PsychCorp

MMPI-2-RF®

Minnesota Multiphasic Personality Inventory-2 Restructured Form®

Auke Tellegen
Yossef S. Ben-Porath

Technical Manual

ALWAYS LEARNING

PEARSON

Published by: University of Minnesota Press
Distributed by: Pearson
P.O. Box 1416
Minneapolis, MN 55440
800.627.7271
www.PearsonClinical.com

MMPI-2-RF (Minnesota Multiphasic Personality Inventory-2-Restructured Form) Technical Manual

PEARSON

Chapter 1
Introduction

The Minnesota Multiphasic Personality Inventory-2-Restructured Form® (MMPI-2-RF®) is a broad-spectrum inventory, the result of a data-based and construct-oriented revision of the MMPI®-2. The overall objective of this revision was to represent the clinically significant substance of the MMPI-2 item pool with a comprehensive set of psychometrically adequate measures. The 338 items of the MMPI-2-RF accommodate a total of 51 scales: 9 Validity Scales and 42 substantive scales.

The 9 Validity Scales consist of 7 revised MMPI-2 validity indicators and two new scales. The 42 substantive scales consist of the 9 previously developed Restructured Clinical (RC) Scales (Tellegen et al., 2003) and 33 new scales. The latter comprise 3 Higher-Order Scales measuring broad areas of dysfunction (Emotional/Internalizing, Thought, and Behavioral/Externalizing Dysfunction), 23 Specific Problems Scales, 2 Interest Scales, and 5 revised Personality Psychopathology Five (PSY-5) Scales. The full set of 51 scales is listed in Table 1-1.

The general rationale and methodology for developing the RC Scales also served to guide construction of the 33 additional substantive measures. We first examine the reasons for this sequence.

Why First Restructure the Clinical Scales?

As the name implies, the RC Scales are the result of an effort to restructure the MMPI-2 Clinical Scales. Two jointly compelling observations motivated this first step: (1) many of the Clinical Scale *items* capture important distinctive features of the major psychopathologies because of the broad range and largely empirical derivation of the Clinical Scales; (2) however, it has long been recognized that as *aggregate measures* the Clinical Scales are not psychometrically optimal. In the following, we elaborate on these observations and on the implications they had for the restructuring process.

1. The Clinical Scales Contain Clinically Informative Items

The special strengths of the MMPI-2 Clinical Scales derive from Hathaway and McKinley's well-known empirical method of selecting items. For each Clinical Scale, they selected from a diverse and clinically relevant item pool those items whose endorsement frequencies differentiated a carefully diagnosed criterion group from a representative non-patient sample. In their own terms, "Every item finally chosen differentiates between criterion groups and normal groups and that is the reason for acceptance or rejection of the items. They are not selected for their content or theoretical import. Frequently the authors can see no possible rationale to an item in a given scale; it is nevertheless accepted if it appears to differentiate" (McKinley & Hathaway, 1944/2000, pp. 31–32). In this radically empirical manner (though with significant exceptions, two of which we consider shortly), Hathaway and McKinley constructed eight Clinical Scales (Hs, D, Hy, Pd, Pa, Pt, Sc, and Ma) targeting major psychiatric disorders.

Over the years, Hathaway and McKinley's item selection procedure has stimulated much debate. Best known among the early assessments is Meehl's vigorous advocacy of empirical keying (e.g., Meehl, 1945). Subsequent commentators, including Jackson (1971) and his associates (Helmes & Reddon, 1993), have countered with persistent criticism and have advocated a "construct-oriented" approach to test construction. Meehl himself, while still strongly

Table 1-1.

