

MMPI-2-RF HIGHER ORDER PATHOLOGY OF PHYSICIANS IN DISTRESS

By

Paul Borden Ingram ^{IV}

Copyright 2015

Dissertation submitted to the graduate degree program in Counseling Psychology and the Graduate Faculty of the University of Kansas in partial fulfillment of the requirements for the degree of Doctor of Philosophy.

Chairperson Karen D. Multon, Ph.D.

Barbara A. Kerr, Ph.D.

Rick Ginsberg, Ph.D.

Sarah B. Kirk, Ph.D., ABPP

Patricia A. Lowe, Ph.D.

Date Defended: October 20th, 2015

The Dissertation Committee for Paul Borden Ingram ^{IV} certifies that this is the approved version of the following dissertation:

MMPI-2-RF HIGHER ORDER PATHOLOGY OF PHYSICIANS IN DISTRESS

Chairperson Karen D. Multon, Ph.D.

Date approved: October 20th, 2015

ABSTRACT

Modern theories of psychopathology incorporate a higher-order approach to conceptualization (e.g., Krueger & Tackett, 2003). These theories are often characterized as having internalizing, externalizing, and thought disorder content as critical elements. This higher order approach has recently been incorporated into the MMPI-2-Restructured Form (MMPI-2-RF; Tellegen & Ben-Porath, 2008/2011). Using individual higher-order content scales and instrument wide interpretive schemas, a hierarchical emphasis on interpretation is part of a critical effort to bring the MMPI in line with current conceptualizations of psychopathology. However, there is a paucity of research available about the utility of scale wide interpretive frameworks, despite their use being recommended for interpretation. Using a population of physicians in distress, this study evaluated the potential portability and generalization of the MMPI-2-RF's interpretive framework into a novel population through structural analysis using several analytic methods (i.e., Confirmatory Factor Analysis, Exploratory Structural Equation Modeling, and Exploratory Factor Analysis). Each of these techniques included different assumptions about scale interrelatedness and broader relationships and, when taken together, provides a comprehensive evaluation of the feasibility of the higher-order themes as an interpretive framework for the MMPI-2-RF in a population of physicians in distress. Although no model was found as entirely appropriate to the proposed framework noted by the authors of the MMPI-2-RF, the three critical areas of psychopathology did emerge. Issues with incorporating these higher order models are related to the need for nested factors, specifically relating to externalization content of anger, hostility, and aggression. Beyond guidance on general interpretability of the higher order themes, specific guidance is offered for use of the MMPI-2-RF in a population of physicians in distress.

Keywords: MMPI-2-RF, Higher-order Pathology, Personality, Personality Assessment

ACKNOWLEDGEMENTS

I owe a great deal of thanks to many people for many things over many years. Many have helped, and continue to help, give me the support, perspectives, and determination needed to pursue my long held dream over my life. There are too many to name and words here will insufficiently describe the gratitude and importance that these many people have played in my life. This has taken many years to accomplish, but I have grown through it. Through this process of growth I have developed my interest and curiosity about the world; with support I have developed the skills to evaluate, learn, and act as an agent of change.

To my parents: Without you I would not be the person I am today. I would not be able to appreciate the world as I do or to think, perceive, and question the things around me. You helped me be curious, adventurous, and stubborn enough to persist. You gave me the tools and the wisdom to use them as I set out on my journey to explore the world.

To Dr. Karen Multon, my committee members, and professional mentors: You have offered me direction, guidance, and support. You have encouraged me as a practitioner, as a scientist, and as a future psychologist. This has shaped my professional perspectives, my goals, and abilities in a way that I am only now beginning to understand. You have enabled me to think, to test out those thoughts, and to incorporate those that work into my schema of the world.

To my friends: Your support and friendship has added richness and meaning to my world.

Our experiences together have been personally invaluable; life without rich relationships is stark and barren. Our connections, no matter how recent or old, have engendered growth, resiliency, and purpose.

To the journey: I have grown by your process. I have stumbled, fallen, and attempted again.

I am enriched and invigorated; I am amazed at the remarkableness of the course of learning itself. Through my daily engagement with my studies I reminded that there is no path; the path is made by walking.

“A little learning is a dang’rous thing,
Drink deep, or taste not the Pierian Spring:
There shallow draughts intoxicate the brain,
And drinking largely sobers us again
Fir’d at first sight with what the Muse imparts,
In fearless youth we tempt the heights of Arts,
While from the bounded level of our mind
Short views we take, nor see the lengths behind;
But more advanc’d, behold with strange surprise
New distant scenes of endless science rise!”

Alexander Pope

An Essay on Criticism

TABLE OF CONTENTS

CHAPTER 1	8
Abstract	8
MMI-2-RF Higher Order Pathology of Physicians in Distress: A Review	9
MMPI-2-Restructured Form Development.....	13
RC Scale Development	14
RC Scale Construction Criticisms	15
RC Scale Construction Defense	17
Higher-Order Content of the MMPI-2-RF.....	19
MMPI-2-RF Substantive Scales	23
Clashing Paradigms Guide Theory	25
References.....	32
CHAPTER 2	38
Abstract	38
MMI-2-RF Higher Order Pathology of Physicians in Distress: A Review	39
Present Study	46
Method	48
Participants.....	48
Instrumentation	49
MMPI-2-Restructured Form	49
Procedure	49
Data Analysis	50
Confirmatory Factor Analysis.....	51
Exploratory Structural Equation Model.....	51
Exploratory Factor Analysis	52
Results.....	53
Descriptive Characteristics	53
Confirmatory Factor Analyses.....	54
Exploratory Structural Equation Models	55
Exploratory Factor Analysis	56
Discussion	59
Issues of Covariation.....	62
Insufficiency of Model Correction	64
Insufficiency of Three Factor Solution.....	65
An Integrated Model of Psychopathology	66
Limitations and Future Directions	70
Conclusion	72
References.....	75
Appendices.....	81
Tables.....	81
Table 1: MMPI-2-RF Scale Names and Descriptions	81
Table 2: Descriptive Statistics for the Validity, Higher Order, and RC Scales	83

Table 3: Descriptive Statistics for the Specific Problem and PSY-5.....	84
Table 4: Frequency of Higher Order, and RC scale Elevation ≥ 65	85
Table 5: Frequency of Specific Problem and PSY-5 Elevation ≥ 65	86
Table 6: MMPI-2-RF Means Contrasted with Comparison Groups.....	87
Table 7: Summary Table of CFA Analyses.....	88
Table 8: Summary Table of ESEM Analyses.....	89
Table 9: Exploratory Factor Analysis Structural Matrix.....	90
Table 10: Inter-domain correlations and Eigenvalues of factors extracted during EFA ...	91
Figures.....	92
Figure 1: Proposed structure for the three higher-order themes of the MMPI-2-RF.....	92
Figure 2: CFA following Data-based and Theory-based Corrections.....	93
Figure 3: Scree Plot for Observed Eigenvalues in Exploratory Factor Analysis.....	94

CHAPTER 1

Abstract

In the years since its recent release, the MMPI-2-Restructured Form (MMPI-2-RF; Tellegen & Ben-Porath, 2008/2011) has amassed a base of research literature showing strong support for its capacity as a clinical instrument. However, during its development, the MMPI-2-RF underwent a number of changes including abandonment to scale construction approaches (both statistical and theoretical approaches) that have long been part of the interpretive and developmental history of the MMPI. As a result of these critical changes, scholars have placed themselves in positions to critique, lambast, or defend the MMPI-2-RF, often labeling it either as the prodigal next step in assessment development or as an under-developed instrument that has abandoned its interpretive core in a manner suggestive of clinical poverty. No matter the position taken, the revisions inherent to the MMPI-2-RF pose a cross-road for understanding of assessment. This paper begins with a review of refinement approaches utilized within the MMPI-2-RF revision efforts while contextualizing contrasting opinions seen in the literature about these decisions. Then, from within the context of these clashing perspectives, the underlying assumptions about the broad development of assessment psychology as a scientific practice are evaluated using a Kuhnian lens, using the MMPI-2-RF as a proxy. The acceptance or rejection of test revisions is proposed as evidence of crisis within the evolving science of assessment psychology.

Keywords: MMPI-2-RF, MMPI-2, Scale Development, Personality Assessment, Kuhn

CHAPTER 1

MMI-2-RF Higher Order Pathology of Physicians in Distress: A Review

The Minnesota Multiphasic Personality Inventory (MMPI; Hathaway & McKinley, 1940a; 1940b; 1942) was published in 1940 as a result of an inter-professional, collaborative effort at the University of Minnesota. It proved a turning point in the professional roles of psychologists, increasing their professional independence and emphasizing a specific area of clinical practice in which they could provide a novel contribution (Buchanan, 1994). The MMPI's development has been suggested as evidence that the history of the modern psychologist is inseparable from assessment and testing (Benjamin, 2005). Devised as a diagnostic tool to reduce time spent on lengthy psychiatric interviews (Buchanan, 1994), the success of the MMPI was the result of an alternative item-inclusion approach to its scales (Benjamin, 2005). Earlier personality tests (e.g., Woodworth, 1917) relied on a rationale scale development effort without an empirical basis for item selection. In contrast to tests devised using this rational approach, the MMPI's goal was to establish a purely empirical technique that could serve effectively in evidence-based, actuarial clinical decision making (Meehl, 1946). Focusing this empirical approach on adjustment-based dysfunction (Gibby & Zickar, 2008), the MMPI was developed to predict aberrant behaviors and outcomes. Largely as a function of its use in World War II in military screenings, the MMPI succeeded in its effort to move personality assessment towards more empirically grounded approaches by unquestionably proving its predictive clinical utility (Butcher & Williams, 2010). Thus, the MMPI supplanted other self-report tests of the era to become the dominant personality assessment tool through the incorporation of an innovative approach to categorizing psychological dysfunction.

The MMPI's comprehensive approach to assessment resulted in prediction by which "no other test was able to compare" (Welsh & Dahlstrom, 1956, p. 17). The MMPI turned the field's

opinions on personality testing to a perspective that embraced personality assessment as essential and clinically useful (Buchanan, 1994). Its methodology and clinical success even rendered the test inscrutable amidst arduous times where public outcry against psychological practice was at a height (Buchanan, 2002). The next half century has served to reinforce the MMPI's popularity as it quickly became the most researched personality measure of its era (e.g., Camara, Nathan, & Puente, 2000; Colligan, 1985).

The techniques employed in the creation of the MMPI have become so central to the interpretive tradition that it is impossible to understand the debate on future structural development without first understanding empirical keying. Much of the MMPI's development hinged on the disregard for the scale development traditions of the era. It elected instead to create this new method (Gough, 1988). Empirical keying was developed as a technique for scale construction in which items were selected solely on the basis of their ability to successfully differentiate group membership. Reliance on this technique was welcomed as supporters did "not have the confidence of the traditional personality-test maker that the relation between the behavior dynamics of a subject and the tendency to respond verbally in a certain way must be psychologically obvious" (Meehl, 1956, p. 8). And to this end, Hathaway has said that no item has been excluded from the MMPI scale creation due to its content since the item's content was secondary to the scale development process (Buchanan, 1994). For example, an item selected for Scale 2 (Depression) would merely have to effectively differentiate known group membership of a non-psychiatrically ill individual from a psychiatric patient with depression while offering discriminant capacity with the other clinical scales. Being able to accomplish this would alone demonstrate criterion sufficiency for MMPI scale inclusion. Thus, items became part of a clinical scale entirely on the basis of their performance, including items with lower face validity if their

ability to differentiate the clinical and non-clinical groups was sufficient. Empirical keying dismissed the importance of active theoretical processes as a basis for item selection. Instead, it relied purely on the identification of an associative relationship that discriminated group membership.

The reliance of the MMPI on an objective approach to differentiating groups provides a way to correct for when clients do not provide accurate information, either intentionally or as a function of pathology (Butcher & Williams, 2011). With Hathaway and McKinley's (1940a, 1940b, 1942) dedicated reliance to actuarial methods, empirical keying became the central method for establishing scales on the MMPI. Although other approaches later augmented this method, such as Welsh's (1956) use of factor analysis or the use of deduction (Butcher et al., 1990; Wiggins, 1966), empirical keying existed as the incontrovertible center of the MMPI. In fact, empirical keying became so widely used that the number of published scales available for use surpassed the number of items on the MMPI (Dahlstrom et al., 1975) resulting in revision of criteria for scale incorporation, which required improved reliability and emphasized again the importance placed on diagnostic accuracy (Butcher & Tellegen, 1978).

Since its publication, the MMPI has undergone two major revisions. The first revision began in 1982 and concluded with the release of the revised form, the MMPI-2 (Butcher, Dahlstrom, Graham, Tellegen, & Kaemmer, 1989). This form included the removal of objectionable item content, the rewriting of several items to increase readability, the renorming of scales, and the development of uniform t-scores. In addition, adolescent norms (previously incorporated into the initial MMPI) were excluded, leading to the later introduction of the MMPI-Adolescent form (Butcher, et al., 1992). Although absolute adherence and blind allegiance to the findings based on empirical keying were already being questioned by ardent

MMPI supporters (Meehl, 1972), the Clinical Scales derived from its use were not refined using alternative statistical methods. The lack of refinement is likely a function of the successful clinical utility of the instrument's scales.

These revisions to the MMPI-2 were all made for the purpose of expanding the clinical applicability of the MMPI and, accordingly, scholars widely welcomed these revisions. Ward (1991) noted that the advances in approach seen within the MMPI-2 corrected well documented errors inherent to the older test. Vincent (1990) commented that the limit of diagnostic accuracy had been reached for the MMPI and that the revisions were likely to address the major and recurrent criticisms of the aging instrument. Even amidst topics of fierce debate, there was seldom disagreement that the revisions of the MMPI-2 provided a better interpretive base and greater opportunity for clinical accuracy (e.g., Humphrey & Dahlstrom, 1995; Munley & Zarantonello, 1990; Tellegen & Ben-Porath, 1996).

Some saw these revisions as somewhat limited, leaving them dissatisfied and yearning to bring to fruition the developmental potential of the MMPI. Helmes and Reddon (1993) noted that the diverse use of the MMPI was likely not predicted when it first premiered since it was designed for a much narrower clinical use. As such, they questioned the appropriateness of its broad use. The MMPI's emphasis on continuity of content and constructive techniques likewise limited its long-term potential. The MMPI's rejection of theory in favor of blind empiricism is problematic when that empiricism has been shown to be flawed (Goldberg, 1971). Issues remaining unresolved in the revisions of the MMPI-2 included problematic criterion groups (Helmes & Reddon, 1993), item redundancy across scales (Norman, 1972), poor performance on complex disorders (Gottesman & Prescott, 1989), and difficulties with incorporating correction indexes (i.e., K correction) into interpretations (Graham, 2006). The MMPI-2 also kept the

problematic factor structure (Costa, Zonderman, Williams, & McCrae, 1985; Reddon, Marceau, & Jackson, 1982) and subtle items (Bagby, Rogers, & Buis, 1992) of its predecessor. This culmination of issues led to the need for another revision aimed directly at addressing the method for assessment and structural form identified as problematic within the MMPI-2. This next revision needed a comprehensive redressing of validity concerns thematic within the MMPI/MMPI-2 (Helmes & Reddon, 1993).

With the introduction of the Revised Clinical (RC) scales for the MMPI-2 in 2003 (Tellegen, Ben-Porath, McNulty, Arbisi, Graham, & Kaemmer, 2003) and the subsequent release of the MMPI-2-Restructured Form in 2008 (MMPI-2-RF; Ben-Porath & Tellegen, 2008a), the development of the MMPI continued to its next stage. The RC scales provided an initial context and a glimpse of the methodological and theoretical underpinnings of the subsequent instrument wide revision. Subsequently, the MMPI-2-RF provided full scale instrument revisions aimed at addressing the long-standing structural and interpretive concerns raised in earlier versions of the MMPI (e.g., Gottesman & Prescott, 1989; Helmes & Reddon, 1993). Given the impact that the MMPI revisions have had on test construction efforts (Rogers & Sewell, 2006), and the fact that the MMPI is a dominant force in personality assessment (Butcher & Williams, 2010), it should come as no surprise that the introduction of the MMPI-2-RF, and its component RC Scales, have received critical attention.

MMPI-2-Restructured Form Development

Acknowledging the importance of the RC scales as the interpretive core of the then developing MMPI-2-RF, Meyer (2006) edited a special edition of the *Journal of Personality Assessment* in order to provide a forum for debate regarding the empirical, theoretical, and rational decisions which led to these scales' development. This special issue invoked an

empirically-based theoretical discussion about the merits of the decisions involved in the revisions leading to the MMPI-2-RF. Much of this debate centered on the structural changes to the instrument and how these changes impacted clinical use. Given the importance that these decisions had in understanding capacities for scale interpretation on the MMPI-2-RF, developmental decisions leading to this revised form are presented here briefly. For a more complete review, readers should refer to the monograph by Tellegen and colleagues (2003).

RC Scale Development. Recognizing the long-established first factor problem (i.e., a pattern of global scale covariation) of the MMPI/MMPI-2 (e.g., Wiggins, 1973), Tellegen and colleagues (2003) undertook the task of parsing out the overshadowing response pattern associated with all MMPI profiles. Their goal was not only to remove the bias frequently influencing clinical scale elevations but also to reduce scales to their singular core- a move in line with modern testing approaches (Nunnally & Bernstein, 1994). To accomplish the task of removing interpretive bias and simplifying assessed constructs, Tellegen and colleagues theorized that the issue causing the greatest interpretive difficulty in the MMPI-2 Clinical Scales was a general sense of distress shared by measurement of all mental illnesses.

Reasoning that this first-factor was characterized by the highest level of affect, the pleasant-versus-unpleasant axis of Watson and Tellegen's (1985) mood model, a seed scale measuring was formed through a joint factor analysis of items on Clinical Scales 2 (Depression) and 7 (Psychasthenia). Scales 2 and 7 were selected as the basis for this seed scale because of the conceptual similarity of those scales to a hedonic, negative emotionality associated with the pleasant/unpleasant emotional axis (Watson & Tellegen, 1985). Tellegen and colleagues termed this seed scale's content *demoralization* which represented a broad experience of negative emotionality. Items from this demoralization seed scale were then temporarily added to each

MMPI-2 Clinical Scale in order to weight the scales in such a way that additional items associated with demoralization could be identified during subsequent factor analysis. Items on each of the Clinical Scales that also had high loadings on demoralization were removed. As demoralization items were identified and extracted from the item pool of the Clinical Scales, the shared variance between scales causing the first-factor issue was reduced. Remaining items for each Clinical Scale were finalized through another factor analysis aimed at ensuring that a singular structure emerged from each parent Clinical Scale. Also as part of the Clinical Scale revisions, each factor analysis included not only the Clinical Scales but also all 338 items intended for inclusion into the MMPI-2-RF to broaden the potential content covered by the scales and ensure a more comprehensive interpretation (Ben-Porath, 2012b). The result of these multiple iterations of factor analytic methods was the formation of the Revised Clinical (RC) scales. Each RC scale represents a single construct drawn from its parent Clinical Scale, without an inflated general factor of distress. The demoralization items were also finalized and included as a separate, standard component of the MMPI-2-RF called RCd; this demoralization content has been described as an over-arching affective saturation which has long influenced MMPI scale elevations (Ben-Porath, 2012b).

