

# 11 The Restructured Form of the MMPI-2 (MMPI-2-RF)

## The MMPI-2-RF

The Restructured Form of the MMPI-2 (MMPI-2-RF; Ben-Porath & Tellegen, 2011; Tellegen & Ben-Porath, 2011) is a 338-item, self-report measure consisting of 51 new and revised Validity and Substantive scales, with the Restructured Clinical (RC) scales as its foundation. The MMPI-2-RF was developed with the support of the University of Minnesota Press, the copyright holder of the MMPI-2, following the publication of the RC scales. Each of the MMPI-2-RF items appears on the 567-item MMPI-2; thus, dual scoring is possible when the respondent was initially administered the MMPI-2 using item-conversion tables available in the *MMPI-2-RF Manual for Administration, Scoring, and Interpretation* (Ben-Porath & Tellegen, 2011). As of the time of this writing, the MMPI-2-RF is being offered as an alternative to, but not a replacement for, the MMPI-2; the MMPI-2 continues to be fully supported by the test publisher (Ben-Porath & Tellegen, 2011).

Tellegen et al. (2003), in concluding their introduction of the RC scales, noted that creation of those scales might instigate research into the creation of additional scales to measure important clinical aspects beyond those captured by the clinical scales, and which may also be confounded by a demoralization component. Such research culminated in the release of the MMPI-2-RF, which Ben-Porath and Tellegen (2011, p. 1) state was “designed to provide an exhaustive and efficient assessment of the clinically relevant variables measureable with the instrument’s item pool.”

As noted above, the core of the instrument consists of the nine RC scales, introduced in Chapter 10. Beyond the RC scales, MMPI-2-RF contains 9 validity scales, 3 higher-order scales, 23 specific problem scales, 2 interest scales, and the revised PSY-5 (PSY-5-r) scales. In total, the MMPI-2-RF comprises 51 scales. Each of these will be described within a section devoted to a general discussion of their group.

In addition to its reduced length and administration time, as compared with the MMPI-2, the 42 substantive scales (i.e. excluding the validity/response style scales) of the MMPI-2-RF, the average scale length is 14 items, and the average proportion of items keyed True for these scales is 73 percent. By contrast, for the 103 substantive (again excluding the validity/response style scales) MMPI-2 scales included in the most recent edition (Butcher et al., 2001) of the MMPI-2 *Manual*, the average scale length is 22 items, and the average proportion of items keyed True for these scales is 63 percent. In brief, the scales of the MMPI-2-RF average 36 percent fewer items and 14 percent more items keyed in the True direction than the comparable substantive scales of the MMPI-2.

As of the time of this writing (the first quarter of 2013), there remains a dearth of research specifically devoted to the interpretation of the MMPI-2-RF, although substantial interpretive guidance can be found in Ben-Porath (2012). Unless otherwise noted, the correlates and interpretive suggestions offered for this instrument are derived from this source and from the two MMPI-2-RF *Manuals* (Ben-Porath & Tellegen, 2011; Tellegen & Ben-Porath, 2011); the reader is encouraged to keep abreast of new reports in the psychological assessment literature that will undoubtedly add to our body of knowledge regarding this relatively new instrument.

## **Administration**

### *Test User Considerations*

Qualifications for administration, scoring, and interpretation of the MMPI-2-RF are the same as for the standard MMPI-2, described in Chapter 2. The reader is urged to review the chapter on administration and scoring the MMPI-2 at this time. The reader is also encouraged to review and abide by the American Psychological Association's "Ethical Principles of Psychologists and Code of Conduct," specifically Standard 9, which pertains to psychological assessment.

### *Test-Taker Considerations*

The MMPI-2-RF is designed to be used with adults, aged 18 years and older. It is not recommended to be used, nor is its use supported, with individuals younger than 18 years.

Ben-Porath and Tellegen (2011) recommend that the test taker has at least a sixth-grade reading level to complete either the booklet or computerized administrations of the MMPI-2-RF. If there is doubt about the test-taker's ability to adequately read and understand the items, they recommend that a standardized test of reading ability be administered. If the test taker does not possess an adequate reading level or if it is not possible to perform an assessment of reading level, they recommend that a standardized audio version of the test be administered using either the audio CD or computerized version available from the test publisher. They further note that if an individual who does not possess adequate reading ability is nonetheless administered a written version of the test, it is very likely that the inconsistency scales will detect the difficulty.

### **Available Formats**

As of the time of this writing, the MMPI-2-RF is available only in English. Translations into other languages are not yet available. As with MMPI-2, both softcover and spiral-bound booklets are available for pencil and paper administration. Both require a separate answer sheet and are reusable. Given the popularity of computer-based administration, the MMPI-2-RF can be used with administrative and scoring software available through Pearson Assessments. Additionally, an audio CD is available for use with individuals who do not meet the minimum reading ability requirements for the test.

## **Scoring**

### *Normative Sample*

The normative sample for the MMPI-2-RF is a subset of the MMPI-2 re-standardization sample of 1,138 men and 1,462 women. Because of the authors' desire to create non-gender specific *T*-scores for MMPI-2-RF, the larger re-standardization female sample was reduced by randomly selecting a subsample of 1,138 women to yield a total normative sample of 2,276 for the MMPI-2-RF. The ethnicity, education, and age distributions of the MMPI-2-RF normative sample very closely resemble those of the MMPI-2 re-standardization sample. The reader is referred to the test manual (Ben-Porath and Tellegen, 2011) for a further description of the normative sample.

### **T-score Conversions**

As noted above, the MMPI-2-RF incorporates only non-gendered norms. The use of non-gendered norms complies with the provision of the Civil Rights Act of 1991 prohibiting the consideration of sex in employment practices. Some (e.g. Butcher & Williams, 2012; Nichols, 2011) have criticized the decision to use non-gendered norms. A comparison of gender-specific versus non-gendered norms for the MMPI-2 scales conducted by Ben-Porath and Forbey (2003) revealed a dearth of significant differences, with most *T*-score differences falling within a range of three points.

Uniform *T*-scores (see Chapter 1) are incorporated for the majority of the MMPI-2-RF scales. The only exceptions to this are for the validity scales and the interest scales, for which linear *T*-score transformations are incorporated.

### **Protocol Scoring**

Protocols can either be scored by hand or by proprietary Q Local computer software available through Pearson Assessments, or their online scoring and report service, Q Global. When scoring by hand, raw scores for each scale are first calculated using individual scoring templates available from the test publisher for use with official scoring sheets. Ben-Porath and Tellegen (2011) note that due to the time commitment involved in scoring 51 scales by hand, clinicians might be tempted to not score the full profile, which they caution against. Once raw scores are calculated for each scale, they are transferred to a profile sheet, enabling their transformation into *T*-scores.

Computerized scoring is available in several formats. If the test taker was administered the test using the Q Local software, scoring is performed automatically and a score report can be printed from the test administration software. If the test responses were administered on paper, they can be manually entered into the Q Local software so that a scored protocol can be generated. Finally, response sheets can be scanned into the Q Local software and scored protocols will be generated from the scanned responses.

### **Validity Scales**

The MMPI-2-RF validity scales (see Table 11.1) consist of eight revised forms of their counterparts on the standard MMPI-2, and one new validity scale created specifically for

Table 11.1 MMPI-2-RF validity scales

<i>Abbreviation</i>	<i>Name</i>
<i>VRIN-r</i>	Variable Response Inconsistency
<i>TRIN-r</i>	True Response Inconsistency
<i>F-r</i>	Infrequent Responses
<i>Fp-r</i>	Infrequent Psychopathology Responses
<i>Fs</i>	Infrequent Somatic Responses
<i>FBS-r</i>	Symptom Validity
<i>RBS</i>	Response Bias Scale
<i>L-r</i>	Uncommon Virtues
<i>K-r</i>	Adjustment Validity

MMPI-2-RF. These scales can be grouped into those measuring response consistency, self-unfavorable reporting, and self-favorable reporting. As the majority (eight of nine) of these scales are similar to their MMPI-2 predecessors, the reader is referred back to the earlier chapters on the validity scales and their interpretation (Chapters 3 and 4) for a more thorough discussion of those earlier scales. This chapter will highlight the revisions that these scales have undergone and provide guidelines for their interpretation on the MMPI-2-RF.

### ***Item Omissions***

In general, respondents who have been adequately prepared to complete the MMPI-2-RF will omit very few items. Still, Greene (2011) points out that even relatively few omitted items can have rather consequential effects on scale *T*-scores, given the relatively few items on the scales of the MMPI-2-RF (an average of 14 items per scale, with some having as few as 4 or 5). Ben-Porath and Tellegen (2011) suggest that even when fewer than 15 items are omitted, some shorter scales may be invalid, although the protocol may still be interpreted with the understanding that some scale scores may underestimate the dimensions or problems measured by that scale. When more than 15 items are omitted, caution should be used when interpreting the profile. If more than 10 percent of a scale's items are omitted, that scale should not be interpreted.

### ***Response Consistency***

The Variable Response Inconsistency (*VRIN-r*) and True Response Inconsistency (*TRIN-r*) scales comprise the measures of response consistency on the MMPI-2-RF. A rigorous, five-criterion methodology was used for selecting items for the response consistency scales, and the reader is urged to review the MMPI-2-RF *Technical Manual* (Tellegen & Ben-Porath, 2011) for a complete explanation of the methodology. Although the method for selecting items for *VRIN-r* and *TRIN-r* was somewhat different from the method used for creating *VRIN* and *TRIN* for MMPI-2, the overall similarity in the scales would suggest that they can be interpreted in much the same manner as

before. This suggestion is supported by limited empirical research (Handel, Ben-Porath, Tellegen, & Archer, 2010) demonstrating that *VRIN-r* and *VRIN* respond similarly to random responses, while *TRIN-r* and *TRIN* respond similarly to fixed response patterns.

