There were 13 psychiatric cases and 11 medical cases. Included were: (a) 5 code-type pairs which were psychiatric, (b) 4 pairs which were medical cases, and (c) 3 pairs which included a psychiatric and a medical case.

Selection and Nature of CBTI Systems

As of December, 1985, the authors were aware of 14 commercially available CBTI systems for the MMPI (see Table 4.6).

Nine of these systems were invited to participate in the project. The selection of the companies was made *largely* on the basis of the company's expression of interest in attending the 1984 APA-sponsored test publishers' meeting which the first author helped to organize. One company did not reply; eight of these companies agreed to participate. However, one company (Prime Focus' Weathers MMPI Report) later withdrew its software from the project. Thus, seven companies, namely half of the companies, participated. All of the older CBTI systems (Behaviordyne: Report No. 7, Detailed Clinical Report; the Caldwell Report; NCS Minnesota Report: Adult System; and the WPS Test Report) were included. In addition, 3 of the 10 new CBTI Systems (Applied Innovations: MMPI Interpretation, NCS FASTTEST, formerly PSYCH SYSTEMS, MMPI, and Psych Lab: The Morris-Tomlinson Report) participated.

Fowler (1985) has described six CBTI systems for the MMPI, including two earlier systems (Behaviordyne and Caldwell) and two later systems (WPS Test Report and Minnesota Report) covered in this study. The authors requested that each participating CBTI company provide manuals or documentation materials provided to CBTI users. Materials from the companies are cited in this section.

Finney's Detailed Clinical Report, Report No. 7, marketed by Behaviordyne (BE), does not provide a copyright date. Its history can be traced back to the 1960s (Dahlstrom et al., 1972; Finney, 1968; Graham, 1977; Wiggins, 1973) and this CBTI service was reviewed by Adair (1978b), Butcher (1978b), and Sund-

TABLE 4.6
 CBTI Systems for the MMPI: Total N, Invitees, and Participants in the WRAMC
 Validity Study

	<i>Total N. CBTI Systems</i>	<i>CBTI System</i>	<i>CBTI Systems</i>
		<i>Invitees</i>	<i>Participants in Study</i>
1	Applied Innovations: MMIP Interpretation	Y	Y
2	Behaviordyne: Report No. 7 (Detailed Clinical Report)	Y	Y
3	Caldwell Report	Y	Y
4	Integrated Professional System: MMPI Software	Y	N
5	International Information Systems: The MMPI Test	N	N
6	NCS Minnesota Report: Adult System	Y	Y
7	Morris-Tomlinson Report (PSYCH LAB)	Y	Y
8	NCS FASTTEST (formerly PSYCH SYSTEMS) MMPI	Y	Y
9	Precision People: MMPI Computer Report	N	N
10	Psychological Assessment Resources: The MMPI Interpretive System	N	N
11	Psychometric Software: MMPI Report Computer Program	N	N
12	Sienna Software: PSYCHSTAR	N	N
13	Prime Focus: Weathers MMPI Report	Y	ya
14	WPS Test Report	Y	Y

Note. Y = yes; N = no.

^a CBTI system software was withdrawn by CBTI company.

berg (1985a). The Behaviordyne Reports are based on Finney's norms, using normalized *T* distributions. Fowler (1985) notes Finney's reports, which are somewhat psychoanalytically oriented, are based on interpretations of the basic clinical scales, configurations and scores from his special scales. Behaviordyne incorporates Finney's idiosyncratic approach to MMPI interpretation. Behaviordyne uses the DSM-III classification system (American Psychiatric Association, 1980). Diagnostic impressions are listed according to the label most likely to fit the subject.

The Caldwell system, marketed as the Caldwell Report, also does not contain a copyright date. According to Fowler (1985, p. 750), this early developed system “is a highly configural simulation of Caldwell’s own interpretive style.” It is based on a large number of code types (A. B. Caldwell, personal communication, April 4, 1986). A single narrative statement “describing someone with a ‘49-94’ code as ‘manipulative, dramatizing, and acting out’ might well have five to ten different validation sources for each of the three terms, and those sets of sources would be partially overlapping.” Caldwell refers CBTI users to studies such as Chase’s (1974) dissertation, which has been reviewed by Moreland (1985). Caldwell’s system has been described and reviewed by Dahlstrom et al. (1972), Graham (1977), Adair (1978c), Butcher (1978c), and Greene (1980). The Caldwell Report features sections on treatment planning, early-childhood correlates of profile types, and alternative diagnoses (which are listed in rank order “in terms of probability of fit”) (A. B. Caldwell, personal communication, April 4, 1986). Caldwell reported that he was converting the Caldwell Report to the American Psychiatric Association’s (1980) *Diagnostic and Statistical Manual* (DSM–III) from its second edition.

The Lachar system, marketed as the WPS Test Report (WPS), is described in great detail in Lachar’s (1974) manual. Its report has a 1979 copyright date. The manual reports on the research samples on which the CBTI descriptions for many code types are based. One section of the manual gives the algorithms used in the CBTI. Lachar reports, for example, on the subroutines used to generate 14 possible narrative statements to interpret the validity of the profiles. Adair (1978a), Butcher (1978a), and Sundberg (1985b) have reviewed the WPS Test Report.

The Minnesota Report (MN), authored by Butcher, was developed in the late 1970s and has a 1982 copyright date (Fowler, 1985). National Computer Systems (1982) has issued a user’s guide which includes descriptions of the scales used and gives some cutoff scores. Butcher bases his interpretations on code types, individual scales, and on special scales. Butcher and Keller (1984, p. 317) describe this system as one which

Tailors interpretive statements according to the subject’s population (mental health outpatient or inpatient, medical, adult, correctional, personnel, or college counseling) and according to demographic data such as education, marital status, and ethnicity, which research has shown to be modifiers of interpretive rules.

The Morris–Tomlinson Report (MT) is a CBTI system with a 1983 copyright date which was prepared by Leon M. Morris and Jack R. Tomlinson and is marketed by Psych Lab. Psych Lab provides CBTI users with a form letter which points out that the CBTI system makes use of the normalized T-scores reported by Colligan et al. (1983). The Morris–Tomlinson Report is based on the DSM–III terminology and the reports “Frequently include statements regarding the

patient's social, vocational, and academic functioning as well as statements related to assertiveness and the forensic implications of test findings."

The NCS FASTTEST (FT) interpretation for the MMPI, copyrighted 1984, is one of 30 assessment tools that Psych Systems originally marketed as part of a system combining hardware and software for interpreting psychological instruments. The assets of Psych Systems, including the FASTTEST, were purchased by National Computer Systems (Fowler, 1985). The users of the NCS FASTTEST (National Computer Systems, undated) received brief documentation materials and reprints such as Miller, Johnson, Klingler, Williams, and Giannetti (1977). The system continues to be available to the original users, but is no longer sold to new users. The FT promotional materials provide the following information:

Psych Systems uses five different interpretive schemes: it first checks to see if the profile generated is a well known code type. If so, it prints an interpretation based on the profile configuration. If a well known code type is not found and the patient is a male, the program checks to see if there are any elevated scale scores. If there are, it uses linear combinations of scale scores to arrive at both predictive and descriptive statements about the patient. . . . If the profile falls within normal limits, regardless of sex, then a series of special scale interpretations are used to generate an interpretive statement. The emphasis with normal profiles is to interpret results in terms of social relationships, vocational issues, and problems of health behavior.

FT makes use of interpretations based on Gilberstadt and Duker (1965), Stelmacher's interpretations of code types (cf. Lachar, 1974), and "linear regression equations developed by Bloch (1983) relating to Johnson, Butcher, Null and Johnson's (1984) MMPI factor scales."

Applied Innovations (AI) has in the past marketed an MMPI CBTI system, developed by Bruce Duthie, copyright 1984. It is still available to interested purchasers. Recently, AI has also marketed the Marks Adult MMPI Report. This company provided CBTI users with a manual (Duthie, 1985) which addresses the operation of the system. Duthie (1985) reported that Applied Innovations:

Consider the MMPI Computerized Interpretation Manual to be an application of artificial intelligence. Specifically it is designed to be an expert system for interpreting the MMPI. One of the major criteria of an expert system is that the decision theory be open to scrutiny. This manual explains the decision theory by which individual statements within the software are included in the report. The clinician can establish the clinical validity of any statement as it relates to a particular patient. See the appendices for a list of all possible statements and trigger codes generated by this software. . . . Our philosophy in the MMPI Computerized

Interpretation Manual is to totally illuminate the contents of the black box. (p. 5, subsection 3.22)

AI based its diagnostic suggestions on the DSM-III.

Selection and Background of Raters

The raters were nominated by the third author, who is familiar with the training and experience of the approximately 130 psychologists engaged in clinical assessment in Army facilities throughout the United States and Europe. Thirteen raters were chosen to participate in the project. All raters completed the ratings. However, rating data completed by Rater 5, who was assigned to rate the 1/9 code type, were lost in the mail (Table 4.3).

The 12 raters from whom rating data were received were generally representative of Army clinicians who use the MMPI. They were stationed throughout the United States and Germany. These clinicians, who were employed in an Army mental health function, were white, non-Hispanic men, who had completed clinical internships approved by the American Psychological Association, and were licensed to practice psychology. They had 7.5 median years of postdoctoral experience in clinical psychology. Half of them had worked at WRAMC. Three-fourths had completed a doctorate in clinical psychology and one held a Diplomate in Clinical Psychology awarded by the American Board of Professional Psychology.

Eleven of the 12 raters listed the MMPI reference sources they used. All of them listed Lachar's 1974 manual, which includes the algorithms for the WPS Test Report, and which is regularly used in the Army's clinical training programs. Seven raters reported using Greene's (1980) book, 4 used the Dahlstrom et al. (1972) text, and three listed Graham's (1977) book.

Eleven of the 12 raters listed their experience in using the seven CBTI systems. Five of 11 had no experience with any of the CBTI systems for the MMPI. Of the 6 raters with CBTI experience, 3 had some experience and 2 had extensive experience in using the Minnesota Report; 3 had some experience in using the NCS FASTTEST (formerly PSYCH SYSTEMS), two listed some experience with Applied Innovations, and one reported some experience in using the WPS Test Report.

Rater Materials and Instructions

Input of Answer Sheet Data. The CBTIs were generated from MMPI hand scored and National Computer Systems (NCS) scannable (mark sense) answer sheets from the subjects' files. In order to minimize scoring errors stemming from erasures and variations in the neatness and darkness of marked answer

sheets (see, e.g., Grayson & Backer, 1972), we developed a list of potentially problematically marked answers and standardized their interpretation.

Our procedures for handling the data input depended on the preferences and procedures employed by the CBTI service companies. We provided an interpretation of ambiguous answers to those handling the input of the data. For the AI, MT, and FT data, the authors themselves keyed the item data into the computer, using the software provided by the company. The item data for the CA system were keyed in by the staff of the Caldwell Report. For the BE and WPS systems, the authors transcribed the data onto the answer sheets used by each CBTI company. Where possible, NCS answer sheets were scanned by optical mark reader by NCS. For the remaining subjects, who had used hand-scored answer sheets, we transcribed their answers onto NCS scannable forms. NCS provided a check on the accuracy of the transcribed data by keying in the item data themselves. (The authors received these backup data after they mailed the CBTIs to the raters.)