RC Scale Construction Criticisms. Despite this empirically sophisticated approach, the methodology has been repeatedly questioned because it dilutes the richness and utility of the MMPI (Butcher & Williams, 2010). Butcher, Hamilton, Rouse, & Cumella (2006) argued that the removal of demoralization from the Clinical Scales is problematic as this content represents important diagnostic information and not simply measurement error impeding scale interpretation. Thus, the unitary nature of the RC scales has been said to have abandoned the historic structure of MMPI scale interpretation. Butcher and colleagues point to changes in the

interpretation of RC3 (Cynicism) over Clinical Scale 3 (Hysteria; HY) as a prime example of this dilution. Similarly, Caldwell (2006) objected to the removal of RCd from the Clinical Scales noting that attempts to separate it may be clinically useless as it overly constricts interpretation by relying heavily on factor analysis since, “a scale built of highly intercorrelated items is inevitably narrow (p. 195).” Expanding these points, Nichols (2006) noted that Tellegen and colleagues’ (2003) methods failed to capture the “syndromal complexities” inherent to the original Clinical Scales through a process Nichols terms “construct drift.” To Nichols, syndromal complexity is the inherent multi-dimensionality of clinical disorders, and construct drift is a shift from historical conceptualizations of these disorders. Nichols points specifically to RC3 (Cynicism) saying that the exclusion of items assessing somatic concerns renders its interpretation flat and contextually meaningless. Nichols argued that RC3’s rejection of contextual richness equated to the abandonment of its usefulness given interpretive approaches to Clinical Scale 3 that emphasized those symptoms. Critics of the RC scales view this interpretive shortcoming as compounded since the RC scales have attempted to hit only “soft targets” of empirical correlations (Nichols, 2006, p. 127), emphasizing only constructs already measured elsewhere on the MMPI-2 (Rouse, Greene, Butcher, Nichols, & Williams, 2008).

Rogers and colleagues (2006) provide a slightly different criticism of the RC scale development. While they note the scale reduction process eliminated clinically important elements (i.e., subtle items) as others have identified, issue was also taken with Tellegen et al.’s efforts to incorporate Jackson’s (1970) methodologies. Rogers and colleagues suggest that while these efforts are laudable, the RC development effort lands short by not maintaining fidelity to Jackson’s sequential system of scale development.

Taken together, these criticisms produce three thematic concerns regarding the construction of the RC scales: (a) the separation of RCd from the Clinical Scales, even if successful, may be clinically useless; (b) the dilution of the Clinical Scales to the “construct drifted” RC scales loses the essential and diagnostic syndromal complexity; and (c) scale revision methodologies do not mirror previously employed approaches on which the RC scales were based. The thematic crux of these three criticisms is an assertion that the RC scales have forgone the rational and diagnostic purpose of their predecessors through a process of misguided scale over-simplification. Responses to these three criticisms have been made and will be summarized below.

RC Scale Construction Defense. While critics of the RC approach to the first factor problem have suggested its failures by narrowing the measured construct to a point of clinical uselessness (e.g., Caldwell, 2006), they have also acknowledged its historical precedent and have suggested that such an approach is worthy of consideration (Nichols, 2006). Despite their alleged failures, research has shown that the RC scales produce clear relationships to psychopathology (Hoelzle & Meyer, 2008; Simms, Casillas, Clark, Watson, & Doebbeling, 2005; van der Heijden, Egger, Rossi, Grundel, & Derksen, 2013) and mirror national prevalence rates (Tarescavage et al., 2013). Moreover, even the methods (such as Welsh’s, 1956 Anxiety or Repression) suggested by Nichols as viable alternative to Tellegen and colleagues (2003)’s solution to the first factor problem have shown a .95 correlation to RCd (Weed, 2006). This suggests that despite urgings to conceptualize the first factor in an alternative way, doing so would have not produced strikingly different results. In a refreshing departure from the large sample data-based arguments predominating this debate, Finn and Kamphuis (2006) provided several case examples

in which demoralization's removal, and RCd's inclusion, resulted in a cleaner and a more interpretable MMPI profile.

Cogent responses to concern over construct drift have been made by Tellegen et al (2006) and Weed (2006) in two ways. First, while Nichols (2006) believes that only items initially comprising the Clinical Scales can act as measures of it the scale's essential core, Tellegen and colleagues offer the perspective that the reshuffling of items from multiple scales has simply reinforced the intended core. Taking Nichols' RC3 example in which 80% of the items appear on the cynicism (CYN) MMPI-2 Content Scale, Tellegen and colleagues note that this overlap is not bi-directional and that only 52% of CYN content appears on the RC3 scale. The remaining CYN items represent self-referential concerns and fit better, empirically and rationally, to RC6. Exemplifying this argument, Ingram, Kelso, and McCord (2011) have noted that RC3 requires a uniquely new contextual metric from which interpretation may be drawn despite its similarity with previous scales. Second, the revision of concepts in multidimensional scales are inevitable over time and, as Weed argues, construct drift may better be described as construct sharpening. Of syndromal complexity for so-called subtle items, Weed notes that these are just poorly performing items which have done little psychometrically. This is likely, in part, why the RC scales have produced strikingly better reliability, as well as convergent and discriminant estimates, than the traditional Clinical Scales (Ben-Porath & Tellegen, 2008b; Simms et al., 2005), and why the MMPI-2's Clinical Scales were historically considered psychometrically suboptimal (Helmes & Reddon, 1993).

Over concerns of fidelity to Jackson's (1970) methods, Tellegen and colleagues (2006) write that "it was not our intent to simply adopt Jackson's (1970) sequential scale construction method... [because of] its requirement that the test developer start with a fully formulated and

collaborated trait model (p. 149).“ In contrast, the RC scales were derived before being further refined using Jacksonian approaches. Thus, although Jackson’s work provided inspiration, the approach employed for the development of the RC scales was one in which blind empiricism was rejected in favor of a balanced theoretical and empirical approach. In line with this, Meehl (1972) has even remarked that the blind allegiance to empirical keying that he held as an author of the MMPI was misguided and an over-representation of the statistical capability of such blind empiricism. Meehl went on to state that the ongoing development of the MMPI will likely incorporate other statistical approaches as the primary techniques for scale development, specifically commenting on factor analysis’ potential. This serves to highlight that critiques of the MMPI-2-RF’s development, which often focus on scale construction methodology, are viewed as misguided techniques represent the long projected future development of the MMPI.

Higher-Order Content of the MMPI-2-RF. The revisions of the RC scales identified their core content and were the first step to establishing cleaner and more concise clinical interpretations. Contemporary approaches to use of the MMPI-2 place emphasis on interpretation of select scale code types, or pattern of scales elevation, dependent upon the respondent’s most pronounced and distinctive elevations. These patterns of elevations are viewed as the descriptive core of the clients’ clinical presentation. Code types are defined as a numerical indicator of the highest Clinical Scale elevations (typically the highest two or three Clinical Scales) on a given profile and are used to describe types of respondents and their subsequent pathologies (Greene, 2000; Graham, 2006). Based upon supposition that this technique provides a diagnostically richer and clinically more useful approach, this method to conceptualization argues that it captures more nuanced and thematic elements inherent to an individual respondent’s presentation.

With the release of the MMPI-2-RF, previous interpretive guidelines and research about the Clinical Scales are gone and there is a dearth of information available guiding interpretation (Butcher, 2010). Although scale equivalence was established between the MMPI and the MMPI-2 code types (Dahlstrom, 1992), the drastic shifts in the scale content on the MMPI-2-RF make efforts to equate code types to this new interpretive context impossible. Not only have the meaning of single scales shifted (e.g., Ingram et al., 2011), so too has the ability to meaningfully combine those scales into code types - an abandonment not seen in the transition to the MMPI-2. The loss of these code types as an interpretive base has been regarded as one of the most difficult aspects of incorporating the MMPI-2-RF into clinical use (Rogers et al., 2006) because code types have a richly defined interpretive history.

However, code types also have substantial limitations that warrant interpretive transition. Code types are exclusive to one another and, thus, it is not possible for competing clinical concerns to be equally represented in a single code type despite concurrence. For example, it is impossible to describe someone as having both a 27-72 code type and a 49-94 code type despite the possibility for dysfunction to present in an individual in both of these areas (Ben-Porath, 2012b). While code types could be argued to exist as conceptual starting points from which individual scale elevations drive the final clinical profile, doing so mirrors the interpretive approach employed by the MMPI-2-RF (i.e., encouraging the clinician to incorporate individual scales into a unified profile distinct to each individual).

Likewise, early factor analyses of the MMPI have shown that using alternative rotation techniques could produce higher order scales on the MMPI that would bring its interpretive structure in line with other personality tests (e.g., Kassebaum, Couch, & Slater, 1959). Consistent with *Overcontrolled/internalizing* and *Undercontrolled/externalizing* dimensions identified

within other popular personality instruments (e.g., Achenbach & Edelbrock, 1978; Krueger & Tackett, 2003), these two dimensions of higher order symptoms have consistently emerged on the MMPI (Blais, 2010). However, until the revisions of the MMPI-2-RF, these broad dimensions had not been formally incorporated. In addition to internalizing and externalizing psychopathologies, thought disordered thinking has also been long conceptualized as a thematic pattern of pathology and a core component of the MMPI. Using the RC scales' clarified content, a higher-order thought dysfunction scale emerged with a distinct structure which emphasized paranoid thinking and aberrant experiences (Ben-Porath, 2012b). Thus, the MMPI-2-RF has implemented a new contextual framework by developing higher order structural measures, composed of RC content that represents broad levels of pathology and that enables a hierarchical MMPI-2-RF interpretation. By doing so, the MMPI-2-RF has become more theory guided and has aligned itself with our current understandings of broad psychopathological presentation.

These three Higher Order (HO) constructs were derived to test the feasibility of measuring three distinct areas of dysfunction concurrently through a hierarchical interpretive design (Ben-Porath, 2012b). Consistent with past interpretative approaches to the MMPI (Meehl, 1946), the three HO constructs described above (i.e., internalizing, externalizing, and thought dysfunction) were able to be derived empirically in a way congruent with theory. Using the combined content from all RC scales, these three distinctive dimensions emerged after analysis of rotated factor solutions (Tellegen & Ben-Porath, 2008/2011). They were labeled Emotional/Internalizing domain (EID), Thought Dysfunction domain (THD), and Behavioral/Externalization domain (BXD) with each representing a common code type (27-72, 68-86, and 49-94 respectively). Independent factor analyses have indeed confirmed that RC6 and RC8 act as the primary markers of the thought disorder dimension (Hoelzle & Meyer, 2008;

Sellbom, Ben-Porath, & Bagby, 2008), that internalizing content was measured best by RC2, RC7, and RCD, and that the externalizing dimension of personality was captured by content on RC4 and RC9 (Sellbom et al., 2008).

Outside of factor analyses used to confirm the proposed HO structures, very little work has been done to examine the interpretive contexts of the higher order scales. While some external correlates have been provided (Tellegen & Ben-Porath, 2008/2011), there have been no explorations detailing the typical, within MMPI-2-RF presentation of clients with elevations across code types. Ben-Porath (2012, p.106) concludes the section on interpreting HO Scales in his book *Interpreting the MMPI-2-RF* by saying merely that the empirical correlates of the HO scales are, "...quite consistent with those identified previously for the 27-72, 68-86, and 49-94 Clinical Scale code types, respectively, providing a dimensional measurement perspective on these clinically relevant phenomena." That is to say, such little work has been done to examine the clinical utility of the HO Scales that MMPI-2-RF interpretation should be based on dated profiles drawn from scales which no longer exist in the MMPI-2-RF. Although the HO scales are able to align with current theoretical perspectives on psychopathology and personality development (e.g., Krueger & Tackett, 2003), very little guides their interpretation. This lack of work on higher order structures is curious since the hierarchical nature of the MMPI-2-RF, with scales covering both the very broad and the very narrow clinical issues, lends itself to the development of comprehensive clinical profiles aligned with these higher order themes.

The refinement of higher order clinical presentation seems particularly important given that the MMPI-2-RF has proposed its structure as one that captures the broadest patterns in pathology. While the HO scales were derived largely from RC content, their conceptualizations are also related to other substantive scales of the MMPI-2-RF. These HO scales have been

suggested as mere indicators of these three broad patterns of pathology, albeit strong representations because their content was drawn largely from the RC scales. HO scales are said to be only one representation of the broader pathology that they measure (Ben-Porath & Tellegen, 2008a, p.22). Although the HO scales were created as distinct measures of internalizing, externalizing, and thought disordered behavior with content drawn only from the RC scales, these three patterns of pathology are also able to be conceptualized as latent constructs which are inclusive of all scale types lending further credence to the interpretive shift required in the adoption of the MMPI-2-RF.

MMPI-2-RF Substantive Scales. In addition to the development of the RC scales, the MMPI-2-RF also incorporates a number of other content-based scales intended to measure more specific areas of clinical concern. The substantive content were developed as part of the Specific Problem and Interest (SP/IS) scales or the revised personality psychopathological-five (PSY-5) scales (Ben-Porath, 2012b), with each offering unique additions to the MMPI-2-RF framework already in place following the RC scales.

The content of the SP and IP scales was identified in one of four ways and refinement methods differed slightly across scales given the differential necessity of researcher judgement (Ben-Porath, 2012b). However, the goal of the SP/IP scales, as with the RC scales, remained focused on the development of empirically validated clinical additions to the MMPI-2-RF. The first way that SP/IP content identified was through the inclusion of content inherent to the Clinical Scales that was excluded from RC revision efforts due to a lack of focus on pathology from RC revision (i.e., Clinical Scale 5 and 0). The second method of identifying SP/IP scales was by refining the excessive component of the heterogeneous MMPI-2 Clinical Scales which were identified, and excluded, from RC development. The third approach taken was the

broadening of multifaceted RC scale content. Lastly, other clinically important aspects that thematically appeared in the item pool were targeted for scale development. For instance, some items in the item pool are related to suicidality, which posed to offer important clinical information to the MMPI-2-RF profile.

Using an iterative approach to development, the SP and IP scales underwent repeated validation cycles and included multiple revisions based on empirical criteria and expert opinion (Ben-Porath, 2012b). As part of these revisions, the SP and IP scales widened from an initial pool of 14 content areas to its final form of 25 scales which are conceptually grouped to measure symptoms related to reports of: (a) somatic and cognitive complaints, (b) internalizing or (c) externalizing symptoms, (d) interpersonal distress, and (e) interests (Tellegen & Ben-Porath, 2008/2011). In conjunction with a substantial number of external correlates (Tellegen & Ben-Porath, 2008/2011) and some promising work showing their incremental value in clinical decisions (e.g., Arbisi, Polusny, Erbes, Thuras, & Reddy, 2011), the SP/IS scales tend to have moderate correlations between one another and with their intended RC parent scale (Ben-Porath, 2012b). Likewise, with only a few of the shorter scales as exceptions, the SP/IS scales have acceptable reliability estimates (Tellegen & Ben-Porath, 2008/2011) consistent with, or above, those typical of psychological measurement research (Peterson, 1994).

The PSY-5 were initially created as components of the MMPI-2 and were intended to describe the dimensional structure of personality disordered individuals (Harkness & McNulty, 1994; Harkness, McNulty, & Ben-Porath, 1995). Often described as having constructs similar to the Five Factor Model (e.g., Bagby, Sellbom, Costa, & Widiger, 2008; Ingram et al., 2011), the PSY-5 dimensions were developed separately but align neatly. However, possessing more than mere incremental value (Bagby et al., 2008), the PSY-5 have consistently shown a great deal of

predictive capacity (Sellbom, Graham, & Schenk, 2006). Unsurprisingly, this capacity and utility led to their revision (Harkness & McNulty, 2007) and subsequent inclusion within the MMPI-2-RF. Building upon their earlier method of replicated rational selection (i.e., a deductive process in which items are selected by trained judges based upon the availability within an item pool and then retained only by agreement of the judges reaches a given threshold), the PSY-5 revised form (PSY-5-RF) scales salvaged a majority of the items from its MMPI-2 form and then underwent reduction efforts in order to reduce cross-correlation and increase scale coherence (Harkness, McNulty, Finn, Reynolds, Shields, & Arbis, 2014). The results of the PSY-5 revisions were a set of scales that were largely congruent with the constructs included in the MMPI-2.

Clashing Paradigms Guide Theory

The MMPI-2-RF differs in its primary construction approach from the MMPI-2; it abandoned empirical keying and emphasizing heterogeneous and higher order scale conceptualization. In doing so, the MMPI-2-RF relies on disparate techniques from those integral to older versions of the MMPI. While scales have been constructed before in these manners (e.g., Harkness & McNulty, 1994; Jackson, 1970; Walsh, 1959) and have found favorable inclusion and common use in the interpretative tradition of the MMPI (Graham, 2006), the critical distinction of the MMPI-2-RF and its predecessors is its sole reliance on these contemporary assumptions. However, the MMPI-2-RF also retained much of the core content and scale structure by simply revising most validity and clinical scales using the same items and normative sample. Thus, the MMPI-2-RF offers a recapitulated and contemporary version of the historic instrument. Its core is retained through measurement of the same thematic aspects of psychopathology through different techniques. Despite maintaining important congruence with

earlier versions of the MMPI, these changes altered interpretation in important ways. Not merely additive components, the changes in scale construction for the MMPI-2-RF replaced previous approaches understood as integral in psychological testing. With these elements no longer included, the ingredients that served to make psychological testing respected (Colligan, 1985) and psychologists successful (Buchanan, 1994) are absent in the future development of the MMPI.