### ***Variable Response Inconsistency (VRIN-r)***

This scale contains 53 item pairs, as opposed to the 67 item pairs that comprise the *VRIN* scale for MMPI-2. Only 13 item pairs correspond between the MMPI-2 and MMPI-2-RF versions of the scale. Although it is an oversimplification of the scoring procedure, it is illustrative to state that item pairs are scored when the test taker responds inconsistently (one item True and one item False) in a particular manner to selected pairs of items with consistent semantic meaning. Scoring can be complicated by the fact that in only six of the item pairs does the particular direction of the inconsistency (i.e. True–False vs. False–True) have no bearing on scoring.

### ***Interpretation***

Scores below a *T* of 39 indicate a very deliberate approach to the assessment procedure, with less inconsistency than is normally expected. Scores in this range suggest that the protocol is interpretable if other validity indicators do not suggest invalidity. Greene (2011) suggests that *T*-scores in this range may be indicative of a respondent making a self-favorable report.

*VRIN-r* *T*-scores between 39 and 69 also suggest consistent reporting. These scores fall within the average range. The protocol is valid as long as other validity indicators do not suggest otherwise.

Scores in the range of *T*-70 to 79 on *VRIN-r* indicate a degree of inconsistent responding beyond what is usually seen. This inconsistency could be due to a variety of reasons, including reading or language difficulties, response errors, carelessness, distraction, or fatigue. The administrator should try to ascertain if the test taker experienced any difficulties in completing the instrument.

*VRIN-r* scores above *T*-80 generally indicate markedly inconsistent responding to the stimulus items. Profiles with scores in this range are considered invalid. If the administrator is able to ascertain the cause for the inconsistent response patterns and take actions to correct the situation (e.g. administering the protocol via audiotape), a re-administration may be attempted.

Greene (2011) has noted that 18–19-year-olds in a clinical sample scored about four *T* points above the mean on *VRIN-r*. This effect was not found in any other age group.

### ***True Response Inconsistency (TRIN-r)***

The *TRIN-r* scale contains 26 item-pairs, whereas the MMPI-2 *TRIN* scale contains 23 pairs. Only five pairs of items overlap the two versions of the scale. Whereas *VRIN-r* assesses the tendency to respond dissimilarly to similar items, *TRIN-r* measures the tendency to respond similarly to dissimilar items. The item pairs were chosen to be essentially reversals in content. The items are scored such that each individual begins with a raw score of 11. One point is added for each item pair that is answered True–True

and one point is subtracted for each pair that is answered False–False. Raw scores above 11 indicate responding in the True direction, whereas raw scores below 11 indicate a tendency to respond in the False direction. As with *TRIN*, *T*-scores are calculated using linear *T*-scores, with the caveat that scores below 50 are “reflected” to indicate a higher *T*-score (i.e. a raw score that would normally result in a *T*-score of 40, or one standard deviation below the mean, is converted to a *T*-score of 60, or one standard deviation above the mean). Response direction is indicated by appending either a T or F to the standardized score.

### *Interpretation*

Profiles with *TRIN-r* *T*-scores  $\geq 80$ F (raw score  $\leq 7$ ) evidence considerable nay-saying; these profiles are considered invalid and should not be interpreted. The MMPI-2-RF can be re-administered after the clinician ascertains the reason for the invalidity and takes corrective action.

*TRIN-r* scores in range of *T*-70 to 79F (raw scores = 8) indicate significant nay-saying and may be invalid. The individual’s response set should be investigated, and the test could be retaken after corrective action, or cautiously interpreted, depending on the setting.

Scores in the range of *T*-50 to 69 (raw scores = 9 to 13) suggest that the respondent was not answering in a fixed manner. If other validity indicators suggest the profile is valid, then *TRIN-r* scores in this range confirm it.

*TRIN-r* scores in range of *T*-70 to 79T (raw scores = 14) indicate significant yea-saying and may be invalid. The clinician should proceed in a manner similar to that described above for scores in this range in the False direction.

Profiles with *TRIN-r* scores  $\geq 80$ T (raw score  $\geq 15$ ) evidence considerable yea-saying and are considered invalid. These protocols should not be interpreted. As with scores in this range in the False direction, the MMPI-2-RF may be re-administered once the clinician ascertains the reason for the invalidity and takes corrective action to guard against it.

### **Self-Unfavorable Reporting**

Scales that assess self-unfavorable reporting on the MMPI-2-RF include the Infrequent Responses (*F-r*), Infrequent Psychopathology Responses (*Fp-r*), Infrequent Somatic Responses (*Fs-r*), and Symptom Validity (*FBS-r*) scales. Of these scales *F-r*, *Fp-r*, and *FBS-r* represent revisions of an extant MMPI-2 scale; *Fs-r*, on the other hand, is a new scale, created for MMPI-2-RF. The broad purpose of these scales is to assess for over-reporting of psychological or other difficulties. Each of these scales will be addressed, in turn.

#### ***Infrequent Responses (F-r)***

The *F-r* scale assesses the degree to which a respondent endorses items that are infrequently endorsed in the normative sample. *F-r* consists of 32 items, whereas its MMPI-2 analogue, the Infrequency (*F*) scale, consists of 60. *F-r* shares 11 items with *F* and 10 items with *FB* (Back Infrequency), and is thus largely a hybrid of these scales on



the MMPI-2. Each of the *F-r* items was endorsed by less than 10 percent of the MMPI-2-RF normative sample; these items refer to a broad range of psychological, somatic, and cognitive symptoms. The items of *F-r* overlap, to varying degrees, with eight of the nine RC scales; three items appear on *RCd*, four items are common to *RC1*, two items overlap *RC2*, three items are shared with *RC4*, six items are shared with *RC6*, two items are common with *RC7*, five items appear on *RC8*, and one item appears on *RC9*.

Up to a certain degree of elevation, *F-r* scores can serve as a rough indicator of the degree of psychological distress that an individual is experiencing. However, the item content of *F-r*, as with *F*, is rather obvious, which can make it relatively easy for individuals to either deny symptoms or to over-report problems. Interpretation for various score ranges is given below.

### *Interpretation*

Low scores on *F-r* ( $T \leq 44$ ) are obtained by individuals who report less socially unacceptable content or unusual experiences than normal. Greene (2011) cautions that such profiles may indicate an individual who is attempting to appear in a positive light, or may simply represent an individual who is very conventional. Other validity scales should be carefully examined in order to ascertain the meaning of low *F-r* scores.

Average scores on *F-r* ( $T=45$  to 60) are seen in individuals who are reporting an average number of unusual experiences. The profile can be safely interpreted if other validity indicators suggest that it is valid.

Moderately high scores on *F-r* ( $T=61$  to 79) occur among individuals who report a slightly higher number of unusual experiences, attitudes, feelings, and behaviors than is the norm. Occasionally, scores in this range occur in individuals with some psychopathology who have adjusted to their chronic problems, and therefore are not in great immediate distress. Scores in this range can serve as an indicator of the degree of clinical distress an individual is experiencing. Rogers, Gillard, Berry, and Granacher (2011) found that scores in this range, or below, were exceedingly rare among individuals feigning mental disorders.

High *F-r* scores ( $T=83$  to 106) often indicate serious psychopathology, especially if other validity indicators do not suggest invalidity. *VRIN-r* and *TRIN-r* should be examined for response consistency and yea-saying; *Fp-r*, in particular (see below), should be examined to determine whether the obtained elevation represents true disturbance or an attempt to present oneself in a negative light. If it is determined that the score reflects a true psychological disturbance, the level of elevation can, to some degree, indicate the severity of disturbance; at the upper end of this range, however, one may see some exaggeration of the level of disturbance, even in a profile that is not necessarily invalid.

Very high scores on *F-r* ( $T \geq 111$ ) are associated with profiles that are generally uninterpretable. Check other validity indicators (e.g. *VRIN-r*, *TRIN-r*, and *Fp-r*) for evidence of invalidity. Occasionally, one will come across a score in the lower portion of this range ( $T=109$  to 119) in an otherwise valid profile. In such cases, the score is likely to be obtained from an individual with an acute psychotic disturbance. However, individuals such as this can usually be readily ascertained in an interview; the resulting profiles are usually so unstable as to be interpretatively meaningless, other than as a reflection of their current experience. The clinician is urged to use extreme caution in choosing to interpret

such profiles, especially if there is any likelihood of forensic involvement. Scores greater than *T*-119 should be considered invalid and the profile should not be interpreted even if the scores on other validity indicators are within acceptable ranges.

### ***Infrequent Psychopathology Responses (Fp-r)***

Whereas *F-r* measures an individual's tendency to endorse items that are infrequently endorsed within the normative sample—and can thus function as a rough measure of psychopathology—*Fp-r* measures an individual's tendency to respond to items that were infrequently endorsed among individuals with significant psychopathology. *Fp-r* is a revised version of the *F(p)* scale which is described in Chapter 3 on MMPI-2 validity scales; the reader is encouraged to refer to that chapter for an explanation of the derivation of *F(p)* and a review of the relevant research surrounding its use.

