INFORMATION TO USERS

This manuscript has been reproduced from the microfilm master. UMI films the text directly from the original or copy submitted. Thus, some thesis and dissertation copies are in typewriter face, while others may be from any type of computer printer.

The quality of this reproduction is dependent upon the quality of the copy submitted. Broken or indistinct print, colored or poor quality illustrations and photographs, print bleedthrough, substandard margins, and improper alignment can adversely affect reproduction.

In the unlikely event that the author did not send UMI a complete manuscript and there are missing pages, these will be noted. Also, if unauthorized copyright material had to be removed, a note will indicate the deletion.

Oversize materials (e.g., maps, drawings, charts) are reproduced by sectioning the original, beginning at the upper left-hand corner and continuing from left to right in equal sections with small overlaps.

ProQuest Information and Learning
300 North Zeeb Road, Ann Arbor, MI 48106-1346 USA
800-521-0600

METHODS FOR EVALUATING CONSISTENCY OF ITEM ENDORSEMENT ON THE MMPI-2

A Dissertation

Presented to the Faculty of

Pacific Graduate School of Psychology

Palo Alto, California

In Partial Fulfillment of the

Requirements for the Degree of

Doctor of Philosophy in Psychology

by

D. Chris Bullard

June, 2002

UMI Number: 3062671

ProQuest Information and Learning Company
300 North Zeeb Road
P.O. Box 1346
Ann Arbor, MI 48106-1346

METHODS FOR EVALUATING CONSISTENCY OF ITEM ENDORSEMENT ON THE MMPI-2

D. Chris Bullard

Pacific Graduate School of Psychology, 2002

The purpose of this study was to develop statistically stable stepwise multiple regression models that can be used by clinicians to predict the probability for completion of a consistent MMPI-2 profile based on an adult forensic outpatient litigant's WAIS-R FSIQ, WRAT-R reading level, and years of completed education before the MMPI-2 is administered.

Using correlational analysis for the complete forensic sample, significant negative correlations were found between the MMPI-2 criterion variables F, F_B, TRIN, and $F + F_B + /F - F_B/$ and the WAIS-R FSIQ, WRAT-R reading level, and years in education. Significant negative correlations were found between VRIN and $VRIN + /F - F_B/$ and the WAIS-R FSIQ and WRAT-R reading level. Finally, a significant negative correlations was found between $/F - F_B/$ and WAIS-R FSIQ.

Next, these forensic litigants were divided into odd and even samples based on odd and even identification numbers in the forensic data set. An independent sample t-test was used to compare mean T-Scores for the first four MMPI-2 criterion variables and raw score for the last three MMPI-2 criterion variables. All seven comparisons between the odd and even forensic samples were non significant.

Then, stepwise multiple regression analyses for each MMPI-2 criterion variable F, F_B, TRIN, $/F - F_B/$, VRIN + $/F - F_B/$, and F + F_B + $/F - F_B/$ regressed on the WAIS-R FSIQ, WRAT-R reading level, and years in education produced significant multiple correlation coefficients for all models except VRIN for forensic litigants in the odd sample. Finally, cross-validation was conducted for all significant MMPI-2 criterion variable models derived from the odd forensic sample and significant multiple correlation coefficients were found for all the models except $/F - F_B/$.

Clinical applications are discussed that will allow clinicians to determine if adult forensic outpatient litigants can endorse MMPI-2 items consistently if their WAIS-R FSIQ scores and years of completed education are known before the test is administered.

METHODS FOR EVALUATING CONSISTENCY OF ITEM ENDORSEMENT ON THE MMPI-2

This dissertation by D. Chris Bullard, directed and approved by the candidate's committee has been accepted and approved by the Faculty of Pacific Graduate School of Psychology in partial fulfillment of the requirements for the degree of

DOCTOR OF PHILOSOPHY

IN PSYCHOLOGY

Linda Garcia-Shelton June 6, 2002
Linda Garcia-Shelton, Ph.D., M.H.S.A.
Vice President for Academic Affairs

Dissertation Committee:

Roger L. Greene
Roger L. Greene, Ph.D.
Chair

Charles E Baker
Charles E. Baker, Ph.D.
Committee Member

Gary T. Miles, Ph.D.
Committee Member

W. Grant Dahlstrom
Grant Dahlstrom, Ph.D.
Committee Member

DEDICATION

To Dr. Roger Greene and Dr. W. Grant Dahlstrom. Dr. Greene, whose guidance, support, and unending patience provided me with the opportunity to complete this dissertation despite several years of personal obstacles. Dr. W. Grant Dahlstrom who through his knowledge, focus, and leadership so generously supported and inspired me to continue towards my ultimate goal. To both mentors, I owe much thanks.

ACKNOWLEDGMENTS

This author gratefully acknowledges the generosity of Dr. Greiffenstein who made the forensic data set used in this dissertation available and who was always available to clarify or discuss any issues as needed. A special debt of gratitude also goes to Dr. Greene, Dr. W. Grant Dahlstrom, Dr. Miles, and Dr. Baker who served on my dissertation committee and whose support I appreciate.

TABLE OF CONTENTS

Page

LIST OF TABLES

CHAPTER I: INTRODUCTION

Description of MMPI/MMPI-2

The Minnesota Multiphasic Personality Inventory (MMPI) was developed in the late 1930s and 1940s by Hathaway and McKinley (1940) and is used widely as a descriptive instrument or criterion measure in a vast array of clinical and research investigations producing more than 8,000 references (Anastasi, 1988). It is the most frequently used clinical test in community mental health centers and Veterans Administration medical centers, and ranks first in frequency of usage across psychiatric hospitals and counseling centers (Lubin, Larsen, Matarazzo, & Seever, 1985). Its easy administration, objective scoring, generally objective interpretive procedures, and established validity as a criterion measure are some of the reasons for its extensive use (Dahlstrom, Welsh, & Dahlstrom, 1972, 1975).

In 1989, the revised MMPI, the Minnesota Multiphasic Personality Inventory-2 (MMPI-2) was developed that implemented new national norms more representative of the current U.S. population, introduced new uniform T-scores for the eight basic clinical scales, modified item wording, deleted items containing objectionable content, and added new content items and supplementary scales (Butcher, Dahlstrom, Graham, Tellegen, & Kaemmer, 1989). In their investigation of the new items, Ben-Porath and Butcher (1989) concluded that the original MMPI items could be improved and objectionable language corrected at no psychometric cost.

Brief Overview of Validity

Both the MMPI and MMPI-2 contain scales that measure profile validity and response consistency. In determining the validity of the MMPI/MMPI-2 profile, the consistency and accuracy of item endorsement must be evaluated. If the person provides a consistent and accurate self-appraisal in response to the items, the profile is considered valid (Greene, 2000). Consistency of item endorsement verifies the reliability of item endorsement and may be considered irrelevant to or independent of item content. Thus, consistency assesses whether there has been a reliable pattern of response to the items throughout the inventory independent of the item content. A determination of consistency of item endorsement is necessary before any assessment of the accuracy of item endorsement is made so that inconsistent patterns of item endorsement are not inappropriately labeled overreporting of psychopathology (Rogers, Dolmetsch, & Cavanaugh, 1983). Accuracy of item endorsement depends on and is related to item content and assesses whether there has been distortion of responses to items in some specific manner (Greene, 2000).

Lack of consistency of item endorsement may result in a random pattern of item endorsement and too much within-person scatter (Tellegen, 1988). In random responding, the person may either respond to items in a clearly random manner or they may adopt an idiosyncratic response pattern (Graham, 1993). Random responding may be suspected when a person

completes the MMPI-2 much too quickly (Greene, 2000) or is unable to comprehend item meaning (Ben-Porath & Tellegen, 1992).

Dahlstrom and Tellegen (1993) stress that the validity of the individual test protocols be carefully assessed and, given a valid test protocol, the MMPI-2 pattern should prove to be useful in appraising the degree and form of an individual's adjustment difficulty.

General Reasons for Inconsistent Endorsement

Greene (2000) identifies several possible reasons for random responding that result in inconsistent item endorsement and invalid protocols: lack of the intellectual or reading ability necessary to complete the inventory appropriately; limited education; psychological or neuropsychological impairment; a degree of toxicity that prevents completion of the inventory appropriately although the person possesses the necessary skills; and an unwillingness to cooperate with the inventory task. Rogers et al. (1983) also include hostility and attempts to malinger as reasons for inconsistent endorsement. This malingering may take the form of inconsistent responding as well as symptom exaggeration (Fox, Gerson, & Lees-Haley, 1995).

MMPI Inconsistency Scales

One measure used to determine consistency of item endorsement on the MMPI is the TR Index. The TR Index is composed of 16 pairs of identically worded items and is the total number of repeated items that have been endorsed inconsistently. The TR Index will easily detect random response sets. The TR Index item pairs are separated by approximately 260 items.

Buechley and Ball (1952) concluded that a score of three or more on the TR Index suggested an invalid profile, while Dahlstrom et al. (1972) suggested that four or more inconsistent responses showed questionable reliability. Gravitz and Gerton (1976) found that normal men and women gave different responses to three or four pairs of items. Most normal persons make four or fewer inconsistent responses and 75% of psychiatric patients make three or fewer inconsistent responses (Greene, 1989).