The MMPI-2-RF Scales

Validity Scales	
VRIN-r	Variable Response Inconsistency – Random responding
TRIN-r	True Response Inconsistency – Fixed responding
F-r	Infrequent Responses – Responses infrequent in the general population
Fp-r	Infrequent Psychopathology Responses – Responses infrequent in psychiatric populations
Fs	Infrequent Somatic Responses – Somatic complaints infrequent in medical patient populations
FBS-r	Symptom Validity – Non-credible somatic and cognitive complaints
RBS	Response Bias Scale – Non-credible memory complaints
L-r	Uncommon Virtues – Rarely claimed moral attributes or activities
K-r	Adjustment Validity – Uncommonly high level of psychological adjustment
Higher-Order (H-O) Scales	
EID	Emotional/Internalizing Dysfunction – Problems associated with mood and affect
THD	Thought Dysfunction – Problems associated with disordered thinking
BXD	Behavioral/Externalizing Dysfunction – Problems associated with under-controlled behavior
Restructured Clinical (RC) Scales	
RCd	Demoralization – General unhappiness and dissatisfaction
RC1	Somatic Complaints – Diffuse physical health complaints
RC2	Low Positive Emotions – A distinctive, core vulnerability factor in depression
RC3	Cynicism – Non-self-referential beliefs that others are bad and not to be trusted
RC4	Antisocial Behavior – Rule breaking and irresponsible behavior
RC6	Ideas of Persecution – Self-referential beliefs that others pose a threat
RC7	Dysfunctional Negative Emotions – Maladaptive anxiety, anger, and irritability
RC8	Aberrant Experiences – Unusual perceptions or thoughts associated with psychosis
RC9	Hypomanic Activation – Over-activation, aggression, impulsivity, and grandiosity
Specific Problems (SP) Scales	
Somatic/Cognitive Scales	
MLS	Malaise – Overall sense of physical debilitation, poor health
GIC	Gastrointestinal Complaints – Nausea, recurring upset stomach, and poor appetite
HPC	Head Pain Complaints – Head and neck pain
NUC	Neurological Complaints – Dizziness, weakness, paralysis, loss of balance, etc.
COG	Cognitive Complaints – Memory problems, difficulties concentrating
Internalizing Scales	
SUI	Suicidal/Death Ideation – Direct reports of suicidal ideation and recent suicide attempts
HLP	Helplessness/Hopelessness – Belief that goals cannot be reached or problems solved
SFD	Self-Doubt – Lack of self-confidence, feelings of uselessness
NFC	Inefficacy – Belief that one is indecisive and inefficacious
STW	Stress/Worry – Preoccupation with disappointments, difficulty with time pressure
AXY	Anxiety – Pervasive anxiety, frights, frequent nightmares
ANP	Anger Proneness – Easily angered, impatient with others
BRF	Behavior-Restricting Fears – Fears that significantly inhibit normal behavior
MSF	Multiple Specific Fears – Fear of blood, fire, thunder, etc.

Table 1-1. (continued)

The MMPI-2-RF Scales

Externalizing Scales	
JCP	Juvenile Conduct Problems – Difficulties at school and at home, stealing
SUB	Substance Abuse – Current and past misuse of alcohol and drugs
AGG	Aggression – Physically aggressive, violent behavior
ACT	Activation – Heightened excitation and energy level
Interpersonal Scales	
FML	Family Problems – Conflictual family relationships
IPP	Interpersonal Passivity – Being unassertive and submissive
SAV	Social Avoidance – Avoiding or not enjoying social events
SHY	Shyness – Feeling uncomfortable and anxious around others
DSF	Disaffiliativeness – Disliking people and being around them
Interest Scales	
AES	Aesthetic-Literary Interests – Literature, music, the theater
MEC	Mechanical-Physical Interests – Fixing and building things, the outdoors, sports
Personality Psychopathology Five (PSY-5) Scales	
AGGR-r	Aggressiveness-Revised – Instrumental, goal-directed aggression
PSYC-r	Psychoticism-Revised – Disconnection from reality
DISC-r	Disconstraint-Revised – Under-controlled behavior
NEGE-r	Negative Emotionality/Neuroticism-Revised – Anxiety, insecurity, worry, and fear
INTR-r	Introversion/Low Positive Emotionality-Revised – Social disengagement and anhedonia