The authors checked the accuracy of the raw scores for the subjects' MMPI by comparing them as they appeared in the printouts for all the CBTI systems, except those from the BE and MT systems. Behaviordyne does not provide raw scores, but does include the publisher's T-scores based on the Minnesota normals. The Morris-Tomlinson Report reports raw scores and normalized T-scores based on the normative group reported by Colligan et al. (1983). In spite of efforts to minimize raw score variations, minor discrepancies did occur. Small raw score differences have been routinely reported in the research literature on the accuracy of computer scoring of the MMPI (cf. Fowler & Coyle, 1968; Grayson & Backer, 1972; Klett, Schaefer, & Plemel, 1985; Weigel & Phillips, 1967).

Rater Instructions. The raters completed research forms given in Appendix A. The entire narrative, with attachments, was used exactly as it was sold to CBTI users with the company's name identified. The format and editorial style of each CBTI was distinctive. The authors numbered every sentence for each CBTI system, with the exception of those in footnotes. Raters were instructed to rate each numbered narrative statement for the cases in a prescribed order. They began by rating the two common-matched 7/2 code type cases. The 7/2 code type was chosen because it is a two-point code type which appears frequently (Greene, 1980); it is considered to be among the most accurate code types for making diagnoses (see, for example, Hathaway & Meehl, 1951, Tables XVI-XIX). This code type has generated numerous external correlates (Greene, 1980).

All cases were presented to the raters in counterbalanced order by race. The raters received the instructions given in Appendix A and their material was arranged in the prescribed order. The data for each subject included his specially developed identification number, age, race, marital status, educational level, and a description of the source of the subject (inpatient, out-patient, or normative

study), as listed in Tables 4.4 and 4.5. The raters were not provided with a listing of the medical, psychiatric, or neuropsychological nature of the cases. A subject was classified as medical if the referral came from a nonpsychiatric physician and psychiatric if referred by a psychiatrist. Cases involving a neuropsychological evaluation were identified by the third author; for most of these cases, test score data from a neuropsychological battery of tests were available to raters.

The seven CBTI systems included in the study are listed alphabetically in Fig. 4.1 and were described earlier. Each rater received the printouts for all seven CBTI systems for each assigned case. The CBTI system printouts used were identical to the ones offered by CBTI companies, with each narrative statement numbered to facilitate the ratings. Although the order in which the 13 raters evaluated the CBTI systems was constant (alphabetical as in Fig. 4.1), they started at different points in the list.

The instructions to raters are summarized in Fig. 4.1 (see Appendix A). The rater started by rating individual narrative statements for a CBTI for his first case. Then he evaluated specific features of the CBTI system for the case: (a) overall accuracy of the diagnosis, (b) overall accuracy of the CBTI, and (c) helpfulness of the CBTI system in the disposition of the case, that is, in diagnostic evaluation and in disposition planning. He then repeated these steps for each CBTI system for the first case. These steps were repeated for each case. After all six cases were rated, the rater completed the Final Rating Sheet (Appendix A), in which he ranked the CBTI systems according to overall accuracy. Then he ranked them in terms of their overall helpfulness to the clinician in disposition planning.

Raters were provided with a description of the Colligan et al. (1983) normative study on which the Morris–Tomlinson Report is based because it reports on a recent restandardization effort.

RESULTS

The thesis of this study is that CBTI systems vary in overall relevancy and accuracy, when case histories (or self-report questionnaire) are used as a rating criterion. We will begin by presenting the overall judgments of accuracy although the raters made these judgments after having evaluated the sentence-by-sentence accuracy of individual narrative statements from the printouts (see Fig. 4.1). Global and specific accuracy ratings and indicators of their reliability are given in Tables 4.7 to 4.10.

The manner of analyzing the relevancy and accuracy of each narrative statement is indicated in Tables 4.11 and 4.12. Descriptions of the pooled data (across CBTI companies and raters) for the common cases are presented in Tables 4.13 and 4.14 and Appendix B. The *N*s given in these tables refer to the number of percentages involved in the pooled data. Specific data on each rater's evaluation

TABLE 4.7
Final Rank-Order Ratings and Coefficient of Concordance for Overall Accuracy of MMPI CBTI System by Raters of All Cases

CBTI System	Rank Order of CBTI System Rater Number												Median Rank for CBTI System
	1	2	3	4	6	7	8	9	10	11	12	13	
AI	2	6	5	6	4	6	5	4	5	2	6	5	(5)
BE	6	7	7	7	6	7	7	7	7	7	7	7	(7)
CA	3	5	2	5	3	4	4	2	3	3	3	1	(3)
MN	1	4	3	1	1	2	1	1	1	1	1	2	(1)
MT	7	1	6	4	7	3	6	6	6	6	2	6	(6)
FT	4	2	4	2	2	1	3	5	4	4	5	3	(4)
WPS	5	3	1	3	5	5	2	3	2	5	4	4	(3.5)

Note. $W = .60$, Chi square = 43.3, *** $df = 6$, *** $p < .001$.

TABLE 4.8
Final Rank Order Ratings and Coefficient of Concordance for all Cases in Overall Helpfulness of MMPI CBTI System by Raters in Case Disposition

CBTI System	Rank Order of CBTI System Rater Number												Median Rank for CBTI System
	1	2	3	4	6	7	8	9	10	11	12	13	
AI	4	6	6	6	4	6	5	5	5	2	4	5	(5)
BE	5	7	7	7	7	7	6	6	7	7	7	7	(7)
CA	3	5	2	5	3	3	2	2	3	3	2	1	(3)
MN	1	4	3	1	1	2	1	1	1	1	1	2	(1)
MT	6.5	1	5	4	6	4	7	7	6	6	6	6	(6)
FT	2	2	4	2	2	1	3	3	4	4	3	3	(3)
WPS	6.5	3	1	3	5	5	4	4	2	5	5	4	(4)

Note. Case disposition, i.e., diagnostic evaluation and disposition planning.
 $W = .66$, Chi square = 47.5***, $df = 6$, *** $p < .001$.

TABLE 4.9
Intraclass Correlation Among 12 Raters for Accuracy Ratings for Each Common Case and Across Each CBTI System

Profile Type and Race	Intraclass Correlation Among 12 Raters Across Each CBTI (r_{cc})	Intraclass Correlation of an Average of 12 Ratings for Each CBTI System (r_{tk})
7/2 White	.49	.92
7/2 Black	.44	.90
Subclinical Normal White	.49	.92
Subclinical Normal Black	.16	.70

Note. Based on 3-point overall accuracy ratings: 1 = generally inaccurate; 2 = somewhat accurate, and 3 = generally accurate.

TABLE 4.10
Frequency of Specific Overall Ratings of Accuracy of Diagnostic Statements by
Code Type for CBTI Systems Across Raters

<i>CBTI Systems</i>	<i>7/2</i>					<i>Subclinical Normal</i>					<i>Unique</i>				
	<i>0</i>	<i>1</i>	<i>2</i>	<i>3</i>	<i>Omit</i>	<i>0</i>	<i>1</i>	<i>2</i>	<i>3</i>	<i>Omit</i>	<i>0</i>	<i>1</i>	<i>2</i>	<i>3</i>	<i>Omit</i>
AI	2	15	4	1	2	6	1	3	14	0	1	10	7	6	0
BE	0	12	11	1	0	0	16	7	1	0	1	10	12	0	1
CA	1	2	11	10	0	3	12	9	0	0	2	8	5	5	4
MN	0	2	11	9	2	9	1	5	9	0	3	3	9	7	2
MT	6	7	8	3	0	13	0	2	8	1	12	2	7	3	0
FT	8	5	9	2	0	10	2	2	9	1	6	5	8	5	0
WPS	3	4	11	6	0	11	7	4	2	0	8	2	10	3	1

Note. 0 = CBTI system does not provide a diagnostic evaluation. 1 = inaccurate; 2 = somewhat accurate; 3 = accurate; omit = item omitted by rater.

TABLE 4.11
Example of Frequency Distribution and Percentages for One Rater Evaluating
One CBTI System Using All Rating Categories for One Subject

N = 41 Narrative Statements

<i>Rating Categories for all Narrative Statements</i>	<i>Frequency</i>	<i>Percentage</i>
1. Data insufficient to make a rating.	1	2
2. Generally applicable or repetitive statement. ^a	5	12
3. Inaccurate narrative statement.	13	32
4. Somewhat accurate narrative statement.	17	41
5. Accurate narrative statement.	5	12
Total	41	99

^a Statement does not contribute to the understanding of the case.

TABLE 4.12
Example of Frequency Distribution and Percentages for One Rater Evaluating
Relevancy and Accuracy for One CBTI System for One Subject

N = 41 Narrative Statements

<i>Rating Categories for Accuracy of CBTI</i>	<i>Frequency</i>	<i>Percentage</i>
<i>Statements Relevant to Case</i>	32 ^a	78
(3) Inaccurate	13	41 ^b
(4) Somewhat Accurate	17	53 ^b
(5) Accurate	2	6 ^b

^a Data reported for rating categories (3), (4), and (5) as in Table 4.11 with three validity profile (VP) statements eliminated from category (5), according to formula: $(\sum N - (1) - (2) - VP)$.

^b Percentage of relevant statements.

TABLE 4.13
Median Percentage for Common Cases and Ratings of Narrative Statements
Across MMPI CBTI Systems and Raters
(N = 168)

<i>Rating Categories</i>	<i>Common Cases</i> <i>7/2</i>	<i>Subclinical Normal</i>
Unratable (1)	14	50
General Repetitive (2)	10	12
Inaccurate (3)	11	00
Somewhat Accurate (4)	27	12
Accurate (5)	26	14

TABLE 4.14
Median Percentage for Common Cases and Ratings of Relevancy and Accuracy
Across MMPI CBTI Systems and Raters
(N = 168)

<i>Rating Categories^a</i>	<i>Common Cases</i> <i>7/2</i>	<i>Subclinical Normal</i>
Relevant to Case ^b	67	25
Inaccurate (3)	19	00
Somewhat Accurate (4)	42	43
Accurate (5)	35	33

^a (3) + (4) + (5) - validity profile statements/ (1) + (2) + (3) + (4) + (5).

^b (3) or (4) or (5)/ (3) + (4) + (5); except for validity profile statements.

of each subject's CBTIs for all cases are provided in Tables 4.15, 4.16, and 4.22. We conclude by reviewing the data related to the relevancy and accuracy of the CBTI systems for all cases (Tables 4.17 to 4.19). The ratings of the extent to which the CBTIs for the common cases were evaluated as relevant and accurate by each rater are used to evaluate their ratings of the unique cases (Tables 4.20, 4.22, and 4.23). Table 4.21 reports similar data for the subnormal clinical cases.