As a result, debate over acceptance of the MMPI-2-RF (Butcher & Williams, 2009, 2010, 2012; Ben-Porath & Flens, 2012) and its component RC scales (Bolisky & Nichols, 2011; Bowden, White, Simpson, & Ben-Porath, 2014; Rouse et al., 2008; Scholte et al., 2012; Wolf et al., 2008) has widely occurred. Scholars of the MMPI are aligning themselves in diametric ways, noting the crossroads and critical choice facing psychologists and psychological assessment. Views of this stage of development as critical are the result of the MMPI-2's nature as the vanguard of clinical assessment practices (Rogers & Sewell, 2006). Given the MMPI's role in the development of assessment practices, contextual understanding of common views about the revisions of the MMPI-2-RF provide insight into contemporary issues underscoring approaches to assessment and scale development. Since the MMPI has played a critical role in the establishment of professional roles of psychologists (Buchanan, 1994), perspectives on the MMPI's development will likely lead to broad understanding of past interpretive traditions (Benjamin, 2005) as well as future directions for the field.

Persistently throughout the development of a science, challenges to beliefs will occur in such a manner as to require substantive re-evaluation of the field's underlying assumptions (Kuhn, 1961). These periods of increased dissonance offer opportunity to address previous theoretical shortcomings, methodological inadequacies, or accepted mythos. Perspectives on the

revisions inherent to the MMPI-2-RF offer such a challenge given its developmental lineage (Benjamin, 2005; Buchanan, 1994) and the revisions offer challenges to important areas within applied psychology. Perspectives on its assumptive use offer a guide to progress and profitable future opportunities for the field (Meyer, 2006). Thus, while other empirical methods and theory-driven models have been used before within the MMPI, the complete abandonment implores critical questions about conceptualization of modern testing and the future of psychological practice (Rogers & Sewell, 2006). Most important are the issues of if psychology has developed enough to enable the use of a theory to drive our broad conceptualization of people and if we are able to trust responses provided during testing to face valid items.

The proponents of the MMPI-2-RF, by emphasizing hierarchical interpretation and rationale scale development, support psychology as entering a phase in which it more accurately understands the gestalt of human behavior. Proponent belief are exemplified by a balanced approach of theoretical grounding and empirical construction that still assesses expected differences between clinical groups with substantial psychometric support. Conversely, MMPI-2 proponents that emphasize the need to retain it (Butcher & Williams, 2010) and its dated methods and interpretive tradition (Finn & Kamphuis, 2006) bring perspective of the field that emphasizes a need for continued differentiation focused more on divergent, rather than convergent, criteria (Rogers & Sewell, 2006). This perspective embraces the idea that the theory is not yet wholly sufficient to over-rule pure empiricism. Meehl's (1972) commentary on actuarial assessment emphasizes this approach; the differentiation of groups is not only important but also requisite to creating instruments capable of diagnostic classification. This emphasis on differentiation is considered critical because doubt has long been cast on the trustworthiness of responses to theory driven items (e.g., Meehl, 1958). Thus, the perspective on the strengths and

utility of these two instruments reflect a broad belief about the current state of psychology's scientific development and our capacity to incorporate theory in a manner that rejects pure empiricism.

The capacity for a dominant paradigm to emerge is required for development and advancement as a science (Kuhn, 1961). A consensus to diagnostic approach, identification of causal mechanisms for pathology, and agreement on definitions of psychological health are needed to create this unification in psychology. These anchors could guide determination about the current state of psychology in its capacity to conceptualize people and form assessments of them. To accurately theorize and assess psychopathology, as the MMPI-2-RF proposes, consensus on the nature and causes of behaviors must exist. Such a consensus must be broad and accurate enough to generalize not just between people, but also between disciplines (Slife & Williams, 1995). Before the revisions of the MMPI-2 and the MMPI-2-RF, Jackson (1972) suggested a need for theoretical inclusion as part of instrument design. Despite being criticized for not embracing a preeminent foundation composed of psychological theory as Jackson had recommended (Rogers et al., 2006), Tellegen and colleagues (2006) note that they have embraced a balanced perspective of theory and empiricism. This involvement of theory is an improvement upon the MMPI-2's intentionally non-theoretical chassis. This step towards a sweeping embrace of theory as a critical, instrument wide aspect of personality measures provides evidence of contemporary perspectives of the field.

Scale construction approaches are used and wed to the era and circumstance in which they live. They are inseparable from the theoretical assumptions and statistical aptitudes of their time. Each progressive step in measurement technique represents the capacity of the field within the confines of the existing scientific paradigm (Kuhn, 1961). The aging MMPI-2 is no exception

(Rogers Sewell, Harrison, & Jordan, 2006). The revisions to the MMPI-2-RF suggest that the changes in methodology (i.e., abandoning empirical keying) and structure (i.e., incorporating a theoretically-based higher order structure) demonstrate a shifting of beliefs about assessment methodology. While the MMPI-2-RF is not the first to introduce theory as critical for test construction (Clark & Watson, 1995), it is the first iteration of the MMPI to reject entirely its exclusion. The friction and dissonance inherent to the revisions of the MMPI-2-RF provide an evaluative lens on the degree to which aged theoretical and methodological assumptions have engrained themselves within the field of psychology.

Similar outcry over revisions have been noted before in the MMPI (Greene, 1991), suggesting that the MMPI-2-RF is merely the latest catalyst for the ongoing discussion on the evolving assumptions of the field. Failure to recognize the limitations of dated measurement techniques may hamper the evolution of assessment practices started long ago by the MMPI (e.g., Colligan, 1985). The evolution of assessment practices has resulted in increased integration of theoretical rationales during scale development. The MMPI-2-RF is the result of recent testing adaptation that emphasizes the importance of theoretical congruence in assessment that has long been recognized (e.g., Jackson, 1972). Similar movement towards theoretical integration has been occurring concurrently within clinical practice, to the point that it is now openly embraced (Goldfriend, Packankis, & Bell, 2005; Norcross & Goldfried, 2005).

The concurrent integration of perspectives across both treatment and assessment techniques requires unification of belief about motivations, symptom presentation, and behavioral reporting. This bridging of behavior to higher order understanding is necessary for integrative practice; these requisite elements are also critical in the development of hierarchical and theoretically driven assessment instruments. Such growth of integrative movements supports

the notion that the assumptions made within the MMPI-2-RF's revisions (i.e., that assessment should be theoretically involved and contain higher order pathology) are timely and suited to the current era of psychological practice. Broad understandings of human behavior are well aligned with the trajectory of the field. Movement towards theoretical integration in assessment may not be as far along as had been envisioned (Jackson, 1970) but is becoming an increasing aspect of the developmental history of the MMPI. The revisions inherent to the MMPI-2-RF hold promise that the MMPI stays relevant and a leader in practices of assessment (Graham, 2015). As Rogers and Sewell (2006) note of the MMPI-2-RF revisions as a critical point in assessment's history, "looking backward, we see a proud tradition and a sound measure... looking forward, we see exciting developments" (p. 178).

As considerations are given to the role of the MMPI-2 and MMPI-2-RF in future of psychological testing, the field should consider the underlying assumptions made by each of these respective instruments. They need to be considered in relation to the integrative trends occurring within clinical practice. The MMPI-2-RF provides a basis for understanding individuals in a manner that aligns with the predominant beliefs of the field. It offers unification and theoretically-driven implications while supplanting good psychometrics with even stronger evidentiary support. The question of assessment psychology possessing the theoretical grounding and empirical techniques needed to predict broad patterns of behavior using empirically testing has been answered. The MMPI-2 provided innovation (Benjamin, 2005; Buchanan, 1994), but those contributions have aged and novelty is required for continued advancement of assessment practices to stay current with perspectives held by professional psychology. Indeed, the MMPI-2-RF aligns to the current paradigm of clinical practice

However, the progress of assessment practices epitomized by the revisions of the MMPI-2-RF represents the current station of psychology's development as a science. However, improved as it may be in its interpretive capacity and alignment with current theoretical approaches, the MMPI-2-RF will age as well; it will require adaption in time as well. Integrative approaches, while increasingly common, are not yet fully grounded and advancements are still needed to expand our ability to conceptualize and incorporate higher order understandings of pathology. Integration is a direction in which the field is growing, but progress towards integration has not been completed. While the capacity of psychology to form a coherent single professional theory has not fully emerged, the steps of the MMPI-2-RF are laudable and serve to represent advances within clinical practice. Continued advancements of theory and measurement techniques are needed in future iterations of the instrument (Graham, 2015). The MMPI-2-RF is likely to see greater acceptance as continued movement toward theoretical integration occurs for personality assessment instruments. Such is the necessity of psychology's development as a science and the MMPI as an instrument and lens for the field.

References

- Arbisi, P.A., Polusny, M.A., Erbes, C.R., Thuras, P., & Reddy, M.K. (2011). The Minnesota Multiphasic Personality Inventory–2 Restructured Form in National Guard Soldiers Screening Positive for Posttraumatic Stress Disorder and Mild Traumatic Brain Injury. *Psychological Assessment, 23*, 203-214. doi: 10.1037/a0021339
- Ben-Porath, Y.S. (2012). *Interpreting the MMPI-2-RF*. Minneapolis, MN. University of Minnesota Press.
- Ben-Porath, Y., & Tellegen, A. (2008a). *Minnesota Multiphasic Personality Inventory–Restructured form: Manual for administration, scoring, and interpretation*. Minneapolis: University of Minnesota Press.
- Ben-Porath, Y., & Tellegen, A. (2008b). Empirical correlates of the MMPI-2 Restructured Clinical (RC) scales in mental health, forensic, and nonclinical settings: An introduction. *Journal of Personality Assessment, 90*, 119-121. doi: 10.1080/00223890701845120
- Brown, T .A. (2006). *Confirmatory factor analysis for applied research*. New York: The Guilford Press.
- Buchanan, R.D. (1994). The development of the Minnesota Multiphasic Personality Inventory. *Journal of the History of the Behavioral Sciences, 30*, 148–61. doi:10.1002/1520-6696(199404)30:23.0.CO;2-9.
- Buchanan, R.D. (2002). On not “giving psychology away”: The Minnesota Multiphasic Personality Inventory and public controversy over testing in the 1960s. *History of Psychology, 5*, 284-309. doi: 10.1037//1093-4510.5.3.284
- Butcher, J.N. (2010). Personality assessment from the nineteenth to the early twenty-first

- century: Past achievements and contemporary challenges. *Annual Review of Clinical Psychology*, 6, 1-20. doi: 10.1146/annurev.clinpsy.121208.131420
- Butcher, J. N. (2011). *MMPI-2: A beginner's guide* (Third edition). Washington, DC: The American Psychological Association.
- Butcher, J. N., Dahlstrom, W.G., Graham, J. R., Tellegen, A., & Kaemmer, B. (1989). *Manual for the restandardized Minnesota Multiphasic Personality Inventory: MMPI-2: An administrative and interpretive guide*. Minneapolis: University of Minnesota Press.
- Butcher, J.N., & Williams, C.L. (2010). Personality assessment with the MMPI-2: Historical roots, international adaptations, and current challenges. *Applied Psychology: Health and Well-being*, 1, 105-135. doi: 10.1111/j.1758-0854.2008.01007.x
- Caldwell, A.B. (2006). Maximal measurement or meaningful measurement: The interpretive challenges of the MMPI-2 Restructured Clinical (RC) scales. *Journal of Personality Assessment*, 87, 202–210. doi: 10.1207/s15327752jpa8702_09
- Finn, S.E., & Kamphuis, J.H. (2006). The MMPI-2 restructured clinical (RC) scales and restraints to innovation, or “What have they done to my song?” *Journal of Personality Assessment*, 87, 202–210. doi: 10.1207/s15327752jpa8702_10
- Gibby, R.E., & Zickar, M.J. (2008). A history of the early days of personality testing in American industry: An obsession with adjustment. *History of Psychology*, 11, 164-184. doi: 10.1037/a0013041
- Goldfried, M.R., Packankis, J.E., & Bell, A.C. (2005). A history of psychotherapy integration. In J.C. Norcross & M.R. Goldfried (Eds.), *Handbook of psychotherapy integration* (2nd ed., pp. 24-60). New York: Oxford University Press.
- Graham, J.R. (2015, June). MMPI Retrospective. 2015 MMPI-2-RF, MMPI-2, MMPI-A

- Symposium. Minneapolis, Minnesota.
- Greene, R. L. (2000). *The MMPI-2: An interpretive manual* (2nd ed.). Needham Heights, MA: Allyn & Bacon.
- Graham, J.R. (2006). *MMPI-2: Assessing personality and psychopathology* (4th ed.). New York: Oxford.
- Harkness, A.R., McNulty, J.L., Finn, J.A., Reynolds, S.M., Shields, S.M. & Arbis, P.A. (2014). The MMPI-2-RF Personality Psychopathology Five (PSY-5-RF) Scales: Development and Validity Research, *Journal of Personality Assessment*, 96, 140-150, doi: 10.1080/00223891.2013.823439
- Hathaway, S.R., & McKinley, J.C. (1940a). A multiphasic personality schedule (Minnesota): I. Construction of the schedule. *Journal of Psychology*, 10, 249-254.
- Hathaway, S.R., & McKinley, J.C. (1940b). A multiphasic personality schedule (Minnesota) III. The measurement of symptomatic depression. *Journal of Psychology*, 14, 73-84.
- Hathaway, S.R., & McKinley, J.C. (1942). *The Minnesota Multiphasic Personality Schedule manual*. Minneapolis, MN: University of Minnesota Press
- Ingram, P.B., Kelso, K., & McCord, D.M. (2011). Empirical Correlates and Expanded Interpretation of the MMPI-2-RF Restructured Clinical Scale 3 (Cynicism). *Assessment*, 18, 95-101. doi: 10.1177/1073191110388147
- Jackson, D. N. (1970). A sequential system for personality scale development. In C. D. Spielberger (Ed.), *Current practices in clinical and community psychology* (Vol. 2, pp. 60-96). New York: Academic.

- Marsh, H.W., Ludtke, O., Muthen, B., Asparohov, T., Morin, A.J., Trautwein, U., & Nagengast, B. (2010). A new look at the big five structure through exploratory structural equation modeling. *Psychological Assessment, 22*, 471-491. doi: : 10.1037/a0019227.
- Meehl, P.E. (1946). Profile analysis of the Minnesota Multiphasic Personality Inventory in differential diagnosis. *Journal of Applied Psychology, 30*, 517-524. doi: 10.1037/h0062318
- Meehl, P.E. (1972). Reactions, reflections, projections.(In J. N. Butcher (Ed.), *Objective personality assessment: Changing perspectives* (pp. 131—183). New York: Academic Press.
- Meyer, G.J. (2006). Special Issue: The MMPI-2 restructured Clinical Scales Introduction. *Journal of Personality Assessment, 87*, 119-120.
- Nichols, D. (2006). The trials of separating bath water from baby: A review and critique of the MMPI–2 Restructured Clinical Scales. *Journal of Personality Assessment, 87*, 121–138.
- Norcross, J. C., & Goldfried, M. R. (Eds.). (2005). *Handbook of psychotherapy integration* (2nd ed.). New York: Basic Books
- Peterson, R.A. (1994). A meta-analysis of Cronbach’s Coefficient Alpha. *Journal of Consumer Research, 21*, 381-391.
- Rogers, R., & Sewell, K.W. (2006). MMPI-2 at the Crossroads: Aging technology or radical retrofitting? *Journal of Personality Assessment, 87*, 175-178. doi: 10.1207/s15327752jpa8702_06
- Rogers, R., Sewell, K.W., Harrison, K. S., & Jordan, M. J. (2006). The MMPI–2 Restructured Clinical Scales: A paradigmatic shift in scale development. *Journal of Personality Assessment, 87*, 139–147.

- Rouse, S.V., Greene, R.L., Butcher, J.N., Nichols, D., & Williams, C.L. (2008). What do the MMPI-2 Restructured Clinical Scales reliably measure? Answers from multiple research settings. *Journal of Personality Assessment*, 90, 435-442. doi: 10.1080/00223890802248695
- Schönbrodt, F.D., Perugini, M. (2013). At what sample size do correlations stabilize. *Journal of Research in Personality*, 47, 609-612. doi: 10.1016/j.jrp.2013.05.009
- Schotle, W., Tiemens, B., Verheul, R., Meerman, A., Egger, J., & Hutschemaekers, G. (2012). The RC scales predict psychotherapy outcomes: The predictive validity of the MMPI-2's restructured clinical scales for psychotherapeutic outcomes. *Personality and Mental Health*, 6, 292-302. doi: 10.1002/pmh.1190
- Sellbom, M., Ben-Porath, Y.S., & Bagby, R.M. (2008). On the hierarchical structure of mood and anxiety disorders: confirmatory evidence and elaboration of a model of temperament markers. *Journal of Abnormal Psychology*, 117, 576-590. doi: 10.1037/a0012536.
- Simms, L. J., Casillas, A., Clark, L. A., Watson, D., & Doebbeling, B. N. (2005). Psychometric evaluation of the Restructured Clinical scales of the MMPI-2. *Psychological Assessment*, 17, 345-358.
- Simms, L.J. (2006). Bridging the divide: Comments on the restricted Clinical Scales of the MMPI-2. *Journal of Personality Assessment*, 87, 211-216.
- Slife, B. D., & Williams, R. N. (1995). What's behind the research: Discovering hidden assumptions in the behavioral sciences. London: Sage.
- Tarescavage, A.M., Marek, R.J., Finn, J.A., Hicks, A., Rapier, J.L., & Ben-Porath, Y.S. (2013).

- Minnesota Multiphasic Personality Inventory-2-Restructured Form (MMPI-2-RF)
Normative Elevation Rates: Comparisons With Epidemiological Prevalence Rates. *The Clinical Neuropsychologist*, 27, 1106-1120. doi: 10.1080/13854046.2013.83286
- Tellegen, A., & Ben-Porath, Y.S. (1992). The new uniform T-scores for the MMPI-2: Rationale, derivation, and appraisal. *Psychological Assessment*, 4, 145-155. doi: 10.1037/1040-3590.4.2.145
- Tellegen, A., Ben-Porath, Y.S. (1993). Code-type comparability of the MMPI and MMPI-2: Analysis of recent findings and criticisms. *Journal of Personality Assessment*, 61, 489-500.
- Tellegen, A., Ben-Porath, Y. S., McNulty, J. L., Arbisi, P. A., Graham, J. R., & Kaemmer, B. (2003). *The MMPI-2 Restructured Clinical scales: Development, validation, and interpretation*. Minneapolis: University of Minnesota Press.
- Tellegen, A., Ben-Porath, Y.S., & Sellbom, M. (2009). Construct validity of the MMPI-2 Restructured Clinical (RC) scales: Reply to Rouse, Greene, Butcher, Nichols, and Williams. *Journal of Personality Assessment*, 91, 211-221.
- 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, 217-222. doi: 10.1207/s15327752jpa8702_12
- Welsh, G. S. (1956). Factor dimensions A and R. In G. S. Welsh & W. G. Dahlstrom (Eds.), *Basic readings on the MMPI in psychology and medicine* (pp. 264–281). Minneapolis: University of Minnesota Press.
- Woodworth, R. S. (1917). Personal data sheet. Chicago: C. H. Stoelting Company.