The TR Index provides an independent estimate of the consistency of item endorsement, since it is not affected by the presence of psychopathology as the F scale. However, an acceptable score on the TR Index suggests only that the person has endorsed the items consistently, and not necessarily accurately, since overreporting and underreporting of psychopathology are possible (Greene, 1989). Tellegen (1988) reported that the TR Index will not detect inconsistent responding if the responding takes the form of indiscriminate wholesale endorsement or nonendorsement of the items; for example, answering all items true or false results in a TR score of 0.

In a study including juveniles and adults, Greene (1979) concluded that the higher frequency of inconsistent responses by juveniles reflected their general uncooperativeness and poor motivation, and concluded that, with juveniles excluded from the results, the TR Index is independent of type or degree of psychopathology. Greene's conclusion has been supported with results showing small mean differences between diagnostic groups and gender (Coché & Steer, 1974; Jones, Neuringer, & Patterson, 1976). In

comparing the TR Index with the F scale, the TR Index has been described as identifying some invalid protocols that would go undetected by the F scale alone and some profiles as valid that the F scale would consider invalid (Greene, 1979). A combination of the TR Index and F scale has been cited as particularly helpful in identifying consistent versus inconsistent patterns of item endorsement. Maloney, Duvall, and Friesen (1980) concluded that random response patterns tended to be more prevalent in protocols with F scale elevations and in those groups who were either psychotic or of low intelligence. In support of this finding, Evans and Dinning (1983) found that inconsistent responders' profiles showed random responding and stressed the use of the TR Index in enhancing interpretation of profiles with a very high F scale score.

The use of a combination rule (i.e., TR Index greater than 4 and F scale raw score greater than 16) has been described as having the greatest clinical utility for correctly classifying random responders (Rogers et al., 1983). They found the TR Index could classify correctly 97.5% of random responders and 77.5% of nonrandom responders using this rule, while Rogers, Harris, and Thatcher (1983) cited a 95% accuracy.

Another measure used to determine item consistency on the MMPI is the Carelessness (CLS) scale. This scale is composed of 12 pairs of items judged to be psychologically opposite in content and involves counting logically inconsistent responses to these items. It is useful in detecting a more sophisticated lack of cooperation and confusion or inability to respond

consistently to items. CLS provides additional information about whether the person is willing or able to complete the MMPI appropriately (Greene, 1978). Haertezen and Hill (1963) concluded that psychologically opposite items are more sensitive than repeated items (such as on the TR Index) in detecting the inability or unwillingness of a person to complete the inventory appropriately. Since the items on CLS are more subtle, it allows identification of inconsistent responding by sophisticated persons who may recognize the existence of identical repeated items and avoid being detected by the TR Index. The CLS item pairs are separated by approximately 127 items. With a maximum of twelve inconsistent responses possible on the CLS scale, a score of four or more deviant responses is considered optimal in identifying an invalid protocol (Greene, 1978). However, Bond (1986) suggested that indecision may be a more important cause of inconsistency than carelessness per se.

The correlations of CLS with the TR Index range from .23 to .57 and from .43 to .53 with the F scale, which suggests that the CLS scale is measuring a related but not identical test-taking behavior. CLS has been cited as identifying an additional 10% of profiles as invalid that the TR Index would consider valid, providing a substantial increment in the ability of clinicians to detect inconsistent item endorsement (Greene, 1978). Nichols, Greene, and Schmolck (1989) found that the sum of the TR Index and CLS scale was effective in discriminating inconsistent protocols in three separate samples of adult psychiatric inpatients. However, two additional studies with forensic outpatients showed no increase in classification accuracy using the CLS scale

above and beyond that obtained from using the F scale and TR Index (Rogers et al., 1983; Rogers et al., 1983). Nichols et al. (1989) noted the F scale, with all of its items occurring among the first 293, would not detect waning compliance in endorsing the items, while the TR Index and CLS scale may identify inconsistent responses after item 300.

MMPI-2 Infrequency Scales

As in the MMPI, the F scale of the MMPI-2 is one measure to determine consistency of item endorsement and is considered one of the most reliable indicators of a random sort (Dahlstrom et al., 1972; Rogers et al., 1983). The F scale of the MMPI-2 has a total of 60 items instead of 64 items included in the MMPI F scale.

In studies comparing MMPI scores with MMPI-2 scores, Munley (1991) found T-scores on the MMPI-2 F scale were higher for men and somewhat lower for women than on the original MMPI F scale. Greene (1991) also reported that higher T-scores on the MMPI-2 F scale were more common in clinical samples.

A newly developed scale for the MMPI-2 that provides an additional measure of consistency of item endorsement is the F_B scale. This scale originally consisted of 64 items but only 40 of them were included in the final version of the MMPI-2. The F_B scale functions like the F scale except that the items in this scale are located in the latter part of the MMPI-2 (e.g., items 281-555).

Developed in the same manner as the F scale, the F_B scale can be used as an adjunct in the interpretation of the F scale. Both the F and F_B scales consist of infrequently endorsed items. If the person is endorsing items consistently, approximately the same number of items on both scales would be endorsed (Graham, 2000). If the F scale shows a valid protocol, an elevated F_B scale may show that the person stopped paying attention to the items that occurred in the latter half of the booklet and began randomly responding to the items. A cutoff raw score of ≥ 9 on F_B has been cited as producing an overall hit rate of 92.5% and with sensitivity and specificity of 91.5% and 92.6%, respectively (Berry et al., 1991).

MMPI-2 Inconsistency Scales

The Variable Response Inconsistency Scale (VRIN) provides a measure of the consistency of item endorsement on the MMPI-2 similar to the MMPI CLS (Greene, 1978). VRIN was developed to detect tendencies to respond indiscriminately to the MMPI-2 items. This scale also functions as a measure of the inability to read and comprehend the items, carelessness, or errors in the recording of responses (Caldwell, 1991). In the 67-item pairs that are similar or opposite in content, 18-item pairs are repeated, leaving the total of 49 unique item pairs so VRIN raw scores can range from 0 to 49. Each time a person answers an item pair inconsistently, one raw score point is added to the VRIN score. Hand scoring is complicated because points can be added for two true responses, two false responses, or a true and false response depending upon the item pair being evaluated (Graham, 2000).

Butcher et al. (1989) reported that a raw score of ≥ 13 indicates inconsistent responding that may invalidate the protocol, while Graham (1993) suggested that raw scores greater than 12 on VRIN will be found with random responding. If a person responds true to all items or false to all items, the raw score will be equal to 5. Berry et al. (1991) found in college students that the maximal hit rate was 97.1% using a cutoff raw score of ≥ 14, with sensitivity and specificity of 41.4% and 100%, respectively. Paolo and Ryan (1992) found when evaluating male veteran psychiatric outpatients that the overall hit rate for VRIN using a cutoff raw score of ≥ 13 was 92%, with sensitivity and specificity 87% and 97%, respectively. VRIN is most sensitive to similar or moderately divergent true and false response percentages (Nichols & Greene, 1997). However, VRIN is ineffective in detecting acquiescent and nonacquiescent response sets. A VRIN raw score of 5 would show response consistency in both acquiescent or nonacquiescent response sets.

Another cause of inconsistent item endorsement may be the waning motivation of the person. Greene cautions that waning motivation is a particular problem with the MMPI-2 because a significant portion of the content scale items are congregated in the last 267 items and, in fact, items that specifically address harm to self or others occur in the last 50 items. He identifies three methods for assessing the possibility of waning motivation to endorse items consistently as examining the measures across blocks of 100 items, especially when VRIN raw scores are between 6 and 16. These three

methods include: 1) determining the absolute value of the difference between raw scores on the F and F_B scales, 2) adding the VRIN raw score plus the absolute value of F - F_B, and 3) summing the raw scores on the F and F_B scales plus the absolute value of F - F_B (Greene, 1991). Greene (1991) found that psychiatric patients were only slightly higher on VRIN and significantly higher on F and F_B scales than the MMPI-2 normative group, and both groups were distinctly lower than random data on all three scales. In analyzing VRIN scores greater than 9 in 37 psychiatric inpatients by blocks of one-hundred items, Greene (1991) found that slightly less than half endorsed the items consistently, one-third stopped endorsing the items consistently before the end of the test, and one-third endorsed the items inconsistently throughout the test. All patients with VRIN scores greater than 16 endorsed the items inconsistently. He implied that scores greater than 16 show an inconsistent pattern of item endorsement while a score less than 6 showed item consistency. Berry et al. (1992) found that self-reports of random responding correlated significantly with the F, F_B, and VRIN scales in samples of college students and volunteers. In addition, Wetter, Baer, Berry, Smith, and Larsen (1992) found that both analog random and malingering responding produced significant elevations on the F and F_B scales, whereas VRIN was affected only by random responding. Therefore, they implied that clinicians and researchers should not assume random responding is absent from protocols and should screen for its presence through a review of the F, F_B, and VRIN scales. Finally, Cramer (1995) found that summing the raw scores on the F

and F_B scales plus the absolute value of F - F_B index could distinguish all levels of profile randomness.