Rater Reliability

The raters assigned a final overall rank order score to each CBTI system after evaluating the overall accuracy for six cases, including two matched Black/white pairs (7/2 and subclinical normal profile) and one unique code-type pair (Fig. 4.1, Instruction step 7; Table 4.7). Kendall's Coefficient of Concordance ($W = .60$, chi square = 43.3, $df = 6$, $p < .001$), a special analysis-of-variance method revealed the highly significant extent to which the 12 raters agreed (see Guilford & Fruchter, 1973, pp. 264–266). The median ranks (across raters) of the CBTI systems showed that the MN Report was rated the highest in overall accuracy and the BE system was the lowest. The raters agreed the most in ranking BE and MN and agreed the least in ranking AT, FT, and WPS. Similar results were found in the ratings of the overall helpfulness of the CBTI system for case disposition, which includes the diagnostic evaluation and disposition planning (Fig. 4.1, Instruction step 7; Table 4.8); here Kendall's Coefficient of Concordance was $W = .66$, chi square = 47.5, $df = 6$, $p < .001$. This again demonstrates that there was significant agreement among raters.

Further evidence of interrater reliability was obtained from specific ratings made for each CBTI system for the common cases, using Specific Answer Sheet item 8 (see Appendix A and Fig. 4.1, Instruction to Raters Step 4). Intraclass correlations for each profile type by race (7/2 white, 7/2 Black, subclinical normal white, and subclinical normal Black) were based on three-point ratings of overall accuracy made by each rater for each CBTI system. The analysis was based on the variance between CBTI systems, using the overall accuracy ratings of 12 raters to compute correlations between raters. Intraclass correlations (Table 4.9) for each profile type, analyzed by race, showed the typical intercorrelation for 12 raters. One can say the typical reliability for a single rater's ratings, for three cases, 7/2 White, 7/2 Black, and subclinical normal White was similar ($r_{cc} .44 - .49$), but lower ($r_{cc} = .16$) for the subclinical normal Black case (Guilford & Fruchter, 1973, pp. 263–264). If we averaged the evaluations of the raters for each CBTI system and could correlate this set of averages with a set of comparable ratings from a similar set of raters, the range of the intraclass correlations would be $r_{kk} = .70 - .92$.

The data on rater reliability show considerable rater agreement on the final overall rank order for evaluating the accuracy of the output of seven CBTI systems. Furthermore, the raters showed significant agreement in rating the overall accuracy for the CBTI systems for each of the following three cases: 7/2

TABLE 4.15
Summary of Chi-Square Data for Rater Evaluations of Relevancy and Accuracy of
Common Cases by CBTI Systems

<i>7/2 White Case</i>				
<i>Chi Square A</i>			<i>Chi Square B</i>	
<i>df = 24 N = 541</i>			<i>df = 18 N = 541</i>	
<i>Rater</i>	<i>Table</i>		<i>Table</i>	
1	1	> 148.5***	2	119.6***
2	3 ^a	110.1***	4	68.3***
3	5	75.7***	6	49.7***
4	7	> 227.9***	8	> 164.7***
6	9	> 151.4***	10	97.5***
7	11	> 138.3***	12	129.7***
8	13	71.8***	14	56.1***
9	15	96.8***	16	64.1***
10	17	83.6***	18	51.7***
11	19	> 138.3***	20	107.1***
12	21 ^a	95.5***	22	62.4***
13	23	69.7***	24	53.2***

<i>7/2 Black Case</i>				
<i>Chi Square A</i>			<i>Chi Square B</i>	
<i>df = 24 N = 502</i>			<i>df = 18 N = 502</i>	
<i>Rater</i>	<i>Table</i>		<i>Table</i>	
1	25 ^a	102.3***	26	76.5***
2	27 ^b	57.1***	28	41.4***
3	29 ^b	104.9***	30	75.2***
4	31 ^b	79.3***	32	65.2***
6	33 ^b	101.1***	34	71.0***
7	35 ^b	133.1***	36	114.2***
8	37 ^c	81.3***	38	60.7***
9	39 ^b	77.1***	40	61.6***
10	41 ^b	67.5***	42	48.8***
11	43 ^c	102.4***	44	71.3***
12	45 ^a	81.0***	46	59.1***
13	47 ^b	108.4***	48	71.1***

(continued)

Subclinical Normal White Case

<i>Rater</i>	<i>Chi Square A</i> <i>df = 24 N = 297</i>		<i>Chi Square B</i> <i>df = 18 N = 297</i>	
	<i>Table</i>		<i>Table</i>	
1	49 ^a	48.4**	50	28.5*
2	51	60.0***	52	33.0**
3	53	90.6***	54	52.6***
4	55 ^a	85.2***	56	48.9***
6	57 ^a	76.5***	58	45.8***
7	59	62.6***	60	40.2**
8	61 ^a	55.6***	62	24.3
9	63	48.5**	64	29.4*
10	65	68.4***	66	36.4**
11	67 ^a	97.1***	68	74.6***
12	69	Analysis not appropriate due to empty cells	70	Analysis not appropriate due to empty cells
13	71	27.1	72	22.2

Subclinical Normal Black Case

<i>Rater</i>	<i>Chi Square A</i> <i>df = 24 N = 313</i>		<i>Chi Square B</i> <i>df = 18 N = 313</i>	
	<i>Table</i>		<i>Table</i>	
1	73 ^a	78.7***	74	55.5***
2	75 ^b	39.8*	76	16.9
3	77 ^a	94.5***	78	59.4***
4	79	37.2*	80	16.5
6	81	66.1***	82	23.2
7	83	64.7***	84	38.5**
8	85 ^a	39.3*	86	22.1
9	87 ^a	31.5	88	10.9
10	89 ^b	58.1***	90	43.9***
11	91	94.9***	92	68.3***
12	93	Analysis not appropriate due to empty cells	94	Analysis not appropriate due to empty cells
13	95	31.6	96	13.8

Note. The rating categories for chi square A are: unratable (1), general repetitive (2), inaccurate (3), somewhat accurate (4), and accurate (5). In Chi Square-B, the three accuracy ratings are: inaccurate (3), somewhat accurate (4), and accurate (5). The irrelevant category includes items evaluated as unratable (1), general/repetitive (2), and validity profile statements.

^a Rater omitted one statement.

^b Rater omitted two statements.

^c Rater omitted three statements.

*** $p < .001$; ** $p < .01$, * $p < .05$.

TABLE 4.16
Summary of Chi-Square Data for Rater Evaluations of Relevancy and Accuracy of Unique Cases by CBTI Systems

<i>Unique White Cases</i>								
<i>Rater</i>	<i>Chi Square A df = 24</i>			<i>Chi Square B df = 18</i>				
	<i>Table</i>	<i>Code</i>	<i>N</i>		<i>Table</i>	<i>Code</i>	<i>N</i>	
1	97	1	400	100.3***	98	1	400	61.2***
2	99 ^a	1/2	395	36.8*	100	1/2	386	20.6
3	101	1/3	540	95.6***	102	1/3	540	65.5***
4	103	8/1	382	81.0***	104	8/1	382	32.3**
6	105	2	410	139.0***	106	2	410	127.8***
7	107 ^c	2/8	596	136.9***	108	2/8	599	104.9***
8	109	3	339	66.1***	110	3	339	33.6***
9	111	4/2	628	104.4***	112	4/2	628	75.0***
10	113	6	319	90.3***	114	6	319	44.4***
11	115 ^a	8/6	487	194.6***	116	8/6	488	150.8***
12	117	9/6	434	93.7***	118	9/6	434	74.5***
13	119	8/9	401	134.5***	120	8/9	401	66.1***

<i>Unique Black Cases</i>								
	<i>Chi Square A df = 24</i>			<i>Chi Square B df = 18</i>				
1	121	1	340	73.8***	122	1	340	40.1**
2	123	1/2	628	116.7***	124	1/2	628	92.2***
3	125	1/3	461	119.2***	126	1/3	461	91.4***
4	127	8/1	499	167.7***	128	8/1	499	115.6***
6	129	2	337	58.8***	130	2	337	20.9
7	131	2/8	638	100.2***	132	2/8	638	67.9***
8	133	3	366	50.3**	134	3	367	33.6**
9	135	4/2	468	114.8***	136	4/2	468	71.5***
10	137	6	390	83.6***	138	6	390	23.0
11	139 ^b	8/6	563	60.2***	140	8/6	565	28.8*
12	141	9/6	653	131.8***	142	9/6	653	91.6***
13	143	8/9	435	83.4***	144	8/9	435	40.8**

Note. The rating categories for Chi Square A are: unratable (1), general/repetitive (2), inaccurate (3), somewhat accurate (4), and accurate (5). In Chi Square B, the three accuracy ratings are: inaccurate (3), somewhat accurate (4), and accurate (5). The irrelevant category includes items evaluated as unratable (1), general/repetitive (2), and validity profile statements.

^a Rater omitted 1 statement.

^b Rater omitted 2 statements.

^c Rater omitted 3 statements.

*** $p < .001$, ** $p < .01$, * $p < .05$.

white, 7/2 Black, and the subclinical white case. An additional group of comparable raters would be likely to show agreement with these raters. Interrater reliability was lower for the subclinical black case.

Overall Accuracy of Diagnostic Statements

In order to interpret these data, it is useful to understand how different CBTI systems present diagnostic statements. We will use the 7/2 white case as an example. Only four CBTI systems had separately identified sections which con-

TABLE 4.17
Median Percentages for 7/2 Code Type for Ratings of Relevancy and Accuracy of
MMPI CBTI Systems Across Raters and Subjects
(N = 24)

<i>CBTI Systems</i>	<i>Relevant^a</i>	<i>Rating Categories</i>		
		<i>Inaccurate (3)^b</i>	<i>Somewhat Accurate (4)^b</i>	<i>Accurate (5)^b</i>
AI	68	38	33	24
BE	55	21	49	26
CA	66	04	50	46
MN	77	06	42	49
MT	60	24	34	40
FT	70	24	42	31
WPS	68	20	46	33

TABLE 4.18
Median Percentages for Subclinical Normal Cases for Ratings of Relevancy and
Accuracy of MMPI CBTI Systems Across Raters
(N = 24)

<i>CBTI Systems</i>	<i>Relevant^a</i>	<i>Rating Categories</i>		
		<i>Inaccurate^b (3)</i>	<i>Somewhat Accurate^b (4)</i>	<i>Accurate^b (5)</i>
AI	26	00	33	50
BE	22	24	44	36
CA	26	26	42	18
MN	19	00	54	38
MT	36	00	33	33
FT	32	00	50	38
WPS	20	10	43	08

^a (3) + (4) + (5) - validity profile statements / (1) + (2) + (3) + (4) + (5).

^b (3) or (4) or (5) / (3) + (4) + (5); except for validity profile statements.

