

RELATIVE VALIDITY OF THE MMPI-2-RF INTERPRETIVE REPORT AND MMPI-2
COMPUTER-BASED TEST INTERPRETATION SYSTEMS

Joseph K. McLaughlan

A dissertation submitted in partial fulfillment of
the requirements for the degree of
Doctor of Philosophy

Department of Psychology

Central Michigan University
Mount Pleasant, Michigan
January 2013

Accepted by the Faculty of the College of Graduate Studies,
Central Michigan University, in partial fulfillment of
the requirements for the doctoral degree

Dissertation Committee:

Nathan Weed, Ph.D.	Committee Chair
Mark Deskovitz, Ph.D.	Faculty Member
Kyunghee Han, Ph.D.	Faculty Member
Reid Skeel, Ph.D.	Faculty Member
January 28, 2013	Date of Defense
Roger Coles, Ed.D.	Dean College of Graduate Studies
June 26, 2013	Approved by the College of Graduate Studies

ACKNOWLEDGEMENTS

I would like to thank the University of Minnesota Press for their generous grant; Dr. Nathan Weed for facilitating my research interests and guidance throughout my undergraduate and graduate education; the dissertation committee members, Drs. Mark Deskovitz, Kyunghye Han, and Reid Skeel for their insights and recommendations; Brian McCabe for completing ratings and always being available to assist with data analysis and interpretation; and the raters, Lauren Bieritz, Siew Li Ng, and Hina Pant. Your contributions are each immensely appreciated.

ABSTRACT

RELATIVE VALIDITY OF THE MMPI-2-RF INTERPRETIVE REPORT AND MMPI-2 COMPUTER-BASED TEST INTERPRETATION SYSTEMS

by Joseph K. McLaughlan

Computer-based test interpretations (CBTIs) are capable of overcoming a number of limitations of MMPI clinician interpretations, such as bias, errors in judgment, and limitations of knowledge. However, it is essential to provide evidence documenting that CBTIs indeed provide accurate information that is interpreted consistently between clinicians if they are to be used in practice. The present study used an external criterion design to examine the extent to which the MMPI-2-RF CBTI system and MMPI-2 CBTI systems provided clinicians with valid client impressions. Therapists operationalized inferences using the Midwestern Q-sort about 30 clients whom they had seen for four to six sessions. Graduate-level and practicing clinicians completed Midwestern Q-Sorts based on the CBTI profiles from five programs (four MMPI-2 and one MMPI-2-RF) for comparison to therapist ratings. Q-correlations were calculated between q-sorts based on the MMPI-2-RF CBTI, the MMPI-2 CBTIs, and therapist ratings for each of 30 cases. On average, the MMPI-2-RF Interpretive Report profiles were slightly more valid than the individual and aggregated MMPI-2 CBTIs, although validity coefficients were generally low and comparable between programs. Each CBTI program tended to show better validity when at least one Clinical and/or Restructured Clinical scale was elevated. Additionally, the greatest determining factor in the prediction of unique variance in the treating therapists' descriptions was found when at least one scale elevation was produced on the MMPI-2 but not the MMPI-2-RF and vice versa. Interpretive considerations and implications for future research are discussed. For example, it is probable that CBTIs will be most useful in settings in which examinees are likely to endorse significant psychopathology and psychological distress, such as community

mental health agencies and hospitals. Future research should include the examination of CBTI validity based on specific elevation patterns to develop a better understanding of the trend found within the present study.

TABLE OF CONTENTS

CHAPTER

I.	INTRODUCTION.....	1
	Relative Validity of the MMPI-2-RF Interpretive Report and MMPI-2 Computer Based Test Interpretation Systems.....	1
	Differences from the MMPI-2 to the MMPI-2-RF.....	2
	Controversy Regarding the MMPI-2-RF.....	5
	Computer-Based Test Interpretations.....	7
	<i>History</i>	7
	<i>Advantages of CBTIs</i>	8
	<i>CBTI Limitations and Concerns</i>	10
	<i>Current Status of CBTIs</i>	11
	Evaluating CBTIs.....	14
	<i>Q-sort Methodology</i>	14
	<i>CBTI Reliability Studies</i>	15
	<i>CBTI Validity Studies</i>	17
	Present Study.....	21
II.	METHOD.....	23
	MMPI-2 Profiles.....	23
	Raters.....	23
	Instruments.....	23
	<i>Midwestern Q-Sort</i>	23
	<i>CBTIs</i>	24
	Procedure.....	25
	Analyses.....	25
III.	RESULTS.....	28
IV.	DISCUSSION.....	49
	Conclusions.....	52
	Limitations and Future Directions.....	53
	REFERENCES.....	57

LIST OF TABLES

TABLE	PAGE
1. <i>Ratings of the MMPI-2-RF CBTI for Profile 1</i>	29
2. <i>Therapist ratings and average item ratings across CBTIs for Profile 1</i>	29
3. <i>Intercorrelations between CBTI Systems and Therapist Ratings.....</i>	31
4. <i>Most and least consensual Midwestern Q-Sort Items</i>	34
5. <i>Most and least consensual items for the aggregated MMPI-2 CBTIs.....</i>	34
6. <i>Most and least consensual items for the MMPI-2-RF Interpretive Report</i>	35
7. <i>Most and least consensual items within the most valid aggregated MMPI-2 CBTIs.....</i>	36
8. <i>Most and least consensual items within the least valid aggregated MMPI-2 CBTIs.....</i>	37
9. <i>Most and least consensual items within the most valid MMPI-2-RF CBTIs</i>	38
10. <i>Most and least consensual items within the least valid MMPI-2-RF CBTIs</i>	39
11. <i>Most and least consensual items within the AAA CBTI.....</i>	39
12. <i>Most and least consensual items within the MR CBTI</i>	40
13. <i>Most and least consensual items within the PAR CBTI.....</i>	40
14. <i>Most and least consensual items within the PS CBTI</i>	41
15. <i>and least consensual items for Clinical Scale WNL MMPI-2 Profiles.....</i>	42
16. <i>Most and least consensual items for MMPI-2 profiles with at least one Clinical Scale elevation</i>	43
17. <i>Most and least consensual items for RC Scale WNL MMPI-2-RF Profiles</i>	44
18. <i>Most and least consensual items for MMPI-2-RF profiles with at least one RC Scale elevation</i>	45
19. <i>Incremental validity indexed by stepwise regression</i>	46

TABLE	PAGE
20. <i>Incremental validity of the MMPI-2-RF Interpretive Report over the block of MMPI-2 Reports</i>	47
21. <i>Incremental validity of the MMPI-2 CBTI block over the MMPI-2-RF Interpretive Report</i>	48

CHAPTER I

INTRODUCTION

Relative Validity of the MMPI-2-RF Interpretive Report and MMPI-2 Computer-Based Test Interpretation Systems

Computer-based test interpretation (CBTI) systems for the MMPI have been commercially available since 1965, when the first system was released by the Roche Psychiatric Service Institute (Fowler, 1987). Since that time, both availability and usage of CBTI systems have risen dramatically, with at least 14 different CBTI programs becoming commercially available for the MMPI and MMPI-2 (Eyde, Kowal, & Fishburne, 1991). In 2004, Williams & Weed were able to identify eight commercially available CBTI programs for the MMPI-2, and 2008 saw the release of the Interpretive Report, the first CBTI system for the MMPI-2-RF (Ben-Porath & Tellegen, 2008a).

Clinicians frequently integrate CBTI reports with other measures, clinical interviews, client history, and personal test interpretations in order to increase the accuracy of their inferences (Graham, 2006). Unfortunately, however, there is little research investigating the validity of the various CBTI systems available for the MMPI-2 and MMPI-2-RF instruments. In fact, the MMPI-2-RF Interpretive Report has yet to be subjected to formal empirical study in this respect. Although the extant literature has often provided evidence of validity for both the MMPI-2 (Butcher, Atlis, & Hahn, 2004) and MMPI-2-RF (Tellegen & Ben-Porath, 2008), these data do not necessarily translate to clinicians' inferences from CBTI reports. For example, clinicians may conceptualize clients differently from different CBTI systems based on report length, clarity, types of information provided, and diagnostic and/or conceptual content provided within each report (Williams & Weed, 2004a; Williams, 2002; Harrington, 1999). Given the

uncertain state of the validity of current CBTI systems for the MMPI instruments, it is crucial that the strengths and limitations of these measures become better understood if they are to be used confidently within the psychological community.

The following section describes the most significant differences between the MMPI-2 and MMPI-2-RF instruments, types of validity, interpretive strategies for the MMPI measures, background information about CBTIs, the current status of MMPI CBTI systems, relevant MMPI validity and reliability research, and the proposed study.

Differences from the MMPI-2 to the MMPI-2-RF

The MMPI has received substantial additions and revisions since its original publication in 1940 (Hathaway & McKinley, 1940), including the restandardization and development of the MMPI-2 in 1989 (Butcher, Dahlstrom, Graham, Tellegen, & Kaemmer, 1989). However, some authors have argued that the MMPI-2 did not correct several of the flaws in the measure that stem from the test construction methods used during the initial development of the MMPI, namely, the presence of subtle items, high interscale correlations, and K-correction (Tellegen, Ben-Porath, McNulty, Arbisi, Graham, & Kaemmer, 2003).

Consistent with empirical keying methodology, the original MMPI scales were created based on statistical relationships between item endorsement trends and predetermined diagnoses. For example, if examinees who had been diagnosed with schizophrenia in the standardization sample tended to endorse a particular item as true or false, this item was included in Scale 8, regardless of the presence or absence of a theoretical link between the item content and the diagnosis. This atheoretical methodology resulted in the inclusion of what came to be known as “subtle items,” meaning that they were not easily detected by test-takers as belonging to a specific scale. Whereas this result was initially interpreted as desirable because it could

potentially control for a tendency to present oneself in an overly favorable or negative light, it was soon recognized that the inclusion of subtle items was likely representative of sampling error rather than a genuine relationship with the characteristics clinicians were attempting to measure (Weed, Ben-Porath, & Butcher, 1990; Butcher, Graham, & Ben-Porath, 1995).

Empirical keying methodology allows for item overlap for each scale between scales as long as a statistical relationship is found between a criterion group and an item. The application of this methodology allowed the MMPI-2 to contain substantial scale intercorrelations (Tellegen et al., 2003). Because the endorsement of single items contributes to elevations of multiple scales, the discriminant validity of the MMPI-2 is limited (Tellegen et al., 2003). It is worth noting, however, that the problem of high interscale correlations has also afflicted other prominent personality measures that have been developed through other methods as well, such as the Personality Assessment Inventory (PAI; Morey, 1996).

Finally, K-correction, applied to MMPI-2 scales as a correction for test defensiveness, has come to be viewed as a major weakness of the instrument. Despite the theoretical advantage of utilizing this strategy, most research has found that the application of K-correction to scores does not improve and may even reduce test validity (Sellbom, Ben-Porath, McNulty, Arbisi, & Graham, 2006). Standard MMPI-2 reports include scale scores both with and without K-correction, but this presents an interpretive risk in that clinicians who are not familiar with the literature disputing the validity of the correction may continue to rely on this technique despite empirical evidence to the contrary.

In an attempt to correct for the limitations of the MMPI-2, the authors of the MMPI-2-RF incorporated modern statistical analyses and theoretical advancements to derive new scales that are more unidimensional, psychometrically valid, and theoretically sound (Tellegen et al., 2003).

This endeavor was initially limited to the creation of the Restructured Clinical (RC) scales. The first step in this process involved the extraction of a common factor, labeled demoralization, which was a major component of the aforementioned high scale intercorrelations (Sellbom et al., 2006). Demoralization has been described as being pervasive across scales despite its lack of relatedness to the core constructs of any of the Clinical scales (Tellegen et al., 2003). Rather, this factor comprises content related to general maladjustment, emotional distress, and common unhappiness (Graham, 2006).

Demoralization items were conceptualized as mapping onto Watson's and Tellegen's (1985) model of Pleasantness-versus-Unpleasantness, which was later conceptualized as the third level of a three-level hierarchy including Positive Activation and Negative Activation at the second level and primary affects at the first (Tellegen, Watson, & Clark, 1999). The second level of this hierarchy is thought to be an indicator of anxiety at the Negative Activation end and reversed Positive Activation is an indicator of depression. Therefore, the authors proposed that using factor analysis to extract items loading on the first factor from MMPI-2 Clinical scales 2 and 7 (described as most resembling depression and anxiety, respectively) would result in the identification, to the largest extent possible, of demoralization-laden items. Although the presence of demoralization within all clinical scales harms discriminant validity, measurement of demoralization maintains clinical value as a stand-alone factor. This led to the authors' decision to identify other markers of maladjustment and malaise by correlating the demoralization items identified from Clinical scales 2 and 7 with all other MMPI-2 items to identify other markers of demoralization that could be used in a Demoralization scale (Graham, 2006).

Next, factor analyses of the Clinical scales, including the Demoralization items, were undertaken to assist with the identification of the core components of the scales. Correlational

analyses were next conducted to increase the distinctiveness of these components and to determine the items that were maximally correlated with the core components and minimally correlated with the other components, thus increasing convergent and discriminant validity of the resultant RC scales (Sellbom, Ben-Porath, McNulty, Arbisi, & Graham, 2006).