CHAPTER 2

Abstract

The Minnesota Multiphasic Personality Inventory-2–Restructured Form (MMPI-2-RF) includes several important and substantive revisions. The most important of which may be an integration of its hierarchical interpretation structure. This is in line with research on psychopathology that suggests distress can be captured in one of the three broad patterns of internalizing, externalizing, and thought disordered content. Along these lines, the Higher Order Scales of the MMPI-2-RF were created as abbreviated and discrete measures of this content. However, it has also been proposed that these Higher Order scales were mere aspects of instrument-wide higher order themes which are captured across all types of scales. Accordingly, interpretive guidance for the MMPI-2-RF states that conceptualizations of clients should follow this integration of all scale content into themes. However, these instrument-wide themes have not yet been examined structurally, leaving it unclear the degree to which the scales that have been proposed to act as components of these all higher-order themes fit within their proposed structure. Using a series of modeling approaches (i.e., Confirmatory, Exploratory, and Exploratory Structural), this study examined the higher order patterns of scale interpretability within a sample of physicians referred for work-interfering behaviors. While the three predicted themes of psychopathology (e.g., internalizing, externalizing, and thought disorder content) emerged consistently in physicians in distress, there were some problems and a well-fitted model was not established. Higher order themes were not fully disparate and contained a great deal of co-variation. They were also suggestive of nested factors that might act as intermediary interpretive structures. These problems were particularly pronounced for the scales assessed behavioral aggression, anger, and trait hostility.

Keywords: MMPI-2-RF, Higher Order, Personality, Psychopathology, Validity

CHAPTER 2

MMI-2-RF Higher Order Pathology of Physicians in Distress: A Study

With its substantial psychometric properties (e.g., Simms, Casillas, Clark, Watson, & Doebbeling, 2005), the Minnesota Multiphasic Personality Inventory-2-Restructured Form (MMPI-2-RF; Ben-Porath & Tellegen, 2008a, 2008b) is poised to continue the MMPI's tradition of wide clinical use amidst a rich research foundation. Despite its clinical promise (Finn & Kamphius, 2006), many areas of the MMPI-2-RF are still developing. As a still burgeoning personality inventory, the MMPI-2-RF has yet to establish the substantial research basis needed to form its own comprehensive, interpretive tradition. This early stage of interpretive development has proved one of the greatest criticisms for the MMPI-2-RF (e.g., Butcher & Williams, 2010). Of areas noted for need in interpretive growth, there are two which are particularly pronounced: the assessment of special populations and exploration of how effectively hierarchical patterns of psychopathology are captured. While examining the MMPI-2-RF in special populations ensures the capacity for interpretive normality, expansion of hierarchical interpretive strategies offer the capacity to interpret the clinical issues across these populations in a broad manner.

While one unique advantage of the MMPI-2-RF is that several distinct comparison groups are offered within its interpretation manual (Tellegen & Ben-Porath, 2008/2011), it is also clear that special populations also produce response patterns that require special considerations (Archer, Hagan, Mason, Handel, & Archer, 2012). Unsurprisingly, evaluations of special populations have long been attended to as a vital part of the research tradition of the MMPI. They provide a context for testing the structural assumptions and generalizability of interpretations drawn from test responses. Evaluations of special populations are particularly

needed in cases involving high stakes decisions given the implicit need for accurate decision making (e.g., Tarescavage, Corey, & Ben-Porath, 2014). As the MMPI-2-RF has proved exceptionally effective in its prediction of aberrant behaviors and clinical pathologies even within groups with clear motivations to avoid such detection (Sellbom & Bagby, 2008), it is no surprise that the MMPI-2-RF is being stressed as a superior instrument for evaluation in special populations that are part of high risk settings (Ben-Porath, 2012a).

The term *physician in distress* is often used to describe an encompassing group of medical students, residents, and doctors who have engaged in problematic and work interfering behaviors (e.g., substance use, boundary violations, or prescription problems). These behaviors have also resulted in referrals for fitness for duty evaluations by employers or medical licensing boards and often follow concern over harm to patients and/or the field. Concern over how to assess and treat these medical providers is a topic of much discussion in academic medicine as these work interfering practices have clear implications for both the individual medical provider and the clients with whom they work (Domino et al., 2005). As a result of these broad-reaching and higher stakes decisions, there has been a recent call to expand the interpretive capacity of clinical instruments used in the assessment of physicians in distress (Finlayson, Dietrich, Neufeld, Roback, & Martin, 2013). Given that physicians face the same myriad of mental health concerns present in general populations (e.g., Hughes et al., 1992; Ruitenbergh, Frings-Dresen, & Sluiter, 2012) and that these concerns can produce consequences in the effective care of others, it is not surprising that the medical field is becoming increasingly concerned with ensuring accuracy during these evaluations (Gastfiend, 2005).

In addition to the higher stakes impact decisions being made during these evaluations, evidence suggests that the physicians who are emotionally compromised present themselves in a

manner that requires careful interpretive attention. Dorr (1981) examined the MMPI profiles of impaired physicians who were patients in an inpatient psychiatric hospital and determined that physicians report a considerably higher amount of emotional distress than the MMPI normative comparison group. Such elevations may be due to the widely recognized higher stress vulnerability observed in the physician population at large (Devi, 2011), to the inflated scores observed in older versions of the MMPI (Graham, 2006), or to a combination of these factors. While important to consider, Dorr's study is limited by two considerable shortcomings. Conclusions relied on information obtained from an older version of the MMPI that used dated scoring norms and questionable item content (Ben-Porath, 2012b), and the scores on which conclusions were based were drawn from profiles of a mere 36, all-male sample. Regardless of these limitations, this study provided evidence of a problem in the assessment of physicians in distress. Responses provided by physicians in distress involve complicated clinical interpretation.

Building upon Dorr's (1981) evidence of a need interpretive clarity, Roback and colleagues (2007) examined a sample of 88 outpatient physicians in distress using the MMPI-2 (Butcher, Dahlstrom, Graham, Tellegen, & Kaemmer, 1989) and found that both the Clinical and Validity Scales were substantially elevated. Based on patterns of elevation, they concluded that distressed physicians are likely to present with interpretively complex profiles characterized by heightened levels of pathology and difficulties obtaining valid and interpretable profiles. These difficulties led Roback and colleagues to conclude that there remains a paucity of research outlining interpretation of assessments conducted on samples of physicians in distress. This conclusion is aligned with the previous supposition suggesting that clinical observation of group distinctions for physicians were common despite no bedrock of quantitative data to support such claims (Dorr, 1981).

Given the novelty of the MMPI-2-RF, researchers have yet to explore how to utilize and interpret profiles obtained from physicians in distress. Changes in rates of elevation between the MMPI-2-RF and earlier versions of the form are common and the MMPI-2-RF often produces elevations lower than those of earlier versions of the MMPI. These lower rates of elevation are often interpreted as the result of the interpretive sharpening of constructs inherent to the RC Scale/MMPI-2-RF (Weed, 2006). Thus, while researchers exploring the MMPI profiles of physicians in distress have previously noted saturation of scale elevations on the MMPI and MMPI-2, the methodological improvements of the MMPI-2-RF provides an opportunity for interpretive clarity. However, the MMPI-2-RF's reliance on items that are highly "face valid" produces a challenge. Profiles may be influenced with easier manipulation, particularly in the case of highly educated respondents (Caldwell, 2006), making the likelihood of substantially lower mean elevations and clinical-level endorsements a stark possibility. The challenge of low frequency clinical elevations appears particularly pronounced for RC9 which measures thematic elements of impulsivity and poor judgment. Although RC9 is conceptually linked to the most common types of problems leading to physician evaluations (Roback et al., 2007), endorsement of that content may be less frequent given the clear motivation for secondary gains present during fitness for duty evaluations. Taken conjunctively, assumptions about the interpretive context of MMPI profiles obtained from physicians in distress continue to pose a challenge in use of the MMPI-2-RF within that population. Consequently, the need for refined interpretative guidelines remains as vital as ever for this iteration of the MMPI (e.g., Roback et al., 2007).

Along with the difficulties of scale elevation and frequencies of clinical endorsement, successful adaptation of the MMPI-2-RF for clinical use in physicians in distress requires a change in interpretive tradition. This change in interpretive tradition is consistent with the

adaptations facing all providers transitioning between the MMPI-2 and the MMPI-2-RF. Earlier versions of the MMPI relied on code types for interpretation for clinical presentation; however, the revisions of the RC scales eliminated this interpretive possibility by removing the component scales that comprised this interpretive tradition. This change in approach resulted in use of code types on the MMPI-2-RF becoming a stark impossibility despite core content of scales being retained (Tellegen et al., 2006).

The removal of this approach to interpretation is not altogether bad since code types face two distinct problems; they are exclusive in nature and they are not representative of our knowledge about the hierarchical nature of psychopathology (Ben-Porath, 2012b). Thus, it is not possible for someone to be described using two code types composed of separate scales despite it being possible for dysfunction to be present across multiple areas. Correcting for this, three Higher Order (HO) scales were constructed as integral components of the MMPI-2-RF. The HO scales were drawn from only the RC item content and labeled as the Emotional/Internalizing (EID), Thought Dysfunction (THD), and Behavioral/Externalization (BXD) domains. They each represent the pathologies inherent to the most common code types. Indeed, factor analyses have confirmed that EID, BXD, and THD do represent the clinical elements for the traditional code types of 27-72, 68-86 (Sellbom, Ben-Porath, & Bagby, 2008), and 49-94 (Hoelzle & Meyer, 2008; Sellbom et al., 2008). However, as physicians demonstrate distinctive patterns and relationships with and between reported symptoms (Dorr, 1981; Roback et al., 2007), the assumption that these higher order patterns of scale responses remain conceptually linked is problematic.

Understanding core patterns of presentation across the clinical core of the MMPI-2-RF is important because these higher order constructs are more than just a replacement for code types;

they bring the MMPI-2-RF into line with current research of psychopathology. Internalizing and externalizing dimensions have been widely accepted as themes of psychopathology and expressive patterns for distress (Blais, 2010; Krueger & Tackett, 2003). Likewise, aspects of thought disorder have also been characterized as key components to functional adaptation (Shelder & Westen, 2004), including a long history of conceptual inclusion in the MMPI (Meehl, 1946).

These three patterns of higher order pathology are interpreted in two ways on the MMPI-2-RF. They are represented in the higher order (HO) scales, in content scales composed solely of items comprising the RC scales, and through a thematic interpretive structure produced in the MMPI-2-RF as a framework incorporating three tiers of hierarchical scales that move from broad (HO Scales) to narrow (Specific Problem Scales; SP) concerns. HO Scales provide scores for structures over RC content but may not represent the most accurate measurement of the broadest pattern of those psychopathology themes (i.e., internalizing, externalizing, thought disorder) given that they exclude items from many other conceptually related scales. For instance, while BXD is said to measure distress externalization, some SP scales associated with common externalizing behaviors (e.g., Aggression) are excluded in the calculation of BXD as that respective SP scale's items are not used in the calculation of the RC scales. The higher order thematic framework, however, promises a comprehensive super structure able to provide the broadest and most complete thematic conceptualization of pathology content on the MMPI-2-RF. This framework is recommended for use in profile interpretation as it reflects broad measurement of accepted higher order patterns of pathology by assuming oblique relationships between scales on broad displays problems (Ben-Porath & Tellegen, 2008a).

While the MMPI-2-RF is said to contain themes which exist as broad structures that incorporate numerous content scales (Ben-Porath & Tellegen, 2008a, p.22), no research to date has tested if the source scales said to comprise the themes of internalizing, externalizing, and thought disorder problems indeed capture these constructs. Structural analysis has focused instead on the HO scales, which are composed solely of the RC scale content. This means that it is yet unknown if these conceptualized higher order themes are able to be incorporated as broader conceptualizations of the extant HO Scales and if the MMPI-2-RF's structure may be able to provide a more complete integrative interpretive profile than is currently available. While RC and HO scales correlate highly with scales measuring narrow concerns (e.g., SP scales), patterns of profile response styles observed in physicians in distress pose a problem. Thus, attempting to align interpretation of the MMPI-2-RF with higher order understandings of pathology is an area in need of study.

Understanding how well, and how broadly, hierarchical themes exist within the MMPI-2-RF may also provide a next step for the interpretation for use with physicians in distress. Given the scale performance variations seen in physicians and the amassing evidence that thematically-based HO scales are able to predict behavioral outcomes (Ben-Porath, 2012b), exploring the broadest structural organization for the MMPI-2-RF would provide a firm grounding in the interpretive approach it endorses. It would provide a context beyond the higher order scales to understand broad patterns of pathology. Although RC9 has suppressed elevations for physicians, other externalizing scales may instead serve to highlight the expression of undercontrolled, behavioral dysfunction. Understanding these broad themes of psychopathology might not only prove useful for crafting of evaluation decisions relative to fitness for duty assessments but also

may help to establish and understand the much needed profiles of physicians in distress (Finlayson et al., 2013).

These higher order themes represent the broadest pathologies observed elsewhere in the literature (e.g., Krueger & Tackett, 2003), but it is only assumption of structure that holds these collections of component scales together on the MMPI-2-RF. Thus, it would be meaningful for structural evaluations to weigh the relative importance of all components within these thematic higher order elements. Doing so provides a needed means to contextualize the MMPI-2-RF's capture of the broad levels of psychopathology and expands understandings of patterns of presentation common for physicians in distress, a distinct population involved in high risk evaluations.

Present Study

Physicians in distress represent a population in need of further study on the MMPI-2-RF. They are characterized by historic problems in measurement on the MMPI (Dorr, 1981; Roback et al., 2007) and substantial levels of distress which place the public at risk when distressed doctors engage in insufficient coping (Devi, 2011). Moreover, the high education of these physicians places them apart from most groups on which research on the MMPI-2-RF has been done, a distinction that has been suggested as impacting profile interpretability (Caldwell, 2006). Because the integration of theory-based conceptualizations of psychopathology represents one of the strengths of the MMPI-2-RF (Ben-Porath, 2012b), evaluations of the structural appropriateness of these higher order thematic pathologies is a critical step to successfully incorporating the MMPI-2-RF into the clinical evaluations of physicians in distress. Moreover, although recommended as an interpretive framework for use across populations, no studies have been located that assess the degree to which higher order themes emerge. Thus, this study has

two distinct goals: (a) examine scale response characteristics of physicians in distress on the MMPI-2-RF and, more importantly, (b) examine the structural composition, and feasibility of interpretive incorporation, of the higher order themes recommended for use on the MMPI-2-RF. While descriptive statistics can outline response styles for physicians in distress, evaluations of the structural form of higher order themes requires more complex modeling procedures.

Three analyses were used to test the hypothesis that content scales of the MMPI-2-RF grouped conceptually in their measurement of higher order patterns of psychopathology. Each analysis tested this hypothesis using increasingly relaxed assumptions about the structural form of higher order pathology. The first analysis used confirmatory factor analysis (CFA) to assess this hypothesized model and incorporated an assumption that no correlations existed of first order content beyond that explained by second order factors. The second analysis utilized exploratory structural equation modeling (ESEM) which allows content scales to load on each of the three conceptual higher order pathologies. The final analysis used oblique exploratory factor analysis (EFA) to examine a best fitted structure of scale content without an assumption that higher order pathology would present in the three factor framework.

Given the substantial literature available about pathology presenting in a hierarchical manner, two findings were hypothesized. The study's first hypothesis was that analysis one (which incorporates scales explicitly according to theory through a confirmatory factor analysis) would produce appropriate levels of fit and be able to act as the guiding structure for clinical interpretations of physicians in distress. The second analysis, utilizing exploratory structural equation modeling, was not expected to substantially alter the pattern of scales inherent to these themes or to produce significantly improved model fit statistics. The study's second hypothesis is that RC9's restriction in range and poorer performance within a sample of physicians in distress

reflects a broader pattern of difficulty associated with measurement of externalizing content for this population. As such, RC9 and those scales most associated conceptually with RC9 [i.e., Aggression (AGG) and Activation (ACT)] will produce poorer factor loadings and more residual error compared to other content scales on the latent construct representing externalization.

Method

Participants

The initial participant pool was comprised of 273 physicians or physicians in training (e.g., residents); however, following exclusionary criteria for elevated validity scales, a total sample of 271 physicians or physicians in training remained eligible for inclusion. This sample (226 male, 83.4%) is composed of individuals who have engaged in evaluation at a private mental health treatment center (PMHTC) located in the Midwestern United States. Although the exact numbers of physicians in training within this sample were not available, they represent the minority of the participant pool as most referrals are for physicians already licensed. They range in age from 23 to 81 ($M=49.25$, $SD=11.97$). Information of ethnicity of study participants were not readily available as it was not entered into the electronic administration software used in the administration of the MMPI-2-RF.

PMHTC provides comprehensive assessment, individual treatment, and forensic evaluation for professionals, primarily physicians, who need assistance in developing personal functioning skills and professional competency. Most of these individuals have been referred in response to serious ethical violations against their respective licenses for issues such as aggressive inter-personal conflict and boundary crossing issues or for legal entanglements which are interfering with their careers and capacity to execute their professional duties. Clients are referred from around the country for an assessment and, if appropriate, may opt to remain

engaged in treatment at the PMHTC facility for a seven week residential program. While physicians in training were included in this sample, they composed a minority of cases as most assessment referrals at the PMHTC are for licensed physicians engaged in medical practice.