TABLE 4.19
Median Percentages for Unique Cases for Ratings of Relevancy and Accuracy of
MMPI CBTI Systems Across Raters
(N = 24)

<i>CBTI Systems</i>	<i>Relevant^a</i>	<i>Rating Categories</i>		
		<i>Inaccurate^b (3)</i>	<i>Somewhat Accurate^b (4)</i>	<i>Accurate^b (5)</i>
AI	60	33	40	22
BE	45	32	48	22
CA	52	12	40	34
MN	64	08	34	40
MT	57	09	50	33
FT	60	17	48	24
WPS	54	13	42	40

Note. Covers following code types: 1, 1/2, 1/3, 8/1, 2, 2/8, 3, 4/2, 6, 8/6, 9/6, and 8/9.

^a (3) + (4) + (5) - validity profile statements / (1) + (2) + (3) + (4) + (5).

^b (3) or (4) or (5) / (3) + (4) + (5); except for validity profile statements.

TABLE 4.20
Median Percentages for 7/2 Code Type for Relevancy and Accuracy by Rater
Across MMPI CBTI Systems
(N = 14)

<i>Rater</i>	<i>Rating Categories</i>			
	<i>Relevant^a</i>	<i>Inaccurate^b</i> (3)	<i>somewhat accurate^b</i> (4)	<i>accurate^b</i> (5)
1	66	43	49	08
2	57	20	44	32
3	81	08	36	50
4	61	26	58	13
6	59	28	44	22
7	71	16	32	42
8	36	20	46	28
9	76	08	41	42
10	64	36	31	30
11	76	14	44	40
12	50	07	44	44
13	64	08	32	54

a (3) + (4) + (5) - validity profile statements/ (1) + (2) + (3).+ (4) + (5).

b (3) or (4) or (5) / (3) + (4) + (5); except for validity profile statements.

TABLE 4.21
Median Percentages for Subclinical Normal Cases for Ratings of Relevancy and
Accuracy by Rater Across MMPI CBTI Systems
(N = 14)

<i>Rater</i>	<i>Rating Categories</i>			
	<i>Relevant^a</i>	<i>Inaccurate^b</i> (3)	<i>somewhat accurate^b</i> (4)	<i>accurate^b</i> (5)
1	40	26	68	04
2	34	00	83	13
3	53	00	36	52
4	28	00	62	32
6	22	24	34	31
7	56	28	25	47
8	06	00	00	00
9	12	00	50	42
10	09	40	00	31
11	46	00	70	16
12	06	00	14	86
13	16	14	00	68

a (3) + (4) + (5) - validity profile statements/ (1) + (2) + (3).+ (4) + (5).

b (3) or (4) or (5) / (3) + (4) + (5); except for validity profile statements.

TABLE 4.22
Median Percentages for Unique Cases for Ratings of Relevancy and Accuracy
by Rater Across MMPI CBTI Systems
(N = 14)

Rater	Code Type	Rating Categories			
		Relevant ^a	Inaccurate ^b (3)	Somewhat Accurate ^b (4)	Accurate ^b (5)
1	1	54	13	63	10
2	1/2	38	12	72	08
3	1/3	64	20	45	36
4	8/1	56	28	65	14
6	2	58	15	48	37
7	2/8	64	08	24	68
8	3	31	39	37	16
9	4/2	74	18	38	42
10	6	52	40	28	28
11	8/6	55	25	40	22
12	9/6	38	04	49	28
13	8/9	64	12	19	69

^a (3) + (4) + (5) - validity profile statements/ (1) + (2) + (3) + (4) + (5).

^b (3) or (4) or (5)/ (3) + (4) + (5); except for validity profile statements.

TABLE 4.23
Rank Order of Median Percentages and Accuracy Ratings for 7/2 and Unique
Code Types by Rater Across CBTI Systems

Rater	Rating Categories							
	Relevant		Inaccurate		Somewhat Accurate		Accurate	
	7/2	Unique	7/2	Unique	7/2	Unique		
1	5	8	1	8	2	3	12	11
2	10	10.5	5.5	9.5	5.5	1	7	12
3	1	3	10	5	9	6	2	5
4	8	6	4	3	1	2	11	10
6	9	5	3	7	5.5	5	10	4
7	4	3	7	11	10.5	11	4.5	2
8	12	12	5.5	2	3	9	9	9
9	2.5	1	10	6	8	8	4.5	3
10	6.5	9	2	1	12	10	8	6.5
11	2.5	7	8	4	5.5	7	6	8
12	11	10.5	12	12	5.5	4	3	6.5
13	6.5	3	10	9.5	10.5	12	1	1

Note. Highest median percentages are assigned rank order of 1. Kendall's tau (Siegel, 1956) for rating categories: Relevant = .60**; Inaccurate = .36*; Somewhat Accurate = .61**; Accurate = .54**.

***p* < .01.
 **p* < .05.

tain diagnostic statements. AI provides a brief section on Alcohol and Drugs; BE has a Diagnostic Impression section which provides alternative DSM-III diagnostic labels; CA provides a Diagnostic Impression section which briefly presents the primary and secondary diagnosis; and MN provides a Diagnostic Considerations section describing possible diagnoses and symptoms. FT has a section on Special Medical Symptoms. MT does not use subheadings and WPS limits its headings to Comments, Critical Items, and Supplemental Scale Interpretation. Thus, raters evaluating diagnostic statements must use considerable judgment in locating these statements and in making overall evaluations of numerous—possibly discrepant—diagnostic statements.

Data from ratings of the overall accuracy of the diagnostic evaluation (see Specific Answer Sheet item 4, Appendix A; Fig. 4.1, Instruction step 4) reveal differences across CBTI systems for the 7/2, subclinical normal, and unique profile types (Table 4.10). With the Black and white cases combined for each profile type, there are altogether 24 evaluations for each type. For CBTI systems which were judged to provide a diagnostic evaluation, accuracy was rated using a three-point scale. The MT system was least likely to provide diagnostic evaluations for all cases.

The accuracy of diagnostic evaluations was determined by analyzing CBTI systems with the highest number of accurate evaluations and the lowest number of inaccurate evaluations. For the 7/2 code type, data for the CA and MN systems show that 9 to 10 evaluations of their diagnostic statements were rated as accurate and only two evaluations for each system were rated as inaccurate. Conversely, the majority of the evaluations of the AI ($N = 15$) and BE ($N = 12$) systems were rated inaccurate for the 7/2 type and only one evaluation for each system was rated as accurate.

For the subclinical normal cases, AI received 14 evaluations rated accurate, and only one was rated as inaccurate; whereas BE received 16 inaccurate evaluations and only one accurate. Three companies (MT, FT, and WPS) received 10 to 13 evaluations that the CBTI system did not provide a diagnostic evaluation.

For the unique code types, MN had the largest number of accurate evaluations ($N = 7$) and relatively few inaccurate ($N = 3$) evaluations. BE received no accurate evaluations and 10 inaccurate evaluations. AI received 6 accurate evaluations; it also received 10 inaccurate evaluations.

These evaluations of diagnostic statements show similarities between the results for the 7/2 and unique cases. MN received the highest number of accuracy ratings and the lowest number of inaccuracy ratings for all of the clinical cases. AI and BE received a low number of accuracy ratings and a high number of inaccuracy ratings for the clinical cases. BE also showed this pattern for the subclinical normal cases. AI, on the other hand, received a large number of accuracy ratings and a low number of inaccuracy ratings for the subclinical normal cases. Three companies (MT, FT, & WPS) were accurate in not providing diagnostic evaluations for the subclinical normal cases.

Fundamental Statistical Units

The fundamental statistical units used are percentages, based on the frequencies with which raters assigned one of five ratings to each narrative statement for each CBTI for each case (see example in Table 4.11). Raters were asked to rate the accuracy of each numbered statement in each CBTI against the data in the subject's file (see General Instructions to Raters). Two types of irrelevancy ratings were available; (1) For data insufficient to make a rating, or (2) Statement generally applicable or repetitive, not contributing to the understanding of the case. Relevant statements were evaluated according to a three-point rating of accuracy. The rating categories for accuracy are labeled throughout the chapter as follows: (3) Inaccurate, (4) Somewhat accurate, and (5) Accurate. Raters were instructed to choose only *one* of the five rating categories for evaluating each narrative statement.

Table 4.11 gives an example of the frequencies and percentages for one subject and Table 4.12 presents the same data, rearranged according to its relevancy and accuracy, with validity profile statements considered irrelevant to the accuracy ratings. (Recall that potential subjects with deviant validity profile scores were omitted from the study.)

Length of CBTI Reports

The data in Appendix B (Tables B-1 and B-2) demonstrate that large differences existed in the number of narrative statements per CBTI system. Therefore, percentages, which use the base of 100, were used for comparison purposes. There are 366 narrative statements for BE's white 4/2 code type, but only 9 narrative statements for the white 2 code type from FT or for the Black 6 code type for the MT system.

The BE printouts were the longest for both common cases (126-225) and for the unique cases (median = 187; range 124-366). The MT Report provided the shortest set of narrative statements for the two common cases (8-21) and unique cases (median = 16, range: 9-34). These data present the range of statements for the particular protocols used in this study and do not necessarily represent all the variations in the computer library of each CBTI system for a wider variety of score combinations.

The length of the CBTI narratives and ratings of their overall accuracy do not show a linear relationship. The median rank in overall accuracy assigned by 12 raters (Table 4.7) was examined in relation to the median number of sentences for each CBTI system for the 24 unique cases (Appendix B-2). The MT system which had the lowest number of sentences (median = 16) was rated sixth in accuracy, whereas BE, the system with the highest number of sentences (median = 187), was rated seventh in overall relative accuracy. MN, CA, and WPS, the three companies with the highest accuracy, had relatively short or middle-range narrative lengths. In other words, narrative length is not directly related to ratings of overall accuracy of CBTI systems.

Examples of Data Format

The data throughout the remainder of the chapter are presented in two formats illustrated in Tables 4.11 and 4.12. Table 4.11 reports the data in the same format used by the raters: five mutually exclusive categories. This format labeled A, which is used in the chi-square summary tables, details the specific data related to rater's assessment of relevancy. The first relevancy category (1) dealt with judgments that the criterion data were insufficient to make ratings. The second rating category pertinent to relevancy (2) was used when the narrative statements were generally applicable or repetitive and did not contribute to the understanding of a case. Categories (3) to (5) represent levels of accuracy: (3) Inaccurate; (4) Somewhat Accurate; and (5) Accurate.

Table 4.12 reports the same data as in Table 4.11, but collapses data from rating categories (1) and (2) and the validity profile (which served to identify test-taking attitudes), as statements irrelevant. The narrative statements referring to the validity profile were eliminated by the authors with guidance from the CBTI companies. The validity profile items were used earlier to ascertain whether the overall profile was valid.

In the examples in Tables 4.11 and 4.12, the CBTI included three validity profile (VP) items, all of which were rated as accurate (5) by the rater. In Table 4.12, the total number of narrative statements presented in rating categories for accuracy (3), (4), and (5) was calculated by the formula: $n - (1) - (2) - VP \div 41 - 1 - 5 - 3 = 32$. Percentages are used in the remainder of the chapter to form a common basis for handling CBTIs which vary in length. In this example the percentage of relevant items thus was 78% (32/41). Of the 32 relevant statements rated, 41% were rated Inaccurate, 53% Somewhat Accurate, and 6% Accurate.