In addition to the creation of the RC scales, the development of the MMPI-2-RF included a number of other substantial changes. Every scale from the MMPI-2 has either been replaced or revised by scales that were developed with the same scrutiny used during the development of the RC scales. The familiar Content scales and Supplementary scales have been replaced by Higher-Order scales, Specific Problems scales, and Interest scales. Scales that have been retained but updated include the Personality Psychopathology Five (PSY-5) scales and the majority of the Validity scales. Additionally, new validity indicators, such as the Infrequent Somatic Responses Scale, have been added. Moreover, the number of test items has been reduced by about 40% from 567 items on the MMPI-2 to 338 items on the MMPI-2-RF (Tellegen & Ben-Porath, 2008). Given the drastic differences between the MMPI-2 and the MMPI-2-RF, authors have raised concerns regarding the advisability of using the MMPI-2-RF for several reasons.

Controversy Regarding the MMPI-2-RF

Some authors have argued that the structure of the MMPI-2 Clinical scales allows for a more accurate representation of the “syndromal complexity” of individuals’ psychological functioning, in that psychological disorder represents dysfunction across domains not exclusive to specific disorders (e.g., Nichols, 2006). For example, anxiety disorders and schizophrenia, which have been associated with Clinical scales 2 and 7, respectively, each often include impaired concentration and agitation as symptoms. Therefore, it has been argued that these symptoms should be reflected across scales that purport, at their core, to represent these

diagnoses in a fashion much like the criteria lists in the *Diagnostic and Statistical Manual of Mental Disorders, Fourth Edition* (DSM-IV; American Psychiatric Association, 1994) (Nichols, 2006). On the other hand, it has been noted that, in addition to allowing for poor psychometric properties and limiting inferential clarity of scales, this interpretation of scale heterogeneity ignores the functioning of the MMPI-2 as an indicator of symptoms and observable qualities rather than a diagnostic instrument (Weed, 2006).

Nichols (2006) also rejected the use of the RC scales due to “construct drift.” This is the idea that restructuring scales pulled their content away from the core clinical features they were originally thought to represent, thereby disallowing the historical interpretations of the meaning of scales elevations. As an alternative approach, Nichols (2006) recommended use of adjustment formulae that allow for the retention of critical items while compensating for their drawbacks. Others (e.g., Weed, 2006) have astutely observed, however, that drift does not necessarily represent attenuated validity. Rather, drift may be indicative of reduced error variance and “noise” within scales and constitute a purer indicator of symptoms (Weed, 2006). Allowing poorly performing items to remain within an item pool and then applying techniques to correct scores for error is essentially asking test-takers to spend additional time completing measures that provides clear benefit to neither the client nor the clinician.

Some argue that the RC scales are redundant with pre-existing MMPI-2 scales. For example, Butcher and colleagues (2006) noted the cynical attitudes represented by an elevated RC3 score are well-represented within the MMPI-2 scales, such as the Cynicism Content scale. Additionally, Rouse and colleagues (2008) documented RC1 as correlating $r = .90$ on average with the Health Concerns Content Scale from the MMPI-2. Proponents of the RC scales have countered these findings with the explanation that the high correlations are inflated due to item

overlap between scales (Tellegen, Ben-Porath, Sellbom, Arbisi, McNulty, & Graham, 2006). Additionally, they claim to be able to measure in the nine RC scales what Nichols (2006) was only able to measure through comparison of these nine scales to 51 other scales, some of which have limited empirical support (Tellegen et al., 2006). Finally, Tellegen and colleagues (2006) found validity coefficients to be consistently higher in a college counseling setting than the scales proposed by Nichols (2006) (Sellbom, Ben-Porath, & Graham, 2006). Regardless of whether the MMPI-2-RF at present represents improved or degraded measurement in comparison to the MMPI-2, new conceptualizations and interpretations of scales based on continued empirical research will certainly be essential to accurate personality assessment practices.

Computer-Based Test Interpretations

History

Personality assessment measures have been the primary area of focus for CBTI systems since their inception, largely due to their structured formats and empirically-driven interpretations (Butcher, Perry, & Dean, 2009). As the most widely used and investigated personality measures, the MMPI and the MMPI-2 have consequently been the most popular instruments for which CBTI programs have been developed (Butcher et al., 2009).

Using the MMPI, the first CBTI system was created by the Mayo Clinic in the early 1960s (Butcher, 1987). The programming for this system included 110 output statements based on empirically-derived single-scale correlates that were provided to clinicians in a statement-by-statement, as opposed to integrated, manner (Rome, Swenson, Mataya, McCarthy, Pearson, Keating, & Hathaway, 1962; Pearson, Rome, Swenson, Matay, & Brannick, 1965). With the progression of time, the programming and output of CBTI systems have become increasingly

complex (Butcher et al., 2009). By the 1970s, CBTI systems had progressed alongside MMPI research and interpretive “cookbooks” to include interpretive statements that were based on code type literature rather than single-scale correlates (Atlis, Hahn, & Butcher, 2006).

The popularity of CBTI systems steadily increased, resulting in at least 14 commercially available programs for the MMPI alone since their introduction (Eyde et al., 1991). This trend was not completely welcome, however. As more systems emerged, concerns were raised regarding the validity and reliability of the various systems, largely fueled by the lack of transparency in the qualifications of system developers and the algorithms on which they based their programming (Butcher, 2006). Consequently, the American Psychological Association released guidelines in 1986 for publishers of CBTI systems that stated that the publishers should provide information about validity studies that describe the accuracy of their system’s interpretive statements, warnings about potential interpretation errors, and an explanation of the origin of the interpretive statements (Conoley, Plake, & Kemmerer, 1991). Although adherence to these standards has been questionable (Williams & Weed, 2004a), several researchers have independently examined the performance of CBTI systems (e.g., Williams & Weed, 2004a; Deskovitz et al., 2006; Harrington, 1999; Eyde et al., 1991). Before deciding whether to use CBTIs, clinicians should first examine the potential benefits and limitations and how they will integrate CBTI interpretations into their client conceptualizations.

Advantages of CBTIs

There are several aspects of CBTIs that enable them potentially to improve the accuracy and consistency of clinicians’ interpretations of MMPI protocols. First, CBTI programs operate with perfect reliability in that a configuration of item responses should result in the same CBTI output statements every time (Butcher, 2009). Subjectivity is unfortunately present in the

programming process, since CBTIs are based on a combination of empirical literature, clinical lore, and clinical expertise rather than purely statistical prediction. However, clinicians can use CBTI programs with the knowledge that programmed algorithms will operate consistently (Williams, 2002).

Another advantage of CBTIs is that important correlates and relationships between patterns are not overlooked due to insufficient human limitations. Whereas clinician interpretations are limited to the knowledge and resources that each individual has accumulated over time, CBTIs can be programmed based on decades of empirical data, interpretive strategies, and multiple expert opinions (Butcher, 2009). Because the CBTI programs are capable of quickly accessing all of this information, the speed of interpretation can also be quickened (Williams, 2002).

CBTIs are also helpful in limiting the effects of human error. For example, clinicians may develop habits of emphasizing particular aspects of their interpretations at the expense of potentially relevant information. In addition, clinicians may simply fail to consider particular pieces of information. Furthermore, clinicians are subject to biases in their inferences, such as halo effects (Butcher, 2009). Although some clinicians may believe their level of expertise places their interpretations above those offered by statistical rules, level of expertise has not been found to be related to performance in comparison to prediction rules (Grove, Zald, Lebow, Snitz, & Nelson, 2000). In fact, experts are often no more accurate than less experienced clinicians (Garb, 1989). Grove and colleagues (2000) conducted a meta-analysis concluding that statistical prediction is approximately 10% more accurate than clinical prediction, on average. Because CBTI reports typically can access more information than clinicians, and at a faster rate, reports are often more thorough than clinician-produced reports (Butcher, 2009).

CBTI Limitations and Concerns

Before and while using CBTI systems, it is important for clinicians to remain cognizant of the limitations and pitfalls of CBTIs. This enables clinicians to have realistic expectations beforehand and can increase the accuracy of the inferences at which they arrive. For example, one danger in using CBTIs is that clinicians can become passive and assume the information they receive is accurate. This problem is worsened by overconfidence stemming from the technological allure of CBTI systems (Butcher, 2009).

One reason that clinicians ought to limit passive reliance on CBTI profiles is that there is limited knowledge of the programming that goes into each CBTI due to intellectual property concerns. It is known that CBTIs are each programmed based on different sets of rules, but the degree to which these rules rely on actuarial prediction is not clear. As described above, clinical prediction places interpretations at greater risk for error (Grove et al., 2000). Therefore, the more that programming relies on clinical lore and expertise, the more questionable the output may be. Although complete actuarial prediction is generally an agreed upon goal, it is far from realized (e.g., Garb, 2000; Butcher, 2009).

Clinicians must also be wary of the Barnum effect (when statements appear to represent an individual but in fact are so vague they could be applied to most anyone) when consulting with CBTI profiles. Even though profiles might seem to be accurate and helpful, clinicians would be wasting the time and effort of themselves and their clients if they were to rely on CBTI systems that provide these types of statements.

Though some research has investigated the accuracy of CBTI interpretations in isolation (e.g., Williams & Weed, 2004a; Harrington, 1999), they are considered to be part of an assessment battery rather than providing a stand-alone assessment. It would be irresponsible for

a clinician to interpret CBTI profiles in the absence of other sources of information (Graham, 2006), such as interviews, observations, and other measures (McMinn, Ellens, & Soref, 1999). This is because interpretations produced by CBTI systems are prototypical descriptions based on a set of scores, configurations, and profile types, therefore incapable of considering unique client factors. This fact is often restated in CBTI reports in warnings that advise clinicians to regard descriptions only as working hypotheses (Williams & Weed, 2004b).

Current Status of CBTIs

The frequency with which CBTIs have been incorporated into assessments has been rising over the course of the past several decades (Butcher, Perry, & Atlis, 2000), a trend that is expected to continue with increasing advancements in and applications of technology (Garb, 2000). It is essential that CBTIs regularly receive updates if they are to keep pace with advancements in personality and psychopathology theory, assessment instruments, and technology. Previous studies have identified at least eight commercially available CBTI systems for the MMPI-2 in the United States alone (e.g., Williams & Weed, 2004a). These reports vary in a number of key areas, such as the scales used for interpretation, the degree to which interpretive statements are programmed based on clinical expertise or actuarial formulae, and whether the reports are structured as a list of descriptive statements or in an integrated manner (Williams & Weed, 2004b). For the purpose of the present study, the current standing of four of the most popular CBTI programs for the MMPI-2 and one program for the MMPI-2-RF, the Interpretive Report, will be reviewed below.

Automated Assessment Associates (AAA). The AAA Adult Interpretive Report – Version 10 incorporates MMPI-2 validity, clinical, content, and MAC-R, APS, and AAS substance abuse scales into an integrated report. Programming is described as being determined by MMPI and

MMPI-2 research and clinical literature and 70 years of clinical and research experience between the system's authors, Donald Strassberg, Ph.D. and Leslie Cooper, Ph.D. The report contains code type and single-scale interpretations, therapeutic implications, and diagnostic considerations. The AAA costs \$525 with unlimited usage on the first computer and \$250 for installation on each additional computer (Automated Assessment Associates, 2008).

Minnesota Report (MR). The Minnesota Report, authored by prominent MMPI-2 authority James N. Butcher, is available in three different versions based on setting, clinical, personnel, or forensic. This system was last revised in 2001. Important features of the MR include narrative sections that provide setting-specific descriptions, interpretive statements that are relevant to mental health settings, and item response percentages to provide clinicians with accurate perceptions of the frequency of particular statements in each setting. The following narrative sections are included in each report: profile validity, symptomatic patterns, profile frequency, profile stability, interpersonal relations, diagnostic considerations, and treatment considerations. MR reports are sold individually with the price based on quantity, ranging from \$44.50 each for less than 5 to \$34.50 each if 2,500 or more reports are ordered (Butcher, 2011).

Psychological Assessment Resources (PAR). The MMPI-2 Adult Interpretive System Version 4 is authored by a well-known MMPI-2 researcher and clinician, Roger Greene (Williams & Weed, 2004a). This system provides clinicians with the option of receiving one of three types of reports, a Codetype Report, a Content-Based Report, or a Codetype/Content-Based Report that combines the features of the first two types. Additionally, the PAR system has recently been updated to allow for interpretation of MMPI-2-RF profiles (not examined in the current study). The Codetype Report is described as an “empirically derived prototypic profile based on ‘goodness-of-fit’” with the client’s MMPI-2 scale scores. (Greene, Brown, & PAR

Staff, 2011). Interpretive statements are based on the client's best match with one of 381 prototypic profiles. In addition, the Codetype Report includes interpretive statements based on individual scale scores, including Validity scales, Clinical scales, Harris and Lingo's scales, RC scales, Content and Content Component scales, Supplementary scales, and PSY-5 scales. The MMPI-2-RF Report includes scale-by-scale interpretive statements for each of the available scales. Finally, PAR reports also include descriptions of treatment prognosis and diagnostic considerations. The PAR system costs \$725 for unlimited uses. Upgrading to the newest system costs \$375 (Green et al., 2011).