Instrumentation

MMPI-2-RF. The MMPI-2-RF (Ben-Porath & Tellegen, 2008a, 2008b) is a 338-item self-report personality inventory with scales derived from items included on the MMPI-2. It is comprised of a total of 51 scales, including 9 validity indexes. Of the remaining 42 substantive scales, 3 are Higher Order (HO) construct scales, 9 are Restructured Clinical (RC) scales, 23 Specific Problem (SP) scales, 2 interest scales, and 5 Personality Psychopathology scales (PSY-5). The SP encapsulate four distinctive themes with 5 scales measuring somatic/cognitive complaints, 5 measuring interpersonal problems, 9 measuring internalizing problems and 4 assessing externalizing problems. One week test-retest and reliability estimates for the RC and Higher Order Scales range from .64 to .90 and .73 to .95, respectively. Specific Problem scales have slightly lessened psychometric properties, with test-retest ranging from .54 to .85 and coefficient alpha estimates ranging from .46 to .92. The Psy-5 scales have a test-retest ranging from .76 to .93 and reliability coefficients between .68 and .88. For more comprehensive discussion about the specific psychometric properties or the development of the MMPI-2-RF, readers are referred to the technical manual (Tellegen & Ben-Porath, 2008/2011). Descriptions of the 51 content scales that comprise the MMPI-2-RF are provided in Table 1.

Procedures

This study used a retrospect database of MMPI-2-RF profiles administered as part of routine evaluations from 2008 to 2014 at the PMHTC. Administration of the MMPI-2-RF during this period was conducted using the Q-local computerized administration system (Q-Local,

2009). Due to the nature of the evaluations, there were no instances of missing data within the collected dataset. Following export of anonymized test scores, profiles were examined for valid responding and excluded profiles on the basis of clinically elevated MMPI-2-RF validity scales. The MMPI-2-RF recommends that a scores not be interpreted or included in research in cases of invalid profiles as defined by the MMPI-2-RF technical manual (e.g., scaled scores of Cannot Say [CNS] > 18; True Response Inconsistency [TRIN-r] or, Variable Response Inconsistency [VRIN-r] > 80; if Infrequent Response [F-r] > 120, Adjustment Validity [K-r] > 60, or Infrequent Pathology [Fp-r] > 100). Given the effectiveness of these cut-offs in past studies on valid responding (Ben-Porath, 2012b), this study employed use of recommended cut scores. Interest scales, given their non-clinical nature, were excluded from all analyses.

Data Analysis

In order to describe performance characteristics of physicians in distress, descriptive statistics were conducted. Additionally, and consistent with initial evaluations of the MMPI-2-RF special populations (e.g., Archer et al., 2012), differences in scale scores between the sexes were evaluated using a multiple analysis of variance (MANOVA). Due to the MANOVA's capacity to calculate an estimate of overall difference for all included dependent variables using the omnibus *F*-statistic, MANOVA provides a robust evaluative approach for testing sample homogeneity.

To evaluate patterned occurrences of higher order, three separate analyses were conducted. Each approach offered distinct information about the structural composition of thematic MMPI-2-RF content by varying their underlying assumptions of content inter-relatedness. The scale indicators that were selected for use were limited to those identified by the

interpretive manual of the MMPI-2-RF (Ben-Porath & Tellegen, 2008a) as representing the broadest and most critical core clinical content.

Confirmatory Factor Analysis. The first analysis employed a confirmatory factor analysis (CFA) and tested the fit of the structural model proposed by Ben-Porath and Tellegen (2008/2011, p.22). It emphasized the higher order critical pathology on the MMPI-2-RF (i.e., internalization, externalization, and thought disorder content), is consistent with interpretive and developmental approaches recommended for the MMPI-2-RF (Archer, Handel, Ben-Porath, & Tellegen, 2015). Hypothesized loadings are located in Figure 1. To determine the appropriateness of this model's fit, Hu and Bentler (1999)'s guidelines were followed for identifying excellent CFA fit when using maximum likelihood estimation. These guidelines include a RMSEA of close to .06 or below as well as a CFI and TLI of .95 or greater. The impact of data correction was also planned using two methods: a data-based approach and an a priori, theory driven method. Model based corrections are planned using data-driven information available in the form of modification indices if fit falls below Hu and Bentler's recommended threshold. Although data-based corrections are not preferred (Little, 2013), they allow for evaluation and improvement of a model with minor problems in fit while retaining an assumption of heterogeneity between included variables, both latent and manifest. An a priori approach, utilizing all expected covariances was planned as a means to test theory driven corrections. This corrective iteration began with the initial CFA model and made all corrections in a single stage as fit statistics for the initial model were not adequate.

Exploratory Structural Equation Modeling. The second analysis utilized exploratory structural equation modeling (ESEM). This approach differs from CFA by using a less restrictive model that does not make an assumption of a non-relationship (e.g., zero-order correlation)

between unidentified model pathways. Instead, all scale indicators are allowed to load onto each latent construct simultaneously. In other words, all of the manifest indicators included in the initial CFA analyses are able to load freely onto each of the higher order theme factors due to expectations of multicollinearity between many types of pathology (Barlow et al., 2013).

Similarly to CFA, however, is ESEM's requirement to specify a number of higher order factors onto which all items load. Thus, ESEM can be conceptualized as a less general application of EFA analyses (Marsh, Morin, Parker, & Kaur, 2014). ESEM has proven useful in the structural evaluation of personality because the content is either highly related or co-occurring (Marsh et al., 2010), and this type of intercorrelation has been widely conceived as probable within higher order constructs on the MMPI-2-RF (Ben-Porath, 2012; Caldwell, 2006). Consistent with the CFA analyses, examination of structural composition utilizing modifications was planned if fit was not acceptable for the baseline ESEM model (i.e., Hu and Bentler, 1999). The ESEM post-hoc analysis plan included the same data-based and theory-based corrections incorporated into the CFA model. The capacity to do such corrections is one of the design strengths of ESEM analyses (Marsh et al., 2010; 2014).

Exploratory Factor Analysis. The third analysis incorporated the loosest set of modeling assumptions possible for exploring potential higher-order factor structures; it did so through use of exploratory factor analysis (EFA). Using the oblique, promax rotation and a maximum likelihood estimation technique that allows for the expected correlations between MMPI-2-RF content, the EFA in this study utilized the observed correlation patterns between inputted variables to determine the number of higher-order factors that appear necessary to explain the data (Brown, 2006). EFA is similar to ESEM in that it allows content to load on multiple higher-order factors; however, it does not require the identification of a set number of

emergent, higher order factors. Maximum likelihood is an estimation technique that is most likely to generalize to CFA methods as they both rely on a common factor model and attempt to account for correlations between items. Such an approach is distinct from principal component analyses (PCA) that rely on explaining total variance; as such, PCA would produce lower fit due to covariance problems. Likewise, oblique rotation techniques are appropriate when there is a relationship expected between emergent factors. Use of the promax rotation is a well suited option for analysis as it allows for a starting assumption of an orthogonal relationship, but breaks that relationship as necessary to fit the data. The high number of post-hoc model corrections identified as necessary during the CFA of the MMPI-2-RF higher order themes supports its inclusion.

Results

Descriptive Characteristics

Prior to exploring structural patterns, descriptive characteristics for the sample were calculated. Table 2 and Table 3 list means, standard deviations, and scale mean scores differentiated by gender. The evaluation of gender differences is a common evaluative practice for the MMPI-2-RF given its move towards non-gender normed scores (e.g., Archer et al., 2012; Wygant et al., 2007). The analysis of gender differences is an important part of evaluating interpretive patterns on the MMPI-2-RF as non-gender scores have been criticized for departing from the tradition of the MMPI (Butcher, 2010). Tables 4 and 5 present percentage of clinical elevation (e.g., T -score > 65) observed for the Validity, Higher Order, and RC scales as well as the content scales, respectively. Finally, comparisons between scale scores (collapsed across sex) with the normative sample and a comparison group in an outpatient treatment setting are presented in Table 6. This information highlights the uniqueness of the clinical profile for

physicians in distress. While distinct from previous trends (Dorr, 1981; Roback et al., 2007), these response patterns emphasize a need for evaluation of the interpretive frameworks surrounding physicians in distress.

Confirmatory Factor Analysis

CFA analyses examining the capacity of the proposed three factor structure of higher order themes to measure thematic content in physicians in distress produced poor fit with all fit indices falling well below any recommended acceptance criteria (Brown, 2006), $\chi^2(347) = 2541.407$, RMSEA = 0.153(0.147-0.158), CFI= 0.693, TLI=0.665. Following poor fit, modification indices (i.e., the estimations of change in χ^2 scores following a data correction) were examined and data based adjustments were conducted sequentially on each of the largest corrections. The poor fit statistics suggested a need for multiple modification iterations. After ten corrective interactions, the model fit was greatly improved and approached an acceptable level of fit despite remaining uninterpretable, $\chi^2(337) = 1547.626$, RMSEA=0.115(0.109-0.121), CFI= 0.831, TLI=0.810. Table 7 provides associated fit statistics for each of the ten data corrected models, and Figure 2a shows the model associated with the tenth iteration of data based corrections.

After the ten corrective interactions, corrected relationships accounted for 35.7% of the 28 total model estimations described in the initial factor solution. Such high numbers of corrections in comparison to numbers of initial solution estimations make interpretation difficult as the theoretical basis necessary for CFA model formulation becomes increasingly abandoned. However, another interpretive option is to incorporate measurement of correlated uniqueness (CU; the a priori covariances between expected elements of a factor). CU approaches (e.g., Marsh, 1989) use theory driven methods within the model they are assessing to handle model

corrections and is an alternative to the data-driven approach of incorporating individual modification indices. After correlating indicators within each of their Higher Order and RC parent factors (Ben-Porath & Tellegen, 2008a), CFA fit showed poor fit, worse even than the model fit observed following ten data based corrections, $\chi^2(314) = 2043.318$, RMSEA=0.143(0.137-0.148), CFI= 0.758, TLI=0.709. The CU corrected model and the model observed following ten data based corrections are presented in Figure 2b.

There was a discrepancy between observed and appropriate estimations of model fit along with a large number of still high modification estimates. Accordingly, it was determined that the strict assumptions of confirmatory analysis are not well suited to understanding highly inter-related constructs (Marsh et al., 2010; Marsh et al., 2014), such as appears in the higher order structural presentation of the MMPI-2-RF. To address this limitation, confirmatory approaches were abandoned for more liberal modeling techniques. In addition to the numerous corrections needed, this abandonment of confirmatory methodology was supported by the tendency of corrections needed within their supposed higher-order theme. If scales had produced consistent elevation patterns within a higher order theme then stronger factor loads and lower portions of variance would have been evident instead of multiple corrections between grouped items.

Exploratory Structural Equation Modeling

An ESEM was then conducted to explore the degree to which a three factor solution was possible for the MMPI-2-RF once scales were able to cross load. This initial ESEM model produced poor fit, $\chi^2(297) = 1940.056$, RMSEA=0.143(0.137-0.149), CFI= 0.770, TLI=0.707. Examination of modification indices revealed numerous correlated scales; consistent with the CFA, ten corrective iterations of data-based, single covariance corrections were taken. Fit

statistics for each of these ten iterations are reported in Table 8. Acceptable fit was reached during the tenth iteration of model corrections, RMSEA=0.077(0.070-0.084), CFI= 0.942, TLI=0.916. However, a non-positive covariance matrix led to rejection of this model as appropriate. Next, as with the CFA model, a priori CU corrections were implemented into the baseline ESEM model as a single step insertion. This corrective stage produced fit that was poor, $\chi^2(264) = 1541.999$, RMSEA=0.134(0.127-0.140), CFI= 0.821, TLI=0.744, and also had a non-positive covariance matrix.

When non-positive covariance matrixes are observed, they often indicate negative variances, correlation above one (e.g., a Haywood case) between variables, or linear dependencies amongst variables. In both the data based and theory based corrections, the problem was related to BXD's negative variance. Thus, ESEM models obtained using both data based and theory based corrections suggest poor fit.

Exploratory Factor Analysis

Next, an EFA analysis using maximum likelihood estimation and a promax rotation was conducted to explore a structurally unrestricted model of psychopathology (e.g., a model where the number of extracted factors as well as relationships between all factors and items are freely estimated as a function of observed correlation matrixes within the data). Evaluation of the EFA began with an exploration to ensure equal variances across the sample. The adequacy of the correlational matrix for the composite scales of the MMPI-2-RF that were entered into the EFA was acceptable: Kaiser–Meyer–Olkin = .886; Bartlett's test of sphericity $\chi^2 = 7212.943$, $p < .001$. Both of these tests confirmed the assumption of the data as normally distributed, ensuring its appropriateness for factor analysis.

Based on eigenvalues generated using a Monte Carlo simulation and visual inspection of the scree plot (see Figure 3), a four-factor solution emerged as most appropriate for the MMPI-2-RF items. These four-factors accounted for 65.9% of variance. Table 8 shows the structure matrix for the extracted factors and Table 10 lists the intercorrelations amongst the four extracted factors. As a result of using an oblique rotation, the potential for high factor loadings on items across multiple domains is common. It leads to a greater difficulties with interpretive translations but also produces the most generalizable descriptions of the data (Brown, 2006). Consistent with research using factor analytic approaches on personality, a factor loading of .4 for a scale was required for consideration as a component of an extracted factor. A second criterion was also identified to maximize the potential for interpretation: for a loading to be conceptually assigned to a given factor, that loading must be the largest on that factor amongst the four extracted factors.

While the first factors appear to generally align to the interpretive model suggested by Tellegen and Ben-Porath (2008/2011), several distinct differences for each theme of psychopathology did emerge. Among items conceptually associated with internalizing disorders, two content scales (AXY and MSF) did not load as expected. One RC scale (RC6) also demonstrated a surprising and strong loading. Representing measurement of trauma-related reactance and proneness of phobias, the excluded content domains did not appear significantly thematic in the measurement of internalizing distress. Conversely, RC6's measurement of distrust, paranoia, and suspiciousness of others associated nearly as strongly with problem internalization as it did with thought disorder. The expected thematic factor of psychotic thinking emerged as the second factor and included a number of unexpected scales beyond those identified as thematically inclusive by Tellegen and Ben-Porath (2008; 2011). Included in this

interpretive model were the content scales of BRF, AGG, and ACT which measure behavior restricting fears, physical aggression, and manic/hypomanic behaviors respectively. The thematic content of behavioral dysfunction emerged within the EFA as the third factor. It showed no loadings with the content scale of ACT, that measures manic/hypomanic symptoms, or the PSY-5 domains of AGGR-r or DISC-r. This exclusion of PSY-5 personality traits suggests a pattern of behavioral disruptions not associated with more severe personality dysfunction. Finally, a novel fourth factor appeared which loaded most strongly with items thematically aligned to aggression, distrust, and psychomotor activation. Items for this factor were drawn primarily from thematic items drawn from the behavioral dysfunction but with content focused more explicitly on aggressive behavior.

While interpretive criteria was been identified for items that were the highest loading on each factor along with a loadings above .4, the high degree of correlation between many of the scales (an average cross factor loading of +.42) led to the decision to add an additional inclusion criteria. For a scale to be included and identified as a discrete part of an extracted factor, a difference of at least .10 between the loadings of factors was decided upon. This decision was taken to maximize discriminate validity of the interpretation model. Three exceptions were made to this rule based on theoretical basis; such decisions are frequently needed in EFA to correct for data-derived information not in line with theory (Brown, 2006). All three had still had at least a .05 magnitude difference in factor loadings. ACT and AGG were assigned to factor four, given its thematic inclusion of aggression and overall hostility. However, AGG's relatively stable loadings across all four factors should be noted, making this theoretical assignment one that may be problematic during replications. RC6, a measure assessing paranoia and distrust, was assigned to factor 2 as that factor most aligns with Tellegen and Ben-Porath's (2008/2011) description of

thought disordered content. The final four factors (e.g., those factors composed of content scale scores denoted with asterisks on Table 8) appear to present patterns of psychopathology responses according to the following general descriptions: *Internalizing Problems* (Factor 1), *Thought Dysfunction* (Factor 2), *Behavioral Disruption* (Factor 3), and *Externalization* (Factor 4); three of these thematically align with the higher order themes proposed by Tellegen and Ben-Porath (2008/2011).

Discussion

This study evaluated the proposed framework of the three theory driven higher-order themes of psychopathology (i.e., emotional, behavioral, and thought dysfunction) incorporated into interpretation of the MMPI-2-RF. Included into the MMPI-2-RF as an effort to align the structure of the MMPI-2-RF to research on broad patterns of distress response, this study is the first to test the structural utility and appropriateness of these higher order themes. Using a sample of physicians in distress, this study's evaluation of response styles provides an expansion of the interpretive tradition of the MMPI in a population referred for high stakes assessments (e.g., Roback et al., 2007) that are prone to increased levels of distress (Devi, 2011; Lee, Stewart, & Brown, 2008). This study's exploration of the MMPI-2-RF expands available research on the interpretive patterns for an important and frequently used clinical instrument (Camara, Nathan, & Puente, 2000) in a sample in need of study (Finlayson et al., 2013). Beyond a basic descriptive evaluation of scale responses, this study use a series of analytical approaches and planned post-hoc analyses with progressively loosened assumptions about model structure. Doing so provides a test of the generalized feasibility of incorporating the structural composition of the proposed interpretive framework into clinical practice.

Descriptive results of this study produced two distinct findings: (a) scale elevations (e.g., Validity, RC, HO, SP, and PSY-5) are generally comparable to the normative group in the MMPI-2-RF technical manual (Tellegen & Ben-Porath, 2008, 2011) and (b) significant enough variability exists on some scales such that it warrants interpretations drawn explicitly from this special population. MMPI-2-RF scores were generally congruent with the normative means and evidenced a marked improvement over measurement efforts in this population using earlier versions of the MMPI (Dorr, 1981; Roback et al., 2007) as scale elevations were significantly less inflated. The ability of the MMPI-2-RF normative sample to effectively reflect respondent performance is exemplified by mean absolute differences between this sample and the normative sample being consistently smaller (i.e., possessing substantially fewer differences exceeding half a standard deviation) when compared to other comparison samples. However, the magnitudes of these differences are still substantial on many scales.

Areas of critical importance to focus on during assessment of physicians are most characterized by feelings of demoralization, hopelessness, sadness, and paranoid thinking. Elevation rates are slightly higher for many RC scales (excluding only RC3, RC4, and RC9), experiences of depressive (RCd/RC2) as is paranoid and distrustful (RC6) thinking. This is not surprising given that elevations of these concerns mirror national occurrences (Tarescavage et al., 2013). Important to note is the strikingly under-elevated score for RC9, falling over a standard deviation below the normative mean. Given its substantially lower mean, use of the recommended cut-off is likely causing an under-assessment of impulsivity, mood instability, risk-taking, and excitability in physicians. This lowered RC9 mean has implications for the use of the BXD scale as well as the fact that RC9 comprises five of its items (21.7%) with that higher order scale. This possibility of decreased measurement sensitivity of externalization for

physicians in distress is particularly likely since so few physicians in distress (1.1%) exceeded BXD scores of clinical significance, despite that being a primary reason for referral (Finlayson et al., 2013). Taken together, the descriptive characteristics observed in this sample supports the general utility of the MMPI-2-RF in special populations while also suggesting unique interpretive challenges.