When these data are presented in the chi-square tables labeled B in Tables 4.15 and 4.16, they are reported in terms of irrelevant rather than relevant statements in order to provide nonoverlapping data in the cells of the tables. In this example, there are nine irrelevant statements ($41 - 32 = 9$).

Pooled Data for Common Cases

Tables 4.13 and 4.14 present pooled data across CBTI systems and raters for the common cases which have the following linear T-scores: (a) 7/2 white, 97 T and 96 T, respectively; (b) 7/2 Black, 89 T and 77 T, respectively; (c) subclinical normal white (Scale 2, 65 T; Scale 9, 58 T); and (d) subclinical normal Black (Scale 7, 66 T; Scale 2, 56 T). The Black and white cases were combined because the Black/white differences were negligible. The data in these tables were pooled across CBTI systems, raters, and race. These data are based on 168 percentages (12 raters \times 7 CBTI systems \times 2 Black/white cases). Table 4.13 reports all five rating categories and Table 4.14 shows the data grouped accord-

ing to the relevancy of the ratings to the case. These composite tables provide the base rates for interpreting the results for individual CBTI companies and for evaluating the response tendencies of raters.

These tables show that there are some differences in the relevancy and the accuracy of CBTI narrative statements for the two kinds of profiles. Half of the statements for the subclinical normal profile were Unratable, whereas for the 7/2 code type, the median percentage unratable was only 14 (Table 4.13). The two kinds of profiles show similar medians for the percentages of General/Repetitive statements.

Table 4.14. shows that the median percentage of Inaccurate narrative statements is greater for the 7/2 profile (19) than for the subclinical normal profile (00). (Half of the Inaccurate percentages for the subclinical normal profile were zero.) Otherwise, the medians for the Somewhat Inaccurate and Accurate ratings for the two profiles are similar. The major difference between the ratings of the 7/2 and the subclinical normal cases are in their relevancy to the case histories. The relevancy ratings for the subclinical normal cases are low. This would be expected since the MMPI was designed for use in clinical diagnosis.

Chi-square Results by Subject and Raters for CBTI Systems

The chi-square tables, which may be obtained from the authors, provide frequencies, percentages, and chi-square data. Each table presents data for one rater, for one case, for all seven CBTIs. Half of these tables involve all five rating categories (Chi Square A), the other half collapse the unratable and General/Repetitive ratings into a single "irrelevant" category (Chi Square B).

Chi-square statistics were computed using Tracy L. Gustafson's EPISTAT software (Wise, 1985). Due to the small number of narrative statements for CBTI systems, such as the Morris-Tomlinson Report, the expected cell frequencies were often less than five. No chi-square statistic was reported when such analysis was inappropriate due to empty cells (see Siegel, 1956, p. 110). Cell frequencies reached reasonable levels when percentages were pooled across raters, profile types, or CBTI systems.

In Table 4.15, the vast majority of chi-square statistics are significant beyond the .001 level.

All chi-square values for the 7/2 white and 7/2 Black cases were statistically significant at the .001 level. For the subclinical normal white case, only one chi-square value, for rater 8, was not significant at the .05 level.

The chi-square results for the Black subclinical normal case were less clear-cut. Less than half of the Chi-Square B values, based on data in which the irrelevancy ratings were collapsed into one category for each of the 12 raters, were found to be statistically significant at the 5% level. The difference between

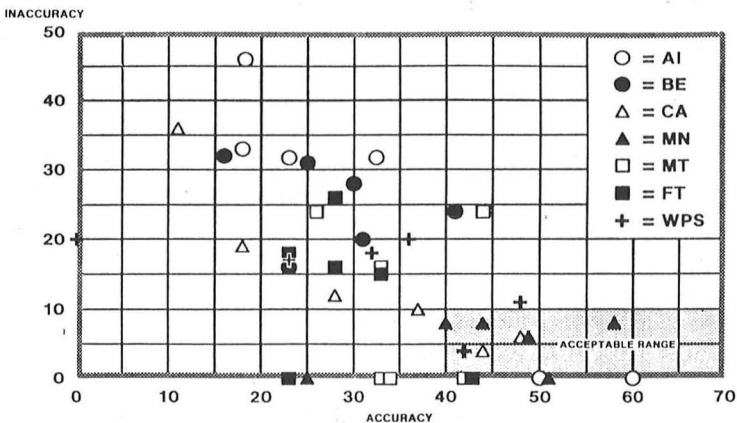
raters was greater for the Black subclinical normal case than for the other cases. This will be commented on in a later paper on the Black/white cases.

The results for the unique cases (Table 4.16), each of which was rated by only one rater, parallel the results for the 7/2 cases which were rated by all 12 raters. For the white unique cases, the Chi-Square B values were statistically significant at the 1% level for 11 raters. For the Black unique cases, the Chi-square B values were statistically significant at the 5% level for 10 raters.

The chi-square results for the Black/white pairs for the 13 code types and for the white subclinical normal code type show that raters differentiated among CBTI systems in their ratings of the relevancy and accuracy of CBTI sentences at a statistically significant level. The results, considered in combination with the overall accuracy ratings reported in Tables 4.7 and 4.8, show that raters consistently differentiated among CBTI systems.

Pooled Data for 3 Profile Types

Tables 4.17, 4.18, and 4.19 summarize the relevancy and accuracy ratings for the CBTI systems across raters and subjects for the 7/2, subclinical normal, and unique code types. For the 7/2 code type, the CA and MN systems were evaluated as having the highest percentage of Accurate sentences relevant to the cases (median = 46% and 49%, respectively). For the subclinical normal cases, AI was rated high in the Accurate sentences relevant to the case (median = 50%).



NOTE: 7/2, Subclinical Normal and Unique Cases (1, 2/1, 1/3, 8/1, 2, 2/8, 3, 4/2, 6, 8/6, 9/6, & 8/9) Rated by 12 Raters.

FIG. 4.2. Scatter diagram of median percentages for hit rate of CBTI systems for 3 MMPI profile types for matched Black/White cases.

For the twelve unique cases, WPS and MN were highest in the Accuracy of relevant sentences (median = 40% for both systems). These data show that raters are relatively consistent in their sentence-by-sentence judgments of the accuracy of the different CBTI systems, which differ significantly from each other in the accuracy of clinical and subclinical profile types.

Variation in the rated Accuracy and Inaccuracy of each profile type for Black and white cases is illustrated in Fig. 4.2. The figure presents these data in scatter diagram format, using 42 coordinates for three profile types for matched Black/white pairs for the seven CBTI systems. There are six coordinates for each CBTI system. This provides CBTI users with the comparative Hit Rates for the CBTI systems. The scatter diagram shows the considerable variability in the Hit Rate for the CBTI systems.

The Hit Rate is illustrated in the lower right quadrant using a very stringent set of cut scores: Accuracy median $\geq 40\%$ and Inaccuracy median $\leq 10\%$. However, CBTI users may set their own cut scores for the Hit Rate. Using these cut scores, we find that one CBTI system had five profile types that met our criteria and one CBTI system had none.

Pooled Data for Clinical Code Types and Individual Raters

Tables 4.20 to 4.22 provide data on the relevancy and accuracy ratings for each rater for the 7/2, subclinical normal, and unique profile types. Code types 2/8 and 8/9 received the highest ratings in sentence-by-sentence Accuracy (median = 68% and 69%, respectively) and relatively low ratings on the Inaccuracy end of the three-point scale (median = 8% and 12%, respectively). Code types 1 and 1/2 were evaluated relatively low on Accuracy (median = 10% and 8%, respectively) and received relatively high Somewhat Accurate ratings (median = 63% and 72%, respectively). Code types Spike 3 and Spike 6 were rated relatively high in Inaccuracy (median = 39% and 40% respectively).

Table 4.23 reports the rank order of sentence-by-sentence ratings for each rater using the data from the two sets of clinical cases (7/2 and unique code types). The table also reports Kendall's rank-order correlation coefficients for each of the four rating categories (Relevant, Inaccurate, Somewhat Accurate, and Accurate), all of which were statistically significant at the 5% or 1% level. The raters showed significant rank-order agreement in their sentence-by-sentence ratings for the two sets of clinical code types.

From these statistics we may infer that raters showed a response style in making their ratings. We define response style as the clinician's application of his internal criteria in a consistent way. For example, raters who rated the CBTI sentences for the 7/2 code type high in Relevancy were also likely to rate their unique code type relatively high in Relevancy. Information from Table 4.23 may also be used to evaluate the response style of individual raters. For example, rater

2 evaluated the Black and white 2/1 code type. He showed considerable rank-order agreement in evaluating the 7/2 and 2/1 code type (Rank order 10 and 10.5, respectively) for Relevancy. However, his rank-order rating of the Accuracy of the relevant sentences for the 2/1 case was relatively lower (rank = 12), compared with his rank order evaluation of the Accuracy of the 7/2 case (rank = 7). On the other hand, rater 9's rank order of the Relevancy and Accuracy of the 7/2 and 4/2 code type was similar (Relevancy rank order: 2.5 and 1 respectively; Accuracy: 4.5 and 3). These data provide information about individual rater's style and skill in applying clinical criteria in rating common and unique cases.

DISCUSSION

This chapter reports on the development and the application of a methodology for the study of comparative validity of the output of the CBTI systems. Experienced clinical psychologists rated the relevancy, accuracy, and usefulness of the output of CBTI systems. They judged the relevancy and accuracy of the CBTI systems for clinical and subclinical normal cases against an external criterion: case histories or self-report questionnaires. Ratings were made at both sentence-by-sentence and global levels. Sentences were first rated according to their relevancy to each case history, that is, determinations were made as to whether relevant data were available in the criterion, whether a sentence was relevant to the case, or whether sentences were overly general or repetitive. The accuracy of sentences was rated only for those sentences relevant to the case. Global ratings were made of CBTI systems after all sentence-by-sentence ratings for a case were completed and finally after sentence-by-sentence ratings were completed for all cases.

The study controlled for test-taking attitudes and gender, and cases (all males) were selected from a limited age range. Systematic procedures for selecting cases from a large sample of patients and normal personnel from a wide geographical area, using prespecified profile codes and clinical cutoff scores ($T \geq 70$ on a clinical scale) minimized sample bias. In spite of the fact that only 28 cases were rated and reported, the care with which this sample was selected from 1,500 existing cases, renders the results generalizable to a much larger male sample. Matched cases (Black/white pairs) and CBTI systems were rated in a counter-balanced order. By having each rater judge cases rated by all raters and by also having cases rated by only one rater, it was possible to obtain interrater reliability data and also to rate a large number of different profile types.

The study evaluated the comparative relevancy and accuracy of the output of seven CBTI systems for the MMPI, representing half of the existing commercially available systems in 1985. All but two of the nine CBTI systems invited took part in the project and data from 12 of 13 Army clinical psychologists were received and reported. Data from 28 cases involved 14 matched Black/white

male cases and represented subclinical normal and clinical (medical, psychiatric, and neuropsychological) inpatient and out-patient cases. Included were frequently occurring neurotic, psychotic, and characterological code types.