PsychScreen (PS). The PsychScreen MMPI-2 Interpreter (1998) is a portion of a larger suite of psychological measures that are integrated into one complete report. The interpretive statements provided for the MMPI-2 are based on the Clinical, Content, Harris-Lingo's, Supplementary, and Subtle-Obvious scales (Williams & Weed, 2004a). The PS CBTI report sections include intellectual functioning, validity of test results, emotional functioning, alcohol and drug use, somatic functioning, interpersonal functioning, self-image, defenses, personality dynamics, diagnostic considerations, and treatment recommendations (PsychScreen, Inc., n.d.). When the MMPI-2 is administered in conjunction with other measures, it is not clear from which test each interpretive statement is derived and if the MMPI-2 is incorporated into each report heading, however. The PS software costs a one-time fee of \$250 for unlimited uses of the MMPI-2, excluding all other measures within the suite (PsychScreen, Inc., n.d.).

Interpretive Report (RF). The MMPI-2-RF Interpretive Report provides interpretive statements derived from each of the MMPI-2-RF scales. The report starts with a brief synopsis, which is followed by a description of protocol validity. The next heading, substantive scale interpretation, provides subheadings that provide interpretive descriptions from integrated scale

scores within key domains. These include: somatic/cognitive dysfunction, emotional dysfunction, thought dysfunction, behavioral dysfunction, interpersonal functioning scales, and interest scales. The report then describes diagnostic considerations and treatment considerations. The RF concludes with a list of critical responses endorsed by the examinee. RF CBTI reports, like the MR CBTIs, are purchased in a per-report manner. Purchased in quantities of less than five, each report costs \$41.50. The least expensive unit price is \$31.50, which occurs when 2,500 or more CBTI reports are purchased at a time (Pearson, 2011).

Evaluating CBTIs

Q-sort Methodology

Q-sort methodology, or the “q-technique,” is a way to operationalize the inferences that are typically considered by qualified professionals when formulating cases. The q-technique also provides a common language across professionals, enabling effective and efficient communication between these professionals. Often times, the true source of disagreement between clinicians is difficult to determine because of differences in vernacular and phraseology; however, applying q-sort methodology bridges this gap by forcing clinicians to use identical, easily understood descriptors. From there, points of agreement and disagreement can easily be identified and investigated further (Block, 1978). In this way, q-sort methodology is well-suited for psychological assessment research applications (Weed, 2006). Ratings on items from two q-sorts can be intercorrelated as an index of agreement, referred to as a q-correlation. Q-correlations that are derived between a q-sort based on a CBTI and a q-sort based on some external criterion (e.g., therapist ratings) can be taken as an index of validity.

The q-sorting procedure entails organizing a set of statements into a designated number of categories with fixed frequencies determined by the degree to which they describe an individual. Items sorted at one end of the continuum represent the characteristics that are least representative of an individual, whereas those at the opposite end are the most salient characteristics of an individual (Block, 1978). One particular q-set, the Midwestern Q-Sort (Williams & Weed, 2003) is comprised of 100 common inferences from the MMPI-2. Within this set, the authors attempted to include only observable behaviors in non-technical language to limit disagreement that may arise from varying interpretations of the meaning of statements rather than the intended content of the statements. In addition, the Midwestern Q-Sort contains a balance of positively and negatively keyed items to control for desirability (Weed, 2006).

To complete the Midwestern Q-Sort, raters sort each of the 100 inferences into one of seven banks, which are arranged in a symmetrical, forced-choice manner that permits fewer extreme descriptors (Weed, 2006), a distribution thought to best represent the real-world structure of personality and psychopathology. The seven categories, arranged in order with 5, 10, 20, 30, 20, 10, and 5 descriptors in each, are quantified from -3 (least descriptive) to +3 (most descriptive) of an individual. After completing the q-set, a rater's sort can then be compared to that of another rater via product-moment correlation (Weed, 2006).

CBTI Reliability Studies

Deskovitz, Weed, and Williams (2005) examined the interpretive reliability of six CBTI programs for the MMPI-2: Automated Assessment Associates (AAA), the Minnesota Report (MR), Psychological Assessment Resources (PAR), Psychometric Software, Inc (PSI), PsychScreen, Inc. (PS), and Western Psychological Services (WPS). Interpretive reliability refers to the consensuality of inferences that are made about an examinee based on the entirety of

a measure rather than examining reliability on a scale-by-scale basis. This approach was selected because it is consistent with how the MMPI-2 is interpreted in clinical settings. The authors calculated both inter-interpretor and inter-program reliability, with the former defined as the level of agreement across raters within individual CBTI profiles and the latter defined as the agreement between the CBTI programs. These authors selected profiles based on the 20 MMPI most frequently occurring code types in inpatient and outpatient settings (Archer, Griffin, & Aiduk, 1995; Graham, 1999; Greene, 2000; Lewandowski & Graham, 1972). These were identified using archival data, and they were then used to produce CBTI interpretive reports that were interpreted by four graduate student raters. With 20 profiles for each CBTI system, a total of 120 reports were interpreted by each rater. The raters used the Midwestern Q-sort to operationalize the interpretations of the reports.

Within each CBTI report, inter-interpretor reliability was calculated as the mean pairwise intercorrelation between the four sorters' ratings for each MMPI-2 profile. Inter-interpretor reliability ranged from .56 to .62 with an average pairwise reliability coefficient across programs of .60. There were no significant differences between programs. At the code type level, the only average pairwise intercorrelations that were significantly less than the overall mean were PS 1/3 code type, PAR 7/8, and Within Normal Limits (WNL) for AAA.

Inter-program reliability was indexed as the average correlation between the mean sort of the four raters of a single profile and the rater means on that profile for the other five CBTIs. The intercorrelations between CBTIs for individual code types ranged from .22 to .93 with a grand mean of .83. There were no significant differences between programs aggregated across all code types. When examining CBTIs at the profile level, several reports were less consensual than the grand mean. PAR 7/8, AAA 2/4, and PAR 2/8 had the lowest correlations with other

programs interpreting the same code type. The three highest inter-program reliabilities were associated with the 4/9 code type interpretations by AAA, PA, and PAR.

McLaughlan and colleagues (2011) used the indices of interpretive reliability from Deskovitz et al.'s (2005) study to examine the convergence of these systems with the MMPI-2-RF Interpretive Report. Inter-interpreter reliabilities for the MMPI-2-RF CBTI were found to be similar or superior to those of the MMPI-2 with the exception of the 1/3 profile. Aggregating across MMPI-2 CBTIs, the MMPI-2-RF Interpretive Report ($\alpha = .85$) displayed slightly better inter-interpreter reliability than the MMPI-2, overall ($\alpha = .80$).

The MMPI-2-RF report was found to possess poorer inter-program reliability than the MMPI-2 reports. This is likely attributable to the differences in the measures from which the CBTIs were programmed and definition of the profiles that were selected. First, because the analyses compared the agreement of six MMPI-2 CBTIs to one MMPI-2-RF report, it should have been expected that more agreement would be discovered between the reports that were based on the same measure. Second, the 20 archival profiles that were used in the study were selected based on MMPI-2 profile definition. When conducting analyses, McLaughlan et al. (2011) recognized that many of the scale elevations shifted from a 2- or 3-point code type to a Within Normal Limits profile. Therefore, the resulting CBTI reports provided dramatically different pictures of the profiles. Because of these findings, it is unclear whether the observed discrepancy between programs is due to the MMPI-2-RF Interpretive Report possessing superior or inferior validity to the MMPI-2.

CBTI Validity Studies

The examination of CBTI program validity is unique in that the validity is limited by the accuracy of the testing instrument on which the program is based. In this way, there are three

variables that can influence CBTI validity: the validity of the parent instrument, the CBTI programming, and how the reports are interpreted by clinicians. If CBTI statements are programmed based on inferences at which one would be expected to arrive based on the proper interpretation of scale patterns, it cannot necessarily be expected that the accuracy of the CBTI program will be similar to the accuracy of the parent instrument. If the presentation of information in a CBTI report is unclear or poorly organized, the validity should be expected to be worse than that of the parent instrument regardless of the accuracy of programming.

Previously, CBTI validity studies have consisted of consumer satisfaction surveys, comparison of clinician to CBTI ratings, and external criteria studies. Each of these methods has strengths and limitations, and the most complete picture of CBTI validity is likely captured by considering all types.

Early validity studies utilizing consumer satisfaction surveys revealed that clinicians were generally pleased with the accuracy, quality, and usefulness of CBTI reports (e.g., Eyde et al., 1991, Green, 1982). However, these studies failed to control for the Barnum effect, in which vague statements of little value are considered accurate due to their broad range of potential interpretations (Graham, 2006). As a result, Williams and Weed (2004a) conducted a CBTI consumer satisfaction survey using novel methodology to overcome this limitation explicitly. Clinicians submitted MMPI-2 answer sheets that were then randomly assigned to one of eight commercially available CBTIs. A CBTI report was generated, and this was returned to the clinicians. However, clinicians sometimes received CBTI reports that were based on modal MMPI-2 reports for the setting from which the answer sheet was submitted. Because little incremental insight into clinical concerns can be gathered from modal profiles, it was thought that high ratings of these profiles reflect the Barnum effect.

Williams and Weed (2004a) asked that clinicians rate CBTI reports based on 10 criteria: conciseness, confirmation of therapist's impressions of client, usefulness for diagnosis and/or treatment, accuracy, provision of new information, presence of contradictory information, omission of important information, organization and clarity, presence of useless information, and appropriateness of diagnostic considerations. Among the eight CBTI programs, Automated Assessment Associates received the highest overall rating and was rated highest in four dimensions (confirmed opinion, clinically useful, accuracy, and appropriate diagnosis), the Caldwell Report was rated highest in two dimensions (inclusion of new and important information, not omitting important information), the Minnesota Report was rated highest in lack of contradictory information and organization and clarity, and Western Psychological Services was rated highest in conciseness and lack of useless information. BehaviorData was rated lowest in nine of the 10 dimensions and was the only system with a negative overall rating. Most of the modal reports were rated poorly by clinicians, suggesting the impact of Barnum statements on high clinician opinions for authentic reports is low.

Epstein and Rotunda (2000) examined the validity of CBTI measures by comparing symptom ratings based on either computer-generated or clinician-generated reports to symptom ratings that were completed by staff from the hospital at which the clients had been seen. Symptom ratings based on the reports were completed by independent psychologists. In addition to reports, these clinicians were provided with basic MMPI-2 scores and demographic information. The results of this study indicated that those who received clinician-generated reports were far more accurate in their symptom ratings than those who received computer-generated reports. However, interpretation of these results is complicated because clinician-generated reports were written with unlimited knowledge of assessment information, such as

biographical information and other assessment measures whereas the computer-generated reports were based solely on MMPI-2 scores. Additionally, it is not clear how much the raters relied on the information available within the computer report as opposed to interpreting the MMPI-2 scores on their own.

McNeal (1997) conducted a CBTI validity study utilizing an external criteria design. Using q-sort methodology, he compared the interpretations of CBTIs (Minnesota Report), MMPI-2 profiles, and therapist ratings. The Minnesota Report and interpretation of the MMPI-2 profile in isolation were found to possess similar validity. Additionally, clinical interpretation of the MMPI-2 profiles did not provide incremental validity over the Minnesota Report. This study was limited by sample size, however ($N = 18$).

Harrington (1999), using a similar methodology, gathered a sample of 26 individuals referred from a university counseling center to compare six MMPI-2 CBTI programs. For this study, external criteria included personality and psychopathology ratings, operationalized using q-sort methodology, collected from the participant's therapist and a significant other, such as a close friend or family member. CBTI ratings were collected from two additional raters. Whereas inter-rater reliability was generally found to be good, low validity coefficients were obtained. This was attributed to two potential factors. First, participants with little psychopathology may have self-selected for the study, thereby creating uncertainty in how to rate various aspects of functioning. Second, the inclusion of an equal number of positive and negative descriptors in the q-set may have contributed to therapists and significant others rating the participants in an overly positive manner.

Overall, the above studies provide mixed results regarding CBTI validity. Therapists have generally expressed satisfaction with their experiences with CBTI reports. Particular CBTI

programs (e.g., AAA, Caldwell Report, Minnesota Report) tended to be rated higher than others (e.g., BehaviorData). Additionally, inferences based on CBTI reports in isolation (i.e., without knowledge of the client aside from the report) have shown limited validity. These results may have been hindered by a number of limitations, such as sample size and characteristics, however.

Present Study

Although previous studies (e.g., Williams & Weed, 2004a; Harrington, 1999) have examined the validity of MMPI-2 CBTIs, these programs have all received revisions since that time. Moreover, the validity of the MMPI-2-RF Interpretive Report has not been examined to date. McLaughlan et al. (2011) examined the interpretive reliability of the MMPI-2-RF Interpretive Report in comparison to eight MMPI-2 CBTI systems and found the Interpretive Report to possess greater inter-rater reliability on average than the MMPI-2 CBTIs. Across programs, the results of this study indicated that agreement between CBTI systems was better for the MMPI-2 CBTIs than for the MMPI-2-RF Interpretive Report. However, it is unclear at this time whether the observed differences stem from the clinician interpretations from the MMPI-2-RF Interpretive Report being either more valid or less valid than those of the MMPI-2 CBTIs. By comparing q-sort ratings based on CBTI outputs to therapist ratings, the present study seeks to provide evidence to assist in answering this question. It was expected that validity coefficients would generally be small. This is consistent with previous findings obtained by Harrington (1999), who used a similar methodology as the current study. Consequently, it was important to examine the relative validity, as opposed to absolute validity, of the CBTI systems.