In addition to a distinct interpretive profile produced through individual scales, analyses of higher order themes provided evidence of patterns of pathology. Observed scale patterns align in many important ways with the proposed higher order themes of the MMPI-2-RF (e.g., Tellegen & Ben-Porath). However, this alignment is not wholly appropriate to broad scale implementation. Given that the problems repeatedly occurred during efforts to fit an interpretive model with observed responses of physicians in distress, these theorized higher order themes might only be appropriate for consideration as a loose framework. Despite problems with the thematic interpretive framework, this developing approach to interpretation that shows promise as an appropriate and useful interpretive structure as its issues become addressed. Thus, the MMPI-2-RF has provided some evidence of alignment with itself and accepted higher-order models pathology (e.g., Achenbach & Edelbrock, 1978; Krueger & Tackett, 2003) while still requiring work for generalized interpretive incorporation within populations of physicians in distress. While a narrow population, attempts to incorporate the MMPI-2-RF higher order themes into generalized use with other groups should consider issues raised here as they may reflect broader patterns of thematic responses.

Three types of problems occurred during the structural evaluation of response profiles for physicians in distress. Each problem occurred thematically across all three analyses (e.g., CFA, ESEM, and EFA). These three thematic issues are related to the (a) covariation that exists

between scales of the MMPI-2-RF result in factors with less than desired discriminate capacity, (b) incorporation of model corrections (e.g., CFA and ESEM) does not increase the interpretability of the MMPI-2-RF higher order themes to a point of acceptable use, and (c) difficulties appropriately fitting expected structural models are likely a result of the assumption of a three factor solution in being the most appropriate for physicians in distress. Taken together, these issues suggest that an approach to interpretation which incorporates the higher order themes of psychopathology is problematic in its current form. These issues are inter-related and provide direction for the evaluation and interpretation of higher order themes on the MMPI-2-RF. The interpretive impact of each of these three distinct problems is discussed separately below. Following that is an attempt to establish procedures that may aid in further development of an interpretive model more appropriate for use in the physicians in distress.

Issues of Covariation

As analyses (i.e., CFA, ESEM, and EFA) explored the emergent structural form of the MMPI-2-RF higher order themes, a pattern of intercorrelation and covariation emerged between scales. While not all scales within each theme demonstrated this pattern, a troubling portion of higher order component scales demonstrated relationships with content beyond that expected. These relationships occurred unsteadily within each theme; some themes showed greater portions of residual correlations relative to their number of component scales. As a result, when higher order themes were assessed using CU corrections, there was a substantially lower fit compared to models incorporating data-based corrections, despite 90% of data-based corrections occurring between component scales within the same higher order theme. This discrepancy of model fit observed between the fitted CU model and most iterations of the data based corrections

suggests that covariation between scales, even within supposed higher order theme areas, are not occurring in a way that would be theorized.

It is important to note that these unexpected relationships were not cross-loadings (e.g., where an item or item parcel loads onto multiple factors). Instead, the model-corrected relationships were correlations between the residual error of scales not accounted for by the higher order theme (i.e., correlated residuals). Such patterns of correlated residuals can be suggestive of either an additional hierarchical order structure or a convergence of concepts. As such, substantial patterns of correlated residuals can cause traditional interpretive models to become either inappropriate or complex beyond a point of interpretation as the relationships existing do not exist in a manner congruent with proposed theory. In the case of Negative Emotionality (NEGE), for instance, there are substantial relationships between it and other scale factors drawn from multiple higher order themes (e.g., RC7 and DISC). Such a relationship demonstrated through patterns of high covariance suggest that a simple three factor structure, as assessed, may be unable to capture the relationship being played by NEGE.

Moreover, even if the structural model had been pruned of post-hoc identified component scales based on poor factor loadings (such as those seen with *Multiple Specific Fears* (MSF) and *Anxiety* (AXY) on *Emotional Dysfunction*, see Figure 2a), such an approach would have been unlikely to address the broader problems inherent to scale correlations. Returning again to the case of NEGE, despite the needed addition of three correlated residuals, NEGE had one of the strongest factor loadings inherent to the higher order theme of *Emotional Dysfunction*. As such, even use of a pruned model would not have been likely to reduce, or eliminate, the need for post-hoc addressing of problematic amounts of correlated residuals.

In addition to the patterns of intercorrelation amongst scales, there was a poor discriminant capacity for higher order themes. Substantially sized correlations between each of the higher order factors are problematic as it suggests that theme structures carry some overlapping measurement content. This intercorrelation is despite higher order theme component scales attending to discrepant content and having been stripped of the first factor that was supposed to have caused such uniformity in measurement (e.g., Tellegen et al., 2003). Thus, highly correlated themes suggested that there are patterns of scale elevation that remain unaddressed and that issues of covariation occur not only within themes but also between them.

Insufficiency of Model Correction

Beyond the notable rates of correlation between the scales of the MMPI-2-RF, the use of approaches implementing data and theory based corrections failed to provide a structural model that was clearly interpretable. During CFA analyses, data-based corrections witnessed noteworthy increases in structural appropriateness by the tenth model iteration. This model, as with the CU theory-based corrections model, failed to reach an acceptable standard of model fit (Bentler & Hu, 1999). This leaves the recommended interpretive framework of the MMPI-2-RF (Ben-Porath, 2012b) as an insufficient guide for evaluating patterns occurring between the instrument's multi-tiered scale structure (e.g., Restructured Clinical and Specific Problem). Even a loosening of the structural approach from CFA to ESEM failed to produce a stable model for interpretation.

However, ESEM did provide evidence of a substantially stronger fit over the more restrictive CFA models. Indeed, following data based corrections, ESEM was even able to achieve what might be considered acceptable fit by some (e.g., Brown, 2006) were it not for the persistent problem of negative variance observed for BXD. While modifications to structural

components of the model may have been able to account for the negative variance during ESEM by removing BXD, the centrality of the BXD scale to the higher order theme of *Behavioral Dysfunction* results in question about the degree to which such an approach would be appropriate. The removal of BXD, which includes some items from *Behavioral Dysfunction scales*, from the structural model may have resolved issues of negative variance but would have done so at a cost. Designed as a single scale measurement of many of the thematic patterns associated with behavioral dysfunction (e.g., aggression, impulsivity, etc.), BXD's removal would have a higher order interpretation of behavioral dysfunction as unable to incorporate a discrete measure of behavioral dysfunction.

Insufficiency of Three Factor Solution

The three factor solution does not work well for the production of clean and easily interpretable structures of psychopathology. Even following various corrections in CFA and ESEM analyses, the three factor model requires a number of cross-loaded corrections suggesting an inter-play between constructs beyond that expected and needed for clear interpretation. In an attempt to address poor model fit, attempts were made to correct for this high degree of correlation utilizing approaches that were both data-based and theory based. However, even following numerous iterations of incorporated model corrections, the generated structural solutions did not provide a satisfactory context for interpretation. That is, the three factor model of higher order psychopathology suggested by Tellegen and Ben-Porath (2008, 2011) as an interpretive framework did not appear well suited to physicians in distress being evaluated for fitness for duty. These corrections have added some important information about patterns of responses, but are limited in their holistic interpretive capacity. Data based model corrections were commonly done both within scales associated with a given higher order theme (as one

might expect) and between scales of differing higher order types of pathology. It is this inter-relationship of scales between the higher order themes that was most strongly suggestive of problems using a three factor model of psychopathology. It required the use of a looser set of model assumptions than was allowed during CFA and ESEM analysis in order to allow for broader measurement of higher order scale relationships.

Given these intercorrelations between higher order theme component scales and the poorly fitted confirmatory structures, it was not surprising to see EFA suggest a structural model utilizing more than three primary interpretive factors. This expanded model of primary factors suggests a need to conceptualize patterns of pathology beyond the three major components recommended by Tellegen and Ben-Porath (2008/2011). Thus, while criticisms of not using hierarchical interpretation models proposed in the interpretive manual (e.g., Tarescavage et al., 2013) may be appropriate in some cases, this study provides evidence that the MMPI-2-RF's measurement of psychopathology does not align neatly with this higher order model in one special population.

This difficulty in identifying a three-factor interpretive core of psychopathology may be due to inherent problems imposing oblique assumptions onto the measurement of personality and psychopathology (e.g., Marsh et al., 2010). This possibility is bolstered as three higher order themes emerged as prominent factors during EFA. However, high rates of cross-factor loadings suggest that some problems will likely occur as interpretation attempts integration of the themes as discrete occurrences.

An Integrated Model of Psychopathology

While research has repeatedly demonstrated the marked strength of the MMPI-2-RF scales at predicting and discriminating various clinical pathologies and behavior outcomes (e.g.,

Simms et al., 2005), there remains areas for improvement within the instrument related to its capacity to incorporate broad, instrument wide interpretive structures. Substantive problems occurred in validation of the thematic interpretive model for physicians in distress using Tellegen and Ben-Porath's (2008/2011) proposed structure. However, the information gleaned from these analyses provides points of interpretive consideration during use of the MMPI-2-RF in populations of physicians in distress. These considerations provide guidance on how higher-order models may require adaptation in order to fit well within the population and evaluative context that are being seen. Until those issues are resolved, however, interpretation of the MMPI-2-RF should rely on individual clinical scales, and not higher order themes, as scales have produced a strong capacity to perform in a heterogeneous manner to predict behaviors and discriminate between clinical syndromes (e.g., Simms et al., 2005).

The first interpretive point is that not all thematically associated higher order theme content areas are likely to load onto their proposed super-structures. Lower factor loadings and greater residual errors provide a common evaluative lens for such content appropriateness (Brown, 2006). Poorly associated content areas may produce elevations, but these elevations should not be considered related to conceptualization of a broad pathology. Instead, elevations for those content areas are likely a function of individual variability and are not part of the thematic patterns of pathology. As prime examples, MSF and AXY are each a Specific Problem scale not clearly associated with an elevation in the higher order theme of *Emotional Dysfunction*, despite conceptual similarity. However, given that a well suited model was unable to be reached, decisions about which scales should be excluded are difficult to make in a definitive sense, but Specific Problem and PSY-5 personality scales appear the ones most likely to cause problems in acceptable loadings. Thus, as higher order interpretive models are

incorporated, extra attention should be paid to the performance of scales measuring either narrow (Specific Problem) or broad (PSY-5) content.

Secondly, higher order measurement of psychopathology appears influenced by nested response factors in a manner similar to first factor problem (e.g., Welsh, 1956). While the first factor was conceptualized as general distress and purportedly removed (Tellegen et al., 2003), nested response patterns appear evident across thematic groups of content scales. As such, nested components, when not accounted for in structural models, can provide influential thematic patterns even within homogeneous higher order themes. For instance, the fourth factor observed during EFA was one characterized by scales measuring aggression, distrust, and psychomotor activation. While conceptually similar to the Behavioral Disruption factor identified during EFA, the thematic doubling of measurement assessing behavioral dysfunction (e.g., the *Aggression* and *Behavioral Disruption* factors) aligns with Roback and colleagues (2007) finding that most doctors in distress are referred for behaviorally based problems. This fourth factor is composed of scales measuring elements of that thematic presentation for each of the three domains. It is likely that this fourth factor measures a response style common to those undergoing evaluation and may be integrated into the three factor higher order model conceptualized for the MMPI-2-RF by using a nested approach. This is particularly likely given the large cross-indicated factor loadings observed between scales of the fourth factor.

Third, efforts to develop concrete higher-order structures within the MMPI-2-RF may face challenge no matter the analysis used. Confirmatory factor analysis was unable to produce more than poor fit, likely as a result of its restrictive factor structure and model assumptions. Given the high covariation and correlation between scales, these assumptions are likely to pose problems in future efforts to incorporate higher levels of psychopathology. Exploratory Factor

Analysis has a similar problem; its highly correlated factor structures make it difficult to discern where discrete scales should load, despite showing interpretive promise. Thus, while the proposed structures that emerged in the EFA are likely to reflect broad patterns observed in physicians in distress and generally align with much of the theoretical model touted by Tellegen and Ben-Porath (2008,2011), there remains much room for interpretive adaptation based on personal choice rather than discrete theory. Thus, evaluations of EFA models should continue to utilize common uniqueness (CU) to correct for apparently cross-loaded items. Finally, although ESEM did not produce acceptable fit, it did improve it substantially, with the cost that discrete theme interpretation is no longer possible. This conclusion is in line with Marsh et al. (2010) who noted substantial improvement over baseline CFA models assessing other theoretical models of personality. However, it leaves much to be desired as interpretive models that allow for greater covariation and correlation are innately more difficult to understand.

Lastly, the greatest difficulty in integrating the higher order themes as an interpretive framework is tied to their poor discriminant performance. While there appears a potential to interpret the higher order patterns, fit proved problematic and interpretation was difficult not only due to item loadings but also because of the frequent co-varied elevation between supposed discrepant higher order themes. The strong EFA cross loadings and CFA latent correlations suggest a general response pattern. This general response pattern may have been similar to that which has long posed problems during attempted removals in the MMPI. While the first factor was removed at an item level within each of the component scales (e.g., Tellegen et al., 2003), thematic saturation of scales is still evident given high correlations. This adds difficulty in the creation of integrative profiles using multiple scales. One possible approach to addressing this difficulty is the use of a *general pathology* factor as a superordinate influence on the higher order

themes. Such an approach would be in line with emerging work in personality (e.g., Barlow et al., 2013; Loehlin & Horn, 2012; Rushton & Irwing, 2009a, 2009b). The incorporation of general pathology into structural models might provide an appropriate interpretive framework and further the MMPI-2-RF's goal of integrating current knowledge on the structure of pathology.

Limitations and Future Directions

Despite using an array of structural analysis techniques, this study has some limitations in its ability to draw inferences about generalizable patterns of structural models for the MMPI-2-RF. This study employed the use of summed scale scores instead of item-level structural models. This approach was taken because of the substantial sample size that would be needed in order to estimate the large number of relationships observed between all scales and all higher-order themes. Such an approach would have likely acted as a prohibitive factor to examining these themes as even liberal estimates of sample size for factor analysis suggest a minimum of 50 participants for each item (Brown, 2006). Likewise, the somewhat repressed rates of scale elevation may have caused some problems with range restriction; however, use of the scale scores over individual items is likely to have corrected for this some as it increases the chance of variability across a specific content area.

Additionally, demographic characteristics of the sample were not fully analyzed. This can pose problems for the generalization of findings as race and age may have played a role in mediating the ability to produce stable structural forms. Likewise, while this study examined differences in scale score performance between the sexes, the portion of female participants was much smaller when compared to males. Thus, sex-related scale performance are more likely to be sample specific than broadly generalizability for females. However, these shortcomings in demographic analyses are offset by the likelihood that patterns observed here are representative

of referral rates for physicians in distress across the country. Given the limited number of locations that these evaluations occur, observed patterns remain likely to be reflective of the physicians in distress population referred for evaluation and treatment.

The greatest limitation, however, is related to the use of a sample composed of physicians in distress. While an important population that is in need of further study (Finlayson et al., 2013), there is a paucity of knowledge available about how patterns observed in this group relate to the broader group of physicians. Not only the patterns of scale elevations unique in many ways from comparison groups, so too are the descriptive characteristics of physicians. For instance, the educational aptitude and achievement observed in this sample is beyond that typically measured during studies on the MMPI-2-RF. Exemplifying this, Ingram and Ternes (in press) note that their meta-analysis of validity scale effectiveness at the detection of malingering was limited in its ability to evaluate the role of education of response style because only one study included a mean education level that included completion of college.

What is unknown is how well the problematic model fits observed here generalize to other populations, both physician and general outpatient. It may be that physicians in distress are distinct in their interpretive profiles from physicians, which would lend credence to the ability of the MMPI-2-RF to differentiate this group. It is also possible that as a field with high levels of stress (Devi, 2011; Lee et al., 2008) and distinctive demographic characteristics, physicians (both those in distress and those functioning well) produce a unified profile. If this were the case, this information could help guide reduction efforts for burnout by highlighting areas of prominent concern. Where such efforts might be most appropriate would dependent upon if profile elevations were due to trait features of those entering the field or state-based reactance to work environment. So, while the generalizability of this study to a broader population is a substantial

limitation, it is one that resonates with a general lack of research in the field. Moreover, the information provided here is likely useful in aiding the interpretive context of physicians in distress by emphasizing individual scale interpretation, and not instrument-wide pathology theme, should be the focus of clinical interpretation. In this way, the evaluation of a descriptively distinct population offers insight into an area in need of further evaluation despite its limited applicability to a narrow sample, at this point.

In total, these limitations highlight areas that need further research; they also offer insight and guidance about response patterns of specific population while attempting a novel evaluation of underlying interpretive structures in the MMPI-2-RF. That these structures did not generalize to this population may be a function of the sample. It may relate to the assumptions of structure inherent to recommended interpretation of the MMPI-2-RF. Either way, these limitations point to areas of needed research so as to bridge this evaluation with a capacity to generalize findings more broadly.

Conclusion

This study evaluated response styles of physicians in distress on the MMPI-2-RF by examining the degree to which higher-order themes emerge. Utilizing a series of analytic method with varying degrees of restrictive structural assumptions, this study is the first to have examined applicability of these higher order themes as emergent structures and not just guides to theoretical conceptualization. As the purpose of development efforts for the MMPI-2-RF were to bridge theory-driven and empirically-based methods (Tellegen et al., 2003), this evaluation of higher order theme structural form provides a test of the ability of the MMPI-2-RF to successfully navigate this new and important integration into the historically entirely non-theoretical MMPI.

In general, themes that evidence internalization, externalization, and thought disorder did consistently appear. However, they did so in a manner that is problematic. Some scales did not load successfully (potentially due to low elevation rates resulting from a great deal of content specificity) or loaded in a general way across multiple themes. Since the higher order structures that emerge as thematic elements in the MMPI-2-RF do not fully align to the interpretive structures suggested by Tellegen and Ben-Porath (2008, 2011), model corrections were made that often included instances of scales with residual correlations to multiple other scales. Despite this, structures continued to have problems with high intercorrelation. This unresolved issue of non-discriminant themes suggests problems in the intended higher order conceptualization of psychopathology, lending need for further development instead of continued use at this time. These patterns of high cross-theme correlation and residual correlations also suggest the potential for intermediary factors and future attempts to define higher order themes of psychopathology should include the evaluation of nested pathology.