VRIN also provides an independent estimate of consistency of item endorsement, since it is relatively unaffected by the type and severity of psychopathology and will detect some profiles with inconsistent responses that would be considered valid by traditional validity indicators such as the F scale (Greene, 1991). As Greene (1991) suggested above, the VRIN scale is most useful when it is used in conjunction with the F and F_B scales. High F scale scores and high VRIN scale scores suggest that the person responded randomly to the MMPI-2. In contrast, a high F scale score and a low or moderate VRIN scale score shows that the protocol did not result from random responding or confusion, and may, instead, suggest that the person was severely disturbed and responded consistently to the items or approached the items with the intention of appearing more disturbed than is really the case (Graham, 1993; Wetter et al., 1992).

The True Response Inconsistency Scale (TRIN) of the MMPI-2 was developed to identify persons who respond to items by giving true responses indiscriminately (acquiescence) or false responses indiscriminately (nonacquiescence), which may lead to a protocol that is invalid and uninterpretable. TRIN consists of 23-item pairs that are opposite in item content, but since three pairs are repeated, it really consists of 20 unique item pairs so TRIN raw scores can range from 0 to 20. Two "true" responses or two "false" responses to the same item pair indicate inconsistent responding

(Graham, 1993). Eleven of the 20-item pairs are scored as being inconsistent if the response is "true" to both items, while "six" of the item pairs are scored as being inconsistent if the response is "false" to both items. Three additional item pairs are scored as being inconsistent if the responses are either both "true" or both "false".

The raw score on TRIN is derived by subtracting the number of item pairs with "false" responses from the number of item pairs with "true" responses, and then adding a constant value of +9 to the difference (Graham, 1993). These 9 points are added to the score so that the total TRIN scores will be positive. TRIN raw scores of 13 or more or 5 or less suggest indiscriminant responding that may invalidate the protocol (Butcher et al., 1989). Higher TRIN scores indicate acquiescence while lower TRIN scores indicate nonacquiescence (Graham, 1993). Very high TRIN or very low scores also may reflect indiscriminate true or false responding. Butcher et al. (1989) indicated that true and false responding to item pairs without reference to item content could indicate lack of reading skills, confusion, or negativistic attitudes toward the assessment process. However, Greene (2000) advised that the TRIN scale should not be used as a measure of response inconsistency. High or low TRIN scores reflect a tendency to indiscriminately answer true "yea-saying" or a false "nay-saying" without regard to the TRIN item pair content.

Factors Affecting Assessment of Inconsistency

The following section will examine three of the cited factors for inconsistent item endorsement (intelligence, reading level, and education) more thoroughly.

Effects of Intelligence

Individuals who score below 80 on either Verbal or Full-Scale IQ on the Wechsler Adult Intelligence Scale (WAIS) or the Wechsler Adult Intelligence Scale-Revised (WAIS-R) may be unable to complete the MMPI (Dahlstrom & Welsh, 1960; Dahlstrom et al., 1972; Williams, 1985). Greene (1991) has suggested that persons who score less than 90 on the WAIS-R may require a taped administration of the MMPI-2.

Mittenberg, Tremont, and Rayls (1996) found nonresponsitivity to item content was probable when outpatients diagnosed with central nervous system dysfunction scored below 70 on the Wechsler Memory Scale-Revised (WMS-R) Memory or Attention/Concentration index or when the WAIS-R IQ fell 20 points or more below premorbid level. They found these effects to be independent of the individual's measured reading or intellectual level.

To date, studies investigating the relationship between IQ and MMPI-2 scale scores have been inconclusive. Using multivariate analysis, Sabine (1993) found a significant effect for IQ on MMPI-2 responding in his outpatient community sample of lower socioeconomic men and women who were administered either a taped or written version of the MMPI-2. In men, significant T-score differences were found on Scales F, K, 2, 6, 7, 8, and 9

when IQ was split at the median to form high and low groups. In women, a significant T-score difference was found on Scale F, 2, and 5 when IQ was split at the median to form high and low groups. He implied that lower IQs for both men and women are associated with higher elevations on the above scales. Using a forced entry multiple regression technique in the above sample, Sabine (1993) found that IQ had a negative correlation with Scale F and Scale F_B but had a positive correlation with Scale K for men. For women, Sabine (1993) found a negative correlation between IQ and Scale F. Conversely, a nonsignificant relationship was found between IQ and Scale F scores in an outpatient sample of adult Caucasian men referred by probation and vocational rehabilitation agencies (Crossman, Casey, & Reilley, 1994).

Effects of Reading Level

It is estimated that from 21% or 1 million to 25% or 100 million Americans are less than functionally proficient or illiterate and another 25% to 27% have limited reading skills (Davis, Michielutte, Askov, Williams, & Weiss, 1998; Irwin, 1991; Weiss, 1998). With such a high rate of limited reading ability within the population, determination of each individual's ability to read and comprehend adequately the items contained in the MMPI-2 is essential.

Reading comprehension refers to a person's ability to understand linguistic materials that are written or spoken (Sternberg & Powell, 1983). With written material, comprehension depends not only on the information provided by the text itself, but also the readers knowledge about the topic so that the information can be conceptualized as a schema, an organized

framework that contains slots to be filled by the information from the text (Beck & Carpenter, 1986). Reading comprehension involves selective encoding, selective combination and selective comparison in which relevant information is separated from irrelevant information, cues are combined into a workable definition, and new information about a word is related to old information already stored in memory. These processes operate on a relatively stable set of cues (e.g., temporal, spatial, value, stative descriptive, functional descriptive, causal/enablement, class membership, and equivalence cues) provided by the context in which new words occur (Sternberg, 1987).

Since words are the semantic units in language, there is evidence for postulating a relationship between reading comprehension and word or reading recognition skills (Smith, Stenner, Horabin, & Smith, 1989). In fact, word recognition or vocabulary is considered highly predictive of level of reading comprehension (Sternberg, 1987). Word recognition is the medium in the reading process that conveys the message and has been found to have the highest correlation with comprehension difficulty (Chall, 1987). In fact, the comprehensibility or difficulty of the written message is considered largely governed by the familiarity of the words or vocabulary and the sentence structures used (Smith et al., 1989).

In the process of word recognition, as readers recognize a word, they have access to a variety of other sources of information including the word's meaning and syntactic possibilities, linguistic patterns, and memory of the

preceding context. These sources of information then interact to allow the reader to construct an ongoing and updated linguistic and conceptual representation. Word recognition skills have been found to correlate highly with reading comprehension until the third grade and play a detectable role in the reading ability of adults. With adults, better readers have been found to be faster at pronouncing words than less skilled readers (Beck & Carpenter, 1986).

Perhaps one of the reasons why word recognition skills are related to reading comprehension is that the various processes individuals use to understand or comprehend written material in which words are known also are used to infer the meaning of unknown words. Individuals then acquire word meanings, whether spoken or written, by making inferences about the syntactic, semantic, and referential properties of the lexical content (Beck & Carpenter, 1986). Therefore, there is a direct interplay between word recognition and word comprehension skills.

The vocabulary and sentence structures used in a passage have been described as governing the difficulty of written material and, therefore, affecting reading comprehension. Word frequency is considered the best operationalization of the semantic component of reading, while sentence length is considered the best predictor of passage difficulty (Smith et al., 1989).

Butcher et al. (1989) suggest that an 8th grade level of reading comprehension is required for valid administration of the MMPI-2 based on

the findings that only 10% of items exceed an 8th grade reading level or a 999 Lexile value. The Lexile value, developed by Stenner, Horbin, Smith, and Smith (1988) as an index of reading difficulty and comprehension, reflects several variables related to reading difficulty including familiarity of vocabulary, sentence complexity, and sentence length. Butcher et al. (1989) provides the Lexile values for all 567 items of the MMPI-2 that range from -389 to 1506. The 567 MMPI-2 Lexile values were averaged by Paolo, Ryan, and Smith (1991) and Dahlstrom, Archer, Hopkins, Jackson, and Dahlstrom (1994) who found a 5th grade reading level was necessary to comprehend items sufficiently to produce a valid protocol. In addition, Paolo et al. (1991) determined that 90% of the items require less than a 9th grade reading level but three items require a high school education or more. They determined that the Amorality (Ma1) subscale has an average reading level greater than 8th grade and that nine scales have 25% of their items requiring more than an 8th grade reading level: Antisocial Practices (ASP), Type A (TPA), Need for Affection (Hy2), Naivete (Pa3), Lack of Ego Mastery, Defective Inhibition (Sc5), Bizarre Sensory Experiences (Sc6), Amorality (Ma1), Imperturbability (Ma3), and Ego Inflation (Ma4). Their findings suggest that for persons with relatively low reading proficiencies, interpretation of these scales should be made cautiously when results are inconsistent with other information.

Readability of the MMPI-2 also has been determined by the Flesch-Kincaid Grade Level Index formula. This Index is based on the average number of words per sentence and a variable related to the number of

syllables per word which yields a reading difficulty score. The Flesch-Kincaid Grade Level Index equals (.39) x (average number of words per sentence) + (11.8) x (average number of syllables per word) - 15.59 (Software References International, 1992). Using GrammaticMac, Butcher (1991) determined that the Flesch-Kincaid Grade Level Index for the MMPI-2 was 5th or 6th grade, while Schinka and Borum (1993) and Dahlstrom et al. (1994) found a 4th or 5th grade level of readability, respectively. Nevertheless, Butcher (1991) implied that the MMPI-2 would require approximately a 6th grade reading level which was consistent with the traditionally required reading level of the MMPI.