The MMPI-2-RF Interpretive Report was expected to display superior validity over the MMPI-2 CBTIs. The revisions from the MMPI-2 to the MMPI-2-RF (e.g., reduced scale heterogeneity, removal of subtle items, application of modern scale construction principles and

theoretical considerations) allow for reduced ambiguity in scale elevation interpretations, thereby allowing for more accurate interpretive statements within the MMPI-2-RF Interpretive Report. Furthermore, the differences between the MMPI-2 and MMPI-2-RF detailed above are also expected to result in MMPI-2-RF CBTI reports that offer fewer conflicting statements than those for the MMPI-2 CBTI systems.. It was therefore expected that the statements provided by the MMPI-2-RF Interpretive Report would augment those of the MMPI-2 CBTIs, thereby contributing to incremental validity of the MMPI-2-RF Interpretive Report over the four MMPI-2 CBTIs. In other words, it was anticipated that the MMPI-2-RF Interpretive Report would provide statements that are different from the statements produced by the MMPI-2 CBTIs and are interpreted with greater validity.

CHAPTER II

METHOD

MMPI-2 Profiles

The MMPI-2 profiles in this study were collected as part of an ongoing project from clients who were undergoing treatment with a therapist in an outpatient setting in a small Midwestern city. Therapists completed Midwestern Q-Sorts based on these clients after they had seen them for four to six sessions. In total, 65 therapist q-sorts were collected. Out of these profiles, 30 were randomly selected based on matching validity status (i.e., all CBTI programs included statements noting the profile should be interpreted as valid).

Raters

Raters included four graduate student and practicing clinicians with advanced training in personality assessment and psychopathology. Each clinician completed one Midwestern Q-sort for 30 client profiles across five different CBTI programs, totaling 150 q-sorts per rater. Raters were compensated with \$200 each for their assistance.

Instruments

Midwestern Q-Sort

The Midwestern Q-sort was used to operationalize CBTI interpretations. This is a web-based Q-sort instrument that contains 100 statements designed to reflect the most common MMPI-2 inferences. Statements are sorted into fixed frequency categories on a 7-point continuum with -3 as least descriptive and +3 as most descriptive of an individual. Ratings are

also forced into a symmetrical distribution. Taken together, these features control rater error by preventing extreme or unbalanced ratings (Weed, 2006).

The original Windows-based computer program on which the Midwestern Q-sort was modeled was developed by Noland and Weed (1994). It features a split screen with 7 numbered bins on the left side with the frequencies of 5, 10, 20, 30, 20, 10, and 5 for each category and a list of 100 interpretive statements on the right side of the screen. These statements are dragged and dropped into bins on the left hand side until all of the bins are full and there are no descriptors remaining on the right-hand side. The web version of the Midwestern Q-sort that was used for this study was designed by Williams and Weed in 2002 to facilitate remote data collection (Deskovitz et al., 2005).

CBTIs

MMPI-2. The Minnesota Multiphasic Personality Inventory – Second Edition (Butcher et al., 1989) is a 567-item true/false self-report inventory. The following four commercially available MMPI-2 CBTI programs were considered for the present study: Automated Assessment Associates [AAA] (Strassberg & Cooper, 1997); Pearson Assessments Systems: The Minnesota Report [MR] (Butcher, 1993); Psychological Assessment Resources [PAR] (Green, Brown, & Kovan, 1998); and PsychScreen, Inc. [PS] (Minnesota Multiphasic Personality Inventory-2 Interpreter, 1998).

MMPI-2-RF. The Minnesota Multiphasic Personality Inventory – Second Edition – Restructured Form (Ben-Porath & Tellegen, 2008), is a 338-item true/false self-report inventory. The MMPI-2-RF Interpretive Report was developed by (Ben-Porath & Tellegen, 2008).

Procedures

Previously, therapists completed Midwestern Q-Sorts based on their clinical impressions of clients after four to six sessions. These same clients completed MMPI-2 protocols during their first session, which were interpreted via the five CBTI programs described above. Each of the four raters completed 30 Midwestern Q-Sorts based on deidentified CBTI reports from each of the five aforementioned programs. For the MMPI-2-RF Interpretive Report, profiles were first converted from the MMPI-2 to the MMPI-2-RF format.

To avoid potential rater biases, reports were deidentified. This entailed removing tables and specific scores and editing any information that may have identified the CBTI program being evaluated. Reports were edited in a word processing program and sent to the four sorters via electronic mail in an order that was random but identical for each. After CBTI-based q-sorts were electronically submitted to the researcher, they were compared to therapist q-sorts in the following manner.

Analyses

The analyses aimed to answer two primary questions. First, how consensual are CBTI q-sorts with therapist q-sorts? Second, are the MMPI-2-RF CBTIs more valid or less valid in comparison to the MMPI-2 CBTIs? Additionally the degree to which the MMPI-2-RF CBTI reports augments the inferences garnered from MMPI-2 CBTI systems was examined. The areas in which the CBTI sorts are most and least consensual with therapist ratings were then examined to identify strengths and weaknesses within each CBTI program. This was evidenced through item-level analyses. Additionally, this study examined the degree to which the MMPI-2-RF CBTI reports augment the inferences garnered from each MMPI-2 CBTI system.

The amount of agreement between CBTI sorts (averaged across the four raters) and the external criterion (i.e., therapist ratings) index CBTI validity. These data were obtained through Pearson product-moment correlations for each of the 150 MMPI reports. Item-level analyses were performed to determine the specific ways in which CBTI sorts varied from each other and therapist sorts. These are indexed by average absolute mean differences of CBTI ratings between the five systems.

To determine the extent to which the MMPI-2-RF Interpretive Report tends to provide incremental improve inferences beyond that of the MMPI-2 systems and vice versa, multiple regressions were conducted. First, a stepwise multiple regression was calculated to examine the amount of interpretive variance accounted for by each CBTI program over and above one another, as indexed by the coefficient of determination. The MMPI-2-RF was expected to be the first predictor entered into the majority of the equations, accounting for the most variance. This prediction stemmed from the hypotheses that 1) the MMPI-2-RF Interpretive Report should be more valid because of the improved psychometric properties of its parent instrument in comparison to the MMPI-2 on which the other CBTIs are based (Tellegen & Ben-Porath, 2008), and 2) the greater inter-rater reliability of the MMPI-2-RF Interpretive Report (McLaughlan et al., 2011) reflects greater report clarity. Ease of interpretation was expected to assist the raters in arriving at more accurate conclusions.

For a more precise examination of the degree to which the MMPI-2-RF Interpretive Report clarifies clinician conceptualizations, four hierarchical regressions were conducted with the MMPI-2 CBTIs entered as a block as the first predictor and the MMPI-2-RF Interpretive Report entered as the second predictor. Using change in the coefficient of determination as an index, differences in conclusions reached by clinicians with the addition of the MMPI-2-RF

CBTI to other CBTIs were displayed. The order was then reversed to determine the extent to which the block of MMPI-2 CBTIs account for therapist ratings over and above the MMPI-2-RF Interpretive Report.

CHAPTER III

RESULTS

Relative validity was indexed as the agreement between the CBTI raters and the treating clinician within each profile. The first step in analyzing CBTI validity was collecting Midwestern Q-Sorts for each CBTI program from each of the four raters. These were entered into 150 (30 profiles x 5 CBTI programs) 4 x 100 (raters x q-set items) matrices. As illustrated in Table 1 below, which displays one of the 150 matrices, these ratings were then averaged across the raters to be compared to the therapist sort for each profile. This was accomplished by completing one data matrix for each profile (see Table 2). Finally, therapist ratings were intercorrelated with the profiles produced by each CBTI system (Table 3). Product-moment correlation mean and median validities and standard deviations are presented to allow for examination of the range and consistency of correlations at an individual profile level and the general trend for particular systems to produce reports that are interpreted with lesser and greater validity. Grand means were obtained by calculating the average correlation of all 30 profiles. To better demonstrate the differences found between the four MMPI-2 CBTIs and the MMPI-2-RF Interpretive Report, the MMPI-2 CBTI validity coefficients were aggregated, thereby allowing for a side-by-side MMPI-2 and MMPI-2-RF CBTI comparison.

Table 1. Ratings of the MMPI-2-RF CBTI for Profile 1

Item	Rater 1	Rater 2	Rater 3	Rater 4	Mean
1	-1	1	1	2	.75
2	1	0	1	-1	.25
3	0	0	-2	-1	-.75
4	0	1	1	0	.50
.....					
100	-1	1	-2	1	-.25

Table 2. Therapist ratings and average item ratings across CBTIs for Profile 1

Item	Therapist	AAA	MR	PAR	PS	RF
1	.00	.50	-1.25	.00	-.75	.75
2	1.00	.75	.75	-.25	-1.25	.25
3	2.00	1.00	1.50	1.75	.50	-.75
4	-3.00	-1.50	-.25	-2.25	-1.50	.50
.....						
100	3.00	.25	2.75	2.25	1.50	-.25

Note. AAA = Automated Assessment Associates, MR = Minnesota Report (Pearson Assessment Systems), PAR = Psychological Assessment Resources, PS = Psych Screen, Inc., RF = RF Interpretive Report (Pearson Assessment Systems).

Validity coefficients are displayed in Table 3. To enable a more thorough interpretation of the data, this table also includes grand mean correlations, median correlations, mean correlations for profiles based on whether one or more Clinical scales were elevated ($T \geq 65$), and standard deviations for each CBTI program. Correlations were generally low and often negative, but positive grand mean correlations were obtained across each program. Overall, the MMPI-2-RF Interpretive Report ($Gr M = .04$, $SD = .11$) exhibited slightly superior validity in comparison to the aggregate MMPI-2 CBTIs ($Gr M = .03$, $SD = .09$). However, the PS system displayed the best average validity based on individual CBTI programs ($Gr M = .07$, $SD = .12$).

In descending order, the next best performing individual MMPI-2 systems, based on mean correlations, included AAA ($r = .02$, $SD = .10$), MR ($r = .01$, $SD = .10$), and PS ($r = .01$, $SD = .10$). When the effects of extreme profiles were reduced by examining median correlations, the best performing programs, in descending order, were PS ($r = .07$), RF ($r = .01$), MR ($r = .01$), AAA ($r = .01$), and PAR ($r = -.01$).

Table 3. Intercorrelations between CBTI Systems and Therapist Ratings

Scale Elevations		Validity Coefficients					
Clinical	RC	AAA	MR	PAR	PS	M	RF
WNL 1=61T	Spike 1	-.02	-.03	-.01	-.06	-.03	-.08
WNL 1/3=64T	WNL 1/8=59T	-.16	-.07	-.04	.10	-.04	-.06
WNL 2=52T	WNL 2=58T	.07	.05	-.07	.00	.01	.03
WNL 3/6=57T	WNL 4=59T	-.08	-.19	-.03	-.25	-.14	-.12
WNL 3=61T	WNL 6=56T	.02	-.02	.06	.05	.03	-.00
WNL 4=64T	WNL 2=64T	-.14	-.06	-.04	-.03	-.07	-.12
WNL 6=52T	WNL 1/8=47T	-.02	-.03	.02	.03	-.00	-.01
WNL 6=56T	WNL 6=56T	-.05	-.01	-.15	-.01	-.05	.17
WNL 6=64T	6/1	-.01	-.01	-.00	.13	.01	.09
WNL 7/9=53T	WNL 4=54T	.06	.02	-.07	.06	.02	.04
WNL 8/9=51T	WNL 6/9=56T	-.03	.03	-.09	.08	-.00	.02
WNL 8=53T	WNL 6=56T	-.02	-.06	-.05	-.05	-.05	-.05
WNL 8=62T	3/8/d/6	.16	.00	.07	.07	.08	-.01
WNL 9=59T	WNL 3=54T	.02	.02	-.01	.06	.02	.04
Spike 1	1/8	.21	.30	.12	.24	.22	.23
Spike 2	WNL 7=62T	-.10	-.19	-.19	-.01	-.15	-.17
Spike 3	WNL 7=62T	.01	-.04	.05	.08	.05	.21
Spike 4	WNL 3=61T	-.01	-.06	-.06	-.01	-.04	-.04
Spike 6	WNL 6=61T	-.10	-.01	-.05	.09	-.02	-.01
3/4	WNL 2=58T	.04	.04	-.01	.13	.05	-.01
4/2	2/4	-.07	.03	-.11	-.01	-.04	.00
4/3	WNL d=53T	-.02	-.04	-.10	-.01	-.04	-.01
4/1/3	6/4	.13	.15	.18	.26	.18	.24
4/3/2	WNL 1=59T	.27	.17	.14	.26	.21	.18
8/6/7	d/4/8/2/7	.21	.21	.22	.25	.22	.21
4/6/8/7/1	6/d/4	-.05	.02	.11	.18	.07	.11
4/7/6/3/8	d/7/2	.11	.08	.05	.13	.09	.14
3/2/1/6/7/4	2/d	.09	.04	.07	.22	.11	.07
3/2/7/1/4/8	Spike 2	.05	.04	.01	.04	.03	-.05
7/3/4/6/2/8/1	7/d/1	.06	.06	.14	.15	.10	.08
	<i>Gr M</i>	.02	.01	.01	.07	.03	.04
	<i>St D</i>	.10	.10	.01	.12	.09	.11
	<i>Mdn</i>	.01	.01	-.01	.07	.02	.01
	<i>WNL Mean</i>	-.01	-.03	-.03	.01	-.02	.00
	<i>Elev Mean</i>	.06	.05	.04	.12	.06	.07

Note. $N = 30$. Values indicate validity coefficients indicated by Pearson product-moment correlations. Scale scores following "WNL" denote highest elevation(s). AAA = Automated Assessment Associates, MR = Minnesota Report (Pearson Assessments Systems), PAR = Psychological Assessment Resources, PS = Psych Screen, Inc., RF = RF Interpretive Report (Pearson Assessment Systems) M = Mean of the 4 MMPI-2 CBTI validity coefficients within profiles, $Gr M$ = mean correlation between therapist and CBTI ratings across all 30 profiles, Mdn = median correlation between therapist and CBTI ratings, SD = standard deviation, $WNL Mean$ = average for WNL profiles, $Elev Mean$ = average for profiles with at least one elevation ($T \geq 65$). Bolded values indicate significant correlations at the $p < .05$ level.