Particularly striking within physicians in distress is evidence of a fourth factor comprising anger, aggression, and hostility. This potential nested component is comprised of cross-loaded items from multiple areas of psychopathology. The loading issues observed within BXD (e.g., a negative variance and a strong EFA cross-loading) are a prime instances of why there is evidence to suggest that the fourth factor may function as a nested component. This fourth component might be a broadly applicable nested factor, or it may exist as the result of the evaluative context for which this physician in distress sample is drawn. If such a factor does not repeat in other populations, it raises an interesting possibility about how nested factors may differ between sub-groups despite relative stability of the three proposed emergent themes.

Research on response patterns for populations of physicians in distress were not limited to evaluations of higher order themes. Observed patterns of descriptive scale characteristics note a general similarity with normative samples beyond those observed for physicians in earlier version of the MMPI (Dorr, 1981; Roback et al., 2007). However, these patterns of scale elevation which raise the possibility of externalizing problems are not being captured using traditional cut-off scores. While it is possible that this is a function of the context inherent to the fitness for duty evaluation since such a presentation holds prospect for secondary gains, cut scores for the most widely noted referral reason should be adapted to address this, if it is indeed a context specific issue. Thus, this study provides evidence of notable distinctions in the observed profiles for the MMPI-2-RF in physicians in distress that may require both specific clinical attention and further research.

As a whole, the observed higher order structural model suggested for the MMPI-2-RF appears theoretically congruent with proposed models (despite not aligning) while still suffering from substantive structural problems that inhibit use. Thematic interpretive schemes are not as discrete as they need to be, and scale performance suggests measurement issues related to behavioral symptoms. Externalizing behaviors may act as nested factors, and they may be under-assessed using current cutoff scores. Thus, future studies into response patterns of physicians and physicians in distress would do well to emphasize measurement and refinement of behaviorally-based symptom sets as those sets were most characteristic of problem within this study. At present, the scales of the MMPI-2-RF, with their strong psychometric history (e.g., Simms et al., 2005), may be the easiest place to start in the attempt to bring the instrument in line with the clinical needs common to fitness for duty evaluations for physicians in distress, leaving higher order interpretation as a point for further development.

References

- Archer, E.M., Hagan, L.D., Mason, J., Handel, R., & Archer, R.P. (2012). MMPI-2-RF characteristics of custody evaluation litigants. *Assessment*, 19, 14-20. doi: 10.1177/107319110397469
- Archer, R.P., Handel, R.W., Ben-Porath, Y.S., & Tellegen, A. (June, 2015). Introduction to the MMPI-A-RF. MMPI-2-RF, MMPI-2, MMPI-A Symposium. Minneapolis, Minnesota.
- Barlow, D.H., Sauer-Zavala, S., Carl, J.R., Bullis, J.R., & Ellard, K.K. (2013). The nature, diagnosis, and treatment of Neuroticism: Back to the future. *Clinical Psychological Science*, 2, 344-365. doi: 10.1177/2167702613505532
- Ben-Porath, Y., & Tellegen, A. (2008a). Minnesota Multiphasic Personality Inventory—Restructured form: Manual for administration, scoring, and interpretation. Minneapolis: University of Minnesota Press.
- Ben-Porath, Y., & Tellegen, A. (2008b). Empirical correlates of the MMPI-2 Restructured Clinical (RC) scales in mental health, forensic, and nonclinical settings: An introduction. *Journal of Personality Assessment*, 90, 119-121. doi: 10.1080/00223890701845120
- Ben-Porath, Y.S. (2012a). Addressing challenges to MMPI-2-RF based testimony: Questions and answers. *Archives of Clinical Neuropsychology*, 27, 691-705. doi: 10.1093/arclin/acs083
- Ben-Porath, Y.S. (2012b). Interpreting the MMPI-2-RF. Minneapolis, MN. University of Minnesota Press.
- Blais, M.A. (2010). The common structure of normal personality and psychopathology: Preliminary exploration in a non-patient sample. *Journal of Personality and Individual Differences*, 48, 322-326. doi: 10.1017/j.paid.2009.10.028
- Brown, T.A. (2015). Confirmatory factor analysis for applied research (2nd ed.). New York:

Guilford Press.

- Butcher, J. N., Dahlstrom, W.G., Graham, J. R., Tellegen, A., & Kaemmer, B. (1989). *Manual for the restandardized Minnesota Multiphasic Personality Inventory: MMPI-2: An administrative and interpretive guide*. Minneapolis: University of Minnesota Press.
- Butcher, J. N., & Williams, C. L. (2010). Personality assessment with the MMPI-2: Historical roots, international adaptations, and current challenges. *Applied Psychology: Health and Well-being*, 1, 105-135. doi: 10.1111/j.1758-0854.2008.01007.x
- Camara, W.J., Nathan, J.S., & Puente, A.E. (2000). Psychological test usage: Implications in professional psychology. *Professional Psychology: Research and Practice*, 31, 141-151. doi: 10.1037/0735-7028.31.2.141
- Caldwell, A.B. (2006). Maximal measurement or meaningful measurement: The interpretive challenges of the MMPI-2 Restructured Clinical (RC) scales. *Journal of Personality Assessment*, 87, 202–210. doi: 10.1207/s15327752jpa8702_09
- Devi, S. (2011). Doctors in distress. *The Lancet*, 377, 454-455. doi: 10.1016/S0140-6736(11)60145-1
- Domino, K.B., Hornbein, T.F., Polissar, N.L., Renner, G., Johnson, J., Alberti, S., & Hanks, L. (2005). Risk Factors for Relapse in Health Care Professionals With Substance Use Disorders. *Journal of the American Medical Association*, 293, 1453-1460. doi: 10.1001/jama.293.12.1453
- Dorr, D. (1981) MMPI profiles of emotionally impaired physicians. *Journal of Clinical Psychology*, 37, 451– 455.
- Finlayson, A.J.R., Dietrich, M.S., Neufeld, R., Roback, H., & Martin, P.R. (2013). Restoring

- professionalism: The physician fitness-for-duty examination. *General Hospital Psychiatry*, 35, 659-663. doi: 10.1016/j.genhosppsy.2013.06.009
- Gastfriend, D.R. (2005). Physician Substance Abuse and Recovery: What Does It Mean for Physicians-and Everyone Else? *Journal of the American Medical Association*, 293, 1513-1515. doi: 10.1001/jama.293.12.1513.
- Graham, J.R. (2006). *MMPI-2: Assessing personality and psychopathology* (4th ed.). New York: Oxford.
- Hoelzle, J.B. & Meyer, G.J. (2008). The factor structure of the MMPI-2 Restructured Clinical (RC) scales. *Journal of Personality Assessment*, 90, 443-455. doi: 10.1080/00223890802248711
- Hu, L., & Bentler, P. M. (1999). Cutoff criteria for fit indexes in covariance structure analysis: Conventional criteria versus new alternatives. *Structural Equation Modeling*, 6, 1-55
- Hughes, P.H., Brandenburg, N., Baldwin, D.C., Storr, C.L., Williams, K.M., Anthony, J.C., & Sheehan, D.V. (1992). Prevalence of substance use among US physicians. *Journal of the American Medical Association*, 267, 2333-2339. doi: 10.1001/jama.267.17.2333
- Ingram, P.B., & Ternes, M.S. (2016). The detection of content-based invalid responding: A meta-analysis of the MMPI-2-Restructured Form's (MMPI-2-RF) over-reporting validity Scales. *The Clinical Neuropsychologist*, 30, 473-496. doi: 10.1080/13854046.2016.1187769
- Krueger, R.F., & Tackett, J.L. (2003). Personality and psychopathology: Working toward the bigger future. *Journal of Personality Disorders*, 17 (2), 109-128.
- Lee, F.J., Stewart, M., & Brown, J.B. (2008). Stress, Burnout, and strategies for reducing them. *Canadian Family Physician*. 54 (2), 234-235.

- Little, T. D. (2013). *Longitudinal structural equation modeling*. New York: Guilford Press.
- Loehlin, J.C., & Horn, J.M. (2012). How general is the “general factor of personality”? Evidence from the Texas Adoption Project. *Journal of Research in Personality*, 46, 655-663. doi: 10.1016/j.jrp.2012.07.004
- Marsh, H. W. (1989). Confirmatory factor analysis of multitrait-multimethod data: Many problems and a few solutions. *Applied Psychological Measurement*, 13, 335–361.
- Marsh, H.W., Ludtke, O., Muthen, B., Asparohov, T., Morin, A.J., Trautwein, U., & Nagengast, B. (2010). A new look at the big five structure through exploratory structural equation modeling. *Psychological Assessment*, 22, 471-491. doi: : 10.1037/a0019227.
- Marsh, H.W., Morin, A.J., Parker, P.D., & Kaur, G. (2014). Exploratory Structural Equation Modeling: An integration of the best features of exploratory and confirmatory factor analysis. *Annual Review of Clinical Psychology*, 10, 85-110. doi: 10.1146/annurev-clinpsy-032813-153700
- Meehl, P.E. (1946). Profile analysis of the Minnesota Multiphasic Personality Inventory in differential diagnosis. *Journal of Applied Psychology*, 30, 517-524. doi: 10.1037/h0062318
- Q-Local. (2009). Computer Software. Minneapolis, MN: Pearson Assessments.
- Ruitenbergh, M.M., Frings-Dresen, M.H.W., & Sluiter, J.K. (2012). The prevalence of common mental disorders among hospital physicians and their association with self-reported work ability: A cross-sectional study. *BMC Health Services Research*, 12, 292-298. doi: 10.1186/1472-6963-12-292
- Rushton, J.P., & Irwing, P. (2009a). A general factor of personality in the Million Clinical

- Multiaxial Inventory-III, the dimensional assessment of personality pathology, and the Personality Assessment Inventory. *Personality and Individual Differences*, 43, 1091-1095. doi:10.1016/j.jrp.2009.06.002
- Rushton, J.P., & Irwing, P. (2009b). A general factor of personality in 16 sets of the Big Five, the Guilford-Zimmerman Temperament Survey, the California Psychology Inventory, and the Temperament and Character inventory. *Personality and Individual Differences*, 47, 558-564. doi:10.1016/j.paid.2009.05.009
- Sellbom, M., & Bagby, R.M. (2008). Validity of the MMPI-2-RF (Restructured Form) L-r and K-r scales in the detecting of underreporting in clinical and nonclinical samples. *Psychological Assessment*, 20, 370-376. doi: 10.1037/a0012952
- Sellbom, M., Ben-Porath, Y.S., & Bagby, R.M. (2008). On the hierarchical structure of mood and anxiety disorders: confirmatory evidence and elaboration of a model of temperament markers. *Journal of Abnormal Psychology*, 117, 576-590. doi: 10.1037/a0012536.
- Shedler, J., & Westen, D. (2004). Dimensions of Personality Pathology: An Alternative to the Five-Factor Model. *American Journal of Psychiatry*, 161, 1743-1754.
- Simms, L. J., Casillas, A., Clark, L. A., Watson, D., & Doebbeling, B. N. (2005). Psychometric evaluation of the Restructured Clinical scales of the MMPI-2. *Psychological Assessment*, 17, 345-358. doi: 10.1037/1040-3590.17.3.345
- Tarescavage, A.M., Corey, D.M., & Ben-Porath, Y.S. (2014). Minnesota Multiphasic Personality Inventory-2-Restructured Form (MMPI-2-RF) predictors of police officer problem behavior. *Assessment*, 28, 116-132. doi: 10.1177/1073191114534885
- Tarescavage, A.M., Marek, R.J., Finn, J.A., Hicks, A., Rapier, J.L., & Ben-Porath, Y.S. (2013).

- Minnesota Multiphasic Personality Inventory-2-Restructured Form (MMPI-2-RF)
Normative Elevation Rates: Comparisons With Epidemiological Prevalence Rates. *The Clinical Neuropsychologist*, 27, 1106-1120. doi: 10.1080/13854046.2013.83286
- Tellegen, A., & Ben-Porath, Y. (2008/2011). Minnesota Multiphasic Personality Inventory–
Restructured Form: Technical manual. Minneapolis, MN: University of Minnesota Press.
- Tellegen, A., Ben-Porath, Y.S., & Sellbom, M. (2006). Construct validity of the MMPI-2
Restructured Clinical (RC) scales: Reply to Rouse, Greene, Butcher, Nichols, and
Williams. *Journal of Personality Assessment*, 91, 211-221.
- 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, 217-222. doi: 10.1207/s15327752jpa8702_12
- Wygant, D.B., Boutacoff, L.I., Arbisi, P.A., Ben-Porath, Y.S., Kelly, P.H., & Rupp, W.M.
(2007). Examination of the MMPI-2 Restructured Clinical (RC) scales in a sample of
bariatric surgery candidates. *Journal of Clinical Psychology in Medical Settings*, 14, 197-
205. doi: 10.1007/s10880-007-9073-8

Table 1. Content Descriptions of MMPI-2-RF Scales

Scale	Description
Validity Scales	
VRIN-r	Random Responding
TRIN-r	Fixed responding
F-r	Infrequent responses observed within general population
Fp-r	Infrequent responses in psychiatric population
Fs	Infrequent responses in medical patient population
FBS-r	Cognitive and somatic issues associated with higher over-reporting
RBS	Response bias scale compared to people who failed effort tests
L-r	Uncommon virtues
K-r	Adjustment validity represents avowal of positive well being
Higher-Order	
EID	Emotional and internalizing dysfunction
THD	Thought disordered dysfunction
BXD	Behavioral and externalizing dysfunction
Restructured Clinical (RC)	
RCd	Demoralization
RC1	Somatic complaints
RC2	Low positive emotion
RC3	Cynicism
RC4	Antisocial behavior
RC6	Ideas of persecution
RC7	Dysfunctional negative emotions
RC8	Aberrant experiences
RC9	Hypomanic activation
Specific Problem (SP)	
<i>Somatic</i>	
MLS	Malaise
GIC	Gastrointestinal complaints
HPC	Head pain complaints
NUC	Neurological complaints
COG	Cognitive complaints
<i>Internalizing</i>	
SUI	Suicidal and death ideation
HLP	Helplessness and hopelessness
SFD	Self-doubt
NFC	Inefficacy
STW	Stress and worry
AXY	Anxiety including pervasive nightmares and hypervigilance
ANP	Anger proneness
BRF	Behavior-restricting fears
MSF	Multiple specific fears

Table 1. *Continued.*

Specific Problem	
<i>Externalizing</i>	
JCP	Juvenile conduct problems
SUB	Substance abuse
AGG	Aggression
ACT	Activation
<i>Interpersonal</i>	
FML	Family problems
IPP	Interpersonal problem with unassertiveness and submissiveness
SAV	Social avoidance
SHY	Shyness
DSF	Disaffiliativeness
Psy-5	
AGGR-r	Aggressiveness and goal directed aggressiveness
PSYC-r	Psychoticism and disassociativeness
DISC-r	Disconstraint
NEGE-r	Negative emotionality and neuroticism
INTR-r	Introversion and low positive emotionality

Table 2. Descriptive Statistics for the Validity, Higher Order, and RC Scales

Scale	Mean T-Score			F-value	Partial η^2
	Male	Female	Combined		
Validity					
<i>TRIN-r</i>	56.0 (6.2)	55.2 (5.4)	55.9 (6.1)	0.822	0.001
<i>VRIN-r</i>	44.1 (8.5)	45.1 (7.7)	44.2 (8.4)	0.587	0.003
<i>F-r</i>	49.3 (13.7)	52.6 (15.3)	49.8 (14.0)	2.556	0.010
<i>Fp-r</i>	45.2 (6.4)	48.8 (13.4)	45.7 (8.0)	8.911**	0.032
<i>Fs-r</i>	47.3 (8.8)	51.1 (15.5)	47.9 (10.2)	6.325*	0.023
<i>FBS-r</i>	54.8 (9.3)	59.8 (13.5)	55.6 (10.2)	11.419**	0.041
<i>RBS</i>	54.6 (11.4)	55.8 (11.4)	55.0 (11.3)	-	-
<i>K-r</i>	59.5 (9.8)	56.5 (11.4)	59.0 (10.1)	1.502	0.006
<i>L-r</i>	56.8 (12.2)	56.0 (10.7)	56.7 (11.9)	0.067	0.000
Higher Order					
<i>EID</i>	47.9 (13.0)	52.5 (15.2)	48.7 (13.5)	1.041	0.004
<i>THD</i>	44.9 (8.5)	45.5 (9.2)	45.0 (8.6)	0.376	0.001
<i>BXD</i>	43.6 (8.0)	40.0 (7.1)	43.0 (8.0)	0.221	0.001
RC					
<i>RCd</i>	48.4 (12.7)	52.0 (14.5)	49.0 (13.0)	3.573	0.013
<i>RC1</i>	46.7 (10.1)	50.6 (13.2)	47.3 (10.7)	0.764	0.003
<i>RC2</i>	54.0 (12.7)	56.7 (13.4)	54.4 (12.8)	2.370	0.009
<i>RC3</i>	41.7 (8.2)	42.3 (8.3)	41.8 (8.2)	0.540	0.002
<i>RC4</i>	45.1 (9.0)	43.6 (8.2)	44.8 (8.9)	0.031	0.000
<i>RC6</i>	50.4 (10.4)	50.4 (11.8)	50.4 (10.6)	0.773	0.003
<i>RC7</i>	42.3 (9.2)	43.4 (6.8)	42.8 (9.7)	0.596	0.002
<i>RC8</i>	42.4 (7.1)	45.2 (11.4)	42.6 (7.0)	0.136	0.001
<i>RC9</i>	37.9 (7.7)	36.0 (9.0)	37.6 (8.0)	0.065	0.001

Note. MANOVA (multivariate analysis of variance) for sex across scales was 2.814 ($p < .001$, overall model partial $\eta^2 = .381$) using Wilks' Lambda. Numbers within parentheses are standard deviations. Bolded scores are significant, $*p < .05$, $**p < .01$. RBS was excluded from the MANOVA as only half the participants ($n = 181$; 159 male) had scoring profiles that incorporated that scale.