Despite the large amount of empirical evidence available for the MMPI and its potential for actuarial prediction, the output of CBTI systems for the MMPI for individuals were found to vary significantly in their rated relevancy, accuracy, and in their usefulness in case disposition, that is, diagnostic evaluation and disposition planning and accuracy. The quality of a CBTI system apparently depends on how the CBTI developer uses the MMPI's research literature and clinical lore.

The raters showed highly significant agreement in evaluating the overall accuracy of the output of the seven systems in their final global ratings. They also showed agreement in their rating of the overall accuracy of CBTI systems for each of the common cases. Interrater reliability was demonstrated even though raters showed significant response tendencies in their sentence-by-sentence ratings of relevancy and accuracy and despite the differences in the raters' graduate school subspecialties, in their employment experience, and in their experience with CBTI systems. Nine raters were trained in clinical and three in counseling psychology. They had performed different mental health functions in the Army, working in hospitals, community mental health centers, and in organizational settings. Furthermore, they differed in experience with CBTI systems for the MMPI. But experience did not show a linear relationship to ratings of global accuracy. For the three CBTIs with the highest Overall Accuracy ratings, raters reported having prior experience with MN, no experience with CA, and little experience with WPS. All but one rater had reported using Lachar's (1974) manual on which the WPS Test Report is based.

The data support the thesis that the output of CBTI systems show significant differences in their accuracy and relevancy. This conclusion is supported by statistically significant data from the final overall rank order of the CBTI systems (Table 4.7) and the chi-square data (see Tables 4.15 and 4.16) from each rater for each case.

The rater results, at both the sentence-by-sentence and global levels, show consistent, but different results for the subclinical normal and the clinical profiles. These results should be expected, since the MMPI was designed as a tool for psychodiagnosis. Furthermore, the research literature for subclinical cases is more limited than for clinical cases. The subclinical normal cases, rated by all raters, were found to have a high percentage (median 50) of Unratable sentences, whereas the Unratable sentences for the 7/2 common cases was low (median 14). Sentence-by-sentence accuracy ratings for these two cases showed that the judged Accuracy for sentences relevant to the cases was similar (median 33% and 35% for normal and 7/2 cases, respectively; see Table 4.14). Different CBTI systems showed high sentence-by-sentence Accuracy ratings for the subclinical normal and clinical cases. AI showed the highest sentence-by-sentence Accuracy

rate (median 50%; Table 4.18) and the most accurate diagnostic evaluations for subclinical normal cases.

The sentence-by-sentence results for the clinical code types, including the 7/2 common cases and the 12 clinical code-type cases, rated by only one rater, were congruent with the final global ratings of the CBTI systems. The three CBTI systems which were assigned the highest rank order for accuracy (MN, CA, & WPS; Table 4.7) showed the highest ratings for sentence-by-sentence accuracy. Their median 7/2 Accuracy rate, pooled across 12 raters, was 49%, 46%, and 33% respectively (Table 4.17). Also their median Accuracy rate, pooled across CBTI systems, for 12 clinical code type cases rated by one rater, was 40%, 34%, and 40% respectively (Table 4.19). On the other hand, AI and BE, respectively receiving final rank-order ratings of 5 and 7, received low sentence-by-sentence ratings. On a three-point scale of Accuracy (Inaccurate, Somewhat Accurate, and Accurate), these two CBTI systems were low in Accuracy for clinical cases (Tables 4.17 and 4.19). These two systems were also evaluated as having less accurate diagnostic evaluations (Table 4.10).

By analyzing the CBTI systems according to their final overall rank order for accuracy and for their Hit Rates for three profile types (7/2, subclinical normal, and unique codes) we find that the output of the higher rated CBTI systems show moderate validity levels.

The results for the 7/2 cases, for which the base rate is constant, parallel those for the unique clinical code types (1, 1/2, 1/3, 8/1, 2, 2/8, 3, 4/2, 6, 8/6, 9/6 and 8/9) for which the base rates may vary. However, the two code types which are the least frequent, spike 3 and 8/1, were rated relatively low in Accuracy (median 16% for spike 3; median 14% for 8/1). But that was also the case for the more frequent 1/2 code type (median 8%).

CBTI systems markedly vary in the length of the narratives. The BE printouts had the most sentences and MT had the fewest. The relationship between length of narratives and the global ratings of accuracy for CBTI systems was not linear: companies with the highest accuracy ratings had relatively short or middle-range length narratives.

CBTI systems also vary in their percentage of relevant sentences for clinical code types. The AI, BE, & CA systems have a relatively high percentage of Unratable sentences for both the 7/2 and the unique cases (Table B-3, and B-5). BE and WPS have the highest percentage of General Repetitive sentences, whereas MN is relatively low in Unratable and in General Repetitive sentences. Because of variations in the length of narratives and in the percentage of sentences relevant to the cases, the data are reported in percentages with a base of 100 and in median percentages for pooled data. By pooling data across raters, profile types or CBTI systems, we were able to base our conclusions on a relatively large number of rater responses.

The study is limited in that it did not focus on evaluating the Barnum effect (O'Dell, 1972) and was limited to the use of existing (file drawer) data available

in a hospital setting. Instead of including a bogus case for evaluation, as recommended by Moreland (1985), the authors chose to maximize the number of matched Black/white cases, covering 14 profile types rated. Twelve raters each rated 6 profile type cases, which placed heavy demands on them. For example, each rater had to rate 1,653 sentences for relevancy and accuracy for the four common cases. Therefore, it seemed unreasonable to add a bogus case to the study.

The study used existing hospital data for the clinical cases which vary in content and in detail. However, 23 of the 26 clinical cases were inpatients for whom, in general, there were more detailed case histories than for the outpatients. Unfortunately, the time interval between the administration of the MMPI and the preparation of the case history varied. MMPI scales 2, 8, and 9 (W. G. Dahlstrom, personal communication, October 2, 1987) are most likely to show changes in acute symptoms over time. In spite of this, the two unique code types which were rated highest in sentence-by-sentence Accuracy were two-point codes involving these scales: 8/9 (median Accuracy 69%) and 2/8 (median Accuracy 68%). Time interval data were available for one of these code types. For the 8/9 white case, the interval was 2 months and for the 8/9 Black case it was 7 weeks.

In summary, the study showed that the output of CBTI systems for the MMPI was found to vary in relevancy, accuracy, and usefulness using file drawer histories or self-report data for subclinical normal, neurotic, characterological, and psychotic profile types. The output of CBTI systems was found to differ in the accuracy of both clinical and subclinical normal code types. Raters showed considerable agreement in their global and sentence-by-sentence ratings of accuracy and relevancy. For the most highly rated CBTI systems, moderate validity levels were found for the narrative output.

FURTHER RESEARCH

In additional papers, the authors will address the clinical implications of the results for the matched Black/white cases, for neurological and nonneurological cases, and will analyze the possible reasons for the results found.

The study may be repeated, using a larger number of raters for the clinical code types. The research design may be applied in different mental health settings, civilian and military. And the research methodology may be adapted to evaluate and modify CBTIs developed for other personality inventories. Research of this kind for inventories with a limited research literature cannot be regarded as a substitute for the test validation process (American Educational Research Association, American Psychological Association, and National Council on Measurement in Education, 1985; American Psychological Association Committee on Professional Standards and Committee on Psychological Tests and

Assessment, 1986). Obviously the accuracy of CBTIs is limited by the reliability and validity of the test on which the interpretation is based.

ACKNOWLEDGMENT

This chapter is based in part on a paper presented at the 1986 convention of the American Psychological Association, chaired by A. David Mangelsdorff, for the symposium on "Computer-Based Clinical Assessment for Children, Adults, and Neuropsychological Cases." The authors acknowledge the technical assistance provided by W. Grant Dahlstrom, Kevin L. Moreland, Lois C. Northrop, Dorothea E. Johannsen, Samuel E. Krug, and Mason N. Crook. They thank the computer-based test interpretation (CBTI) companies for donating their CBTIs and greatly appreciate the dedication of the following raters: Phillip Appel, Frank Edwards, Randall C. Epperson, David H. Gillooly, Dennis J. Grill, Gregory Hollis, Lawrence Klusman, James E. McCarroll, John Powell, Robert Rankin, Frank Rath, David Schaefer and Thomas R. Waddell. The authors thank Iris Hepburn and Sandra Stewart for clerical assistance.

The opinions expressed are those of the authors and are not necessarily official policy statements.

REFERENCES

- Adair, F. L. (1978a). Review of the Minnesota Multiphasic Personality Inventory. Automated Psychological Assessment. In O. K. Buros (Ed.), *The Eighth Mental Measurements Yearbook*, pp. 945–947.
- Adair, F. L. (1978b). Review of the Minnesota Multiphasic Personality Inventory. Behaviordyne Psychodiagnostic Laboratory Service. In O. K. Buros (Ed.), *The Eighth Mental Measurements Yearbook*, pp. 948–949.
- Adair, F. L. (1978c). Review of the Minnesota Multiphasic Personality Inventory. Caldwell Report: An MMPI Interpretation. In O. K. Buros (Ed.), *The Eighth Mental Measurements Yearbook*, pp. 952–953.
- American Educational Research Association, American Psychological Association, and 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* (3rd ed.). Washington, DC: Author.
- American Psychological Association Committee on Professional Standards and Committee on Psychological Tests and Assessment. (1986). *Guidelines for computer-based tests and interpretations*. Washington, DC: Author.
- Anastasi, A. (1988). *Psychological testing* (6th ed.). New York: Macmillan.
- Black, J. Cited in NCS FASTTEST. (undated). The Minnesota Multiphasic Personality Inventory. Minneapolis: National Computer Systems.
- Butcher, J. N. (1978a). Review of the Minnesota Multiphasic Personality Inventory. Automated Psychological Assessment. In O. K. Buros (Ed.), *The Eighth Mental Measurements Yearbook*, pp. 947–948.