The highest aggregated correlations between the raters and the treating clinicians for the MMPI-2 CBTIs included those with Clinical and RC scale elevations including 8/6/7 and d/4/8/2/7 ($r = .22$), Spike 1 and 1/8 ($r = .22$), 4/3/2 and WNL ($r = .21$), and 4/1/3 and 6/4 ($r = .18$). The highest MMPI-2-RF Interpretive Report profile correlations, which were significant at the $p = .05$ level were 4/1/3 and 6/4 ($r = .24, p = .02$), Spike 1 and 1/8 ($r = .23, p = .02$), 8/6/7 and d/4/8/2/7 ($r = .21, p = .04$), and Spike 3 and WNL ($r = .21, p = .04$).

The lowest correlations for the aggregated MMPI-2 CBTIs were for profiles including Spike 2 ($r = -.15$) and two profiles that were WNL ($r = -.14$ and $-.07$). Regarding the MMPI-2-RF CBTI, the lowest correlations included three WNL profiles ($r = -.17, -.12$, and $-.12$). Each of the three worst correlations for the MMPI-2-RF CBTI overlapped with those for the MMPI-2 CBTIs.

Similar to previous research (e.g., McLaughlan et al., 2011), MMPI-2 CBTIs and MMPI-2-RF CBTIs were typically highly correlated. Consequently, correlations between therapists and MMPI-2 CBTIs and correlations between therapists and MMPI-2-RF CBTIs were typically similar in magnitude. Several exceptions are noteworthy, however. The two profiles displaying the largest discrepancies between the aggregated MMPI-2 CBTIs and the MMPI-2-RF program each favored the RF system. These had Clinical and RC scale elevations of WNL (MMPI-2-CBTI $r = -.05$) and WNL (MMPI-2-RF CBTI $r = .17$) and also Spike 3 (MMPI-2 CBTI $r = .05$) and WNL (MMPI-2-RF CBTI $r = .21$). The third largest discrepancy occurred among the profile with a Clinical scale elevation of WNL (MMPI-2 CBTI $r = .08$) and an RC scale elevation of 3/8/d/6 (MMPI-2-RF CBTI $r = -.01$).

After aggregating the validity coefficients for the MMPI-2 CBTI programs, the average correlation between CBTI raters and treating clinicians' sorts for WNL profiles was $r = -.02$,

whereas the average correlation was $r = .07$ for profiles with one or more Clinical scale elevations. Comparatively, the average validity coefficient for the RF Interpretive Report was $r = -.01$ for WNL profiles and $r = .09$ for profiles with one or more RC scale elevations.

To gain a better understanding of the specific content that was most and least consensual between treating clinicians and CBTI raters, item-level analyses were conducted using absolute mean differences between these two groups. First, absolute mean differences between therapist and CBTI ratings were obtained for each item within each profile of each CBTI program. This resulted in 150 tables (30 profiles x 5 CBTI systems). Next, therapist ratings were compared to the aggregate CBTI sort. There were 5 tables created to complete this step (one for each CBTI program). For ease of comparison, these averages were placed in a table alongside each of the CBTI programs. In addition, the four MMPI-2 CBTI programs were aggregated for a direct comparison to the MMPI-2-RF Interpretive Report item-level performance. For example, it can be observed whether particular items tended to improve or harm validity.

The statements that were sorted most and least consensually are displayed below in tables four through 18. Overall, the items displaying the greatest validity included "is comfortable with his/her sexual orientation" ($|MD| = .40$), "worries about own sexual feelings" ($|MD| = .48$), "values intellectual activities" ($|MD| = .50$), "is easy to impress" ($|MD| = .66$), and "rarely daydreams" ($|MD| = .68$). The least valid items were "has severe family problems" ($|MD| = 2.28$), "describes self as having few problems" ($|MD| = 2.05$), discusses problems openly" ($|MD| = 2.05$), "acts relaxed" ($|MD| = 2.05$), and "finds it easy to put aside troubles" ($|MD| = 2.04$).

Table 4. *Most and least consensual Midwestern Q-Sort Items*

Least		
Item	Absolute Mean Difference	Statement
42	2.28	Has severe family problems
21	2.05	Describes self as having few problems
24	2.05	Discusses problems openly
7	2.05	Acts relaxed
35	2.04	Finds it easy to put aside troubles
Most		
Item	Absolute Mean Difference	Statement
48	.40	Is comfortable with his/her sexual orientation
99	.48	Worries about own sexual feelings
97	.50	Values intellectual activities
49	.66	Is easy to impress
83	.68	Rarely daydreams

Table 5. *Most and least consensual items for the aggregated MMPI-2 CBTIs*

Least		
Item	Absolute Mean Difference	Statement
42	2.26	Has severe family problems
21	2.12	Describes self as having few problems
35	2.10	Finds it easy to put aside troubles
7	2.08	Acts relaxed
24	2.03	Discusses problems openly
Most		
Item	Absolute Mean Difference	Statement
48	.40	Is comfortable with his/her sexual orientation
99	.47	Worries about own sexual feelings
97	.50	Values intellectual activities
49	.65	Is easy to impress
83	.69	Rarely daydreams

Table 6. *Most and least consensual items for the MMPI-2-RF Interpretive Report*

Least		
Item	Absolute Mean Difference	Statement
42	2.33	Has severe family problems
24	2.14	Discusses problems openly
30	2.11	Feels capable of overcoming difficult life situations
5	2.04	Acts depressed
86	2.02	Reports good physical health
Most		
Item	Absolute Mean Difference	Statement
48	.40	Is comfortable with his/her sexual orientation
99	.51	Worries about own sexual feelings
97	.52	Values intellectual activities
83	.60	Rarely daydreams
60	.66	Is perfectionistic

It was also important to examine the items that contributed the most to the CBTI profiles that were found to be most and least valid. This was done by calculating the average absolute mean difference between the three most valid profiles and three least valid profiles for the aggregated MMPI-2 CBTIs and the MMPI-2-RF Interpretive Report. As displayed in Table 7, the most consensual aggregated MMPI-2 CBTI items within profiles displaying the most validity included "values intellectual activities," ($|MD| = .25$), "is comfortable with his/her sexual orientation," ($|MD| = .27$), "fakes illness to receive attention or other benefits" ($|MD| = .58$), "is easy to impress" ($|MD| = .60$), and "worries about own sexual feelings" ($|MD| = .60$). The least consensual items were "sees or hears things that do not exist" ($|MD| = 2.56$), "believes things that obviously are not true" ($|MD| = 2.29$), "generally tells the truth" ($|MD| = 2.29$), "is likely to develop trust in a counselor" ($|MD| = 2.23$), and "has severe family problems" ($|MD| = 2.19$).

Table 7. *Most and least consensual items within the most valid aggregated MMPI-2 CBTIs*

Least		
Item	Absolute Mean Difference	Statement
88	2.56	Sees or hears things that do not exist
14	2.29	Believes things that obviously are not true
36	2.29	Generally tells the truth
54	2.23	Is likely to develop trust in a counselor
42	2.19	Has severe family problems
Most		
Item	Absolute Mean Difference	Statement
97	.25	Values intellectual activities
48	.27	Is comfortable with his/her sexual orientation
29	.58	Fakes illness to receive attention or other benefits
49	.60	Is easy to impress
99	.60	Worries about own sexual feelings

The most consensual items within the least valid aggregated MMPI-2 CBTIs (Table 8) were "values intellectual activities" ($|MD| = .12$), "rarely daydreams" ($|MD| = .29$), "resists change" ($|MD| = .29$), "worries about own sexual feelings" ($|MD| = .40$), and "sees projects through to completion" ($|MD| = .48$). The items displaying the least agreement between CBTI raters and treating clinicians within these profiles were "sees or hears things that do not exist" ($|MD| = 2.56$), "believes things that obviously are not true" ($|MD| = 2.29$), "generally tells the truth" ($|MD| = 2.29$), "is likely to develop trust in a counselor" ($|MD| = 2.22$), and "has severe family problems" ($|MD| = 2.19$).

Table 8. *Most and least consensual items within the least valid aggregated MMPI-2 CBTIs*

Least		
Item	Absolute Mean Difference	Statement
35	3.60	Finds it easy to put aside troubles
7	3.40	Acts relaxed
21	3.29	Describes self as having few problems
2	3.13	Acts anxious
42	3.13	Has severe family problems
Most		
Item	Absolute Mean Difference	Statement
97	.12	Values intellectual activities
83	.29	Rarely daydreams
87	.29	Resists change
99	.40	Worries about own sexual feelings
89	.48	Sees projects through to completion

The most consensual MMPI-2-RF Interpretive Report items among the best performing CBTI profiles (Table 9) included "values intellectual activities" ($|MD| = .08$), "is comfortable with her/his sexual orientation" ($|MD| = .25$), "worries about own sexual feelings" ($|MD| = .33$), "is perfectionistic" ($|MD| = .50$), and "needs to achieve" ($|MD| = 0.58$). The least consensual items were "has severe family problems" ($|MD| = 3.08$), "believes things that obviously are not true" ($|MD| = 2.92$), "takes responsibility for own problems" ($|MD| = 2.83$), "acts depressed" ($|MD| = 2.75$), and "is aggressive" ($|MD| = 2.67$).

Table 9. *Most and least consensual items within the most valid MMPI-2-RF CBTIs*

Least		
Item	Absolute Mean Difference	Statement
42	3.08	Has severe family problems
14	2.92	Believes things that obviously are not true
92	2.83	Takes responsibility for own problems
5	2.75	Acts depressed
46	2.67	Is aggressive
Most		
Item	Absolute Mean Difference	Statement
97	.08	Values intellectual activities
48	.25	Is comfortable with his/her sexual orientation
99	.33	Worries about own sexual feelings
60	.50	Is perfectionistic
78	.58	Needs to achieve

Within the MMPI-2-RF Interpretive Report profiles found to be least valid, the items rated most similarly between treating therapists and CBTI raters were "is willing to discuss most personal problems" ($|MD| = .33$), "rarely daydreams" ($|MD| = .33$), "values intellectual activities" ($|MD| = .33$), "sees projects through to completion" ($|MD| = .42$), and "worries about own sexual feelings" ($|MD| = .42$). The least consensual items among these profiles were "acts relaxed" ($|MD| = 3.75$), "finds it easy to put aside troubles" ($|MD| = 3.75$), "acts anxious" ($|MD| = 3.17$), "acts depressed" ($|MD| = 3.17$), and "describes self as having few problems" ($|MD| = 3.17$). These values are displayed in Table 10.

Table 10. *Most and least consensual items within the least valid MMPI-2-RF CBTIs*

Least		
Item	Absolute Mean Difference	Statement
7	3.75	Acts relaxed
35	3.75	Finds it easy to put aside troubles
2	3.17	Acts anxious
5	3.17	Acts depressed
21	3.17	Describes self as having few problems
Most		
Item	Absolute Mean Difference	Statement
72	.33	Is willing to discuss most personal problems
83	.33	Rarely daydreams
97	.33	Values intellectual activities
89	.42	Sees projects through to completion
99	.42	Worries about own sexual feelings

Among the CBTI profiles displaying the best validity, common items that performed well on both the MMPI-2 and MMPI-2-RF CBTIs included "is comfortable with his/her sexual orientation," "worries about own sexual feelings," and "values intellectual activities." Those that consistently performed the worst across both instruments included "has severe family problems" and "discusses problems openly."