The Flesch-Kincaid Reading Ease Index also has been used to determine readability of the MMPI-2. The Flesch-Kincaid Reading Ease Index is based on the number of syllables per 100 words and a variable related to the average sentence length which yields a reading ease score. The Reading Ease Index equals 206.835 - .846 x (number of syllables per 100 words) + 1.015 (average sentence length) (Kincaid, Fishburne, Rogers, & Chissom, 1975). Using this index, Dahlstrom et al. (1994) found a 6th grade level of readability for the complete MMPI-2 item set.

Finally, the Gunning's Fog Index has been used to determine the readability of the MMPI-2. The Gunning's Fog Index is based on the average sentence length and a variable related to the percentage of words with three or more syllables. The Gunning's Fog Index equals .4 (average sentence length + percentage of words with three or more syllables) (Gunning, 1952).

Using this index, Dahlstrom et al. (1994) found an 8th grade level of readability for the MMPI-2 items.

Although differences in the readability of the MMPI-2 items have been found among the various readability indices mentioned above, the Lexile index appears to be the most workable because of its sufficient range of application and its relatively smooth progression in scaling of passage readability from both traditional and recently developed tests of reading comprehension. In fact, Smith (1992) found a mean correlation of .871 between the Lexile Index and Peabody Individual Achievement Test (PIAT), Woodcock Reading Mastery Test (WRMT), Woodcock-Johnson Psychoeducational Battery (WJPB), and the SRA Achievement Series.

To date, studies investigating the relationship between reading level and the F, F_B, and VRIN scales of the MMPI-2 have yielded inconsistent results. Sabine (1991) found that reading level had a strong effect on Scale F, Scale F_B, and VRIN in their lower socioeconomic outpatient community sample of adult men who were administered either a taped or written version of the MMPI-2. They found that reading level and Scale F, Scale F_B, and VRIN were correlated negatively, and that VRIN contributed substantially to the obtained effects. Conversely, Crossman et al. (1994) in their outpatient sample of adult Caucasian men who were referred by probation and vocational rehabilitation agencies found a nonsignificant relationship between reading level and Scale F. In partial support of this finding, Sabine (1993) also found that reading level

had a nonsignificant relationship with Scale F, Scale F_B, and VRIN by forced entry multiple regression.

Effects of Education

A reading level between the 6th and 8th grade is necessary to read the MMPI-2 items adequately (Butcher, 1990). In developing the MMPI-2, the normative group had a higher level of education than the normative group for the MMPI and had completed more years of schooling than had the typical American adult (Dahlstrom & Tellegen, 1993). In fact, the MMPI-2 normative adults reported a higher average educational and occupational level than did adults in the 1980 U.S. census (Bogue, 1985). In comparing the normative groups used for the MMPI and MMPI-2, Dahlstrom & Tellegen (1993) suggest that the original MMPI normative group may have overrepresented lower educational levels. The mean level of schooling in the MMPI normative group completed was 9.7 years for men and 10.0 years for women, while the mean level of schooling in the MMPI-2 normative group was 15.0 for men and 14.4 for women. It has been suggested that individuals with lower educational backgrounds may not have volunteered or were eliminated from the MMPI-2 normative sample because of their inability to read well enough to complete the instrument validly (Butcher, 1990).

Studies investigating the relationship between education and the MMPI-2 scales are mixed. Butcher (1990) found small, positive correlations between Scale K and Scale 5 in the MMPI-2 normative sample. Sabine (1993) also found a significant effect for education on MMPI-2 responding in his lower

socioeconomic outpatient community sample of adult men and women who were administered either a taped or written version of the MMPI-2. In men, significant T-score differences were found on Scales F, 5, 6, 7, 8, and 0 when education was split at the median to form high and low groups. In women, significant T-score differences were found on Scales 1, 3, and 0 when education was split at the median to form high and low groups. These results suggest that lower educational levels for both men and women were associated with higher elevations on the above scales. Conversely, Sabine (1993) found that education had nonsignificant relationships with Scale L, Scale F, Scale K, Scale F_B, and VRIN by using a forced entry multiple regression technique. However, Crossman et al. (1994) found significant negative correlations between education and Scale 9 and the McAndrews Alcoholism-Revised (MAC-R) scale in their outpatient sample of adult Caucasian men who were referred by probation and vocational rehabilitation agencies by correlational analyses.

Finally, Long, Graham, and Timbrook (1994) found education had a significant effect on MMPI-2 responding in the MMPI-2 normative sample. They found that education and Scales L, F, 1, and 0 were correlated negatively, but Scales K and 5 were correlated positively in men. For women, education was correlated negatively with Scales L, F, 1, 2, 5, 9, and 0 but positively correlated with Scale K.

Caldwell (1997) found that individuals with fewer years of education are more likely to endorse the items inconsistently and that there is a steady

decline on the VRIN and TRIN scales in the MMPI-2 as the years of education increase.

In considering the two measures of consistency of item endorsement (e.g., VRIN and TRIN), Greene (2000) showed there is a steady decline on these measures as the years of education decrease resulting in individuals with fewer years of education being more likely to endorse the items inconsistently. He underscored the importance of making sure that persons with less than a high school education have adequate reading skills.

Forensic research also has shown a possible relationship between educational level and protocol validity. Noukki (1995) found that forensic inpatients producing valid profiles had significantly more years of education than forensic inpatients producing invalid profiles. He suggested the possibility that some less educated forensic inpatients may have produced invalid profiles due to reduced ability to read and understand the items. In a study with 258 court-ordered outpatient evaluations, highly educated men scored higher on Scale 5 (Shondrick, Ben-Porath, & Stafford, 1992).

Statement of the Problem

To date, there remain discrepancies in the literature regarding the relationships between intelligence, reading level, and education and the MMPI-2 measures of response consistency. The studies reviewed above indicate that intelligence and reading level have effects on Scales K, F, and F_B, while education has a significant effect on Scales K, F, F_B, 1, VRIN, and TRIN. In addition, readability estimates of the MMPI-2 have ranged from the 4th to

the 8th grade and have been identified as one of the reasons for inconsistent item endorsement that may be responsible for higher elevations on the infrequency and inconsistency scales.

Because intelligence, reading level, and education have been identified as potential reasons for inconsistent item endorsement, this study will investigate the influence of these factors on the MMPI-2 measures of response consistency. With the increasing use of the MMPI-2 in forensic investigations, this study will examine the relationships between intelligence, reading level, and education on the MMPI-2 infrequency and inconsistency measures for forensic litigants.

Hypothesis

There will be no correlation between intelligence, education, and reading level and F scale scores, F_B scale scores, VRIN scale scores, TRIN scale scores, absolute value of F - F_B, VRIN + the absolute value of F - F_B, and F + F_B + the absolute value of F - F_B.

CHAPTER II: METHODS

Participants

A consecutive sample of 385 adult forensic outpatients was examined with the MMPI-2, the WAIS-R, and the WRAT-R from 1991 to 2001. The majority of adults were tested in 1995. Most adults had been referred by an attorney due to personal injury litigation; the majority of adults claimed minor traumatic brain injury. These adults were referred for outpatient neuropsychological testing by physicians, attorneys, insurers/claims adjusters, state agencies, or other community agencies. Inclusion criteria used in this study were 18 years of age or older. The only exclusionary criteria were an adult's age of less than 18 and a WAIS-R FSIQ of less than 70. Ten adults were excluded due to age, 25 adults due to WAIS-R FSIQ, and 31 adults due to missing a value for gender. A summary of descriptive characteristics of the forensic sample and predictor variables is presented in Tables 1 and 2.

Instruments

The Wechsler Adult Intelligence Scale-Revised (WAIS-R) Wechsler (1981) was designed to estimate the intellectual ability of individuals between the ages of 16 years, 0 months to 74 years, 11 months. It contains 11 subtests six comprising the Verbal Scale and five comprising the Performance Scale. This instrument provides three IQ scores: Verbal IQ (VIQ), Performance IQ (PIQ), and Full-Scale IQ (FSIQ). Each of the three IQ scores has a mean of 100 (SD = 15). The standardization sample consisted of 1,880 persons between

Table 1

Descriptive Characteristics of the Forensic Sample

Variable	%	N	M	SD	Range
Age		385	35.90	11.54	18 – 75
Gender					
Men	54	208			
Women	46	177			
Civil Status					
Single	41	158			
Married	26.8	103			
Separated	22.9	88			
Divorced	8.3	32			
Widowed	.8	3			
Common Law	.3	1			
Head Injury Severity					
Benign	33	113			
Mild TBI	26.6	91			
Severe Impairment In > 2 Neuropsychological Tests					
No	55	205			
Yes	45	168			
Records Contradict Claims					
No	82	306			
Yes	18	67			
Litigation Status					
Third Party	72.9	223			

Table 1

Descriptive Characteristics of the Forensic Sample

Variable	%	N	M	SD	Range
GED Completion if < Than H.S. Education					
No	89.3	334			
Yes	10.7	40			
Years of Community College					
Zero Years	79.3	295			
Years of College or Graduate School					
Zero Years	80.6	300			
Outpatient Mental Health Counseling In the Past					
No	71	260			
Yes	29	106			

the ages of 16 and 74 years meeting the seven demographic strata approximating the 1970 U.S. census data. The WAIS-R manual reports split-half and test-retest reliability coefficients greater than or equal to .93. Previous research has shown that the WAIS-R has satisfactory construct validity in that its correlations with other intelligence tests ranged from .43 to .94 (Wechsler, 1981).