Table 3. Descriptive Statistics for the Specific Problem and PSY-5

Scale	Mean <i>T</i> -Score			F-value	<i>Partial</i> η^2
	Male (<i>n</i> = 226)	Female (<i>n</i> = 45)	Combined		
Somatic/Cognitive					
<i>MLS</i>	45.2 (8.9)	50.3 (13.3)	46.0 (9.8)	10.030**	0.036
<i>GIC</i>	42.2 (8.9)	47.6 (19.0)	43.1 (12.2)	7.111**	0.026
<i>HPC</i>	49.9 (10.3)	50.4 (10.9)	50.0 (10.4)	.059	0.000
<i>NUC</i>	47.3 (8.1)	50.4 (10.4)	47.8 (8.6)	4.924*	0.018
<i>COG</i>	48.9 (10.8)	50.2 (12.2)	49.1 (11.0)	.507	0.002
Internalizing					
<i>SUI</i>	48.9 (10.8)	50.2 (12.2)	49.1 (11.0)	.736	0.003
<i>HLP</i>	48.4 (11.7)	49.9 (13.4)	48.6 (11.9)	.550	0.002
<i>SFD</i>	47.3 (11.2)	51.1 (12.8)	47.9 (11.5)	3.875*	0.014
<i>NFC</i>	47.2 (11.1)	49.7 (11.5)	47.6 (11.1)	1.773	0.007
<i>STW</i>	45.7 (9.9)	49.7 (11.2)	46.3 (10.2)	5.644*	0.021
<i>AXY</i>	49.3 (28.5)	49.2 (13.0)	49.2 (26.6)	0.0003	0.000
<i>ANP</i>	47.7 (10.7)	51.1 (11.4)	48.2 (10.9)	3.422	0.130
<i>BRF</i>	46.9 (8.6)	51.1 (13.7)	48.7 (9.7)	9.589**	0.035
<i>MSF</i>	45.4 (9.4)	47.1 (9.9)	45.8 (9.5)	1.114	0.004
Externalizing					
<i>JCP</i>	45.0 (9.4)	47.1 (9.9)	45.7 (9.4)	0.126	0.000
<i>SUB</i>	45.0 (8.4)	46.7 (11.2)	45.2 (8.9)	1.294	0.005
<i>AGG</i>	44.4 (8.4)	42.7 (6.7)	44.1 (8.3)	1.355	0.005
<i>ACT</i>	44.6 (9.6)	41.6 (7.8)	44.1 (9.4)	3.494	0.062
Interpersonal					
<i>FML</i>	44.1 (21.2)	43.9 (9.0)	44.0 (19.8)	0.003	0.000
<i>IPP</i>	42.0 (10.4)	45.4 (12.8)	42.5 (10.8)	3.499	0.013
<i>SAV</i>	48.4 (23.7)	50.5 (12.8)	48.8 (22.3)	2.750	0.001
<i>SHY</i>	50.2 (11.3)	51.0 (10.6)	50.3 (11.2)	0.142	0.001
<i>DSF</i>	49.9 (10.6)	52.8 (12.5)	50.4 (10.9)	2.444	0.009
PSY-5					
<i>AGGR-r</i>	45.4 (7.6)	44.4 (7.1)	45.5 (7.5)	7.599	0.028
<i>PSYC-r</i>	43.4 (8.5)	44.7 (10.3)	43.5 (8.8)	14.007***	0.050
<i>DISC-r</i>	45.0 (7.7)	40.7 (7.1)	44.4 (7.7)	0.858	0.000
<i>NEGE-r</i>	46.2 (11.2)	47.9 (12.6)	46.4 (11.4)	10.178**	0.037
<i>INTR-r</i>	57.5 (11.7)	60.0 (12.9)	57.8 (11.9)	0.012	0.000

Note. MANOVA (multivariate analysis of variance) for sex across scales was 2.814 ($p < .001$, overall model partial $\eta^2 = .381$) using Wilks' Lambda. Numbers within parentheses are standard deviations. *D* = Cohen's *d* test for effect size. Bolded scores are significant, * $p < .05$, ** $p < .01$.

Table 4. Frequency of Higher Order and RC scale Elevation ≥ 65

Scale	Combined (%)	Males (%)	Females (%)
Higher Order			
<i>EID</i>	12.9	11.9	17.8
<i>THD</i>	3	3.1	2.2
<i>BXD</i>	1.1	1.3	*
RC			
<i>RCd</i>	14.1	13.3	18.2
<i>RC1</i>	5.6	4.0	13.6
<i>RC2</i>	17.8	16.8	22.7
<i>RC3</i>	2.6	2.2	4.5
<i>RC4</i>	3.7	4.4	*
<i>RC6</i>	12.6	12.8	11.4
<i>RC7</i>	4.4	4.0	6.8
<i>RC8</i>	2.6	2.2	4.5
<i>RC9</i>	0.7	0.4	2.3

Note. Percentages reflect the frequency of scale elevation of a T-score ≥ 65 and * denotes no clinical elevations.

Table 5. Frequency of Specific Problem and PSY-5 Elevation ≥ 65

Scale	Combined (%)	Males (%)	Females (%)
Somatic/Cognitive			
<i>MLS</i>	6.3	3.9	19
<i>GIC</i>	4.8	2.2	19
<i>HPC</i>	9.3	8.3	11.9
<i>NUC</i>	2.6	1.8	7.1
<i>COG</i>	6.3	5.3	11.1
Internalizing			
<i>SUI</i>	5.9	5.7	7.1
<i>HLP</i>	6.7	4.1	9.5
<i>SFD</i>	10.4	9.2	16.7
<i>NFC</i>	8.5	8.3	9.5
<i>STW</i>	4.8	4.4	7.1
<i>AXY</i>	6.7	6.1	7.1
<i>ANP</i>	7.5	7.0	11.9
<i>BRF</i>	1.5	1.3	7.1
<i>MSF</i>	6.7	6.6	7.1
Externalizing			
<i>JCP</i>	0.7	0.4	2.4
<i>SUB</i>	4.4	3.9	4.8
<i>AGG</i>	3.0	3.5	*
<i>ACT</i>	3.3	3.7	2.4
Interpersonal			
<i>FML</i>	3.0	3.1	2.4
<i>IPP</i>	6.3	5.3	11.9
<i>SAV</i>	9.3	8.8	19.0
<i>SHY</i>	13.7	13.2	16.7
<i>DSF</i>	9.6	8.3	16.7
PSY-5			
<i>AGGR-r</i>	1.1	1.3	*
<i>PSYC-r</i>	4.8	5.7	*
<i>DISC-r</i>	3.3	3.5	2.4
<i>NEGE-r</i>	5.6	4.4	11.9
<i>INTR-r</i>	10.0	8.8	16.7

Note. Percentages reflect the frequency of scale elevation of a T-score ≥ 65 and * denotes no clinical elevations.

Table 6. MMPI-2-RF Means Contrasted between this Sample and Comparison Groups

Scale	Sample Mean T-Scores			Absolute Difference	
	Current Sample	Independent Outpatient	Normative Sample	Independent Outpatient	Normative Sample
TRIN-r	55.9	50.0	50	5.9	5.9
VRIN-r	44.2	50.5	50	6.3	5.8
F-r	49.8	59.0	50	9.2	0.2
Fp-r	45.7	52.0	50	6.3	4.3
Fs-r	47.9	55.0	50	7.1	2.1
FBS-r	55.6	61.5	50	5.9	5.6
K-r	59.0	48.0	50	11.0	9.0
L-r	56.7	55.0	50	1.7	6.7
EID	48.7	58.0	50	9.3	1.3
THD	45.0	55.0	50	10.0	5.0
BXD	43.0	51.5	50	8.5	7.0
RCd	49.0	59.0	50	10.0	1.0
RC1	47.3	56.5	50	9.2	2.7
RC2	54.4	56.5	50	2.1	4.4
RC3	41.8	49.5	50	7.7	8.2
RC4	44.8	58.0	52	13.2	7.2
RC6	50.4	54.0	51	3.6	0.6
RC7	42.8	53.5	48	10.7	5.2
RC8	42.6	50.0	50	7.4	7.4
RC9	37.6	47.0	51	9.4	13.4
Mean (SD)	48.1 (5.8)	54 (4.1)	50.1 (0.7)	7.7 (3)	5.2 (3.3)

Note. Mean differences are reported using absolute values. Normative samples are drawn, and reported according to, information provided in the MMPI-2-RF technical manual (Tellegen & Ben-Porath, 2008/2011).

Table 7. Summary Table of CFA Analyses

Corrective Iteration	Correction Type	(df) χ^2	RMSEA	CFI	TLI
0	Initial Model	(347) 2541.407	.153 (.147-.158)	.693	.665
1	INTR with RC2	(346) 2292.295	.144 (.138-.150)	.728	.702
2	RC6 with THD	(345) 2207.126	.141 (.136-.147)	.739	.714
3	NEGE with STW	(344) 2096.368	.137 (.131-.143)	.755	.730
4	NEGE with DISC	(343) 1986.413	.133 (.127-.139)	.770	.747
5	SUB with RC4	(342) 1895.884	.129 (.124-.135)	.783	.760
6	AGGR with RC9	(341) 1820.047	.127 (.121-.132)	.793	.771
7	ACT with RC9	(340) 1720.676	.122 (.117-.128)	.807	.785
8	NEGE with RC7	(339) 1649.279	.119 (.114-.125)	.817	.796
9	RC6 with RC8	(338) 1597.290	.117 (.111-.123)	.824	.803
10	PSYC with THD	(337) 1547.626	.115 (.109-.121)	.831	.810
11	Initial + CU	(314) 2043.318	.143 (.137-.148)	.758	.709

Table 8. Summary Table of ESEM Analyses

Corrective Iteration	Correction Type	(df) χ^2	RMSEA	CFI	TLI
0	Initial Model				
1	INTR with RC2	(271) 1208.413	.113 (.107 - .120)	.869	.817
2	NEGE with STW	(270) 1130.040	.108 (.102 - .115)	.880	.831
3	NEGE with DISC	(269) 1021.105	.102 (.095 - .108)	.895	.852
4	RC6 with THD	(268) 925.531	.095 (.088 - .102)	.908	.870
5	RC6 with RC8	(267) 858.228	.090 (.084 - .097)	.917	.883
6	RC4 with BXD	(266) 801.855	.086 (.079 - .093)	.925	.893
7	ACT with RC9	(265) 775.924	.084 (.077 - .091)	.928	.898
8	PSYC with THD	(264) 750.502	.082 (.076 - .089)	.932	.903
9	PSYC with RC6	(263) 708.078	.079 (.072 - .086)	.938	.910
10	NEGE with RC7	(262) 679.189	.077 (.070 - .084)	.942	.916
11	Initial + CU	(264) 1541.999	.134 (.127 - .140)	.821	.744

Note. All models observed during ESEM analysis had a negative variance for BXD

Table 9. Exploratory Factor Analysis Structural Matrix

MMPI-2-RF Scale	Extracted Factor			
	1	2	3	4
<i>EID</i>	*.981	.523	.435	.053
<i>THD</i>	.432	.285	*.726	.349
<i>BXD</i>	.455	*.888	.506	.622
<i>RCd</i>	*.942	.542	.456	.097
<i>RC2</i>	*.729	.359	.126	-.350
<i>RC4</i>	.588	*.936	.482	.320
<i>RC6</i>	.446	.276	*.498	.318
<i>RC7</i>	*.865	.505	.680	.368
<i>RC8</i>	.515	.353	*.871	.273
<i>RC9</i>	.406	.467	.674	*.854
<i>SUI</i>	*.686	.382	.447	.080
<i>HLP</i>	*.667	.333	.415	.071
<i>SFD</i>	*.787	.409	.406	.056
<i>NFC</i>	*.751	.440	.402	.105
<i>STW</i>	*.822	.421	.579	.299
<i>AXY</i>	.259	.182	.250	.046
<i>ANP</i>	*.654	.369	.458	.456
<i>BRF</i>	.519	.269	.551	.143
<i>MSF</i>	.259	.177	.344	.190
<i>JCP</i>	.351	*.729	.426	.351
<i>SUB</i>	.407	*.713	.219	.092
<i>AGG</i>	.550	.474	.551	*.552
<i>ACT</i>	.455	.372	.733	*.505
<i>AGGR-r</i>	-.151	.056	.132	*.647
<i>PSYC-r</i>	.485	.321	*.791	.320
<i>DISC-r</i>	.127	.299	.004	.001
<i>NEGE-r</i>	*.872	.478	.652	.388
<i>INTR-r</i>	*.529	.209	-.028	-.476

Note. Bolded Items indicated their expected factor placement within the three factor model proposed as the interpretive structure by Tellegen and Ben-Porath (2008/2011). Items with a star are those aligned independently for a given factor.

Table 10. Inter-domain correlations and Eigenvalues of factors extracted during EFA

Factor	1	2	3	Eigenvalues	
				Extracted	Simulation
1	-			11.569	1.653
2	.54	-		3.302	1.551
3	.54	.43	-	2.179	1.475
4	.17	.24	.54	1.426	1.417
5	-	-	-	1.104	1.362

Note. Correlations unreported for fifth factor as it failed to meet extraction criteria. Significance levels are not reported for factor correlations because EFA calculates correlations using a weighted component sum for each score. These factors include all possible component items using associated loading weights, which are different between factors (Table 9).

Figure 1. Proposed structure for the three higher-order themes of the MMPI-2-RF

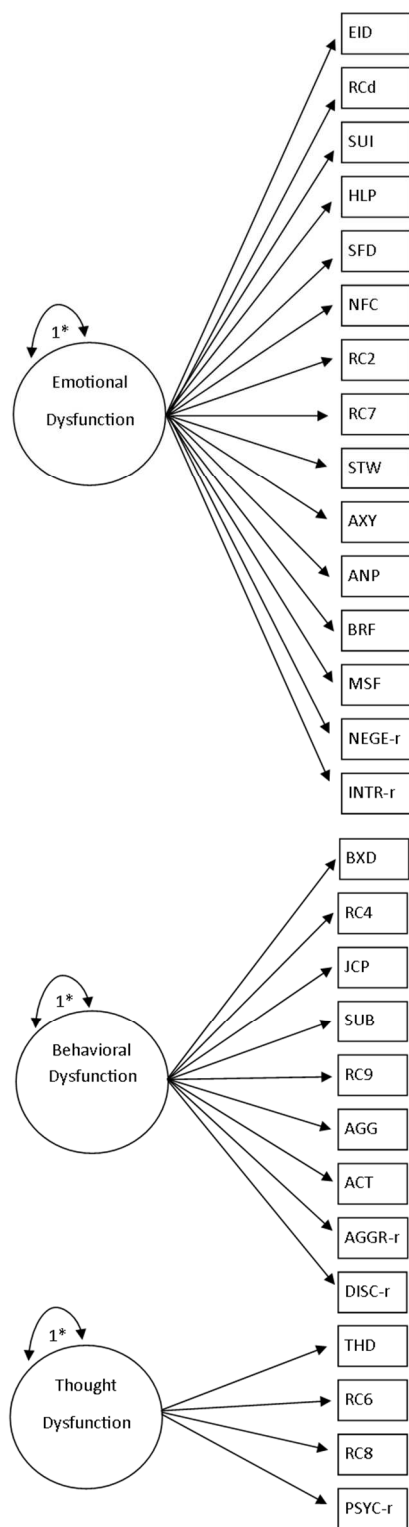


Figure 2. CFA following Data-based and Theory-based Corrections

Figure 2a

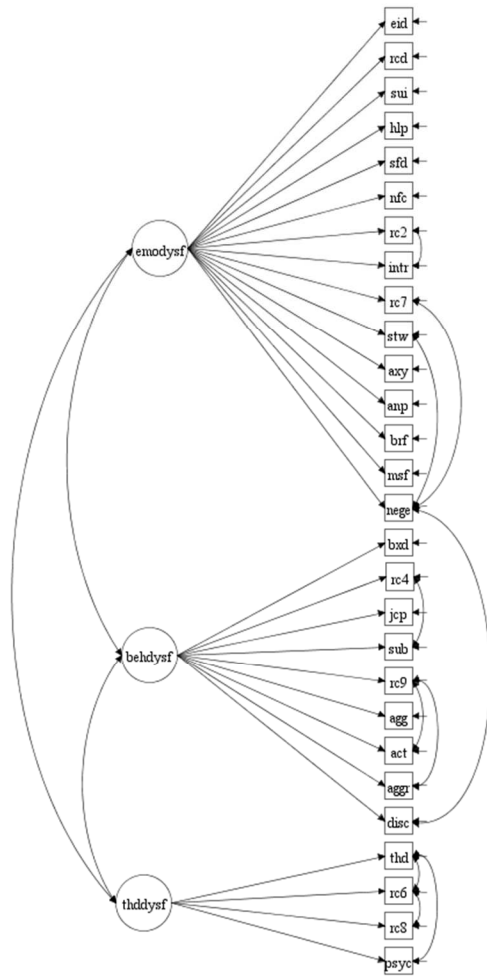
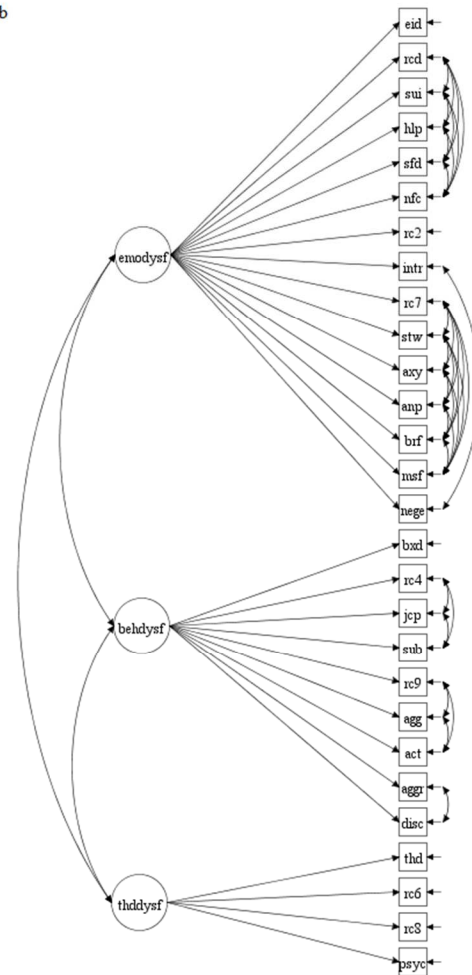


Figure 2b



Note. Data based corrections are in Figure 2a and theory based CU corrects are in Figure 2b.

Figure 3. Scree Plot for Observed Eigenvalues in Exploratory Factor Analysis