Table 11. *Most and least consensual items within the AAA CBTI*

Least		
Item	Absolute Mean Difference	Statement
42	2.37	Has severe family problems
7	2.19	Acts relaxed
37	2.16	Has close friendships
21	2.13	Describes self as having few problems
88	2.13	Sees or hears things that do not exist
Most		
Item	Absolute Mean Difference	Statement
48	.39	Is comfortable with his/her sexual orientation
99	.44	Worries about own sexual feelings
97	.50	Values intellectual activities
83	.56	Rarely daydreams
49	.63	Is easy to impress

Table 12. *Most and least consensual items within the MR CBTI*

Least		
Item	Absolute Mean Difference	Statement
42	2.29	Has severe family problems
21	2.28	Describes self as having few problems
7	2.28	Acts relaxed
35	2.23	Finds it easy to put aside troubles
5	2.22	Acts depressed
Most		
Item	Absolute Mean Difference	Statement
48	.38	Is comfortable with his/her sexual orientation
99	.45	Worries about own sexual feelings
97	.47	Values intellectual activities
49	.62	Is easy to impress
83	.63	Rarely daydreams

Table 13. *Most and least consensual items within the PAR CBTI*

Least		
Item	Absolute Mean Difference	Statement
42	2.30	Has severe family problems
35	2.26	Finds it easy to put aside troubles
2	2.13	Acts anxious
21	2.12	Describes self as having few problems
63	2.08	Is seeking help
Most		
Item	Absolute Mean Difference	Statement
48	.43	Is comfortable with his/her sexual orientation
97	.51	Values intellectual activities
99	.53	Worries about own sexual feelings
49	.68	Is easy to impress
83	.74	Rarely daydreams

Table 14. *Most and least consensual items within the PS CBTI*

Least		
Item	Absolute Mean Difference	Statement
42	2.09	Has severe family problems
86	2.04	Reports good physical health
21	1.97	Describes self as having few problems
88	1.94	Sees or hears things that do not exist
24	1.89	Discusses problems openly
Most		
Item	Absolute Mean Difference	Statement
48	.38	Is comfortable with his/her sexual orientation
99	.47	Worries about own sexual feelings
97	.51	Values intellectual activities
49	.66	Is easy to impress
60	.72	Is perfectionistic

Item-level analyses were also completed while grouping profiles based on scale elevations. For the aggregated MMPI-2 CBTI programs, the most and least valid items were examined based on whether profiles had either 1) no scale elevations on the Clinical scales (WNL) or 2) one or more Clinical scale elevations. MMPI-2-RF Interpretive Report profiles were distinguished and analyzed based on whether they were WNL or had one or more elevations on the Restructured Clinical scales. These steps were taken to develop a better understanding of general accuracy of specific inferences that can be drawn from these profile patterns within MMPI CBTIs.

The most consensual aggregated MMPI-2 WNL profile items (Table 15) were "values intellectual activities" ($|MD| = .46$), "is comfortable with his/her sexual orientation" ($|MD| = .51$), "rarely daydreams" ($|MD| = .56$), "worries about own sexual feelings" ($|MD| = .58$), and "is perfectionistic" ($|MD| = .63$). The least consensual items were "acts relaxed" ($|MD| = 2.91$), "expresses emotion in healthy ways" ($|MD| = 2.61$), "describes self as having few

problems" ($|MD| = 2.58$), "finds it easy to put aside troubles" ($|MD| = 2.52$), and "reports good physical health" ($|MD| = 2.47$).

Table 15. *Most and least consensual items for Clinical Scale WNL MMPI-2 Profiles*

Least		
Item	Absolute Mean Difference	Statement
7	2.91	Acts relaxed
28	2.61	Expresses emotion in healthy ways
21	2.58	Describes self as having few problems
35	2.52	Finds it easy to put aside troubles
86	2.47	Reports good physical health
Most		
Item	Absolute Mean Difference	Statement
97	.46	Values intellectual activities
48	.51	Is comfortable with his/her sexual orientation
83	.56	Rarely daydreams
99	.58	Worries about own sexual feelings
60	.63	Is perfectionistic

Among the aggregated MMPI-2 CBTI profiles with at least one Clinical Scale elevation (Table 16), the items rated most similarly by treating clinicians and CBTI raters were "is comfortable with his/her sexual orientation" ($|MD| = .30$), "worries about own sexual feelings" ($|MD| = .38$), "values intellectual activities" ($|MD| = .51$), "is easy to impress" ($|MD| = .52$), and "sees projects through to completion" ($|MD| = .67$). The least similar item ratings, on average, were "has severe family problems" ($|MD| = 2.17$), discusses problems openly" ($|MD| = 1.90$), "sees or hears things that do not exist" ($|MD| = 1.88$), is seeking help" ($|MD| = 1.87$), and "feels capable of overcoming difficult life situations" ($|MD| = 1.85$).

Table 16. *Most and least consensual items for MMPI-2 profiles with at least one Clinical Scale elevation*

Least		
Item	Absolute Mean Difference	Statement
42	2.17	Has severe family problems
24	1.90	Discusses problems openly
88	1.88	Sees or hears things that do not exist
63	1.87	Is seeking help
30	1.85	Feels capable of overcoming difficult life situations
Most		
Item	Absolute Mean Difference	Statement
48	.30	Is comfortable with his/her sexual orientation
99	.38	Worries about own sexual feelings
97	.51	Values intellectual activities
49	.52	Is easy to impress
89	.67	Sees projects through to completion

Table 17 provides the most extreme item validities for MMPI-2-RF Interpretive Reports with WNL RC scales. The most valid items were "rarely daydreams" ($|MD| = .41$), "values intellectual activities" ($|MD| = .42$), "is comfortable with his/her sexual orientation" ($|MD| = .42$), "is easy to impress" ($|MD| = .43$), and "worries about own sexual feelings" ($|MD| = .44$). The least valid items included "has severe family problems" ($|MD| = 2.69$), "acts relaxed" ($|MD| = 2.59$), "discusses problems openly" ($|MD| = 2.54$), "acts depressed" ($|MD| = 2.48$), and "expresses emotion in healthy ways" ($|MD| = 2.44$).

Table 17. *Most and least consensual items for RC Scale WNL MMPI-2-RF Profiles*

Least		
Item	Absolute Mean Difference	Statement
42	2.69	Has severe family problems
7	2.59	Acts relaxed
24	2.54	Discusses problems openly
5	2.48	Acts depressed
28	2.44	Expresses emotion in healthy ways
Most		
Item	Absolute Mean Difference	Statement
83	.41	Rarely daydreams
97	.42	Values intellectual activities
48	.42	Is comfortable with his/her sexual orientation
49	.43	Is easy to impress
99	.44	Worries about own sexual feelings

The most consensual items within the MMPI-2-RF Interpretive Reports for profiles with at least one RC scale elevation (Table 18) were "is comfortable with his/her sexual orientation" ($|MD| = .35$), "is perfectionistic" ($|MD| = .48$), "worries about own sexual feelings" ($|MD| = .52$), "values intellectual activities" ($|MD| = .65$), and "is organized" ($|MD| = .65$). The least consensual items were "behaves oddly" ($|MD| = 1.98$), "discusses committing suicide" ($|MD| = 1.96$), "accepts advice and suggestions" ($|MD| = 1.92$), "believes things that obviously are not true" ($|MD| = 1.92$), and "reports good physical health" ($|MD| = 1.88$).

Table 18. *Most and least consensual items for MMPI-2-RF profiles with at least one RC Scale elevation*

Least		
Item	Absolute Mean Difference	Statement
12	1.98	Behaves oddly
23	1.96	Discusses committing suicide
1	1.92	Accepts advice and suggestions
14	1.92	Believes things that obviously are not true
86	1.88	Reports good physical health
Most		
Item	Absolute Mean Difference	Statement
48	.35	Is comfortable with his/her sexual orientation
60	.48	Is perfectionistic
99	.52	Worries about own sexual feelings
97	.65	Values intellectual activities
58	.65	Is organized

After analyzing profile and item validities, stepwise regressions were conducted to allow for the examination of the amount of unique interpretive variance accounted for between each CBTI program. Because of the limited correlations between the treating clinicians' sorts and the CBTI raters' sorts and similarity of correlations between CBTI programs, variance accounted for by each program was extremely limited. Therefore, the inclusion criterion was raised to a value that would allow for every CBTI to be included in each model. Additionally, regressions were not conducted for profiles with which sorts for every CBTI program resulted in negative zero-order correlations. The values for each program are displayed in Table 19. The profiles to which the CBTIs contributed the most variance in the first step had Clinical/RC scale elevations including Spike 1 and 1/8 (MR $\Delta R^2 = .09$), 4/3/2 and WNL (AAA $\Delta R^2 = .07$), 4/1/3 and 6/4 (PS $\Delta R^2 = .07$), 8/6/7 and d/4/8/2/7 (PS $\Delta R^2 = .06$), and WNL on Clinical and RC scales (PS $\Delta R^2 = .06$). The system that contributed the most variance was most frequently PS (8 times), followed by RF and AAA (3 times), and MR (once). The PAR program never accounted for the most variance with clinician sorts.

Table 19. *Incremental validity indexed by stepwise regression*

Scale Elevations		Incremental ΔR^2				
Clinical	RC	CBTI 1	CBTI 2	CBTI 3	CBTI 4	CBTI 5
WNL 1=61T	1	---	---	---	---	---
WNL 1/3=64T	WNL 1/8=59T	---	---	---	---	---
WNL 2=52T	WNL 2=58T	AAA .01	---	---	---	---
WNL 3/6=57T	WNL 4=59T	---	---	---	---	---
WNL 3=61T	WNL 6=56T	---	---	---	---	---
WNL 4=64T	WNL 2=64T	---	---	---	---	---
WNL 6=52T	WNL 1/8=47T	---	---	---	---	---
WNL 6=56T	WNL 6=56T	RF .03	---	---	---	---
WNL 6=64T	6/1	PS .02	---	---	---	---
WNL 7/9=53T	WNL 4=54T	---	---	---	---	---
WNL 8/9=51T	WNL 6/9=56T	---	---	---	---	---
WNL 8=53T	WNL 6=56T	---	---	---	---	---
WNL 8=62T	3/8/d/6	AAA .03	---	---	---	---
WNL 9=59T	WNL 3=54T	PS .00	---	---	---	---
1	1/8	MR .09	AAA .01	---	---	---
2	WNL 7=62T	---	---	---	---	---
3	WNL 7=62T	RF .04	---	---	---	---
4	WNL 3=61T	---	---	---	---	---
6	WNL 6=61T	---	---	---	---	---
3/4	WNL 2=58T	PS .02	---	---	---	---
4/2	2/4	---	---	---	---	---
4/3	WNL d=53T	---	---	---	---	---
4/1/3	6/4	PS .07	---	---	---	---
4/3/2	WNL 1=59T	AAA .07	PS .02	---	---	---
8/6/7	d/4/8/2/7	PS .06	---	---	---	---
4/6/8/7/1	6/d/4	PS .03	---	---	---	---
4/7/6/3/8	d/7/2	RF .02	PS .00	---	---	---
3/2/1/6/7/4	2/d	PS .05	---	---	---	---
3/2/7/1/4/8	2	---	---	---	---	---
7/3/4/6/2/8/1	7/d/1	PS .02	---	---	---	---

Note. $N = 30$. Values represent ΔR^2 . CBTI programs are listed in the order in which they entered into the equation. AAA = Automated Assessment Associates, MR = Minnesota Report (Pearson Assessments Systems), PAR = Psychological Assessment Resources, PS = Psych Screen, Inc., RF = RF Interpretive Report (Pearson Assessment Systems). --- = profiles with sorts that resulted in negative beta weights.

Finally, two hierarchical regressions were calculated with the MMPI-2 CBTI reports entered as one block, and the MMPI-2-RF Interpretive Report entered as the other block. In the first regression, the MMPI-2 block was entered first to determine the amount of variance

accounted for by the MMPI-2-RF CBTI over and above the MMPI-2 CBTIs. Results can be found in Table 20. On average, the MMPI-2-RF Interpretive Report contributed 0.7% of unique variance. The RF profiles that contributed the most unique variance had Clinical/RC scale elevations of WNL and 3/8/d/6 ($\Delta R^2 = .04$), 4/3/2 and WNL ($\Delta R^2 = .06$), and 3/2/7/1/4/8 and Spike 2 ($\Delta R^2 = .01$).

Table 20. *Incremental validity of the MMPI-2-RF Interpretive Report over the block of MMPI-2 Reports*

Scale Elevations		Incremental ΔR^2			
Clinical	RC	MMPI-2	RF	ΔF	Sig. ΔF
WNL 2=52T	WNL 2=58T	.01	.00	.06	.81
WNL 6=64T	6/1	.02	.00	.00	.99
WNL 8=62T	3/8/d/6	.03	.04	3.78	.06
WNL 9=59T	WNL 3=54T	.00	.00	.15	.70
1	1/8	.10	.00	.08	.78
3/4	WNL 2=58T	.07	.00	.15	.70
4/1/3	6/4	.07	.01	.60	.44
4/3/2	WNL 1=59T	.09	.02	2.43	.12
8/6/7	d/4/8/2/7	.06	.00	.03	.87
4/6/8/7/1	6/d/4	.03	.00	.06	.80
4/7/6/3/8	d/7/2	.02	.01	.69	.41
3/2/1/6/7/4	2/d	.05	.01	.97	.33
7/3/4/6/2/8/1	7/d/1	.02	.000	.03	.85
	<i>Gr M</i>	.04	.01	.70	.60
	<i>WNL M</i>	.01	.01	1.00	.64
	<i>Elevated M</i>	.05	.01	.56	.59

Note. $N = 30$. MMPI-2 and RF Values represent ΔR^2 . MMPI-2 CBTIs will be entered as the first predictor and RF as the second. *Gr M* = average variance accounted for by the MMPI-2-RF Interpretive Report over and above the block of MMPI-2 CBTIs. *WNL M* = average variance accounted for profiles with WNL Clinical Scales. *Elevated M* = average variance accounted for within profiles with one or more elevated Clinical Scales. MMPI-2 block analyses did not include profiles with negative beta weights, as determined by stepwise regressions (Table 19).