Whereas the items of *F-r* overlap several RC scales, *Fp-r* shows overlap with only RC6, with which it shares five items. *Fp-r* contains 21 items, whereas *F(p)* contains 27; 17 items are common to both scales. Tellegen and Ben-Porath (2011) explained that four items were dropped from the revised version of the scale due to their concomitant appearance on the MMPI-2 *L* scale—which conflicted with their stated goal of creating scales of maximal distinction—and noted that some moderate elevations on *F(p)* could be due to endorsement of only those items (see also, Gass & Luis, 2001). Likewise, three *F(p)* items were dropped from *Fp-r*, due their appearance on the *Fs-r* scale, which is discussed below. Two additional items were dropped due to analyses that suggested that they did not perform as well as other scale items. Three new items were added to the revised version, as their inclusion improved performance of the scale. Overall, *Fp-r* appears to be a distinct improvement over its MMPI-2 counterpart.

Several recent studies have examined the effectiveness of the *Fp-r* scale. Marion, Sellbom, and Bagby (2011) reported that *Fp-r* was the most effective of the MMPI-2-RF over-reporting scales at discriminating individuals instructed to simulate psychopathology (including a sophisticated simulators group) from patients. Likewise, Sellbom and Bagby (2010) reported that *Fp-r* was the most effective of the validity scales in differentiating the profiles of individuals instructed to feign psychopathology from a group of psychiatric inpatients. Rogers et al. (2011) reported that an *Fp-r* *T*-score above 90 was slightly more effective than *F-r* in distinguishing individuals feigning mental disorders from those with psychopathology while reducing the rate of false-positives for feigning. Purdon, Purser, and Goddard (2011) reported that *Fp-r* elevations were associated with clinician ratings of positive symptoms among individuals admitted to a first-episode psychosis clinic; they were unsure whether this represented over-reporting on the part of the patients or an effect of clinical symptoms on their scores. Greene (2011) cautions that individuals scoring above *T*-81 on *Fp-r* are likely to prematurely end treatment despite having serious psychopathology; he suggests discussing this issue with clients obtaining scores in this range.

### ***Interpretation***

Average scores on *Fp-r* ( $T \leq 59$ ) are seen in individuals who endorse an average number of infrequently-endorsed items pertaining to odd or unusual experiences. Greene



(2011), however, cautions that scores below *T*-44 may indicate a tendency toward a self-favorable report; other validity indicators should be examined to determine if this is the case. The profile can be safely interpreted if other validity indicators suggest that it is valid.

Moderately high scores on *Fp-r* (*T*-60 to 77) are seen in individuals who may be accurately endorsing a slightly higher number of unusual experiences, attitudes, feelings, and behaviors than is the norm, especially toward the lower end of this range. Scores above 70 are associated with clinical distress, as well as the tendency for a test taker to respond in self-unfavorable manner. The client's clinical history and presenting problem should help to clarify whether the profile represents over-reporting. Other validity scales should be examined to determine profile validity; if other scales suggest validity, a score in this range would not necessarily invalidate the profile.

High *Fp-r* scores (*T*-85 to 94) often indicate serious psychopathology, especially if other validity indicators do not suggest invalidity. *VRIN-r* and *TRIN-r* should be examined for response consistency and yea-saying. Even if it is determined that the score reflects a true psychological disturbance, the level of elevation can, to some degree, indicate possible exaggeration of the extent of psychological disturbance, even in a profile that is not necessarily invalid.

Very high scores on *Fp-r* (*T*  $\geq$  102) should be considered invalid and the profile should not be interpreted even if other validity indicators suggest that the profile is valid.

### ***Infrequent Somatic Responses (Fs-r)***

The *Fs-r* scale was developed by Wygant and colleagues (Wygant, 2007; Wygant, Ben-Porath, & Arbisi, 2004) to aid in the identification of individuals endorsing infrequent somatic complaints. The scale consists of 16 items, with 12 keyed True. The items chosen for inclusion in the scale were endorsed by no more than 25 percent of the medical patients comprising several large samples. Tellegen and Ben-Porath (2011, p. 15) suggest that the *Fs-r* scale works in a complementary manner to *FBS-r* (described below), with "each scale outperforming the other in some settings and tasks." Greene (2011) has noted that it was unusual for individuals in any of five samples (including the MMPI-2 re-standardization sample, a clinical sample, a pain sample, and two samples of personal injury litigants) to attain *T*-scores above 74 on *Fs-r*. Thus, he suggested that scores higher than this should raise concerns that the respondent is endorsing a number of somatic symptoms that are not generally reported by individuals being treated for known medical issues. Schroeder et al. (2012) have suggested that a cut score of *T*  $\geq$  83 be used to minimize the risk of false positives with regard to symptom exaggeration among those undergoing neuropsychological examination. Sellbom, Wygant, and Bagby (2012) reported that among the MMPI-2-RF over-reporting scales, *Fs-r* was most sensitive to somatic malingering within three samples consisting of individuals who had been instructed to feign physical health problems as if they were participating in a disability evaluation, medical patients who were not involved in litigation, and individuals who had been carefully diagnosed with a somatoform disorder. Wygant et al. (2009) found *Fs-r* effective in identifying over-reporting among individuals in both medical simulation and head injury simulation samples.

*Interpretation*

Low to moderately high scores ( $T < 74$ ) show no evidence of invalidity; the profile can be interpreted if other indicators suggest that it is valid. Toward the upper end of this range ( $T > 58$ ), individuals are endorsing more physical complaints than is the norm, but this may well reflect genuine physical issues in individuals with medical conditions.

High scores ( $T=74$  to 99) may reflect inconsistent responding (check *VRIN-r* and *TRIN-r*), over-reporting of somatic complaints, or significant medical symptoms. Scores in this range likely reflect over-reporting of somatic complaints when they occur in an individual without a history or evidence of physical health problems. Even in individuals with genuine medical issues, scores at the upper end of this range likely reflect some exaggeration. Scores on the somatic scales should be interpreted cautiously if the clinician chooses to proceed with interpretation.

Very high scores ( $T \geq 100$ ) usually indicate inconsistent responding or over-reporting of somatic complaints. The other validity scales should be examined for evidence of inconsistent responding. If there is no evidence of inconsistent responding, one can proceed with interpretation, although scores on the somatic scales should be interpreted in light of this likely exaggeration of symptoms.

*Symptom Validity (FBS-r)*

The *FBS-r* scale for MMPI-2-RF is a revised version of the *FBS* scale that was added to the standard MMPI-2 form in 2007. The scale was originally designed to assist in identifying malingering in personal injury litigation. *FBS-r* contains 30 items, all of which overlap the 43-item version on the MMPI-2. The original version of the scale is discussed in detail in Chapter 3 on MMPI-2 validity scales and the reader is encouraged to review that section for further information on the development of *FBS*, its use, and the controversy surrounding its use. As compared with the original MMPI-2 version of the scale, *FBS-r* contains 19 items (63 percent of *FBS-r*) that overlap with the symptoms described in Lees-Haley's (1988) Litigation Response Syndrome, a proportion greater than that found on the original scale (53 percent). Additionally, the proportion of the *GM* and *GF* items on *FBS-r*, one positively for women and four negatively for women, is greater at 17 percent, versus 14 percent for *FBS*, thereby slightly increasing the gender bias against women for the RF version. In sum, the risk for false positives for malingering, especially for women, appears, if anything, to be greater for the revised *FBS-r* than for the original *FBS*; thus, its interpretation should be avoided.

*Response Bias Scale (RBS)*

The RBS was developed by Gervais, Ben-Porath, Wygant, and Green (2007) to detect a tendency toward over-reporting of symptoms in forensic neuropsychological or disability evaluation settings. The scale contains 28 items, 19 keyed True, and was added to the MMPI-2-RF in 2011. It overlaps *F-r*, *FBS-r*, *RC1*, and Cognitive Complaints (*COG*) by four items each, *Fs* by two items, and Neurological Complaints (*NUC*) and Head Pain (*HPC*) by one item each.

Gervais et al. (2007) reported that *RBS* consistently outperformed *F*, *F(p)*, and *FBS* at identifying individuals in forensic neuropsychological and disability evaluation settings who failed other commonly-used tests of response bias or symptom malingering-neuropsychological symptom validity tests (SVTs). Gervais, Ben-Porath, Wygant, and Green (2008), incorporating archival MMPI-2 and Memory Complaints Inventory (MCI; Green, 2004) data from non-head-injury, disability-related evaluations, found that the *RBS* was a better predictor of the mean memory complaints score than the *F*, *FB*, *F(p)*, or *FBS*. Gervais, Ben-Porath, Wygant, and Sellbom (2010) demonstrated the incremental validity of *RBS* in assessing memory complaints. Wygant et al. (2010) reported that *RBS* was effective at detecting cognitive response bias in separate forensic samples composed of disability claimants and criminal defendants. They further provided evidence of the scale's incremental validity beyond the traditional MMPI-2 and MMPI-2-RF over-reporting validity scales in their disability sample. Schroeder et al. (2012) reported that a cut score of  $T \geq 92$  was sensitive to symptom exaggeration in 43 percent of neuropsychological cases, with a specificity rate of 92 percent, whereas Wygant et al. (2011) reported 70 percent sensitivity and 76 percent specificity for a score of  $T \geq 90$  in individuals undergoing compensation evaluations.

### *Interpretation*

Low scores ( $T < 50$ ) are generally considered to indicate a lack of conscious over-reporting of memory or other cognitive symptoms. The clinician should consider denial or an overly positive presentation, if cognitive testing reveals deficits. If other validity indicators suggest validity, the clinician can proceed with interpretation.

Average scores ( $T$ -50 to 63) are associated with reports of minor memory or cognitive symptoms that are consistent with cognitive test results. If other validity scales suggest possible exaggeration, symptom exaggeration is possible. If other validity indicators do not suggest exaggeration and cognitive test results do not indicate any difficulties, the clinician should consider the possibility of emotional factors (e.g. depression) contributing to memory complaints.