- Butcher, J. N. (1978b). Review of the Minnesota Multiphasic Personality Inventory. Behaviordyne Psychodiagnostic Laboratory Service. In O. K. Buros (Ed.), *The Eighth Mental Measurements Yearbook*, pp. 951–952.
- Butcher, J. N. (1978c). Review of the Minnesota Multiphasic Personality Inventory: Caldwell Report: An MMPI Interpretation. In O. K. Buros (Ed.), *The Eighth Mental Measurements Yearbook*, pp. 953–955.
- Butcher, J. N., & Keller, L. S. (1984). Objective personality assessment. In G. Goldstein, & M. Hersen (Eds.), *Handbook of psychological assessment*. New York: Pergamon Press.
- Chase, L. L. S. (1974). An evaluation of MMPI interpretation systems. *Dissertation Abstracts International*, 35, 6090B. (University Microfilms No. 75–11, 982).
- Colligan, R. C., Osborne, D., Swenson, W. M., & Offord, K. P. (1983). *The MMPI: A contemporary normative study*. New York: Praeger.
- Dahlstrom, W. G., Welsh, G. S., & Dahlstrom, L. E. (1972). *An MMPI Handbook: Vol. 1. Clinical interpretation* (rev. ed.). Minneapolis: University of Minnesota Press.
- Duthie, B. (1985). *MMPI computerized interpretation manual: Subsection 3.2*. Wakefield, RI: Applied Innovations.
- Eyde, L. D. (1985). Review of the Minnesota Multiphasic Personality Inventory. The Minnesota Report: Personnel Selection System. In J. V. Mitchell, Jr. (Ed.), *The Ninth Mental Measurements Yearbook*, pp. 1005–1008.
- Eyde, L. D., & Kowal, D. M. (1985). Psychological decision support software for the public: Pros, cons, and guidelines. *Computers in Human Behavior*, 1, 321–336.
- Eyde, L. D., Kowal, D. M., & Fishburne, F. J. (1987). *Clinical implications of validity research on computer-based test interpretations of the MMPI*. Paper presented at 1987 Convention of the American Psychological Association, A. D. Mangelsdorff, Chair of Symposium on "Practical Test User Problems Facing Psychologists in Private Practice," New York City.
- Eyde, L. D., & Kowal, D. M. (1987). Computerized test interpretation services: Ethical and professional concern regarding U.S.A. producers and users. In L. D. Eyde (Ed.), *Computerised Psychological Testing*. London: Lawrence Erlbaum Associates.
- Finney, J. C. (1968). Normative data on some MMPI scales. *Psychological Reports*, 23, 219–229.
- Fishburne, F. J., & Parkison, S. C. (1984). Age effects on active duty Army MMPI profiles. *Proceedings, Psychology in the Department of Defense Ninth Symposium* (USAF-TR-84-2). Colorado Springs, CO: USAF Department of Behavioral Sciences and Leadership, pp. 575–579.
- Fishburne, F. J., Eyde, L. D., & Kowal, D. M. (1988). *Computer-based test interpretations of the Minnesota Multiphasic Personality Inventory with neurologically impaired patients*. Paper presented at 1988 convention of the American Psychological Association. A. D. Mangelsdorff, Chair. Atlanta, Georgia.
- Fowler, R. D. (1985). Landmarks in computer-assisted psychological assessment. *Journal of Consulting and Clinical Psychology*, 53, 748–759.
- Fowler, R. D., & Butcher, J. N. (1986). Critique of Matarazzo's views of computerized testing: All sigma and no meaning. *American Psychologist* 41, 94–96.
- Fowler, R. D., & Coyle, F. A., Jr. (1968). Scoring error on the MMPI. *Journal of Clinical Psychology*, 24, 59–69.
- Gilberstadt, H., & Duker, J. (1965). *A handbook for clinical and actuarial MMPI interpretation*. Philadelphia: Saunders.
- Graham, J. R. (1977). *The MMPI: A practical guide*. New York: Oxford University Press.
- Graham, J. R. (1978). The Minnesota Multiphasic Personality Inventory (MMPI). In B. B. Wolman (Ed.), *Clinical diagnosis of mental disorders: A handbook*. New York: Plenum Press.
- Graham, J. R. (1987). *The MMPI: A practical guide* (2nd ed.). New York: Oxford Press.
- Graham, J. R., & Lilly, R. S. (1984). *Psychological testing*. Englewood, Cliffs, NJ: Prentice-Hall.
- Grayson, H. M., & Backer, T. E. (1972). Scoring accuracy of four automated MMPI interpretation report agencies. *Journal of Clinical Psychology*, 28(3), 366–370.
- Greene, R. L. (1980). *The MMPI: An interpretive manual*. New York: Grune & Stratton.

- Guilford, J. P., & Fruchter, B. (1973). *Fundamental statistics in psychology and education* (5th ed.). New York: McGraw-Hill.
- Harris, W. G. (1987). Computer-based test interpretations: Some development and application issues. In L. D. Eyde (Ed.), *Computerised Psychological Testing*. London: Lawrence Erlbaum Associates.
- Hathaway, S. R., & Meehl, P. E. (1951). *An atlas for clinical use of the MMPI*. Minneapolis: University of Minnesota Press.
- Holden, C. (1986, September 19). Researchers grapple with problems of updating classic psychological test. *Science*, *233*, 1249–1251.
- Johnson, J. H., Butcher, J. N., Null, C., & Johnson, K. N. (1984). Replicated item level factor analysis of the full MMPI. *Journal of Personality and Social Psychology*, *47*, 105–114.
- Klett, B., Schaefer, A., & Plemel, D. (1985). Just how accurate are computer-scored tests? *The VA Chief Psychologist*, *8*, 7.
- Krug, S. (1987). *Psychware Sourcebook 1987–1988: A reference guide to computer-based products for behavioral assessment in psychology, education, and business* (2nd ed.). Kansas City, MO: Test Corporation of America.
- Lachar, D. (1974). *The MMPI: Clinical assessment and automated interpretation*. Los Angeles: Western Psychological Services.
- Lanyon, R. I. (1984). Personality assessment. In M. R. Rosenzweig & L. W. Porter (Eds.), *Annual Review of Psychology*, *35*, 667–701.
- Matarazzo, J. D. (1986). Computerized clinical psychological test interpretations: Unvalidated plus all mean and no sigma. *American Psychologist*, *41*, 14–24.
- Miller, D. A., Johnson, J. H., Klingler, D. E., Williams, T. A., & Giannetti, R. A. (1977). Design for an on-line computerized system for MMPI interpretation. *Behavior Research Methods and Instrumentation*, *9*, 117–122.
- Mitchell, J. V., Jr. (1984). Computer-based interpretation and the public interest. *CPA Highlights* (Canadian Psychological Association), *6*(4), 4–6, (ERIC, E D 249 286).
- Moreland, K. L. (1985). Validation of computer-based test interpretations: Problems and prospects. *Journal of Consulting and Clinical Psychology*, *53*, 816–825.
- Moreland, K. L. (1987). Computer-based test interpretations: Advice to the consumer. In L. D. Eyde (Ed.), *Computerised Psychological Testing*. London: Lawrence Erlbaum Associates.
- National Computer Systems. (1982). *Minnesota Multiphasic Personality Inventory: User's Guide for the Minnesota Report*. Minneapolis: University of Minnesota Press.
- NCS FASTTEST. (undated). *The Minnesota Multiphasic Personality Inventory*. Minneapolis: National Computer Systems.
- O'Dell, J. W. (1972). P. T. Barnum explores the computer. *Journal of Consulting and Clinical Psychology*, *38*, 270–273.
- Parkison, S. C., & Fishburne, F. J. (1984). MMPI normative data for a male active duty Army population. *Proceedings, Psychology in the Department of Defense, Ninth Symposium* (USAFATR-84-2). Colorado Springs, CO: USAF Department of Behavioral Sciences and Leadership.
- Siegel, S. (1956). *Nonparametric statistics for the behavioral sciences*. New York: McGraw-Hill.
- Sundberg, N. D. (1985a). Review of the Minnesota Multiphasic Personality Inventory. Behavior-dyne Psychodiagnostic Laboratory Service. In J. V. Mitchell, Jr. (Ed.), *The Ninth Mental Measurements Yearbook*. Lincoln, Nebraska: Buros Institute of Mental Measurements. pp. 1003–1005.
- Sundberg, N. D. (1985b). Review of the Minnesota Multiphasic Personality Inventory. WPS Test Report. In J. V. Mitchell, Jr. (Ed.), *The Ninth Mental Measurements Yearbook*, Lincoln, Nebraska: Buros Institute of Mental Measurements. pp. 1009–1011.
- Weigel, R. G., & Phillips, M. (1967). An evaluation of MMPI scoring accuracy by two national scoring agencies. *Journal of Clinical Psychology*, *23*, 101–103.
- Wiggins, J. A. (1973). *Personality and prediction: Principles of personality assessment*. Reading, MA: Addison-Wesley.

Wise, S. L. (1985). Software review of EPISTAT by Tracy L. Gustafson. Round Rock, TX: Author, Undated. *Computers in Human Behavior*, 1, 199–202.

APPENDIX A

Rating Forms and Instructions General Instructions to Raters

There are now 14 companies that offer CBTIs for the MMPI. Seven of these companies are included in this study, in which you will be evaluating the MMPI interpretations for six cases in terms of their case files. In other words, you are rating the validity of the CBTIs, using case histories as a criterion.

Your data will be reported in summary form only and we will provide you with our resulting paper.

1. Please begin by completing the Background Data Form for Raters.
2. Next study the case file for your first subject. Note that for purposes of this study our consultant, Dr. W. Grant Dahlstrom, has set these raw score criteria for declaring MMPI invalid: (a) Can't say scores of 50 or greater; (b) L or Lie scores of 11 or greater; and (c) F or Frequency scores of 22 or greater.
3. Read over everything in each of the seven CBTI reports for this subject, in the order in which you have been instructed to use them (see individualized instructions).
4. Now you are ready to begin rating the individually numbered narratives for the first CBTI.
5. You are to rate the accuracy of each numbered statement in each CBTI against the data in the subject's case file. Rate each statement's accuracy by using *one* of these *five* rating categories:
 - 0 = Data insufficient to make a rating.
 - 9 = Generally applicable or repetitive statement which does not contribute to the understanding of the particular case.
 - 1 = Narrative statement is inaccurate.
 - 2 = Narrative statement is somewhat accurate.
 - 3 = Narrative statement is accurate.
6. Use the general answer sheet to record your rating (0, 9, 1, 2, or 3) of each narrative statement. On the general answer sheet, the narrative statement numbers appear on the left. Column headings identify each of the seven CBTIs.
7. Begin by rating the first CBTI on your list, rating each numbered narrative statement. Complete all statement ratings before going on to the Special Answer Sheet for this CBTI.
8. Repeat instruction 7 for each of the remaining 6 CBTIs for your first subject.

9. After completing all ratings for your first subject, follow instructions 2 to 8 for each of your remaining five subjects.
10. Now turn to your Final Rating Sheet and complete these overall ratings for all seven CBTIs for all six subjects.
11. When you have completed all ratings for all subjects, mail all the materials, *using the most rapid available mailing procedure*, to:

Dr. Lorraine D. Eyde
 2400 S. Arlington Ridge Rd.
 Arlington, VA 22202

We thank you for your assistance. You will be hearing more from us at a later date after we finish our papers.