supporting empirical keying, described his earlier position as "overly 'dust-bowl empiricist'" (Meehl, 1971). The contemporary consensus, it seems fair to say, is that a program of test construction based on truly "blind" empiricism, deliberately stripped of significant conceptual inputs regarding the content and structure of one's measures and relying on superficial and ephemeral criteria, is not viable in the long run. It is now well recognized that such an approach jeopardizes the broader and longer-term personological and clinical significance and usefulness (the construct validity) of these measures (e.g., Jackson, 1971; Loevinger, 1957; Travers, 1951).

However, some reflection on how the Clinical Scales were actually developed shows that Hathaway and McKinley's grand scale construction strategy, while granting little weight to psychological speculation, could not have been "blindly" empirical. "External [i.e., empirical-keying based] scale construction does not take place in a conceptual vacuum. It requires choosing reputational, diagnostic, or life-record criteria, and assembling an item pool. Presumably, these choices are guided by some conception of the criterion variable: why it is important and how it relates to self-reports" (Tellegen & Waller, 2008).

Hathaway and McKinley's scale derivation program was clearly anchored to psychiatric nosology, even as they recognized its imperfections. If they had been unable to assume that disorders such as schizophrenia and depression are "real" (rather than, say, social constructions), the MMPI would never have gotten off the ground. Consistent with their nosological orientation, the MMPI item pool was not a random collection of self-descriptive statements but was assembled from clinically relevant sources, such as psychiatric textbooks and examination guidelines. One can also assume that included items were not selected "blindly" as to targeted disorders. Undoubtedly items with persecutory content were selected as promising candidates for a paranoia scale and not as equally promising candidates for a depression scale.

Having developed a diagnostically and clinically informed item pool, Hathaway and McKinley did

rely on empirical keying for the final assignments of items to scales. We believe that their empiricism in this regard was well taken. It made good sense to let the observed self-portrayal differences between normative and criterion groups have the final say when criterion groups were made up of patients whose affective, cognitive, and behavioral characteristics, including their verbal behavior, were (and still are) incompletely understood.

The item-assignment policy described for the MMPI Clinical Scales is actually analogous to the "empirical evaluation of theoretically defined items" of Jackson's (1970) well-known and avowedly construct-oriented "sequential system for personality scale development," with two crucial exceptions. Jackson's empirical evaluation is based not on external but on internal correlations, which are used to select items with the best internally discriminant and convergent features. And his system only permits the empirically based assignment of an item to a scale if the item is among those already included (on theoretical grounds) in the pool of candidate items for that particular scale (nor do we find provisions for reversing on empirical grounds the original theoretically decided positive or negative keying of an item). In other words, conceptual considerations enter into both Jackson's and Hathaway's approaches to scale construction. However, Jackson's method does not include external criteria and prohibits item selections that conflict with prior theory, while Hathaway's allows external correlates to override tentative prior conjectures.

In conclusion, we consider Hathaway's method of item selection through empirical keying to be both the strongest and the most distinctive component of his scale construction strategy. Because of it, each of the eight Clinical Scales is a uniquely valuable item repository—a specialized pool certain to include clinically discriminating indicators of its targeted disorder.

2. The Clinical Scales Are Not Optimal Aggregate Measures

The total raw score of each Clinical Scale was obtained using the common and simple additive algorithm of summing the keyed item responses (binary in the case of the MMPI). Although the Clinical Scales contain discriminating items, it has been long recognized that as aggregate measures they are not optimal (e.g., Jackson, 1971; Norman, 1972). Two features of the Clinical Scales stand out as the most problematic: (a) *higher than theoretically expected intercorrelations*, and (b) *overly heterogeneous item content* (Tellegen et al., 2003). The extensive item overlap between the Clinical Scales is also often mentioned (e.g., Helmes & Reddon, 1993) but is best seen as a concomitant phenomenon, a secondary problem that is readily addressed once the other two have been resolved. How did these two features come about, and why are they problematic?