Moderately high scores ( $T$ -67 to 76) are associated with increasing memory complaints. If other over-reporting validity scales do not suggest exaggeration, these complaints are likely related to emotional factors. If other over-reporting validity scales suggest exaggeration, the clinician should consider the possibility of an intentional effort to exaggerate symptoms.

High scores ( $T$ -80 to 97) reflect a much higher than average number of non-credible memory complaints than is normally endorsed by individuals with genuine memory problems. This could be due to inconsistent responding, significant medical symptoms, or over-reporting. Other validity indicators should be examined to assess validity. If other indicators indicate the profile is valid, interpretation can proceed, although the test manual encourages the clinician to be alert to the possibility of over-reporting when interpreting the cognitive complaints scales.

Very high scores ( $T \geq 101$ ) are generally considered to reflect either inconsistent responding or over-reporting of memory complaints. Other validity indicators should be examined for evidence of an inconsistent response pattern or over-reporting. Even if *VRIN-r* and *TRIN-r* do not suggest inconsistent responding and other validity indicators

do not suggest over-responding, the clinician is cautioned that this level of symptom report is extremely unusual among those with substantial emotional dysfunction who report credible symptoms. If a clinician chooses to proceed with interpretation, they should be mindful of the substantial possibility of over-reporting of memory symptoms when interpreting the cognitive scales.

## **Self-Favorable Reporting**

Scales that assess self-favorable reporting on the MMPI-2-RF include the Uncommon Virtues (*L-r*) and Adjustment Validity (*K-r*) scales. Although each of these scales represents, to some degree, a revision of an extant MMPI-2 scale, Tellegen and Ben-Porath (2011) state that the scales were created by factor-analyzing the three self-favorable validity scales of the MMPI-2 (i.e. *L*, *K*, and *S*), along with Wiggins' (1959) Social Desirability (*Sd*) scale and creating non-overlapping scales representing the two primary factors. The broad purpose of these scales is to assess for under-reporting of psychological or other difficulties. The two scales will be addressed, individually.

### ***Uncommon Virtues (L-r)***

The *L-r* scale contains 14 items, 11 of which are shared with *L* and keyed False, and 3 of which appear on *Sd* and are keyed True. Briefly, the scale is designed to assess the tendency to engage in positive self-presentation to the degree that one is unwilling to admit to even common shortcomings. Elevations on this scale often reflect a naïve or obvious attempt by the respondent to appear unusually virtuous, culturally conservative, overly conscientious, and above moral reproach.

### ***Interpretation***

Low scores ( $T \leq 42$ ) are generally considered to indicate candor and a willingness to admit to common shortcomings. However, if the other validity scales suggest over-reporting of psychopathology, a score in this range could be part of an attempt to present oneself as disturbed and without moral resources.

Average scores ( $T=47$  to  $57$ ) are associated with individuals who display a balance of admitting to and denying common shortcomings. The upper end of this range of scores may be indicative of a mildly cautious, defensive and/or moralistic individual, especially as the education level of the respondent increases. Scores at the upper end of this range are not uncommon among psychologically unsophisticated individuals.

Moderately high scores ( $T=62$  to  $66$ ) suggest some concern and rigidity over matters of self-control and moral values, as well as a tendency to be conforming and conventional. Individuals with scores in this range may fear that knowledge of their shortcomings may cause others to find them unacceptable or unworthy. A lack of psychological insight and self-awareness is common among individuals with *L-r* elevations in this range. In an educated individual, or in a custody or job evaluation setting, an elevation in this range may represent an attempt to look one's best and to deny unacceptable human impulses.

High scores ( $T=71$  to  $76$ ) are fairly uncommon. Check other validity scales for inconsistent responding. If inconsistent responding can be ruled out, these scores are

associated with the denial of many of the most common and obvious human failings. Such individuals have intense needs to present a good front. They are typically seen as self-righteous and uncompromising and are sensitive to social disapproval. Some see psychological problems as a sign of moral weakness. A score in this range can reflect extreme naïveté in a person from a culturally restricted environment. Other times it reflects a conscious attempt to distort the MMPI-2-RF results.

Very high scores ( $T \geq 81$ ) are generally considered invalid. Even if inconsistent responding can be ruled out, the respondent is engaged in positive impression management to such a degree that interpretation of the substantive scales is not recommended. Even if scores on the substantive scales are elevated, the degree of under-reporting suggested by  $L-r$  would likely mean that they greatly underestimate the respondent's true symptom level.

### ***Adjustment Validity (K-r)***

The  $K-r$  scale contains 14 items, 12 keyed False, each appearing on the original  $K$  scale; five of these items also appear on  $S$ . One item (item 202) is scored in the opposite direction from its scoring on  $K$ . Tellegen and Ben-Porath (2011) note that this was one of the correction items added to the original  $K$  scale as a means of addressing concerns that individuals with psychopathology produced low scores on  $K$ , thus lowering their  $K$ -corrected scores on other scales. Essentially,  $K-r$  is a measure of the respondent's self-reported level of adjustment, with higher scores representing a higher level of self-reported adjustment.

### ***Interpretation***

Very low scores ( $T < 35$ ) occur for several reasons. If there is the possibility of secondary gain from being diagnosed with a psychiatric disorder, the clinician should consider the possibility that the profile may be intentionally exaggerated. Check the self-unfavorable validity scales for evidence of exaggeration or over-reporting. In valid profiles in which the clinical scales are elevated, a  $K-r$  score in this range would suggest a severe disturbance, with little capacity to modulate emotions and behaviors.

Low scores ( $T=35$  to 42) suggest that the test-taker's coping skills are somewhat compromised. Individuals with scores in this range tend to be self-critical, have poor self-esteem, and lack confidence in their own ability to successfully deal with their problems. If substantive scales are elevated, the client is likely to admit to the symptoms and complaints associated with the profile. Greene (2011) suggests that in clients of lower SES, scores in this range reflect a moderate disturbance; in clients of higher SES, however, scores in this range reflect poor coping skills and a more serious disturbance. In cases in which no substantive scales are elevated, scores in this range would suggest candor and openness.

Average scores ( $T=45$  to 55) are seen in individuals who exhibit culturally-appropriate emotional restraint and generally feel in control of their emotions. Even if the substantive scales are elevated, scores in this range, especially toward the upper end of the range, suggest that the respondent maintains some coping skills. The prognosis is better for successful therapeutic interventions when  $K-r$  is in this range. Greene (2011) suggests that, even in this range, some degree of distress will be evidenced by respondents of higher SES.



Moderately high scores ( $T=59$  to  $62$ ) are associated with two differing interpretations. In individuals of lower SES, elevations in this range may already be associated with defensiveness and denial. Among individuals of higher SES, however, scores in this range are associated with descriptions of individuals as independent, enterprising and resourceful, who feel in control of their lives.

High scores ( $T > 65$ ) in clinical samples are generally associated with individuals who are particularly defensive and have a serious lack of insight into their psychological problems. Because of their unwillingness to acknowledge their own difficulties, prognosis for therapy can be poor.  $K-r$  scores in this high range that are obtained in forensic and personnel evaluation situations may reflect the defensiveness that is somewhat expected, given the situation. However, the resultant profile is likely to underestimate psychopathology. Finally, some individuals, usually from higher socioeconomic groups, may score at, or above, a  $T$ -score of  $65$  on  $K-r$  and obtain a non-elevated profile that accurately reflects a lack of psychopathology. In these cases, the  $K-r$  score may reflect

### Higher-Order Scales

The higher-order scales were created by Tellegen and Ben-Porath (2011) to capture a set of clinically-meaningful dimensions that could provide an organizational structure in which to integrate the conceptualization of the substantive scales of the MMPI-2-RF. Among the first attempts to extract higher-order dimensions in the MMPI were those of Welsh (1956) in his creation of the  $A$  and  $R$  scales, which were discussed in Chapter 8. Tellegen and Ben-Porath suggest that  $A$  has been conceptualized as being a similar construct to Eysenck's neuroticism dimension, while  $R$  has been conceptualized as a reciprocal measure of extraversion, although Greene (2011, p. 267) has provided evidence that calls this conceptualization into question. Tellegen and Ben-Porath (2011) further suggest that a "third" dimension of psychoticism has been missing in factor studies of the MMPI, due to the heterogeneous nature of the scales that best measured psychotic experience (clinical Scales 6 and 8). However, such dimensions have been prominent in several previous factor studies (cf. Costa et al., 1985; Eichman, 1961; Johnson et al., 1984; Waller, 1999; Welsh, 1952).

Tellegen and Ben-Porath (2011) repeatedly extracted three dimensions representing emotional/internalizing, thought, and externalizing dysfunctions within three separate clinical samples. The RC scales showing the highest loadings on each of the three dimensions were as follows: for the emotional/internalizing dysfunction factor, the highest scale loadings were found for  $RCd$ ,  $RC2$ , and  $RC7$ ; for the thought dysfunction factor, the highest scale loadings were found for  $RC6$  and  $RC8$ ; and for the behavioral/externalizing factor, the highest scale loadings were found for  $RC4$  and  $RC9$ .

Scales for each of these dimensions were created by obtaining a three-factor structure and corresponding factor scores from the combined items of each of the scales listed above within three samples. These obtained factor scores were correlated with each of the 567 items in the MMPI-2 pool. From those correlations, a set of non-overlapping (among the three higher-order scales) items was chosen for each of the three dimensions; these items make up the three higher-order scales. The clinician is urged to remember that these scales represent broad, domain-level function; an absence of an elevation on



a higher-order scale does not negate the possibility of specific dysfunction within that domain. Each of the scales will be discussed individually.