For the second hierarchical regression, the order of the blocks were reversed so that the MMPI-2-RF Interpretive Report was entered as the first predictor and the MMPI-2 CBTIs were entered as the second predictor (Table 21). Overall, the average variance accounted for by the

MMPI-2 CBTIs over and above the MMPI-2-RF CBTI was 2.7%. Among these profiles, the MMPI-2 CBTIs blocks that contributed the most variance had Clinical/RC scale elevations including 4/3/2 and WNL ($\Delta R^2 = .08$), WNL and 3/8/d/6 ($\Delta R^2 = .06$), and 3/2/1/6/7/4 and 2/d ($\Delta R^2 = .06$).

Table 21. *Incremental validity of the MMPI-2 CBTI block over the MMPI-2-RF Interpretive Report*

Scale Elevations		Incremental ΔR^2			
Clinical	RC	RF	MMPI-2	ΔF	Sig. ΔF
WNL 2=52T	WNL 2=58T	.00	.01	.45	.50
WNL 6=64T	6/1	.01	.01	.91	.34
WNL 8=62T	3/8/d/6	.00	.06	6.32	.01
WNL 9=59T	WNL 3=54T	.00	.00	.30	.58
1	1/8	.05	.05	2.65	.08
3/4	WNL 2=58T	.00	.02	1.70	.20
4/1/3	6/4	.06	.01	1.46	.23
4/3/2	WNL 1=59T	.03	.08	4.47	.01
8/6/7	d/4/8/2/7	.04	.02	2.01	.16
4/6/8/7/1	6/d/4	.01	.02	1.98	.16
4/7/6/3/8	d/7/2	.02	.00	.21	.65
3/2/1/6/7/4	2/d	.00	.06	5.68	.02
7/3/4/6/2/8/1	7/d/1	.01	.02	1.59	.21
	<i>Gr M</i>	.02	.03	2.29	.24
	<i>WNL M</i>	.00	.02	2.00	.36
	<i>Elevated M</i>	.03	.03	2.42	.19

Note. $N = 30$. MMPI-2 and RF Values represent ΔR^2 . The MMPI-2-RF CBTIs were entered as the first predictor and MMPI-2 CBTIs as the second. *Gr M* = average variance accounted for by the MMPI-2 CBTIs over and above the block of MMPI-2-RF Interpretive Report. *WNL M* = average variance accounted for profiles with WNL Clinical Scales. *Elevated M* = average variance accounted for within profiles with one or more elevated Clinical Scales. MMPI-2 block analyses did not include profiles with negative beta weights, as determined by stepwise regressions (Table 19).

CHAPTER IV

DISCUSSION

On average, the MMPI-2-RF Interpretive Report exhibited very mildly superior validity over the MMPI-2 CBTI systems ($M_D = .010$). This was most evident when at least one scale elevation was found on both the Clinical scales and RC scales. In total, the MMPI-2-RF Interpretive Report validity coefficient exceeded that of the MMPI-2 CBTIs in five of the nine times that at least one scale elevation was found on both the Clinical scales and RC scales. The average for these correlations were $r = .11$ for the MMPI-2-RF CBTI and $r = .11$ for the aggregated MMPI-2 CBTIs. When an elevation occurred on one or more Clinical or RC scales but not on both sets of scales, the average MMPI-2-RF validity coefficient was $r = .02$, and the aggregated MMPI-2 CBTI validity coefficient was $r = .01$. Finally, when both the Clinical and RC scales were WNL, the mean therapist and MMPI-2-RF CBTI correlation was $r = -.01$. The mean therapist and aggregated MMPI-2 CBTI correlation was $r = -.03$ in this scenario. Overall, meaningful differences could not be discerned between the aggregated MMPI-2 CBTI programs and the MMPI-2-RF Interpretive Report; however, there was a meaningful tendency for CBTI profiles to be interpreted with greater validity when the Clinical and/or RC scales contained at least one significant elevation.

The trend for WNL profiles to be interpreted with less validity than those that included one or more elevations on the Clinical or RC scales is likely related to several factors. For instance, raters may be provided with more relevant and detailed information from CBTI reports when examinees endorse substantial psychopathology than when examinees deny significant distress. Among WNL reports, interpretive statements are likely to include less nuanced information because there are comparatively less behavioral correlates for these profiles. The

examinees, who were voluntarily enrolled in psychotherapy, were likely experiencing some form of life stressors, but these were self-reported on the MMPI instruments as circumscribed or limited in scope. It is probable that the CBTI programs lack sensitivity to these types of concerns in the absence of extra-test data. Therefore, CBTI reports could only provide interpretive statements that are less specific to the individual and perhaps more vague than statements that are programmed for profiles including scale elevations.

Functionally, this highlights the importance of considering multiple sources of information when conceptualizing individuals with the assistance of CBTI programs, as recommended by various sources (e.g., Graham, 2006). In addition, CBTI reports are potentially more useful to clinicians' conceptualizations when used in settings in which examinees are more likely to produce elevated scales, such as Community Mental Health agencies or hospitals where severe and persistent mental health concerns are more frequent as opposed to private practices in which clients are more likely to be seen for acute psychopathology and adjustment disorders.

Item-level analyses often revealed similar items performing well (e.g., "is comfortable with his/her sexual orientation," "worries about own sexual feelings," "values intellectual activities") or poorly (e.g., "has severe family problems," "discusses problems openly") regardless of CBTI program. Some differences emerged after examining item validity based on scale elevations (WNL or one or more) between the MMPI-2 CBTIs and the MMPI-2-RF Interpretive Report. Between MMPI-2 WNL profiles and MMPI-2-RF WNL profiles, the most consensual items tended to overlap (e.g., "values intellectual activities," "is comfortable with his/her sexual orientation," "rarely daydreams," "worries about own sexual feelings"), but the least consensual items were typically different. Among the five items displaying the least validity, only "acts relaxed" overlapped. A similar pattern was found when comparing MMPI-2

CBTI and MMPI-2-RF CBTI profiles with at least one scale elevation to each other. The items with the greatest agreement between the CBTI raters and the treating clinicians were often the same (e.g., "is comfortable with his/her sexual orientation," "worries about own sexual feelings," "values intellectual activities"), but none of the five least consensual items were the same.

The specific cause of this trend is unclear. Ideally, the most consensual items between CBTI raters and treating clinicians reflect effective statements provided by the CBTI programs; however, it may be that the ratings were similar because the statements were perceived by both the CBTI raters and the clinicians as low base rate events, making it easier to place each statement at an extreme end of the q-sort distribution. In addition, the most consensual items may have been rated similarly based on their lack of transparency. When these inferences were not described within CBTI profiles, raters were more likely to avoid making extreme ratings.

Disparate ratings also have several potential interpretations. For instance, CBTI programs may have provided either no information or conflicting statements regarding these items. Alternatively, and more favorably for the CBTI programs, the statements provided within the CBTIs were possibly accurate as applied to the profiles, but the information was not gathered within the therapeutic context. Therefore, incongruent ratings may reflect appropriate inferences that could benefit therapists' conceptualizations if they were to utilize these programs.

Item-level analyses were also examined within the most and least valid CBTI profiles. The most consensual items for the MMPI-2 CBTIs (e.g., "values intellectual activities," "worries about own sexual feelings") appear to again possibly represent items that are considered exceptionally high or low base rate events, making them readily identifiable by CBTI raters and clinicians. In addition, some of the items (e.g., "sees projects through to completion") may have

been rated similarly due to lack of transparency, forcing raters to again rely on their understanding of prevalence among individuals.

The items with the poorest agreement between therapists and CBTI raters among the most and least valid CBTI profiles seemingly represent inferences that are likely to be more clearly displayed within one of the two available domains (i.e., therapy and MMPI profiles). For example, some of the items with the least amount of agreement between therapists and CBTI raters included "has severe family problems" and "discusses problems openly." In these scenarios, it is probable that a therapist has a clearly developed impression, whereas CBTI raters only relied on statements produced by the programs. Thus, the raters may not have been able to give these items an extreme rating due to less confidence in the inferences, particularly when profiles were WNL. This consideration further highlights the importance of integrating various sources of data to arrive at valid inferences and conceptualizations.

The findings from stepwise and hierarchical regressions were limited by low correlations between therapists and CBTI raters and high correlations between each profile. The Psych Screen, Inc. program tended to contribute the most unique variance, and the MMPI-2-RF Interpretive Report tended to contribute the least, although differences were non-significant. Within hierarchical regressions, the MMPI-2 CBTIs typically accounted for more variance over and above the MMPI-2-RF CBTI ($Mean \Delta R^2 = 0.051$) than the reverse ($Mean \Delta R^2 = 0.010$). This can at least partially be explained by the extra weight applied to MMPI-2 CBTIs because they outnumbered the MMPI-2-RF CBTI reports by a four-to-one ratio.

Conclusions

In clinical practice, the results of the present study appear to offer the most benefit to interpretive decisions in general rather than as they apply to one specific CBTI program.

Whereas some programs performed mildly better than others, the significance of these differences for applied purposes is not clear. However, there was a consistent trend for greater agreement between treating therapists and CBTI raters when there was at least one scale elevation greater than $T = 65$ as opposed to WNL profiles. This may stem from several sources, such as reduced clarity of interpretive statements, fewer statements, more conflicting information, or information that is inaccurate for the examinee. To overcome this limitation, clinicians would benefit from using CBTIs in settings in which greater levels of psychological distress and/or psychopathology are probable. For instance, CBTIs may provide the most valuable information to clinicians who practice in hospital or community mental health settings. Clinicians whose clients are typically seeking services for acute, circumscribed concerns or individuals who use the MMPI for assessments such as police officer screenings or custody evaluations, in which psychopathology is unlikely to be endorsed, are less likely to benefit from the use of the reviewed CBTIs.

Limitations and Future Directions

Several caveats should be considered when interpreting the findings within the present study. First, it is important to recognize that the data represent *relative* rather than *absolute* validity of the CBTIs that were investigated. Validity coefficients document the level of agreement between the raters, who were only provided access to CBTI reports, and the treating clinicians, who did not have access to MMPI results until after their Midwestern Q-Sort ratings were completed. Although numerous studies have used a similar external criterion (i.e., treating clinicians) to document validity (e.g., Sellbom, Ben-Porath, Patrick, Wygant, Gartland, & Stafford, 2012), this methodology is inexorably linked to various sources of error. Whereas clinicians are expected to thoroughly investigate the biopsychosocial functioning of their clients,

they are not privy to the full spectrum of cognitive, affective, physical, and behavioral features that are correlated with MMPI-2 and MMPI-2-RF response profiles. This is especially true in the early stages of treatment when less information has been revealed by the clients.

The treating clinicians within the current study had seen each client four to six times. Though a clinician ought to have developed a conceptualization and treatment plan by this stage, it is common practice for these to be revised continuously throughout the provision of services. For example, this is particularly true of clients who are deeply concerned with appearing well-adjusted and moralistic to others. These clients may take longer to divulge some information to their therapist lest they be judged negatively. This type of approach to treatment could be described in CBTI reports as a result of elevated scales such as L, L-r, K, and K-r, but a therapist may not be aware of this style and could be missing critical information for accurately completing q-sorts. Thus, the clients' approaches to completing MMPI-2 questionnaires and interacting with their therapists introduce sources of error that vary based on the individual. Furthermore, the unknown reliability of the clinicians' sorts may represent a flaw in the external criterion.

The limitations of using therapist q-sorts to document validity are somewhat overcome by interpreting the findings in a relative sense. The method error remained consistent across the five CBTIs within this study, allowing for examination of which CBTI programs tended to be most and least accurate in comparison to each other. Future studies would benefit from additional data sources that can be compared to q-sort ratings based on CBTI profiles, such as q-sorts completed by significant others or therapist ratings collected toward the end of treatment.

The raters who completed sorts within the present study ranged from first year graduate students in a non-terminal clinical psychology doctoral program to practicing psychologists who

were educated within the same program. Consequently, raters had varying levels of experience with MMPI interpretation and clinical practice in general. The degree to which this suppressed correlations between raters and treating clinicians within the current study is not clear due to the limited number of raters. Future studies would benefit from examining the impact of training experiences, clinical experiences, and familiarity with the MMPI instruments on interpretive validity of CBTI profiles. For example, several raters with similar experiences could be grouped and compared to raters with disparate experiences.

Within the present study, CBTI profiles were selected based on concordant statements indicating valid profiles between programs. This method resulted in a large percentage of the selected profiles having no scale elevations. To examine a fuller range of profile types, future studies should use profile selection criteria based on scale elevations. Using this strategy, the relationship between number of scale elevations and interpretive validity can be elucidated. Additionally, common code types should be selected to improve the applicability of findings to clinical practice. These goals could be completed by examining the validity of CBTIs with a sample in which more severe psychopathology is commonly endorsed. Validity coefficients may differ significantly when derived from hospital, Community Mental Health, or prison samples, for example.