A major factor increasing Clinical Scale intercorrelations is the emotionally colored MMPI-2 dimension we have called *Demoralization*. Persons high on Demoralization describe themselves as discouraged, helpless, having low self-esteem, expecting to fail or having failed in various aspects of their lives, and despairing. This pervasive latent variable is a strong representative of the so-called first factor (Tellegen et al., 2006) and an ingredient of many MMPI-2 scales, including the Clinical Scales.

The pervasiveness of this factor is most likely an unintended consequence of Hathaway and McKinley's method of identifying discriminating items. The probable result of comparing each diagnostic criterion group with a sample of "Minnesota normals" was not just to select for each Clinical Scale those items that were specifically associated with belonging to that scale's particular criterion group. Also likely to have been included were Demoralization markers, items that distinguish psychiatric patients in general from non-patients and are indicative of a "sense of patienthood" (Tellegen et al., 2003).

We adopted the Demoralization label from Jerome Frank (1973, 1974a, 1974b, 1985), who used it to characterize self-appraisals that closely match our own descriptors. Frank asserted that Demoralization is common in medical and psychiatric illnesses and is the main and common problem of patients seeking psychotherapy. Our "sense of patienthood" interpretation (Tellegen et al., 2003) is obviously congruent with Frank's view. More broadly, the constructs of "nonspecific psychological distress" (Dohrenwend, Shrout, Egri, & Mendelsohn, 1980), "suffering" (Cassell, 1991), "giving up" (Engel, 1967), "internalization" (Krueger, 1999; Markon, Krueger, & Watson, 2005), and "emotional disorders" (Watson, 2005) refer to the same or nearly the same affective/experiential domain. There is particularly compelling evidence that Demoralization parallels the highest-order Pleasantness-versus-Unpleasantness (Happiness-versus-Unhappiness) dimension of self-report affect inventories (Tellegen, Watson, & Clark, 1999a, 1999b) and the corresponding same-named dimension in the Watson and Tellegen (1985) mood model.

Demoralization has clearly emerged as a clinically important variable, and the MMPI-2-RF appropriately includes a Demoralization measure (RCd) to allow for its separate assessment. However, the substantial presence of Demoralization in MMPI-2 scales targeting non-Demoralization attributes creates problems. It inflates the intercorrelations between these scales and attenuates their convergent-versus-discriminant validity differences. For example, one would have expected two of the Clinical Scales, Pt and Sc, to be distinctive indicators of neuroticism (emotional dysfunction) and psychoticism (thought dysfunction), respectively. However, Pt and Sc correlate .9 in clinical samples (Tellegen et al., 2003) because both are saturated with Demoralization, to the detriment of their differential validities. One could expect a reduction of such saturations to improve the scales.

Hathaway himself observed that items from the Depression Scale tended "naturally to recur in most other scales." He acknowledged the necessity of omitting some of these items if one wanted to prevent scale intercorrelations from being "undesirably high" (McKinley & Hathaway, 1944/2000, p. 32), including the correlation between Pt and Sc (Hathaway, 1956/2000, p. 49). He attributed the recurrence of depression items to the co-occurrence (comorbidity) of depression and other psychiatric disorders, a link he characterized as easily understood (McKinley & Hathaway, 1944/2000, p. 32). This concern prompted him to make exceptions to his empirical item selection rule and remove several of these initially selected items from the non-depression scales. What Hathaway referred to as depression we have reinterpreted as Demoralization, an important component of Clinical Scale D, but not its distinctive core, which we have identified as Low Positive Emotionality. The differentiation between Demoralization and Low Positive Emotionality is in essential agreement with contemporary two-component models of depression (see Tellegen et al., 2006, pp. 156–161).