### ***Emotional/Internalizing Dysfunction (EID)***

The *EID* scale contains 41 items, with 23 keyed True. Thirty-two of the items appear on *RCd*, *RC2*, or *RC7*, with the remaining nine items appearing on other scales. Tellegen and Ben-Porath (2011) have suggested that this scale represents, in a broad form, the basic character of the 2-7 MMPI-2 codetype. Specific dysfunctions associated with *EID* can be assessed by examining *RCd*, *RC2*, *RC7*, Negative Emotionality/Neuroticism-Revised (*NEGE-r*), Introversion/Low Positive Emotionality-Revised (*INTR-r*), and the internalizing scales on the MMPI-2-RF.

Greene (2011) states that *EID* is redundant with *RCd*, being correlated at .95 in a large ( $N = 161,239$ ) clinical sample. Tellegen and Ben-Porath (2011) cite correlations between *EID* and *RCd* of .88 to .95 across a variety of samples. In addition to the high correlations with *RCd*, the scale is highly positively correlated with *RC2* (.61 to .85, with higher correlations in clinical samples), *RC7* (.73 to .81), Self-Doubt (*SFD*; .72 to .87), Inefficacy (*NFC*; .60 to .82); Stress/Worry (*STW*; .65 to .75); and Negative Emotionality/Neuroticism-Revised (*NEGE-r*; .73 to .81). It is negatively correlated with *K-r* (–.69 to –.76).

Tellegen and Ben-Porath report a one-week test–retest correlation of .90 within a subset of the normative sample. Internal consistency (Cronbach's alpha) estimates ranged from .86 to .95 across a variety of samples.

### ***Interpretation***

Low scores ( $T \leq 43$ ) are seen in individuals who report less subjective distress than is usually seen, especially in a clinical setting. Greene (2011) suggests that individuals scoring in this range may be extraverted or impulsive.

Average scores ( $T=45$  to 57) are associated with an average degree of subjective distress and emotional discomfort.

Moderately high scores ( $T=58$  to 64) are associated with slightly more reports of subjective distress and emotional discomfort than average. Their distress may lead them to seek treatment.

High scores ( $T=65$  to 79) are associated with significant emotional distress. Among the correlates of scores in this range are depression, sleep disturbance, hopelessness, and pessimism for both men and women. Suicidal ideation has been associated with scores in this range for women.

Very high scores ( $T > 80$ ) may be associated with symptom exaggeration. If validity indicators do not indicate over-reporting or exaggeration, scores in this range may represent an emotional response to a crisis.

### ***Thought Dysfunction (THD)***

The *THD* scale represents a dimensional measure of thought dysfunction associated with the 8-6 codetype. The scale contains 26 items, with 24 keyed True, of which 13 overlap *RC6* and 13 overlap *RC8*. Tellegen and Ben-Porath (2011) report a one-week test–retest

correlation of .71 within a subset of the normative sample. Internal consistency estimates ranged from .69 to .95 across a variety of samples, with nominally higher estimates of internal consistency observed within clinical samples.

The clinician is strongly urged to note that 22 items also overlap the PSY-5 PSYC-*r* scale, which contains 26 items, as well. Because of this *extremely* high percentage of item overlap between *THD* and PSYC-*r*, scores on these scales will highly correlate ( $r = .85$ )<sup>1</sup> even in the case of random responding. Greene (2011) reports an actual correlation of .96 between these two scales, and correlations between *THD* and *BIZ* and PSYC of .90 and .89, respectively, in a large clinical sample; Tellegen and Ben-Porath (2011) reported correlations ranging from .95 to .98 in both men and women, across a variety of treatment settings and within the normative sample. Thus, interpretive information is essentially identical and it is unnecessary to interpret both scales. Specific dysfunctions associated with *THD* can be assessed by examining *RC6* and *RC8*.

### *Interpretation*

Average scores ( $T \leq 63$ ) are seen in individuals who are not reporting any type of persecutory thought or perceptual disturbances.

High scores ( $T$ -67 to 77) are seen among individuals who are experiencing significant dysfunction in their thought processes. Scores on *RC6* and *RC8* should be examined for additional information regarding how this dysfunction might be manifested. Greene (2011) suggests that *T*-scores in this range are seen in individuals in the early stages of psychotic processes, as well as in individuals with chronic psychoses who have adjusted to their dysfunction.

Very high scores ( $T \geq 81$ ) are associated with serious thought dysfunction if validity indicators do not indicate over-reporting or exaggeration. Ideas of reference, odd thinking, paranoid ideation, and auditory and/or visual hallucinations are seen in individuals scoring in this range. The clinician is urged to assess the respondent for a psychotic disorder or associated personality disorder.

### ***Behavioral/Externalizing Dysfunction (BXD)***

The *BXD* scale was designed to provide an overall gauge of an individual's behavioral acting-out tendencies, and represents a dimensional measure of the 4-9 codetype. *BXD* contains 23 items, 20 keyed True, 13 of which appear on *RC4*, and 9 on *RC9*. With regard to overlap with the PSY-5 scales, 15 items are shared with Disconstraint-Revised (*DISC-r*) and 4 overlap Aggressiveness-Revised (*AGGR-r*). Greene (2011) reports correlations of .91 between *BXD* and *DISC-r*, and of .72 between *BXD* and *DISC*, in a large clinical sample. Tellegen and Ben-Porath (2011) report a one-week test-retest correlation of .71 within a subset of the normative sample. Cronbach's alpha estimates of internal consistency ranged from .74 to .84 across a variety of samples.

Individuals who score high on this scale present with a broad range of difficulties, with an emphasis on poor impulse control. A history of criminal behavior, as well as violent and abusive behavior is correlated with elevated scores. Mattson, Powers, Halfaker, Akeson, and Ben-Porath (2012) have reported that in a sample of individuals identified as being at risk for failure to complete a court-ordered drug treatment program, elevated

scores on *BXD* and its associated scales (especially if  $T > 75$ ) were associated with such failure. The specific dysfunctions associated with elevated *BXD* scores can be assessed by examining *RC4*, *RC9*, and the externalizing scales, as well as *DISC-r* and *AGGR-r*.

### *Interpretation*

Low scores ( $T \leq 43$ ) are seen in individuals who are fairly unlikely to engage in acting-out behaviors. These individuals are often described as passive and inhibited.

Average scores ( $T=46$  to  $63$ ) are seen in individuals who report an average number of externalizing behaviors. These individuals could be described as maintaining adequate behavioral control.

High scores ( $T=65$  to  $78$ ) are seen among individuals who are likely to have engaged in significant acting-out behavior; they are likely to have experienced some type of repercussions as a result of their behaviors. Scores on *RC4*, *RC9*, and the externalizing scales should be examined for additional information regarding how this dysfunction might be manifested.

Very high scores ( $T \geq 81$ ) are seen in individuals who are reporting substantial acting-out, externalizing behaviors. They are very likely to have gotten into trouble as a result of their behaviors. Substance abuse and illegal or criminal behaviors become more likely as scores elevate into this range. Clinicians should be alert to the possibility of antisocial, borderline, or narcissistic traits in individuals scoring in this range. Clients should be assessed for substance abuse issues.

### ***Restructured Clinical (RC) Scales***

The RC scales were discussed in Chapter 10. The reader is referred to that chapter for information regarding derivation of the scales, as well as interpretive statements. The RC scales appear on the MMPI-2-RF in essentially the same form as they appear on MMPI-2, with one exception: Gender-based norms are not available on MMPI-2-RF. All norms on MMPI-2-RF are non-gendered and a single set of  $T$ -score transformations are incorporated.

### **Specific Problems (SP) Scales**

The SP scales (see Table 11.2) were created as a way to delineate important individual issues and characteristics that comprise the population of problems assessed by one of the broader RC scales (such as substance abuse, which is subsumed by the broader *RC4*), as well as clinically-relevant issues that are *not* directly measured by any of the RC scales, such as suicidal ideation or shyness.

Tellegen and Ben-Porath do not fully describe the methodology used to derive the SP scales, which were created conjointly with the interest scales. As noted in the previous chapter devoted to the RC scales, although RC scales were not created for clinical Scales 5 and 0, as the core components of these scales were not judged to reflect psychopathology, seed scales were created for two core components for clinical Scale 5 and one core component of clinical Scale 0. These components later formed the basis of three of the SP and interest scales.

Table 11.2 MMPI-2-RF Specific Problem Scales

<i>Somatic/Cognitive Scales</i>	
MLS	Malaise
GIC	Gastrointestinal Complaints
HPC	Head Pain Complaints
NUC	Neurological Complaints
COG	Cognitive Complaints
<i>Internalizing Scales</i>	
SUI	Suicidal/Death Ideation
HLP	Helplessness/Hopelessness
SFD	Self-Doubt
NFC	Inefficacy
STW	Stress/Worry
AXY	Anxiety
ANP	Anger Proneness
BRF	Behavior-Restricting Fears
MSF	Multiple Specific Fears
<i>Externalizing Scales</i>	
JCP	Juvenile Conduct Problems
SUB	Substance Abuse
AGG	Aggression
ACT	Activation
<i>Interpersonal Scales</i>	
FML	Family Problems
IPP	Interpersonal Passivity
SAV	Social Avoidance
SHY	Shyness
DSF	Disaffiliativeness

According to Tellegen and Ben-Porath (2011, p. 18), “a systematic examination of other MMPI-2 scales yielded a substantial number of additional targets for scale construction.” Initially, 14 additional scales were created, beyond the three identified above. Three additional scales were added based on initial feedback and these 20 scales were then reviewed by unidentified “experts” who suggested additional clinically significant content that they felt should be assessed. Ultimately, 25 new scales, 23 SP scales, and 2 interest scales were created. These scales do not overlap with one another, although they do overlap with higher-order, RC, and PSY-5 scales. Each of the scales is fairly short, averaging 7.52 items each and ranging from 4 to 10 items. Because of their brief length, Cronbach’s alpha estimates are fairly low for some of the shorter scales; average inter-item correlations, which may provide a more appropriate measure

of internal consistency with shorter scales (Clark & Watson, 1995), were not reported. Tellegen and Ben-Porath, however, report that standard errors of measurement are in the acceptable range.