BACKGROUND DATA FORM FOR RATERS

Rater # _____

1. My Ph.D. is in Clinical Psychology: ___yes ___no.
 a. If "no" state specialty area _____.
2. I have completed an APA-approved Clinical Psychology Internship: ___yes ___no.
3. I am licensed to practice psychology: ___yes ___no.
4. I have had the following number of years (full-time or equivalent) of post-doctoral experience in clinical psychology: _____years.
5. I hold a diplomate, issued by the American Board of Professional Psychology: ___yes ___no.
 a. If "yes," state the specialty _____.
6. My race is: _____Caucasian (White) _____Black _____Asian (Oriental)
 or _____other.
7. My ethnicity is: _____Hispanic _____Nonhispanic.
8. My gender is: _____Male _____Female.
9. Do you currently use the MMPI in your practice? ___yes ___no..
10. What interpretative references or sources do you presently use in your practice? Please give references:

11. How much experience have you had in using each kind of computer-based test interpretations (CBTIs) of the MMPI?

	No Experience	Some Experience	Extensive Experience
a. Applied innovations	_____	_____	_____
b. Behaviordyne	_____	_____	_____
c. Caldwell Report	_____	_____	_____
d. Minnesota Report: (Adult System)	_____	_____	_____
e. Morris-Tomlinson Report	_____	_____	_____
f. Psych Systems	_____	_____	_____
g. WPS Test Report	_____	_____	_____

SPECIAL ANSWER SHEET FOR APPLIED INNOVATIONS

Rater # _____

Subject # _____

1. Now that you have rated each narrative statement of the CBTI against the case file for your subject, please list the salient aspects of the case history identified by this particular CBTI.

2. Now list the significant omissions for this case history that this CBTI did not pick up.

3. On the basis of your evaluation of the subject's case file, how would you characterize the mental status of this subject?

Psychotic Neurotic Personality Disorder Normal

4. Now rate the overall accuracy of the diagnostic evaluation described in the numbered narratives, offered by this CBTI compared with the data in the case history, by placing an X in one of these boxes.

CBTI does not provide a diagnostic evaluation.
 CBTI's diagnostic evaluation is inaccurate.
 CBTI's diagnostic evaluation is somewhat accurate.
 CBTI's diagnostic evaluation is accurate.

5. Did the CBTI recommend chemotherapy for this subject? yes no.

If "yes," how appropriate was the recommendation?

not appropriate somewhat appropriate appropriate.

6. Did the CBTI suggest that the subject may have a neurological/organic problem?

yes no.

If "yes," how accurate was the evaluation?

inaccurate somewhat accurate accurate

7. How do you evaluate the adequacy of the special scales, reesearch scales, and critical item listings used in this CBTI system?

not enough listings
 adequate listings
 more listings than needed

8. Rate the overall accuracy of the CBTI System.

The CBTI System is generally inaccurate.
 The CBTI System is somewhat accurate.
 The CBTI System is generally accurate.

9. Rank the overall helpfulness of the CBTI system in the disposition of the case, i.e., in the diagnostic evaluation and disposition planning.

The CBTI System is not helpful.
 The CBTI System is somewhat helpful.
 The CBTI System is quite helpful.

10. General comments on this CBTI System.

FINAL RATING SHEET

RATER # _____

Now that you have completed all of your ratings for six subjects, covering seven CBTIs, please make overall ratings across your subjects.

1. First, place the following seven CBTI systems in rank order in terms of the overall accuracy of their CBTIs for all of the subjects you have rated. Place a "1" next to the CBTI system that produced, on the average, the most accurate overall CBTI. Then, place a "2" next to the CBTI system with the second most accurate overall CBTI. Continue doing so, until you have assigned a "7" rating to the system that produced the least accurate overall CBTI.

- _____ Applied Innovations
- _____ Behaviordyne
- _____ Caldwell Report
- _____ Minnesota Report: Adult System
- _____ Morris-Tomlinson Reports
- _____ Psych Systems
- _____ WPS Test Report

2. Now apply the same overall ranking system to ratings for all subjects in the overall helpfulness of the CBTI system in the disposition of the case, i.e., the diagnostic evaluation and disposition planning. Assign "1" to "7" ratings to these CBTIs.

- _____ Applied Innovations
- _____ Behaviordyne
- _____ Caldwell Report
- _____ Minnesota Report: Adult System
- _____ Morris-Tomlinson Reports
- _____ Psych System
- _____ WPS Test Report

APPENDIX B

Pooled Data Across All Rating Categories of MMPI CBTI Systems

TABLE B.1
Number of Narrative Statements for Common Cases Rated by All Raters of MMPI CBTI System

Code Type Race	Applied Innovations	Behaviordyne	Caldwell Report	Minn. Report Adult System	Morris- Tomlinson	NCS FAST TEST	WPS Test Report
7/2 White	123	225	45	32	21	54	41
7/2 Black	123	188	41	45	14 ^a	52	39
Subclinical Normal White	13 ^a	149	51	22	9 ^a	32	21
Subclinical Normal Black	23	126	77	16 ^a	8 ^a	36	27
Range for CBTI system	13-123	126-225	41-77	16-45	8-21	32-54	21-41

^a Caution should be applied when interpreting percentages based on frequencies < 20.

TABLE B.2
Number of Narrative Statements for Code Types of Unique Cases Rated by One Rater for MMPI CBTI Systems

<i>Number of Statements for CBTI Systems</i>								
<i>Code Type</i>	<i>Rater</i>	<i>AI</i>	<i>BE</i>	<i>CA</i>	<i>MN</i>	<i>MT</i>	<i>FT</i>	<i>WPS</i>
1 W	1	22	198	74	25	10	48	23
1 B	1	16	164	62	28	15	30	25
2/1 W	2	56	124	58	38	16	63	31
2/1 B	2	114	226	57	54	34	72	71
1/3 W	3	106	229	84	38	22	34	27
1/3 B	3	83	139	87	38	28	47	39
8/1 W	4	29	185	73	25	16	35	19
8/1 B	4	119	189	59	33	30	42	27
2 W	6	37	232	65	24	13	9	30
2 B	6	31	151	57	19	12	50	17
2/8 W	7	110	252	49	60	32	51	45
2/8 B	7	115	248	67	56	31	59	62
3 W	8	23	170	53	28	10	36	19
3 B	8	20	183	71	27	14	31	21
4/2 W	9	74	366	53	29	13	62	31
4/2 B	9	104	169	52	30	21	58	34
6 W	10	17	164	70	14	10	30	14
6 B	10	35	166	84	26	9	50	20
8/6 W	11	74	235	48	32	16	42	41
8/6 B	11	107	209	63	37	25	61	63
9/6 W	12	28	192	80	33	15	51	35
9/6 B	12	90	289	83	53	25	52	61
8/9 W	13	66	140	51	31	12	52	49
8/9 B	13	76	176	53	28	17	51	34
Median		70	187	62.5	30.5	16	50	31
Range		16-119	124-366	49-87	14-60	9-34	9-72	14-71

Note. W = White; B = Black

TABLE B.3
Median Percentages for 7/2 Code Type for Ratings of Narrative Statements of MMPI CBTI Systems Across Raters (N = 24)

<i>CBTI System</i>	<i>Rating Categories</i>				
	<i>Unratable (1)</i>	<i>General Repetitive (2)</i>	<i>Inaccurate (3)</i>	<i>Somewhat Accurate (4)</i>	<i>Accurate (5)</i>
AI	18	09	24	23	18
BE	20	20	14	24	14
CA	18	07	02	31	26
MN	10	03	05	28	48
MT	14	07	10	19	29
FT	14	06	12	28	22
WPS	10	18	11	30	30

TABLE B.4
Median Percentages for Subclinical Normal Cases for Ratings of Narrative
Statements of MMPI CBTI Systems Across Raters
(N = 24)

<i>CBTI System</i>	<i>Rating Categories</i>				
	<i>Unratable (1)</i>	<i>General Repetitive (2)</i>	<i>Inaccurate (3)</i>	<i>Somewhat Accurate (4)</i>	<i>Accurate (5)</i>
AI	54	04	00	12	23
BE	56	16	04	08	06
CA	54	14	06	09	08
MN	55	05	00	13	19
MT	25	17	00	12	12
FT	54	06	00	16	14
WPS	33	25	04	11	14

TABLE B.5
Median Percentages for Unique Cases for Ratings of Narrative Statements of
MMPI CBTI Systems Across Raters
(N = 24)

<i>CBTI Systems</i>	<i>Rating Categories</i>				
	<i>Unratable (1)</i>	<i>General Repetitive (2)</i>	<i>Inaccurate (3)</i>	<i>Somewhat Accurate (4)</i>	<i>Accurate (5)</i>
AI	30	04	20	24	16
BE	24	20	13	20	11
CA	30	08	04	22	22
MN	10	07	06	20	35
MT	07	12	08	34	29
FT	26	07	10	30	18
WPS	08	20	10	24	30

Note. Covers following code types: 1, 1/2, 1/3, 8/1, 2, 2/8, 3, 4/2, 6, 8/6, 9/6, and 8/9..

TABLE B.6
Median Percentage for 7/2 Code Type Ratings of Narrative Statements by
Raters Across MMPI CBTI Systems
(N = 14)

<i>Rater</i>	<i>Rating Categories</i>				
	<i>Unratable (1)</i>	<i>General Repetitive (2)</i>	<i>Inaccurate (3)</i>	<i>Somewhat Accurate (4)</i>	<i>Accurate (5)</i>
1	02	22	27	29	10
2	26	04	10	27	26
3	08	04	08	30	46
4	16	16	14	34	12
6	28	10	16	23	17
7	12	14	10	23	35
8	34	19	07	18	16
9	06	11	06	31	37
10	20	05	22	17	28
11	16	02	10	33	28
12	23	06	04	24	28
13	18	11	05	16	36

Note. Rater 5's data lost in the mail.

TABLE B.7
Median Percentages for Subclinical Normal Case Ratings of Narrative
Statements by Raters Across MMPI CBTI Systems
(N = 14)

<i>Rater</i>	<i>Rating Categories</i>				
	<i>Unrated</i> <i>(1)</i>	<i>General</i> <i>Repetitive</i> <i>(2)</i>	<i>Inaccurate</i> <i>(3)</i>	<i>Somewhat</i> <i>Accurate</i> <i>(4)</i>	<i>Accurate</i> <i>(5)</i>
1	06	38	13	32	08
2	44	07	00	28	10
3	24	10	01	18	34
4	46	25	02	21	10
6	54	12	08	08	14
7	08	26	14	14	32
8	55	34	00	00	03
9	60	14	00	08	12
10	74	06	02	00	06
11	25	02	00	38	15
12	82	05	00	04	10
13	68	13	02	00	16

Note. Rater 5's data lost in the mail.

TABLE B.8
Median Percentage for Unique Case Rating of Narrative Statements by Raters
Across CBTI Systems
(N = 14)

<i>Rater</i>	<i>Code</i> <i>Type</i>	<i>Rating Categories</i>				
		<i>Unratable</i> <i>(1)</i>	<i>General</i> <i>Repetitive</i> <i>(2)</i>	<i>Inaccurate</i> <i>(3)</i>	<i>Somewhat</i> <i>Accurate</i> <i>(4)</i>	<i>Accurate</i> <i>(5)</i>
1	1	05	21	10	40	12
2	1/2	39	08	06	32	05
3	1/3	20	08	12	26	27
4	8/1	14	18	14	41	09
6	2	10	04	08	32	35
7	2/8	06	22	05	14	46
8	3	35	23	10	12	12
9	4/2	07	12	14	26	36
10	6	28	03	19	14	22
11	8/6	26	08	16	24	19
12	9/6	50	04	02	19	12
13	8/9	15	06	07	12	43