Finally, the landscape of commercially available CBTI systems has undergone meaningful changes since the beginning of the present study. Most notable, the AAA program is no longer available, and Greene, Brown, and the PAR staff have released a CBTI system for the MMPI-2-RF through Psychological Assessment Resources (2012). Future studies should include ratings for this program in analyses. This would provide users with information regarding the performance of this system in comparison to others as well as allowing for the

examination of an MMPI-2-RF-based system other than the MMPI-2-RF Interpretive Report. If these two programs were shown to converge better with each other than MMPI-2 CBTI systems, this would suggest differences in the parent instruments are the greatest contributor to MMPI-2 and MMPI-2-RF CBTI interpretation differences. However, if an MMPI-2-RF system were to align more closely with MMPI-2 CBTIs rather than the other MMPI-2-RF CBTI, this would likely indicate differences attributable to programming decisions and report styles.

REFERENCES

- Atlis, M. M., Hahn, J., & Butcher, J. N. (2006). Computer-based assessment with the MMPI-2. In J. N. Butcher (Ed.), *The MMPI-2: A practitioner's guide* (pp. 445-476). Washington, DC: American Psychological Association.
- Archer, R., Griffin, R., & Aiduk, R. (1995). MMPI-2 clinical correlates for ten common codes. *Journal of Personality Assessment*, 65(3), 391-407.
- Automated Assessment Associates (2008). Retrieved September 25, 2011, from <http://www.psychsoftware.net/mmpi2report.html>
- Ben-Porath, Y. S., & Tellegen, A. (2008a). *MMPI-2-RF User's Guide for Reports*. Minneapolis, MN: University of Minnesota Press.
- Ben-Porath, Y. S., & Tellegen, A. (2008b). *Minnesota Multiphasic Personality Inventory-2-Restructured Form (MMPI-2-RF): Manual for administration, scoring, and interpretation*. Minneapolis, MN: University of Minnesota Press.
- Butcher, J. N. (1993). *User's Guide for The Minnesota Report: Adult Clinical System – Revised [Computer Software Manual]*. Minneapolis, MN: The Regents of the University of Minnesota.
- Butcher, J. N. (2009). How to use computer-based reports. . In J. N. Butcher (Ed.), *Oxford Handbook of Personality Assessment* (pp. 693-706). New York, NY: Oxford University Press.
- Butcher, J. N. (2011). Retrieved September 25, 2011, from <http://psychcorp.pearsonassessments.com/haiweb/cultures/en-us/productdetail.htm?pid=p 24034>
- Butcher, J. N., Atlis, M. M., & Hahn, J. (2004). The Minnesota Multiphasic Personality Inventory-2 (MMPI-2). In M. Hersen (Ed.), *Comprehensive handbook of psychological assessment: Personality assessment* (pp. 30-38). Hoboken, NJ: John Wiley & Sons, Inc.
- Butcher, J. N., Dahlstrom, W. G., Graham, J. R., Tellegen, A. M., & Kaemmer, B. (1989). *Minnesota Multiphasic Personality Inventory-2 (MMPI-2): Manual for administration and scoring* . Minneapolis: University of Minnesota Press.
- Butcher, J. N., Graham, J. R., & Ben-Porath, Y. S. (1995). Methodological problems and issues in MMPI, MMPI-2, and MMPI-A research. *Psychological Assessment*, 7(3), 320-329. DOI: 10.1037/1040-3590.7.3.320
- Butcher, J. N., Hamilton, C. K., Rouse, S. V., & Cumella, E. J. (2006). The deconstruction of the Hy scale of MMPI-2: Failure of RC3 in measuring somatic symptom expression. *Journal of Personality Assessment*, 87, 199–205. DOI:10.1207/s15327752jpa8702_08

Butcher, J. N., Perry, J. N., & Atlis, M. M. (2000). Validity and utility of computer-based test interpretation. *Psychological Assessment, 12*, 6-18.

Butcher, J. N., Perry, J. N., & Dean, B. L. (2009). Computer-based assessment. In J. N. Butcher (Ed.), *Oxford Handbook of Personality Assessment* (pp. 163-182). New York, NY: Oxford University Press.

Conoley, C. W., Plake, B. S., & Kemmerer, B. E. (1991). Issues in computer-based test interpretation systems. *Computers in Human Behavior, 7*, 97-101.

Cox, A. C., Weed, N. C., & Butcher, J. N. (2009). The MMPI-2: History, interpretation, and clinical issues. In J. N. Butcher (Ed.), *Oxford Handbook of Personality Assessment* (pp. 250-276). New York, NY: Oxford University Press.

Deskovitz, M. A., Weed, N. C., & Williams, J. E. (2006). *Interpretive reliability of six MMPI-2 computer-based test interpretations programs*. Unpublished Manuscript, Central Michigan University, Mount Pleasant, MI.

Epstein, J., & Rotunda, R. J. (2000). The utility of computer versus clinician-authored assessments in aiding the prediction of patient symptomatology. *Computer in Human Behavior, 16*, 519-536.

Eyde, L., Kowal, D. M., & Fishburne, F. J. (1991). The validity of computer-based test interpretations of the MMPI. In T. B. Gutkin & S. L. Wise (Eds.), *The computer and the decision-making process* (pp. 75-123). Hillsdale, NJ: Lawrence Erlbaum Associates.

Fowler, R. D. (1987). Developing a computer-based test interpretation system. In J. N. Butcher (Ed.), *Computerized psychological assessment*. (pp. 50-63). New York: Basic Books, Inc.

Garb, H. N. (1989). Clinical judgment, clinical training, and professional experience. *Psychological Bulletin, 105*, 387-396.

Garb, H. N. (2000). Computers will become increasingly important for psychological assessment: Not that there's anything wrong with that!. *Psychological Assessment, 12*(1), 31-39.

Graham, J. R. (1999). *MMPI-2: Assessing Personality and Psychopathology* (3rd ed.). New York: Oxford University Press.

Graham, J. R. (2006). *MMPI-2: Assessing Personality and Psychopathology* (4th ed.). New York: Oxford University Press.

Green., C. J. (1982). The diagnostic accuracy and utility of MMPI and MCMI computer interpretive reports. *Journal of Personality Assessment, 46*, 359-365.

Greene, R. L. (2000). *The MMPI-2/MMPI: An Interpretive Manual* (2nd ed.). Needham Heights, MA: Allyn & Bacon.

Greene, R. L., Brown, R. C., & Kovan, R. E. (1998). *MMPI-2 Adult Interpretive System for Windows, Professional Manual* (Version 2.0) [Computer Software Manual]. Odessa, FL: Psychological Assessment Resources, Inc.

Greene, R. L., Brown, R. C., & PAR Staff (2012). Retrieved December 8, 2012, from http://www4.parinc.com/Products/Product.aspx?ProductID=MMPI2_AIS_V4

Grove, W. M., Zald, D. H., Lebow, B. S., Snitz, B. E., & Nelson, C. (2000). Clinical versus mechanical prediction: A meta-analysis. *Psychological Assessment, 12*, 19-30.

Harrington, A. (1999). *The relative validity of computer-based interpretations of the MMPI-2*, Unpublished doctoral dissertation, University of Mississippi, Oxford, MS.

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

Lewandowski, D., & Graham, J. (1972). Empirical correlates of frequently occurring two-point MMPI code types: A replicated study. *Journal of Consulting and Clinical Psychology, 39*(2), 467-472.

McLaughlan, J. K., Weed, N. C., & Deskovitz, M. A. (2011). *Convergence of the MMPI-2-RF Interpretive Report with MMPI-2 computer-based test interpretation systems*. Unpublished master's thesis, Central Michigan University, Mount Pleasant, MI.

McMinn, M. R., Ellens, B. M., & Soref, E. (1999). Ethical perspectives and practice behaviors involving computer-based test interpretation. *Assessment, 9*, 71-77.

McNeal, T. P. (1997). *Examining the validity of computer-based test interpretation with interpretive q-sorts*. Unpublished master's thesis, University of Mississippi, Oxford, MS. Minnesota Multiphasic Personality Inventory–2 Interpreter (Version 3.0) [Computer Software]. (1998). Milton, WI: Psych Screen, Inc.

Minnesota Multiphasic Personality Inventory–2 Interpreter (Version 3.0) [Computer Software]. (1998). Milton, WI: Psych Screen, Inc.

Morey, L. C. (1996). *Personality Assessment Inventory: Professional Manual*. Odessa, FL: Psychological Assessment Resources.

Nichols, D. (2006). The trials of separating bathwater from baby: A review and critique of the MMPI–2 Restructured Clinical scales. *Journal of Personality Assessment, 87*, 121–138. DOI:10.1207/s15327752jpa8702_02

Noland, K. A., & Weed, N. C. (1994). *The Mississippi Q Sort* [Computer software]. Oxford, MS: Author.

Pearson (2011). Retrieved September 25, 2011, from <http://www.pearsonassessments.com/hai/Templates/Products/ProductDetail.aspx?NRMODE=Published&NRNODEGUID=%7b4DB6CC4B-D482-4DFD-A1BB-F7E2657F3EB2%7d&NRORIGINALURL=%2fHAIWEB%2fCultures%2fen-us%2fProductdetail%2ehtm%3fPid%3dPAg523&NRCACHEHINT=NoModifyGuest&Pid=PAg523#reports>

Pearson, J. S., Rome, H. P., Swenson, W. M., Mataya, P., & Brannick, T. L. (1965). Development of computer system for scoring and interpretation of MMPI in a medical clinic. *Annals of New York Academy of Sciences*, *126*, 684-692.

PsychScreen, Inc. (n.d.). Retrieved September 25, 2011, from <http://www.psychscreen.com>

Ranson, M. B., Nichols, D. S., Rouse, S. V., & Harrington, J. L. (2009). Changing or replacing an established psychological assessment standard: Issues, goals, and problems with special reference to recent developments in the MMPI-2. In J. N. Butcher (Ed.), *Oxford Handbook of Personality Assessment* (pp. 250-276). New York, NY: Oxford University Press.

Rome, H. P., Swenson, W. M., Mataya, P., McCarthy, C. E., Perason, J. S., Keating, F. R., & Hathaway, S. R. (1962). Symposium on automation techniques in personality assessment. *Proceedings of the Staff Meetings of the Mayo Clinic*, *37*, 61-82.

Sellbom, M., Ben-Porath, Y. S., & Graham, J. R. (2006). Correlates of the MMPI-2 restructured clinical (RC) scales in a college counseling setting. *Journal of Personality Assessment*, *86*(1), 89-99. doi: 10.1207/s15327752jpa8601_10

Sellbom, M., Ben-Porath, Y. S., McNulty, J. L., Arbisi, P. A., & Graham, J. R. (2006). Elevation differences between MMPI-2 clinical and restructured clinical (RC) scales: Frequency, origins, and interpretive implications. *Assessment*, *13*(4), 430-441.

Sellbom, M., Ben-Porath, Y. S., Patrick, C. J., Wygant, D. B., Gartland, D. M., & Stafford, K. P. (2012). Development and construct validation of MMPI-2-RF indices of global psychopathy, fearless-dominance, and impulsive-antisociality. *Personality Disorders: Theory, Research, and Treatment*, *3*(1), 17-38. doi: 10.1037/a0023888

Strassburg, D. S., & Cooper, L. M. (1997). *MMPI/MMPI-2 Adult Interpretive Report* (Version 6.0P) [Computer Software]. Salt Lake City, UT: Automated Assessment Associates.

Tellegen, A., & Ben-Porath, Y. S. (2008). *Minnesota Multiphasic Personality Inventory-2 Restructured Form: Technical manual*. Minneapolis: University of Minnesota Press.

Tellegen, A., Ben-Porath, Y. S., McNulty, J. C., Arbisi, P. A., Graham, J. R., & Kaemmer, B. (2003). *MMPI-2 restructured clinical (RC) scales: Development, validation, and interpretation*. Minneapolis: University of Minnesota Press.

Tellegen, A., Ben-Porath, Y. S., Sellbom, M., Arbisi, P. A., McNulty, J. L., & Graham, J. R. (2006). Further evidence on the validity of the MMPI-2 Restructured Clinical (RC) scales: Addressing questions raised by Rogers, Sewell, Harrison, and Jordan and Nichols. *Journal of Personality Assessment*, *87*(2), 148-171.

Weed, N. C. (2006). Syndromal complexity, paradigm shifts, and the future of validation research: Comments on Nichols and Rogers, Sewell, Harrison, and Jordan. *Journal of Personality Assessment*, *87*(2), 217-222. doi: 10.1207/s15327752jpa8702_12

Weed, N. C., Ben-Porath, Y. S., and Butcher, J. N. (1990). Failure of Wiener and Harmon Minnesota Multiphasic Personality Inventory (MMPI) Subtle Scales as personality descriptors and as validity indicators. *Psychological Assessment*, *2*(3), 281-285. doi: 10.1037/1040-3590.2.3.281

Williams, J. E. (2002). *The relative validity of eight computer-based test interpretations for the MMPI-2*, Unpublished doctoral dissertation, University of Mississippi, Oxford, MS.

Williams, J. E., & Weed, N. C. (2003). *Using a web-based Q-sort program to teach MMPI interpretation*. Paper presented at the 38th Annual Symposium on Recent Developments in the Use of the MMPI, MMPI-2, and MMPI-A. Minneapolis, MN.

Williams, J. E., & Weed, N. C. (2004a). Relative user ratings of MMPI-2 computer-based test interpretations. *Assessment*, *11*(4), 316-329.

Williams, J. E., & Weed, N. C. (2004b). Review of computer-based test interpretation software for the MMPI-2. *Journal of Personality Assessment*, *83*, 78-83.