The SP scales are grouped into four problem areas: somatic, internalizing, externalizing, and interpersonal. Discussion of the individual scales will follow this sequence.

### ***Somatic/Cognitive Scales***

The somatic scales consist of five individual scales that were designed to assess preoccupation with health functioning, somatic symptoms, and cognitive complaints. Graham (2012) notes that the very brevity of each of the scales and the homogeneity of their item content essentially means that high scores on any of the scales reflect likely endorsement of item content consistent with the name of the scale. Ben-Porath and Tellegen (2011) caution that each of the scales should be interpreted in light of the respondent's scores on *Fs-r*, *FBS-r*, and *RBS*, though with the caveat that elevated scores on these validity scales do not necessarily imply that the respondent is intentionally over-reporting symptoms. However, they suggest that when scores on one of these validity scales exceed *T*-100, the clinician should incorporate content-based descriptors to characterize the respondent's subjective symptom presentation, but avoid the use of empirical correlates.

#### *Malaise (MLS)*

The *MLS* scale is described as measuring a general sense of poor health and physical debilitation. The scale contains eight items, all of which overlap *Hy3*; seven are keyed False. Tellegen and Ben-Porath (2011) report a one-week test-retest correlation of .82. Internal consistency (Cronbach's alpha) estimates ranged from .59 to .82 across a variety of samples, with nominally higher estimates obtained within clinical samples. Youngjohn, Wershba, Stevenson, Sturgeon, and Thomas (2011) found that *MLS* was the single best predictor, among all MMPI-2-RF validity and somatic/cognitive scales, of failure on cognitive effort tests among individuals seeking compensation because of a reported traumatic brain injury. Elevated scores are associated with a preoccupation with health problems, multiple somatic complaints, reports of sleep disturbance, and depression in both men and women.

#### INTERPRETATION

Low scores (*T*-38) are associated with a general sense of physical well-being. Elevated scores (*T* ≥ 65) are associated with increasing complaints of poor health, along with feelings of tiredness, decreased energy, and weakness. Reports of sleep problems and sexual dysfunction are associated with elevated scores. As *T* increases above 80, the reports of poor health become more prominent and preoccupying. The respondent may report feeling incapacitated by some physical malady. If a physical cause for the somatic complaints has been ruled out, the clinician should consider a diagnosis of a somatoform disorder. Individuals with elevations on this scale may have difficulty participating in therapy due to their malaise.

*Gastrointestinal Complaints (GIC)*

The *GIC* scale contains five items, four of which are keyed True. *GIC* overlaps *HEA* by four items, three on *HEA1*. Tellegen and Ben-Porath (2011) report a one-week test-retest correlation of .75. Internal consistency (Cronbach's alpha) estimates ranged from .64 to .79 across a variety of samples.

The *GIC* items describe problems related to upset stomach, nausea, vomiting, and poor appetite. High scores have been associated with a preoccupation with health problems and complaints of depression in both men and women. In addition, high scores in men are associated with complaints of sleep disturbance, hopelessness, and difficulty with concentration. High scores in women have been associated with poor coping abilities, suicidal ideation, and multiple somatic complaints.

## INTERPRETATION

Non-elevated ( $T < 65$ ) scores are not interpreted. Elevated scores ( $T \geq 65$ ) are associated with a greater than average number of gastrointestinal complaints. Individuals with elevated scores often have a history of gastrointestinal problems and are preoccupied with their health. As  $T$  increases above 90, the reports of gastrointestinal problems increase. They may complain of multiple physical ailments and see little hope for future improvement. If a physical cause for the somatic complaints has been ruled out, the clinician should consider a diagnosis of a somatoform disorder.

*Head Pain Complaints (HPC)*

The *HPC* scale contains six items, equally balanced between True and False. *HPC* overlaps *HEA* by five items. Tellegen and Ben-Porath (2011) report a test-retest correlation of .78. Internal consistency estimates ranged from .59 to .77 across a variety of samples.

The *HPC* items refer broadly to head and neck pain, as well as to the tendency to develop head pain when upset. Both men and women with elevated scores tend to be preoccupied with their physical health and may present with multiple somatic complaints; they often report feelings of hopelessness. In addition, high scores in men have been associated with complaints of sleep disturbance, depression, and anxiety. High scores in women have been associated with poor coping skills, decreased energy, and suicidal ideation.

*Interpretation*

Non-elevated scores ( $T < 65$ ) are not interpreted. Elevated scores ( $T \geq 65$ ) are associated with reports of head pain. As the  $T$ -score increases above 80, there are increasing reports of diffuse head and neck pain, and head pain associated with stress. They may present with multiple somatic complaints and be preoccupied with their physical functioning. They may have poor coping skills and see the future as not improving. If a physical cause for the complaints has been ruled out, the clinician should consider a diagnosis of a somatoform disorder.



*Neurological Complaints (NUC)*

The *NUC* scale contains 10 items, with 7 keyed False. *NUC* overlaps *Sc6* by seven items, and *HEA2* by six. This scale may reflect state-like problems, rather than trait-level concerns, as Tellegen and Ben-Porath (2011) report a one-week test–retest correlation of .54 within a subset of the normative sample. Cronbach’s alpha estimates ranged from .52 to .75 across a variety of samples, with nominally higher internal consistency observed within clinical samples.

The items refer to a broad range of complaints, such as dizziness, numbness, muscle weakness, paralysis, and loss of motor control. Elevated scores have been associated with multiple somatic complaints, a preoccupation with health concerns, developing physical symptoms in response to stress, and complaints of fatigue and reports of depression in both men and women. Locke et al. (2010) found that a *T*-score  $\geq 65$  identified approximately 91 percent of individuals with psychogenic non-epileptic seizures, but also misclassified 73 percent of individuals with epilepsy as having psychogenic non-epileptic seizures, leading to an overall accuracy of 59 percent. Increasing the cut score to  $T \geq 85$  increased overall accuracy to 67 percent, but decreased sensitivity to 53 percent while increasing specificity to 81 percent.

## INTERPRETATION

Non-elevated scores ( $T < 65$ ) are not interpreted. Elevated scores ( $T=65$  to 91) are associated with vague reports of neurological symptoms. Individuals with elevated scores tend to be preoccupied with their physical health and may present with multiple somatic complaints. They are likely to be experiencing some psychological stress that is expressed through physical complaints. As the *T*-score increases above 92, there are increasing reports of neurological symptoms, such as those described above. If a physical cause for the complaints has been ruled out, the clinician should consider a diagnosis of a somatoform disorder or a neurological/neuropsychological referral.

*Cognitive Complaints (COG)*

The *COG* scale contains 10 items, with 8 keyed True. *COG* overlaps *Sc3* by six items and *D4* by three. The items refer to a broad range of cognitive complaints, including memory and concentration difficulties, confusion, and intellectual limitations. Tellegen and Ben-Porath (2011) report a one-week test–retest correlation of .74. Internal consistency estimates ranged from .64 to .82 across a variety of samples, with nominally higher estimates obtained within clinical samples. High scores are associated with concentration difficulties, low frustration tolerance, memory complaints, a preoccupation with health concerns, and stress and worry.

Gervais, Ben-Porath, and Wygant (2009) report that elevated scores on *COG* are associated with *complaints* of memory problems or other cognitive difficulties, but not necessarily with objectively assessed cognitive deficits. They note that *COG* scores are not an effective predictor of objectively assessed cognitive function and suggest that *T*-scores should be interpreted as a measure of subjective complaints, primarily associated with emotional distress rather than with neurological symptoms.

## INTERPRETATION

Non-elevated scores ( $T < 65$ ) are not interpreted. Elevated scores ( $T=65$  to 80) are associated with reports of cognitive complaints. Individuals with elevated scores tend to be preoccupied with their physical health and may present with multiple somatic complaints. As the  $T$ -score increases above 82, there are increasing reports of cognitive complaints, such as those described above. High scores should alert the clinician to the possibility of memory or other cognitive difficulties, which might require formal neuropsychological assessment.

***Internalizing Scales***

The internalizing scales were created to address specific areas of interest related to elevations on the *EID* Higher-Order scale, as well as *RCd*, *RC2*, and *RC7* among the *RC* scales. The nine internalizing scales range in length from four to nine items. Graham (2012) notes that the brevity and homogeneity of each scale's content means that high scores on any of the scales reflect likely endorsement of item content consistent with the name of the scale. Ben-Porath and Tellegen (2011) suggest that although the correlations among the scales are quite high, the presence of unique empirical correlates allows the scales to be used as substantive measures on their own, rather than merely as interpretive aids for elevations on other scales.

***Suicidal/Death Ideation (SUI)***

The *SUI* contains five items, all keyed True. Tellegen and Ben-Porath (2011) report a test-retest correlation of .68 within a subset of the normative sample. Cronbach's alpha estimates ranged from .41 to .81 across a variety of samples, with higher internal consistency estimates observed within clinical samples.

Four of the scale's five items appear on *DEP4* and on Greene and Nichols' (1995) Structural Summary for the MMPI-2 among five specific MMPI-2 items serving as markers of depressed ideation and attitudes. Additionally, two are those that Sepaher, Bongar, and Greene (1999) identified as the "I mean business" suicide items. High scores have been associated with suicidal ideation, depression, hopelessness, and complaints of sleep disturbance in both men and women.