The Wide Range Achievement Test-Revised (WRAT-R) is the most frequently used instrument within the achievement domain and ranks sixth in

Table 2

Descriptive Characteristics of Predictor Variables for the Forensic Sample

Variable	N	Mean	Standard Deviation	Range
WAIS-R FSIQ	385	89.82	11.02	70 - 133
WRAT-R Reading Level	372	87.90	15.21	45 - 122[a] 4th - 12+[b]
Education in Years	384	11.57	1.61	2 - 21

Note. (Superscripts [a] and [b] equal standard scores and grade level, respectively).

providing important information for persons who assess adults (Harrison, Kaufman, Hickman, Kaufman, 1988).

In addition, the reading subtest from the Wide Range Achievement Test also is the most frequently used test by neuropsychologists to estimate the reading level ability of adults (Stevens & Price, 1999).

The WRAT-R (Jastak & Wilkinson, 1984) was designed to estimate reading ability of individuals between the ages of 12 years, 0 months to 74 years, 11 months and takes about five minutes to administer. It is suggested that Standard Scores be used for the interpretation of test data. Standard Scores are scaled much like Intelligence Quotients, with a mean of 100 (SD =

15). The normative sample included 5,600 individuals stratified by age, sex, race, geographic region, and community of residence. Reliability of the WRAT-R was established with test-retest coefficients ranging from .79 to .97. Its concurrent validity is supported by correlation coefficients of .97 and .74 between the WRAT-R Level 2 Reading subtest and the Reading Recognition subtest of the Peabody Individual Achievement Test (PIAT) and the Reading Comprehension subtest of the PIAT (Jastak & Wilkinson, 1984), respectively.

Research Design & Data Analysis

The predictor variables in this investigation include WAIS-R FSIQ scores, WRAT-R Level 2 Reading subtest Standard Scores, and the total number of years in education completed by each participant. Criterion variables include the F, F_B, VRIN scale T-scores, and the absolute value of the difference between raw scores on the F and F_B, adding the VRIN raw score plus the absolute value of F - F_B, and summing the raw scores on the F and F_B scales plus the absolute value of F - F_B.

The sample of forensic outpatients was divided into two groups: one comprised with off identification numbers and the other group with even numbers based on their identification number in the data set. An independent samples t-test was used in a between subjects design to assess whether the means of the two groups were statistically different from each other on each criterion variable.

A correlational design was used to compare the multiple correlations between intelligence, reading level, and education and each criterion

variable listed above using a stepwise multiple regression procedure for every odd numbered participant. Then, the resulting beta weights were applied to the sample with the even numbered participants resulting in a set of composite scores to be correlated with the values of the criterion variable for the sample. This resulting multiple correlation was compared to the multiple correlation calculated on the odd numbered participants to ensure that a statistically stable relationship between the predictor variables and the criterion variables was measured.

CHAPTER III: RESULTS

Data Analysis and Results

Correlational analysis conducted on the forensic sample found significant negative correlations between Scale F, Scale F_B, Scale TRIN, and F + F_B + /F – F_B/ and the WAIS-R FSIQ, WRAT-R reading level, and years in education. These significant negative correlations suggest that lower WAIS-R FSIQ scores, WRAT-R reading level scores, and years in education were associated with higher F, F_B, TRIN, and F + F_B + /F – F_B/ scores. Significant negative correlations were found between VRIN and VRIN + /F - F_B/ and the WAIS-R FSIQ and WRAT-R reading level scores. These significant negative correlations suggest that lower WAIS-R FSIQ and WRAT-R reading level scores were associated with higher VRIN and VRIN + /F – F_B/ scores. Finally, a significant negative correlation was found between /F – F_B/ and the WAIS-R FSIQ scores. This significant negative correlation suggests that lower WAIS-R FSIQ scores were associated with higher /F – F_B/ scores. A summary of descriptive statistics and intercorrelations between MMPI-2 criterion variables and predictor variables for the forensic sample is presented in Table 3.

An independent sample t-test was used to compare mean T-scores for the first four MMPI-2 criterion variables and raw scores for the last three MMPI-2 criterion variables for the odd and even samples. All seven comparisons between the odd and even forensic samples were nonsignificant. That is, MMPI-2 criterion variable mean scores were

Table 3

Descriptive Statistics and Intercorrelations Between MMPI-2 Criterion Variables and Predictor Variables for the Forensic Sample

Criterion Variable	N	M	SD	Predictor Variables: WAIS-R FSIQ	WRAT-R Reading Level	Education in Years
F	369	62.81	15.15	-.300**	-.205**	-.200**
F_B	222	63.25	19.60	-.244**	-.137*	-.143*
VRIN	215	54.86	11.95	-.177**	-.141*	-.123
TRIN	143	58.51	8.47	-.302**	-.233**	-.301**
$/F - F_B/$	218	3.57	2.50	-.215**	-.122	-.075
VRIN + $/F - F_B/$	215	10.18	4.42	-.285**	-.201**	-.118
$F + F_B +$ $/F - F_B/$	214	16.23	9.21	-.317**	-.148*	-.188**

* $p < .05$. ** $p < .01$. (Two-tailed test).

similar between both forensic samples. A summary of MMPI-2 criterion variable differences between the odd and even forensic samples is presented in Table 4.

Stepwise multiple regression analysis for these MMPI-2 criterion variables F, F_B, VRIN, TRIN, $/F - F_B/$, VRIN + $/F - F_B/$, and $F + F_B + /F - F_B/$ were regressed on the WAIS-R FSIQ, WRAT-R reading level, and education in years in the odd forensic sample. Significant models were derived for Scale F, Scale

Table 4

MMPI-2 Criterion Variable Differences Between the Odd and Even Forensic Samples

Criterion Variable	N	M	SD	df	t
F					
Odd	189	62.89	15.53	367	.105
Even	180	62.73	14.79		
F_B					
Odd	110	62.50	19.37	220	-.566
Even	112	63.99	19.88		
VRIN					
Odd	105	54.65	11.90	213	-.255
Even	110	55.06	12.04		
TRIN					
Odd	71	58.46	8.37	141	-.064
Even	72	58.56	8.62		
/F - F_B/					
Odd	110	3.66	2.48	216	.538
Even	108	3.48	2.52		
VRIN + /F – FB/					
Odd	107	10.08	4.38	213	-.321
Even	108	10.28	4.48		
F + F_B + /F – FB/					
Odd	108	16.50	9.09	212	.426
Even	106	15.96	9.37		

Note. (All two-tailed tests were nonsignificant).

F_B, Scale TRIN, /F – F_B/, VRIN + /F – F_B/, and F + F_B + /F – F_B/.

For example, the Beta coefficient in this equation of model 1: [F = 106.153 + -.343 (WAIS-R FSIQ)] showed that an increase of one standard deviation on the WAIS-R produced a decrease of -.343 on the F scale. The Beta coefficient in the first equation of model 1: [TRIN = 80.380 + -.359 (Education)] showed that an increase of one standard deviation in years of education produced a decrease of -.359 on TRIN. Beta coefficients in the second equation of model 2: [TRIN = 96.934 + -.312 (Education) + -.245 (WAIS-R FSIQ)] showed that an increase of one standard deviation in years of education and the WAIS-R produced a decrease of -.312 and -.245 on TRIN, respectively. A summary of stepwise multiple regression analyses for MMPI-2 criterion variables regressed on predictor variables in the odd forensic sample is presented in Table 5.

The results of stepwise multiple regression analyses for Scale F of model one showed that the WAIS-R FSIQ explained 11.8% of the variance in the odd sample and 6.8% of the variance in the even sample after cross-validation. Education and WAIS-R FSIQ best predicted TRIN scores for model two, accounting for 18.7% of the variance in the odd sample and 11.1% of the variance in the even sample after cross-validation. For TRIN in model one, education accounted for 12.9% of the variance in the odd sample and 5.8% of the variance in the even sample after cross-validation. A summary of MMPI-2 criterion model cross-validation for the even forensic sample is presented in Table 6.