Although the influence of Demoralization was the most widespread, it was not the only source of inflated scale intercorrelations to come to Hathaway's attention and to be dealt with by him, albeit with only partial success. For example, he also removed a number of initially selected somatic items from the Hy Scale to lower an excessively high correlation between it and the Hs Scale.

The Clinical Scales tend to be heterogeneous, if in varying degrees. Clinical Scale Hs is relatively univocal, but even beyond the Demoralization component the multidimensionality and structural fissures within some Clinical Scales, such as Hy, Pd, and Pa, are generally acknowledged to be quite prominent (e.g., Nichols, 2001, pp. 117, 143).

One source of heterogeneity is noise. As a result of inevitable sampling or other errors, the Clinical Scales undoubtedly include questionable items. Many so-called subtle items appear to belong in this category (e.g., Weed, Ben-Porath, & Butcher, 1990). A lack of systematic cross-validation as part of the scale development process (owing to the limited resources available to Hathaway and McKinley at the time) undoubtedly contributed to these occurrences (Graham, 2011). One of the reasons to revisit the Clinical Scales was simply to follow up on Hathaway and McKinley's item choices and identify robust indicators among those included in each Clinical Scale item set.

What if, thanks to rigorous screening, sampling error could be undone and spuriously "subtle" items eliminated? The MMPI authors themselves maintained that "where the clinically recognized diagnosis is impure, the scales will be impure" and often contain "deliberately diverse types of items" (McKinley & Hathaway, 1944/2000, p. 31). Put somewhat differently, one would still expect the Clinical Scales to be heterogeneous by virtue of "containing a number of items that correspond to important but disparate features that characterize the major psychiatric syndromes" (Nichols, 2001, p. 82). In general, we can expect empirical keying to yield items reflecting the complexity and heterogeneity of the criterion (provided, of course, that the range and sensitivity of the item pool can accommodate its varied features).

The Hy Scale is often singled out as an especially apt example of a heterogeneous scale because of its striking mix of items expressing demoralization, somatic complaints, denial of social anxiety or shyness, and of aggressive impulses, and disavowal of cynicism, many of which are even negatively intercorrelated as keyed. These items clearly reflect the paradoxical character of the hysterical syndrome itself (as defined by Hathaway and McKinley). They do not form a unitary structure in a general patient or non-patient population any more than do the actual syndrome features. Nonetheless, as mentioned earlier, the main principle guiding construction of each of the original MMPI scales was to include in a single additive scale all available items predicting a particular syndrome. At the time, this "one-scale-one-syndrome" strategy (Tellegen et al., 2006) was acceptable and could be

viewed as pragmatic. In light of what are now recognized as basic psychometric principles and methodology, this approach can no longer be recommended.

Forty years ago Nunnally (1967) called it a "fallacy" to assume that a heterogeneous variable could "be predicted with one test, which to be effective must be heterogeneous in content" (p. 248). In the case of the Clinical Scales, each scale, viewed factor analytically, would have to implicitly model a factorially complex syndrome as a particular weighted sum of the item factor scales representing the different syndrome features.

However, as Nunnally observed, with a purely criterion-oriented approach, the factors underlying the items of an empirically keyed scale and the weights these factors carry in this composite scale remain unknown. They depend on how many items marking each factor happened to be in the chosen item pool and were therefore included in the scale. These unknown weights of unknown factors are almost certainly non-optimal, which means that empirically keyed scales such as the Clinical Scales should be less than optimal even for predicting the designated criterion for each. It is interesting that Hathaway acknowledged a limited aspect of this problem when he discussed the case of groups of near-identical valid items that would be overweighted if all were included in a scale (Hathaway, 1956/2000, pp. 50–51). In these cases of quasi-redundancy he permitted himself to take content into account, albeit subjectively, without recourse to objective weighting solutions.