Ben-Porath and Tellegen (2011) note that because these items were endorsed so infrequently in the normative sample, endorsement of only one item results in an elevated score and indicates the need for an immediate suicide risk assessment. This recommendation parallels that found in Sepaher et al. (1999). The item content of the *SUI* scale is considered critical; thus, the automated scoring program available from Pearson Assessments prints the items endorsed on this scale. In the case of a clinician choosing to engage in hand scoring, it is recommended that they carefully review the responses for any items on this scale that are endorsed by the respondent. In the case of either automated or hand scoring, this content should be reviewed with the respondent as part of a suicide risk assessment.

#### INTERPRETATION

As noted above, an endorsement of any item results in an elevated ( $T \geq 65$ ) score. Elevated scores are associated with a preoccupation with suicide and death. Individuals with elevated *SUI* scores may have recently attempted suicide or be contemplating an attempt. They are described as feeling helpless and hopeless. As *T*-scores increase above 100, individuals are likely reporting a history of suicidal ideation and/or attempts, along with current suicidal ideation. Ben-Porath and Tellegen (2011) note that suicidal risk is elevated if there is evidence of poor impulse control or substance abuse (see scales *BXD*, *RC4*, *RC9*, *DISC-r*, and Substance Abuse [*SUB*]). The clinician should conduct an immediate and thorough suicide risk assessment in the presence of an elevated score on *SUI*.

#### *Helplessness/Hopelessness (HLP)*

The items of the *HLP* scale reflect the belief that one lacks the ability to make the necessary changes in their life to help them overcome their current difficulties and achieve their goals. *HLP* contains five items, of which four are keyed True. Four items overlap *TRT*, three on *TRT1*. Elevated scores are associated with feelings of depression, hopelessness, and pessimism about the future in both men and women. In addition, high scores in men are associated with sleep disturbance, feeling overwhelmed, and feeling like a failure. High scores in women are associated with poor sexual adjustment, low energy, and suicidal ideation.

Tellegen and Ben-Porath (2011) report a one-week test–retest correlation of .65. Internal consistency (Cronbach's alpha) estimates ranged from .39 to .75 across a variety of samples, with nominally higher estimates obtained within clinical samples.

#### INTERPRETATION

Non-elevated ( $T < 65$ ) scores are not interpreted. Elevated scores ( $T \geq 65$ ) are seen in individuals who are reporting a belief that the future will be unpleasant. They feel overwhelmed by their current situation and believe that life treats them unfairly. As *T*-scores increase above 80, individuals are reporting that they believe they are powerless to effect change in their lives. They tend to see negative outcomes as perhaps inevitable. Given their sense of powerlessness and the inevitability of negative outcomes, they are often lacking in motivation to attempt change.

#### *Self-Doubt (SFD)*

The *SFD* scale is the shortest of the internalizing scales, consisting of only four items, all keyed True. All overlap *LSE1*, and two overlap *DEP3*. It is impossible to obtain a *T*-score greater than 76 on this scale; endorsement of three items results in a *T*-score of 65. Tellegen and Ben-Porath (2011) report a test–retest correlation of .81. Cronbach's alpha estimates ranged from .67 to .84 across a variety of samples. The theme of the *SFD* items reflects a lack of confidence and a feeling of uselessness. Elevated scores are associated with self-degradation, as well as feelings of depression, hopelessness, self-doubt, and failure in both men and women.

## INTERPRETATION

Non-elevated ( $T < 65$ ) scores on *SFD* are not interpreted. Elevated scores ( $T \geq 65$ ) are seen in individuals who may report feeling insecure or inferior. They tend to lack self-confidence and may blame themselves for their shortcomings. They tend to ruminate over their perceived failings. A  $T$ -score of 76 indicates that these concerns are more prominent.

*Inefficacy (NFC)*

The *NFC* scale reflects a belief that one lacks the ability to effectively deal with both major and minor crises. *NFC* contains nine items, all keyed True. Four *NFC* items overlap each *Si*, *OBS*, and *TRT*; three items overlap *TRT*, with two of these on *TRT1*. Tellegen and Ben-Porath (2011) report a one-week test–retest correlation of .82 within a subset of the clinical sample. Internal consistency estimates ranged from .69 to .83 across a variety of samples.

It is one of only three internalizing scales on which low scores are interpreted. High scores are associated with hopelessness, low energy, a lack of self-reliance, and passivity in both males and females.

## INTERPRETATION

Low scores ( $T < 36$ ) are associated with self-reliance and an orientation toward power. Scores in the average range ( $T < 43$  to 64) reflect a relatively healthy balance of confidence and self-doubt. Elevated scores ( $T > 65$ ) are seen in individuals who report being passive. They have difficulty making decisions and doubt their abilities to effectively deal with life's unpleasant situations. They tend to give up easily when confronted with difficulties. When the *NFC*  $T$ -score reaches 80, these concerns are more prominent. Individuals scoring at this level report even greater difficulty with decision making and report being unable to effectively deal with even small crises.

*Stress/Worry (STW)*

The *STW* scale contains seven items, of which five are keyed True. *STW* overlaps *ANX* by five items, *NEGE* by four, and *TPA* by two items. Tellegen and Ben-Porath (2011) report a test–retest correlation of .77. Cronbach's alpha estimates ranged from .52 to .69 across a variety of samples.

The item content of *STW* includes financial concerns, time pressure worries, and a preoccupation with disappointments. Elevated scores have been associated with worry, anxiousness, depression, and feeling overwhelmed in both men and women. High scores in men are also associated with multiple somatic complaints and the development of physical symptoms in response to stress. High scores in women are associated with feeling as if life is a strain, complaints of sleep disturbance, and reports of suicidal ideation.

## INTERPRETATION

Low scores ( $T < 36$ ) are associated with less stress and worry than is normally reported. Scores in the average range ( $T < 43$  to 57) reflect unremarkable levels of stress. Elevated

scores ( $T > 65$ ) are seen in individuals who report an above-average amount of stress and worry. These individuals could be described as anxious or nervous. They may feel a time pressure to accomplish tasks. They may be concerned about financial matters. These individuals are prone to worry about situations and may ruminate over their concerns. They may develop physical symptoms in response to psychological stressors. If the *STW*  $T$ -score = 81, these concerns are more prominent and may involve multiple stressors. Individuals with elevated scores should be assessed for suicidal ideation.

### *Anxiety (AXY)*

The *AXY* scale contains five items, all keyed True, that are “clearly indicative of anxiety” (Ben-Porath & Tellegen, 2011, p. 52). *AXY* overlaps *ANX* by three items, and *FRS1* and *NEGE* by two items each. Tellegen and Ben-Porath (2011) report a one-week test–retest correlation of .71. Internal consistency estimates ranged from .42 to .71 across a variety of samples, with nominally higher estimates observed within clinical samples.

Each of the scale’s items was infrequently endorsed in the normative sample; thus, endorsement of two items produces an elevated  $T$ -score. Because the item content of the *AXY* scale is considered critical, the automated scoring program prints endorsed items from this scale. If the clinician chooses to engage in hand scoring, it is recommended that they carefully check the *ANX* items endorsed by the respondent. In the case of either automated or hand scoring, this content should be reviewed with the respondent.

Elevated scores are associated with post-traumatic stress disorder in trauma victims, but are not necessarily indicative of post-traumatic stress disorder. High scores have been associated with suicidal ideation, complaints of sleep disturbance, nightmares, hopelessness, worry, and depression in both men and women.

### INTERPRETATION

Non-elevated ( $T < 65$ ) scores are not interpreted. Elevated scores ( $T \geq 65$ ) are associated with anxiousness. Individuals with elevated scores may be experiencing sleep disturbances or nightmares. They do not cope well with stress and report significant anxiety, as well as problems associated with anxiety. Reports of suicidal ideation are associated with elevated  $T$ -scores on *AXY*. They may experience intrusive thoughts. When  $T$  reaches 100, reports of these problems escalate. At this level of elevation, respondents may be described as being anxious almost all the time. They may report having a sense of foreboding regarding some dire consequence. Sleep disturbances and nightmares are common at this level of elevation.

### *Anger Proneness (ANP)*

The items of the *ANP* scale focus on the negative emotional experience and expression of irritability, anger, and impatience with others, as well as the tendency to be easily upset or angered. *ANP* contains seven items, with five keyed True. *ANP* overlaps *NEGE* by four items, *ANG* by three, and *TPA1* by two. Tellegen and Ben-Porath (2011) report a one-week test–retest correlation of .81. Cronbach’s alpha estimates ranged from .71 to .80 across a variety of samples.

The correlates for men and women are quite different. Elevated scores in men have been associated with sleep disturbance, temper tantrums, low frustration tolerance, anger, agitation, and resentment, as well as the development of physical symptoms in response to stress and a preoccupation with physical concerns. Elevated scores in women are associated with poor coping abilities when faced with stress.

#### INTERPRETATION

Non-elevated scores ( $T < 65$ ) are not interpreted. Elevated scores ( $T \geq 65$ ) are associated with individuals who are argumentative and hold grudges. They are often described as being irritable and having low frustration tolerance. They tend to be quick to anger and are likely to have temper tantrums when angry, especially if male. As  $T$  elevates to a score of 80, these problems become more pronounced. Individuals scoring at this level may report being overcome by their anger.

#### *Behavior-Restricting Fears (BRF)*

The *BRF* scale contains nine items, all of which appear on *FRS1*. Eight of the items are keyed True. Tellegen and Ben-Porath (2011) report a one-week test–retest correlation of .67. Internal consistency estimates ranged from .44 to .63 across a variety of samples.