Table 5

Stepwise Multiple Regression Analyses for MMPI-2 Criterion Variables Regressed on Predictor Variables in the Odd forensic Sample

Criterion Variable	N	B	SE B	Beta	t
Model 1: F	182				
Constant		106.153			
WAIS-R FSIQ			.098	-.343***	- 4.901
Model 1: F_s	106				
Constant		111.007			
WAIS-R FSIQ			.181	-.282**	- 3.002
Model 1: TRIN	68				
Constant		80.380			
Education			.616	-.359**	- 3.128
Model 2: TRIN	68				
Constant		96.934			
Education			.611	-.312**	- 2.739
WAIS-R FSIQ			.101	-.245*	- 2.154
Model 1: $/F - F_s/$	106	9.486			
Constant			.023		
WAIS-R FSIQ				-.269**	- 2.851
Model 1: VRIN + $/F - F_s/$	103				
Constant		23.289			
WAIS-R FSIQ			.041	-.337***	- 3.600
Model 1: $F + F_s + /F - F_s/$	104				
Constant		46.278			
WAIS-R FSIQ			.084	-.367***	- 3.986

* $p < .05$. ** $p < .01$. *** $p < .001$. (One-tailed tests).

The following tables illustrate the number of forensic outpatients

Table 6

MMPI-2 Criterion Model Cross-Validation for the Even Forensic Sample

Criterion Model	Odd Sample R	Odd Sample R^2	Even Sample R	Even Sample R^2
Model 1: F = 106.153 + -.343 (FSIQ)	.343***	.12	.260**	.07
Model 1: F_B = 111.007 + -.282 (FSIQ)	.282**	.08	.214*	.05
Model 1: TRIN = 80.380 + -.359 (Education)	.359**	.13	.240*	.06
Model 2: TRIN = 96.934 + -.312 (Education) + -.245 (FSIQ)	.433***	.19	.333**	.11
Model 1: VRIN + /F – F_B/ = 23.289 + -.337 (FSIQ)	.337***	.11	.234*	.06
Model 1: F + F_B + /F – F_B/ = 46.278 + -.367 (FSIQ)	.367***	.14	.275**	.08

* $p < .05$. ** $p < .01$. *** $p < .001$. (One-tailed tests for odd group)

(Two-tailed tests for even group)

scoring beyond the cutoff score for each MMPI-2 scale/index by WAIS-R FSIQ levels, educational levels, and WRAT-R reading levels. The majority of forensic outpatients scored beyond a T-score of 90 on Scale F_B and beyond a raw score of six on /F – F_B/ representing 15.32% and 16.51% across all WAIS-R FSIQ levels and educational levels. Across all WRAT-R reading levels, 16.04% and 16.35% scored beyond a T-score of 90 on Scale F_B and beyond a raw score of six on /F – F_B/. In contrast, no forensic outpatient scored beyond a T-score of 80 on TRIN, across all levels of WAIS-R FSIQ,

WRAT-R, and education. No forensic outpatient whose WAIS-R FSIQ ranged from 110 to 130 and above and years in education from 17 to 19 and above scored beyond the cutoff score for any MMPI-2 scale/index. No forensic outpatient scored beyond the cutoff scores on VRIN, TRIN, /F – F_B/, VRIN + /F – F_B/, and F + F_B + /F – F_B/ whose WRAT-R reading level was below 4th grade. A summary of forensic outpatients scoring beyond the cutoff scores for each MMPI-2 scale/index by levels of WAIS-R FSIQ, WRAT-R reading level, and education in years is presented in Tables 7, 8, and 9.

Table 7

Summary of Forensic Outpatients Scoring Beyond the MMPI-2 Scale/Index Score by WAIS-R FSIQ Level

Scales/Index		WAIS-R FSIQ 70-79	80-89	90-109	110-119	120-129	130+	%
F > 90 T-Score	N =	11	10	4	0	0	0	6.78%
F_B > 90 T-Score	N =	9	14	11	0	0	0	15.32%
VRIN > 80 T-Score	N =	0	3	1	0	0	0	1.86%
TRIN > 80 T-Score	N =	0	0	0	0	0	0	.00%
$/F - F_B/ > 6$ (Raw Score)	N =	6	19	11	0	0	0	16.51%
VRIN + $/F - F_B/ > 20$ (Raw Score)	N =	3	2	2	0	0	0	3.26%
$F + F_B +$ $/F - F_B/ > 36$ (Raw Score)	N =	2	2	1	0	0	0	2.34%

Table 8

Summary of Forensic Outpatients Scoring Beyond the MMPI-2 Scale/Index Score by WRAT-R Reading Level

Scales/Index		WRAT-R Reading Level Below 4	4-5	6-7	8-9	10-11	12+	%
F > 90 T-score	N =	2	6	5	4	3	5	7.06%
F_B > 90 T-score	N =	1	5	9	7	6	6	16.04%
VRIN > 80 T-score	N =	0	1	1	0	1	0	1.46%
TRIN > 80 T-score	N =	0	0	0	0	0	0	.00%
$/F - F_B/ > 6$ (Raw score)	N =	0	2	12	5	4	11	16.35%
VRIN + $/F - F_B/ > 20$ (Raw score	N =	0	1	4	0	1	1	3.41%
$F + F_B +$ $/F - F_B/ > 36$ (Raw score)	N =	0	1	1	2	0	1	2.45%

Table 9

Summary of Forensic Outpatients Scoring Beyond the MMPI-2 Scale/Index Score by Educational Level

Scales/Index		Educational Level 2-8	9-10	11-12	13-14	15-16	17-18	19+	%
F > 90 T-score	N =	2	5	18	0	0	0	0	6.79%
F_B > 90 T-score	N =	1	10	20	1	2	0	0	15.32%
VRIN > 80 T-score	N =	0	1	3	1	0	0	0	1.86%
TRIN > 80 T-score	N =	0	0	0	0	0	0	0	.00%
$/F - F_B/ > 6$ (Raw score)	N =	2	6	26	1	1	0	0	16.51%
VRIN + $/F - F_B/ > 20$ (Raw score)	N =	0	2	5	0	0	0	0	3.26%
$F + F_B +$ $/F - F_B/ > 36$ (Raw score)	N =	0	2	3	0	0	0	0	2.34%

CHAPTER IV: DISSCUSSION

The purpose of this study was to develop statistically stable stepwise multiple regression models that can be used by clinicians to predict the probability for completion of a consistent MMPI-2 profile based on a forensic outpatient's WAIS-R FSIQ, WRAT-R reading level, and education in years before the MMPI-2 is administered.

Using correlational analysis for the complete forensic sample, significant negative correlations were found between the MMPI-2 criterion variables F, F_B, TRIN, and $F + F_B + /F - F_B/$ and the WAIS-R FSIQ, WRAT-R reading level, and years in education. Significant negative correlations were found between VRIN and $VRIN + /F - F_B/$ and the WAIS-R FSIQ and WRAT-R reading level. Finally, a significant negative correlations was found between $/F - F_B/$ and WAIS-R FSIQ. For forensic outpatients in the odd sample, stepwise multiple regression analyses for each MMPI-2 criterion variable F, F_B, TRIN, $/F - F_B/$, VRIN + $/F - F_B/$, and $F + F_B + /F - F_B/$ regressed on the WAIS-R FSIQ, WRAT-R reading level, and education in years produced significant multiple correlation coefficients for all the models except VRIN. For forensic outpatients in the even sample, cross-validation was conducted for all significant MMPI-2 criterion variable models developed from the odd sample above and significant multiple correlation coefficients were found for all the models except $/F - F_B/$.

The results of this study are consistent with Sabine's (1993) findings that IQ has significant negative relationships with the F and F_B scales. However,

results of this study are inconsistent with their findings that reading level had a nonsignificant relationship with Scales F, F_B, and VRIN. No correlation was found between reading level and the F, F_B, and VRIN scales perhaps due to restriction of reading level range to 6th grade and above.

There were more inconsistencies with lower education and that there is a decline in VRIN and TRIN as education increases as Caldwell (1997) and (Greene, 2000) found. The same decline in consistency of item endorsement was found in F as education increases (Butcher, 1990; Long, Graham, Timbrook, 1994).

These results can be used by clinicians to predict the probability for completion of a consistent MMPI-2 profile based on a forensic outpatient's WAIS-R FSIQ score and/or years of completed education. For example, if the clinician knows the forensic outpatient's WAIS-R FSIQ score and years of completed education, then, using the derived regression equations, the clinician can predict F, F_B, TRIN, VRIN + /F - F_B/, and F + F_B + /F - F_B/ scores that will most probably result before the forensic outpatient takes the MMPI-2. This procedure would allow the clinician to anticipate whether a forensic outpatient can produce a consistent MMPI-2 profile; therefore, saving valuable time and effort in this era of managed care. Using this information, the clinician may be better able to determine inconsistent profiles, if the forensic outpatient is just answering a lot of the items with true or false bias or just randomly responding to items, or malingering. In forensic settings, this information can be important since forensic outpatients may already be

affected by possible secondary gains resulting in certain MMPI-2 profile types.