As an alternative to developing scales through empirical keying, Nunnally recommended that the predictor domain be represented dimensionally by meaningful factor scales. In his view it should be possible to predict effectively not just one designated criterion but a wide variety of different criteria by combining these scales in different multiple regression formulas. In Hathaway's time, hysteria would presumably have been among these criteria. Yet, in seeming contradiction of Nunnally's plausible argument in favor of multiple regression methods, empirical analyses and simulation studies appear to suggest otherwise. Dana and Dawes (2004) report finding that non-optimized prediction weights, including equal weights, are on cross-validation superior to multiple regression weights when predictive power is modest, as is often true in psychology. They conclude that "regression is rarely useful for prediction in most social science contexts" (p. 317).

Dana and Dawes's (2004) empirical analyses focus on how best to predict a single criterion using measures selected for that specific purpose. However, in several comprehensive MMPI-2 studies (e.g., Tellegen et al., 2003), predictor scales and criteria are diverse, scale-criterion correlations vary substantially in size, and the research samples are large. In such cases, some form of differential weighting is surely called for. For example, in analyses using several MMPI-2 scales as predictors of psychopathy and its two facets, Sellbom, Ben-Porath, Lilienfeld, Patrick, and Graham (2005) found linear multiple regression results to cross-validate well. Continued examination of alternative methods of choosing predictors and weighting the ones chosen would be useful in this area.

All the multivariate formulas considered so far conform to an *additive* model, also referred to as *linear* or *compensatory*. The term *compensatory* refers to the idea that high scores on some predictors can make up for low scores on others. For example, suppose that aptitude measure A and motivation measure M correlate positively and equally with task performance measure P, and that A and M are standardized and summed to predict P. Then three very different groups of people—those who score average on A and M, those who score well above the mean on A and equally far below the mean on M, and those with the opposite pattern—will be predicted to score average on P. In other words, the additive formula predicts that a high A score can completely compensate for a correspondingly low M score and vice versa.

However, a simple additive model may not be adequate. Rather, a *configural* model may be called for, allowing M and A to *interact* in relation to P. Of the three groups just described, the group scoring average on A and M may outperform the other two because very low motivation may prevent the expression of outstanding aptitude, and very limited aptitude may render even outstanding motivation ineffectual. In other words, a *conjunction* of attributes is needed to ensure adequate performance since the contribution of M "moderates" (in this case, potentiates) the contribution of A and vice versa. Given this kind of interaction, a regression equation including the product (or interaction) term MA, in addition to the M and A terms, should improve prediction. A non-additive model may similarly be needed to predict clinical criterion variables.

A categorical form of configural assessment is the *multiple-hurdles model*. In our example, this model

might require a simple conjunction of adequate levels of both A and M as a condition for predicting, say, a "satisfactory" instead of an "unsatisfactory" performance level. In the clinical domain, the rules for diagnosing psychiatric syndromes are examples of multiple-hurdles algorithms (provided the syndrome elements are not all markers of the same latent dimension). MMPI-2 Clinical Scale profile classifications illustrate (variants of) the same approach.

Despite the intuitive plausibility or at least the natural appeal of configural patterns, we are not aware of convincing demonstrations. Goldberg's (1969) exhaustive analyses of how best to differentiate neurotic from psychotic patients demonstrated that a linear combination of MMPI Clinical Scales outperformed or matched a variety of configural algorithms. However, the high intercorrelations, item overlap, heterogeneity, and noisiness of the Clinical Scales, and the restricted range of criteria investigated, may have worked against detecting genuine configural effects. For practical and theoretical reasons it would be informative to re-examine non-linear estimates using the better delineated RC Scales and other MMPI-2-RF scales as predictors and a wide range of criteria as targets.

Development of the RC Scales

Each Clinical Scale was originally intended to predict one particular syndrome. As noted earlier, to the extent this intention was realized, each scale in effect represents a factorially complex syndromal target as a particular weighted sum of item factor scales corresponding to the different syndromal features. For reasons we just discussed, the aim of the restructuring effort was to represent salient components of these heterogeneous indices with a set of separate and substantively as well as structurally coherent and distinctive measures.