A concrete example is provided in which the derived regression equations can be used to obtain predicted MMPI-2 scores for Mr. B. Mr. B. is a 35 year old Caucasian, male who was referred by his attorney for neuropsychological testing following a motor vehicle accident. He claims to have a mild traumatic brain injury. Mr. B. completed the 10^{th} grade. He was administered the WAIS-R and WRAT-R. Results of the WAIS-R indicated a FSIQ of 80 and a standard score of 45 (below 4^{th} grade reading level) on the WRAT-R. By substituting Mr. B's FSIQ score and educational level into the derived stable multiple regression equations developed from the odd sample of forensic outpatients, Mr. B. would be expected to produce a T-score of 78.71 on the F Scale, a F_B T-score of 88.35, a TRIN T-score of 76.79 on model one and a TRIN T-score of 74.21 on model two, a raw score of -3.67 on VRIN + $/F - F_B/$, and a raw score of 16.92 on $F + F_B + /F - F_B/$. These predicted scores would allow the clinician to anticipate whether Mr. B. with a mild traumatic brain injury could produce a consistent and valid MMPI-2 profile before the test is administered. A summary of these predicted MMPI-2 criterion variable scores is presented in Table 10.

There are several limitations of this study that may affect the generalizability of the results. The forensic outpatients in this study were from 18 to 76 years of age, all from the Midwest, and a majority referred by an attorney or seeking compensation following traumatic brain injuries.

Table 10

Summary of MMPI-2 Predicted Criterion Scores for Mr. B.

Model 1:	$\underline{F} = 106.153 + -.343\ (FSIQ)$ $\underline{F} = 106.153 + -.343\ (80)$ $\underline{F} = 78.71$
Model 1:	$\underline{F_B} = 111.007 + -.282\ (FSIQ)$ $\underline{F_B} = 111.007 + -.282\ (80)$ $\underline{F_B} = 57.82$
Model 1:	$\underline{TRIN} = 80.380 + -.359\ (Education)$ $\underline{TRIN} = 80.380 + -.359\ (10)$ $\underline{TRIN} = 76.79$
Model 2:	$\underline{TRIN} = 96.934 + -.312\ (Education) + -.245\ (FSIQ)$ $\underline{TRIN} = 96.934 + -.312\ (10) + -.245\ (80)$ $\underline{TRIN} = 74.21$
Model 1:	$\underline{VRIN} + /\underline{F} - \underline{F_B}/ = 23.289 + -.337\ (FSIQ)$ $\underline{VRIN} = /\underline{F} - \underline{F_B}/ = 31.289 + -.337\ (80)$ $\underline{VRIN} = 4.33$
Model 1:	$\underline{F} + \underline{F_B} + /\underline{F} - \underline{F_B}/ = 46.278 + -.367\ (FSIQ)$ $\underline{F} + \underline{F_B} + /\underline{F} - \underline{F_B}/ = 46.278 + -.367\ (80)$ $\underline{F} + \underline{F_B} + /\underline{F} - \underline{F_B}/ = 16.92$

This design was correlational and as such, cannot determine causal relationships. Therefore, there may be other variables that could account for more of the total variance explained by the predictor variables used in this study. In addition, the direction of causality is difficult to determine with this type of research design.

Another possible limitation may involve item comprehension. Therefore, it is important to determine each client's ability not only to read but also to comprehend the MMPI-2 items. Adequate fluency in oral reading

may mask serious limitations that a client may have in item comprehension. In the data set used in this study, the taped administration of the MMPI-2 was used for all clients with reading levels less than 6th grade. However, even using the taped version, their reading comprehension may have been compromised.

Another possible limitation may be the item content of Scale F_B that emphasizes suicidal ideation, helplessness, and problems with relations (Greene, 2000). Since a majority of clients had F_B T-scores greater than 90 across all levels of WAIS-R FSIQ and education and across all levels of the WRAT-R, this suggested that these forensic outpatients may be experiencing severe affective upheaval or had chosen to simulate a mood disturbance which reflects actual panic anxiety or depressive symptoms, when F_B is elevated and F is in the valid range (Nichols, 2001). As a rule of thumb, scores are generally valid if two out of four MMPI-2 Content Scales, i.e., Depression (DEP), Low Self-Esteem (LSE), Work Interference (WRK), and Negative Treatment Indicators (TRT) were above scores on Bizarre Mentation (BIZ) or if scores on Anxiety (ANX) and Fears (FRS) were greater than BIZ. If these conditions are unmet, then F_B may suggest exaggeration of problems or carelessness in responding to the last 100 items of the MMPI-2.

In addition, since a majority of clients had a raw scores beyond seven on the /F – F_B/ index across all levels of WAIS-R FSIQ and education and across all levels of WRAT-R, this may suggest that the MMPI-2 items were answered inconsistently.

Recommendations for future research would include administering a brief measure of reading comprehension (e.g., The Ohio Literacy Test) along with a brief measure of reading recognition (e.g., WRAT-3) to all clients before administering the MMPI-2. Since the MMPI-2 instructions have been found to be more difficult than the test items themselves, Dahlstrom et al. (1994), Nichols (2001), and Schinka & Borum (1993) recommend that having the client read and explain the instructions or sample items from several levels of difficulty could help evaluate a client's level of reading comprehension. However, administering a brief measure of reading comprehension (e.g., The Ohio Literacy Test) would appear to be more reliable and valid than asking clients to read and explain the instructions or test items.

Regarding the elevated $\underline{F_B}$ scale scores when the $\underline{F}$ scores are in the acceptable range, you could compare two different forensic outpatient samples to determine if there were statistically significantly when the Content Scale scores on DEP, LSE, WORK, and TRT were greater than BIZ or if the scores on ANX and FRS were greater than BIZ.

In reference to the number of inconsistencies found on the $/\underline{F} - \underline{F_B}/$ index across all levels of WAIS-R FSIQ and education and across all levels of the WRAT-R, you also could compare two different forensic samples to determine if there were statistically significant.

Finally, replication of this study with other forensic outpatient samples is recommended to ensure that these results can be generalized to other forensic samples and to other geographic regions of the United States.

REFERENCES

Anastasi, A. (1988). *Psychological Testing* (6th ed.). New York: Macmillan.

Beck, I. L., & Carpenter, P. A. (1986). Cognitive approaches to Understanding reading. *American Psychologist*, 41, 1098-1105.

Ben-Porath, Y. S. & Butcher, J. N. (1989). Psychometric stability of rewritten MMPI items. *Journal of Personality Assessment*, 53, 645-653.

Ben-Porath, Y. S., & Tellegen, A. (1992). Continuity and changes in MMPI-2 Validity indicators: Points of clarification. *MMPI-2 News and Profiles*, 3, 6-8.

Berry, D. T. R., Wetter, M. W., Baer, R. A., Larsen, L., Clark, C., & Monroe, K. (1992). MMPI-2 random responding indices: Validation using a self-report methodology. *Psychological Assessment*, 4, 340-345.

Berry, D. T. R., Wetter, M. W., Baer, R. A., Widiger, T. A., Sumpter, J. C., Reynolds, S. K., & Hallam, R. A. (1991). Detection of random responding on the MMPI-2: Utility of F, back F, and VRIN scales. *Psychological Assessment*, 3, 418-423.

Bogue, D. (1985). *The population of the United States: Historical trends and future projections*. New York: The Free Press.

Bond, J. A. (1986). Inconsistent responding to repeated MMPI items: Is its major cause really carelessness? *Journal of Personality Assessment*, 50, 50-64.

Buechley, R., & Ball, H. (1952). A new test of validity for the group MMPI. *Journal of Consulting Psychology*, 16, 299-301.

Butcher, J. N. (1990). Education level and MMPI-2 measured psychopathology: A case of negligible influence. *MMPI-2 News and Profiles*, 1, 2.

Butcher, J. N. (1990). What interpretive changes do the educational differences between the original MMPI and MMPI-2 normative samples require? *MMPI-2 News and Profiles*, 1, 4.

Butcher, J. N. (1991). Note: How readable is the MMPI-2 anyway? *MMPI-2 News and Profiles*, 2, 3.

Butcher, J. N., Dahlstrom, W. G., Graham, J. R., Tellegen, A., & Kaemmer, B. (1989). *Minnesota Multiphasic Personality Inventory-2 (MMPI-2): Manual for administration and scoring*. Minneapolis, MN: University of Minnesota Press.

Caldwell, A. B. (1991). Commentary on "The Minnesota Multiphasic Personality Inventory-2: A review". *Journal of Counseling and Development*, 69, 568-569.

Caldwell, A. B. (1997). *Forensic questions and answers in the MMPI/MMPI-2*. Los Angeles: Caldwell Report.

Chall, J. S. (1987). Two vocabularies for reading: Recognition and meaning. In M. G. McKeown and M. E. Curtis (Eds.), *The nature of vocabulary acquisition* (pp. 7-17). Hillsdale, NJ: Lawrence Erlbaum Associates.

Coché, E., & Steer, R. A. (1974). The MMPI response consistencies of normal, neurotic, and psychotic women. *Journal of Clinical Psychology*, 30, 194-195.

Cramer, K. M. (1995). Comparing three new MMPI-2 randomness indices in a novel procedure for random profile derivation. *Journal of Personality Assessment*, 65, 514-520.

Crossman, L. L., Casey, T. A., & Reilley, R. R. (1994). Influence of cognitive variables on MMPI-2 scale scores. *Measurement and Evaluation in Counseling and Development*, 27, 151-157.

Dahlstrom, W. G., & Tellegen, A. (1993). *Manual supplement: Socioeconomic status and the MMPI-2. The relation of MMPI-2 patterns to levels of education and occupation*. Minneapolis: University of Minnesota Press.