As reported previously (Tellegen et al., 2003), the derivation of the RC Scales did not conform to one particular test construction method or recipe. With respect to content, the Demoralization construct was introduced and examined within the framework of Tellegen et al.'s (1999a, 1999b) model of affect, additional concepts were developed as required to interpret emerging empirical findings within a contemporary framework of personality and psychopathology, and specific item content considerations influenced decisions throughout. Methodologically, exploratory factor analyses guided dimensional choices and delineations toward parsimony and pattern simplicity, and additional correlational analyses served to monitor and optimize the internal and intercorrelational structures and external validities of the evolving scales. The interplay between ideas and empirical analyses that eventuated in the RC Scales can be broken down into the following four-step sequence:

1. **Capturing Demoralization.** In a series of analyses, the Demoralization construct, hypothesized to parallel the general bipolar "Pleasantness-versus-Unpleasantness" dimension of self-reported mood (Tellegen et al., 1999a, 1999b; Watson & Tellegen, 1985), was empirically tested and (as noted earlier) corroborated. On the basis of these analyses, a set of Demoralization marker items was identified.

2. **Identifying major distinctive components of the Clinical Scales.** The items of each Clinical Scale were combined with the Demoralization items, and each of these item sets was separately factor analyzed. These analyses were performed on the eight designated Clinical Scales, as well as on the Mf and Si Scales, each augmented with Demoralization items. A Demoralization factor was recovered from each item set, defined, as expected, by not only the original Demoralization markers but by significant numbers of Clinical Scale items as well. In addition, two- or three-factor rotations allowed identification of a meaningful core non-Demoralization factor in 9 of the 10 item sets, and a four-factor rotation uncovered 2 such factors in the Mf set, such that the 11 identified core factors were mutually distinctive.

3. **Constructing seed scales.** Carefully delineated *seed scales,* designed to be optimally distinctive, were developed to represent Demoralization and the 11 identified core factors.

4. **Deriving final RC Scales.** Exhaustive correlational analyses were conducted, in which the 12 seed scales were correlated with all 567 MMPI-2 items (with due corrections for item overlap), to identify a set of adequately converging and discriminating candidate members for each of the nine RC Scales (i.e., the scales intended to represent Demoralization and the eight core factors of the eight designated Clinical Scales). These analyses, combined with evaluations of item content, item-scale correlations, and available external item validity data, resulted in the final RC Scale set.

A detailed description of this four-step development, proceeding from the Demoralization analyses to the final RC Scales, is presented in the Tellegen et al. (2003) monograph.

Psychometric Findings with the RC Scales

Tellegen et al. (2003) reported the results of comprehensive reliability and validity analyses of the RC Scales based on data from several thousand subjects who completed the MMPI-2 at mental health inpatient and outpatient facilities. In comparisons with their Clinical Scale counterparts, the considerably shorter RC Scales were found to be about equally or more reliable, to be less saturated overall with Demoralization, to be less highly intercorrelated, and to achieve comparable to improved convergent validities and substantially improved discriminant validities.