Dahlstrom, W. G., Archer, R. P., Hopkins, D. G., Jackson, E., & Dahlstrom, L. E. (1994). *Assessing the readability of the Minnesota Multiphasic Personality Inventory instruments -- MMPI, MMPI-2, and MMPI-A.* (Tech. Rep. No. 2). Minneapolis, MN: University of Minnesota Press.

Dahlstrom, W. G., & Welsh, G. S. (1960). *An MMPI handbook: A guide to clinical practice and research*. Minneapolis: University of Minnesota Press.

Dahlstrom, W. G., Welsh, G. S., & Dahlstrom, L. E. (1972). *An MMPI handbook: Volume I. Clinical interpretation*. Minneapolis: University of Minnesota Press.

Dahlstrom, W. G., Welsh, G. S., & Dahlstrom, L. E. (1975). *An MMPI handbook: Volume II. Research applications*. Minneapolis, MN: University of Minnesota Press.

Davis, T. C., Michielutte, R., Askov, E. N., Williams, M. V., & Weiss, B. D. (1998). Practical assessment of adult literacy in health care. *Health Education and Behavior*, 25, 613-624.

Evans, R. G., & Dinning, W. D. (1983). Response consistency among high F scale scores on the MMPI. *Journal of Clinical Psychology*, 39, 246-248.

Fox, D. D., Gerson, A., & Lees-Haley, P. R. (1995). Interrelationship of MMPI-2 validity scales in personal injury claims. *Journal of Clinical Psychology*, 51, 42-47.

Graham, J. R. (1993). *MMPI-2: Assessing personality and psychopathology* (2nd ed.). New York: Oxford University Press.

Graham, J. R. (2000). MMPI-2: *Assessing personality and psychopathology* (3rd ed.). New York: Oxford University Press.

Grammatik V, *The easiest way to improve your writing: User's guide* (1992). San Francisco: Reference Software International.

Gravitz, M. A., & Gerton, M. I. (1976). An empirical study of internal consistency in the MMPI. *Journal of Personality Assessment*, 32, 567-568.

Greene, R. L. (1978). An empirically derived MMPI carelessness scale. *Journal of Personality Assessment*, 43, 69-71.

Greene, R. L. (1979). Response consistency of the MMPI: The TR index. *Journal of Personality Assessment*, 43, 69-71.

Greene, R. L. (1989). *Assessing the validity of the MMPI in clinical settings.* Clinical notes on the MMPI (No. 11). Minneapolis: National Computer systems.

Greene, R. L. (1991, July). *Some issues in assessing consistency of item endorsement on the MMPI-2*. Paper presented at the MMPI-2 Summer Institute, Colorado Springs, CO.

Greene, R. L. (1991). *The MMPI/MMPI-2: An interpretive manual*. Boston: Allyn & Bacon.

Greene, R. L. (2000). *The MMPI-2: An interpretive manual* (2nd ed.). Boston: Allyn & Bacon.

Gunning, R. (1952). *The technique of clear writing*. New York: McGraw-Hill.

Haertzen, C. A., & Hill, H. E. (1963). Assessing subjective effects of drugs: An index of carelessness and confusion for use with the addictions research center inventory (ARCI). *Journal of Clinical Psychology*, 19, 407-412.

Harrison, P., Kaufman, A., Hickman, J., & Kaufman, N. (1988). A survey of tests used for adult assessment. *Journal of Psychoeducational Assessment*, 6, 188-198.

Hathaway, S. R., & McKinley, J. C. (1940). A multiphasic personality schedule (Minnesota): I. Construction of the schedule. *Journal of Psychology*, 10, 249-254.

Irvin, P. (1991). *Adult literacy issues, programs, and options*. Washington: Congressional Research Service.

Jastak, S., & Wilkinson, G. (1984). *Wide Range Achievement Test-Revised*. Wilmington, DE: Jastak Associates.

Jones, F. W., Neuringer, C., & Patterson, T. W. (1976). An evaluation of an MMPI response consistency measure. *Journal of Personality Assessment*, 40, 419-421.

Kincaid, J. P., Fishburne, R. P., Rogers, R. L., & Chissom, B. S. (1975). Derivation of new readability formula (automatic readability index, fog count, and Flesch reading ease formula) for Navy enlisted personnel. *Navy Training Command Reasearch Branch Report*, pp. 8-75.

Long, K. A., Graham, J. R., & Timbrook, R. E. (1994). Socioeconomic status and MMPI-2 interpretation. *Measurement and Evaluation in Counseling and Development*, 27, 158-177.

Lubin, B., Larson, R. M., Matarazzo, J. D., & Seever, M. F. (1985). Psychological test usage patterns in five professional settings. *American Psychologist*, 40, 857-861.

Maloney, M. P., Duvall, S. W., & Friesen, J. (1980). Evaluation of response consistency of the MMPI. *Psychological Reports*, 46, 295-298.

Mittenberg, W., Tremont, G., & Rayls, K. R. (1996). Impact of cognitive functions on the MMPI-2 validity in neurologically impaired patients. *Assessment*, 3, 157-163.

Munley, P. H. (1991). A comparison of MMPI-2 T-scores for men and women. *Journal of Clinical Psychology*, 47, 87-91.

Nichols, D. S. (2001). *Essentials of MMPI-2 assessment*. New York: John Wiley & Sons, Inc.

Nichols, D. S., & Greene, R. L. (1997). Dimensions of deception in personality assessment: The example of the MMPI-2. *Journal of Personality Assessment*, 68, 251-266.

Nichols, D. S., Greene, R. L., & Schmolck, P. (1989). Criteria for assessing inconsistent patterns of item endorsement on the MMPI: Rationale, development, and empirical trials. *Journal of Clinical Psychology*, 45, 239-250.

Noukki, A. L. (1995). Profile validity of the MMPI-2 in forensic settings. *MMPI-2 News and Profiles*, 6, 1.

Paolo, A. M., & Ryan, J. J. (1992). Detection of random response sets on the MMPI-2. *Psychotherapy in Private Practice*, 11, 1-8.

Paolo, A. M., Ryan, J. J., & Smith, A. J. (1991). Reading difficulty of MMPI-2 subscales. *Journal of Clinical Psychology*, 47, 529-532.

Rogers, R., Dolmetsch, R., & Cavanaugh, J. L., (1983). Identification of random responders on MMPI protocols. *Journal of Personality Assessment*, 47, 364-368.

Rogers, R., Harris, M., & Thatcher, A. A. (1983). Identification of random responders on the MMPI: An actuarial approach. *Psychological Reports*, 53, 1171-1174.

Sabine, D. M. (1993). *The use of the MMPI-2 in the low socioeconomic status group*. Doctoral Dissertation, University of Louisville, Louisville, Kentucky.

Sabine, D. M. (1991). *The utility of the MMPI-2 taped version in a low socioeconomic status group*. Unpublished Master's thesis, University of Louisville, Louisville, Kentucky.

Schinka, J. A., & Borum, R. (1993). Readability of adult psychopathology inventories. *Psychological Assessment*, 5, 384-386.

Shondrick, D. D., Ben-Porath, Y. S., & Stafford, K. P. (1992). Forensic applications of the MMPI-2. *MMPI-2 News and Profiles*, 3, 2.

Smith, D. R. (1992). A descriptive versus a prescriptive use of readability formulas. *Virginia English Bulletin*, 42, 90-93.

Smith, D. R., Stenner, A. J., Horabin, I., & Smith, M. (1989). *The Lexile scale in theory and practice: Final report* (Report No. HD-19448). Washington, DC: National Institutes of health. (ERIC Document Reproduction Service No. ED 307577).

Stenner, A. J., Horbin, I., Smith, D. R., & Smith, M. (1988). *The lexile framework*. Durham: MetaMetrics.

Sternberg, R. J. (1987). Most vocabulary is learned from context. In M. G. McKeown and M. E. Curtis (Eds.), *The nature of vocabulary acquisition* (pp. 89-105). Hillsdale, NJ: Lawrence Erlbaum Associates.

Sternberg, R. S., & Powell, J. S. (1983). Comprehending verbal comprehension. *American Psychologist*, 38, 878-892.

Stevens, K. B., & Price, J. R. (1999). Adult reading assessment: Are we doing the best with what we have? *Applied Neuropsychology*, 6, 68-78.

Tellegen, A. (1988). The analysis of consistency in personality assessment. *Journal of Personality*, 55, 621-663.

Wechsler, D. (1981). *Manual for the Wechsler Adult Intelligence Scale-Revised*. New York: Psychological Corporation.

Weiss, B. D. (1998). Communicating with patients who have limited literacy skills. *Journal of Family Practice*, 46, 168-176.

Wetter, M. W., Baer, R. A., Berry, D. T. R., Smith, G. T., & Larsen, L. H. (1992). Sensitivity of MMPI-2 validity scales to random responding and malingering. *Psychological Assessment*, 4, 369-374.

Williams, C. L. (1985). Use of the MMPI with adolescents. In J. N. Butcher & Graham (Eds.), *Clinical applications of the MMPI* (pp. 37-39). Minneapolis: University of Minnesota Department of Conferences.