Additional RC Scale findings have subsequently been reported on the basis of data collected in a wide range of settings. In mental health settings, RC Scale results have been reported for outpatients (Wallace & Liljequist, 2005), psychiatric inpatients (Arbisi, Sellbom, & Ben-Porath, 2008; Handel & Archer, 2008), private practice outpatients (Sellbom, Graham, & Schenk, 2006), college counseling clients (Castro, Gordon, Brown, Anetsis, & Joiner, 2008; Sellbom, Ben-Porath, & Graham, 2006), individuals receiving substance abuse treatment (Forbey & Ben-Porath, 2007), and high-functioning adults with autism spectrum disorders (Ozonoff, Garcia, Clark, & Lainhart, 2005). In a medical setting, the correlates and implications of bariatric surgery candidates' RC Scale scores (Wygant, Boutacoff, et al., 2007) have been examined. Parental competency examinees (Stredny, Archer, & Mason, 2006), personal injury litigants (Downing, Denney, Spray, Houston, & Halfaker, 2008), criminal defendants (Downing et al., 2008; Sellbom, Ben-Porath, & Stafford, 2007), domestic violence offenders (Sellbom, Ben-Porath, Baum, Erez, & Gregory, 2008), and prison inmates (Megargee, 2006) have been examined in RC Scale studies conducted in forensic and correctional settings. RC Scale findings have also been reported for community-dwelling veterans (Simms, Casillas, Clark, Watson, & Doebbeling, 2005), law enforcement candidates (Sellbom, Fischler, & Ben-Porath, 2007), and college students (Forbey & Ben-Porath, 2008; Osberg, Haseley, & Kamas, 2008). Castro et al. (2008) reported comparable validities for the RC Scales in samples of African American and Caucasian mental health outpatients.

As we noted earlier, one of the objectives in restructuring the Clinical Scales was to arrive at measures that can be linked to current concepts and models of personality and psychopathology. Such connections help us integrate the known content, structure, and criterion validities of the RC Scales into a construct validity framework and thus broaden our understanding of a person's test results. Studies by Sellbom and Ben-Porath (2005) and Sellbom, Ben-Porath, and Bagby (2008b) indicate that the RC Scales are related to normal personality measures in expected and conceptually meaningful ways. Sellbom, Ben-Porath, & Bagby (2008a) show that an elaborated version of Watson's (2005) proposed re-conceptualization of the diagnostic categories for mood and anxiety disorders for the *DSM-5*, which includes Demoralization, improves the ability of this model to differentiate between Distress and Fear disorders.

In a special issue of the *Journal of Personality Assessment*, critics of the RC Scales questioned whether Demoralization adequately represents the MMPI "first factor" (Nichols, 2006), expressed concern that at least some of the RC Scales have "drifted" away from the intended constructs and are redundant with already existing scales (Butcher, Hamilton, Rouse, & Cumella, 2006; Nichols, 2006), argued that the Clinical Scales are better suited than their restructured counterparts to assess complex syndromal variables such as psychiatric diagnoses (Caldwell, 2006; Nichols, 2006), and suggested that a higher than expected proportion of individuals tested in clinical settings produce non-elevated scores on the scales (Rogers, Sewell, Harrison, & Jordan, 2006). Conceptual and empirical analyses conducted by Sellbom, Ben-Porath, and Graham (2006), Sellbom, Ben-Porath, and Stafford (2007), and most extensively by Tellegen et al. (2006) refute each of these assertions.

Conclusion

A growing body of empirical research documents the criterion validity and strengthens the construct validity of the RC Scales in a variety of settings in which the MMPI-2-RF is likely to be used. In the concluding paragraph of the monograph that introduced the RC Scales, Tellegen et al. (2003) observed:

> The introduction of the RC Scales may stimulate additional scale development. It may prove worthwhile to search for and measure the distinctive core features of important MMPI-2 scales other than the MMPI-2 Clinical

Scales, some of which may also be confounded with a strong Demoralization component. Investigations along these lines may lead to additional scales that are incrementally informative beyond the RC Scales. Through such efforts it may be possible eventually to capture the full range of core attributes represented by the large body of MMPI-2 constructs with a set of new scales more transparent and effective than those currently available. (pp. 85–86)

Efforts of the sort just described have resulted in the additional measures that, with the RC Scales, comprise the MMPI-2-RF (see Table 1-1). In Chapter 2 of this manual we describe the development of these additional scales. Extensive psychometric findings on the 51 scales of the MMPI-2-RF are summarized in Chapter 3.

3 4 5 6 7 8 9 10 11 12 B C D E

PEARSON

800.627.7271
www.PearsonClinical.com

PsychCorp

Product Number 25051
510H198-F 07/15