The item content of *BRF* generally refers to fears that restrict one's involvement in activities both within and outside of the home. Ben-Porath and Tellegen (2011) note that elevated scores are associated with agoraphobia, as well as with general fearfulness. High scores in men have been associated with low competitiveness, low achievement needs, and low self-reliance. High scores in women have been associated with multiple fears, nightmares, nervousness, and suicidal ideation.

#### INTERPRETATION

Non-elevated ( $T < 65$ ) scores are not interpreted. Elevated ( $T \geq 65$ ) scores on *BRF* are associated with fearfulness to the degree that it restricts one's activities. Individuals with elevated scores may report generalized anxiety and subjective depression. They may feel anxious when away from home. They may be uncompetitive and have low needs to achieve; such individuals are usually not work-oriented. As the  $T$ -score increases above 90, these reports become more widespread. Individuals scoring in this range may report multiple fears that interfere with their daily lives.

#### *Multiple Specific Fears (MSF)*

*MSF* contains nine items, with five keyed False. All overlap *FRS*, with eight on *FRS2*. Tellegen and Ben-Porath (2011) report a one-week test–retest correlation of .85. Cronbach's alpha estimates ranged from .69 to .72 across a variety of samples. The item content of the *MSF* scale describes unrelated fears of various animals and natural phenomena, such as floods, fire, and blood and item endorsement may indicate an elevated risk for specific phobias. No empirical correlates have been found for elevated scores in men; thus, we recommend interpreting elevations for men only in terms of their



self-report. High scores in women are associated with low aspirations and achievement needs, low competitiveness, stereotypical gender interests, and low energy.

#### INTERPRETATION

Low scores ( $T < 39$ ) are seen in individuals who report fewer specific fears than average. Scores in the average range ( $T=42$  to  $59$ ) reflect an unremarkable number of specific fears and are not generally interpreted. Elevated scores ( $T \geq 65$ ) are seen in individuals who may be described as risk-averse. They report multiple fears regarding various animals and acts of nature, including thunder, natural disasters, and fire. Women may tend to be passive and harm-avoidant.

#### *Externalizing Scales*

According to Ben-Porath and Tellegen (2011), the externalizing scales were created to address specific facets of *RC4* and *RC9* on the RC scales. Two scales, Juvenile Conduct Problems (*JCP*) and Substance Abuse (*SUB*) specifically address the two broad content areas of *RC4*. Likewise, two scales, Aggression (*AGG*) and Activation (*ACT*) measure the two specific areas of interest assessed by *RC9*; low scores are interpretable on these latter two scales. Thus, these scales may be useful in helping the clinician clarify the specific problems leading to elevations on *RC4* or *RC9*, which has been suggested as an area of concern, given the breadth of the item content in the two scales (cf. Bolinsky & Nichols, 2011; Nichols, 2006). Ben-Porath and Tellegen suggest that the externalizing scales can be interpreted even in the absence of elevations on *RC4* or *RC9*. However, Graham (2012) suggests that the discriminant validity among the externalizing scales is not very good.

#### *Juvenile Conduct Problems (JCP)*

The items of the *JCP* scale refer to a history of legal trouble and conduct problems when young. *JCP* contains six items, all keyed True. Five items overlap with *DISC*, four each with *ASP2* and *MAC-R*, and three with *Pd* (two of these on *Pd2*). Tellegen and Ben-Porath (2011) report a one-week test-retest correlation of .85 within a subset of the normative sample. Internal consistency (Cronbach's alpha) estimates ranged from .56 to .75 across a variety of samples.

Elevated scores are associated with antisocial behavior, holding grudges, and feeling that one's family lacks love among both men and women. Additional correlates of elevated scores in men are being physically abusive, angry, and aggressive, temper tantrums, and a history of stormy interpersonal relationships. High scores in women are also associated with trust difficulties, superficial relationships, deception, and low frustration tolerance.

#### INTERPRETATION

Non-elevated ( $T < 65$ ) scores are not interpreted. Elevated ( $T \geq 65$ ) scores on *JCP* are associated with reports of conduct behaviors in school. They may have histories of

illegal behavior as adolescents, as well as additional legal trouble as adults. They tend to have problems with authority figures, in general, and may especially report conflictual relationships with members of their family. Physical abuse of others is more common among men. As the *T*-score increases above 80, these reports become more widespread. It is impossible to score above *T*-84 on this scale.

### *Substance Abuse (SUB)*

The *SUB* scale contains seven items, all of which overlap *AAS*, with six keyed True. Tellegen and Ben-Porath (2011) report a test–retest correlation of .87. Internal consistency estimates ranged from .62 to .77 across a variety of samples. The items broadly refer to significant abuse of substances—either currently or in the past—with alcohol-related items being predominant. Elevations on *SUB* have been associated with sensation-seeking, risk for substance-use problems, difficulty trusting others, and self-defeating behaviors.

The item content of *SUB* has been identified as possibly requiring immediate attention by the tests' authors. Because of this, the automated scoring program prints the responses answered in the keyed direction when the scale is elevated. When not using an automated scoring program, it is suggested that the clinician manually review the endorsed items when the scale is elevated. These items should then be reviewed with the respondent.

### INTERPRETATION

Non-elevated ( $T < 65$ ) scores are not interpreted. Elevated ( $T \geq 65$ ) scores on *SUB* are associated with frank admissions of past and/or current substance use. As the score elevates above *T*-80, the possibility of current substance abuse is greater. These individuals are reporting more frequent use of substances and may find it difficult to relax without the use of substances. Thus, they may experience more difficulties in their interpersonal relationships. They may be described as argumentative and may be physically aggressive with others, especially if the respondent is male.

### *Aggression (AGG)*

The nine items comprising the *AGG* scale reflect physically aggressive behavior; all are keyed True. Six of the *AGG* items overlap *AGGR* and *ANG*, five on *ANG1*. Tellegen and Ben-Porath (2011) report a one-week test–retest correlation of .78. Cronbach's alpha estimates ranged from .58 to .76 across a variety of samples.

Temper tantrums are associated with elevated scores in both men and women. In addition, high scores in men are associated with holding grudges, resentment, stormy interpersonal relationships, and physical abuse of others. Elevated scores in women are associated with trust difficulties, grouchiness, and cynicism. As elevated scores may be associated with a history of violence and abusive behavior, this scale is deemed to contain critical content. Thus, the automated scoring program will print endorsed items if the scale score is elevated. The clinician is urged to manually check for endorsed item content if hand scoring is used. This is one of two Externalizing scales for which low scores are interpreted.

#### INTERPRETATION

Low ( $T < 39$ ) scores are associated with a below average level of aggressive behavior, as long as the self-favorable validity scales ( $L-r$ ,  $K-r$ ) do not suggest a tendency to under-report problems. Average elevations ( $T=45$  to  $61$ ) are not interpreted. Elevated ( $T \geq 65$ ) scores are seen in individuals who report acting aggressively toward others. They may be physically aggressive and violent. They may have histories of legal difficulties and behavioral problems in school. As scores elevate above  $T=80$ , these behaviors become more prominent. Such individuals may enjoy intimidating others or causing others to fear them.

#### *Activation (ACT)*

The *ACT* scale contains eight items, all keyed True. *ACT* overlaps *Pt*, *Ma* (two of these on *Ma2*), and *APS* by three items each; it overlaps *Sc5* by two items. Tellegen and Ben-Porath (2011) report a test-retest correlation of .77 within a subset of the normative sample. Internal consistency (Cronbach's alpha) estimates ranged from .59 to .75 across a variety of samples.

The items address broad aspects associated with hypomanic activation, including a decreased need for sleep, mood swings, and heightened excitement and energy. Tellegen and Ben-Porath (2011) caution that the possibility that elevations on this scale may reflect substance-induced problems should be evaluated by the clinician; the clinician may find it useful, then, to interpret elevations on this scale in light of scores on *SUB*. Tellegen and Ben-Porath reported no empirical correlates above  $|.20|$  for elevations on *ACT*. Thus, elevations on this scale should only be interpreted as reflecting the respondent's self-report.

#### INTERPRETATION

Low ( $T < 39$ ) scores indicate that the respondent endorsed a below-average number of items reflecting increased energy or excitement. The clinician is encouraged to consider scores on *RCd*, *RC2*, and the internalizing scales to assess for the possibility of depression. Average elevations ( $T=44$  to  $59$ ) are not interpreted. These scores reflect the endorsement of an average number of items reflecting increased energy or excitement. Elevated scores ( $T \geq 65$ ) indicate that the respondent endorsed items reflecting increased excitement and energy. They may be reporting a decreased need for sleep. As scores elevate above  $T=80$ , these reports become more predominant. When scores reach this level, the respondent is indicating that they experience uncontrollable mood swings. They are reporting an increased energy level and a decreased need for sleep, such that the clinician would be well-advised to assess for other signs of mania or hypomania, including pressured speech or flight of ideas, expansive mood, and an increased focus in pleasurable activities.

**Interpersonal Scales**

The interpersonal scales consist of five scales whose primary focus is on interpersonal functioning. They are Family Problems (*FML*), Interpersonal Passivity (*IPP*), Social Avoidance (*SAV*), Shyness (*SHY*), and Disaffiliativeness (*DSF*). Tellegen and Ben-Porath (2011) report low to moderate correlations among these scales within the normative sample.

*Family Problems (FML)*

The *FML* scale includes 10 items, 7 keyed True. Nine of these items overlap *FAM* (five on *FAM1*, two on *FAM2*), and two each with *Mf*, *Pd1*, and *Sc1*. Tellegen and Ben-Porath (2011) report a one-week test–retest correlation of .78. Cronbach’s alpha estimates ranged from .64 to .78 across a variety of samples.

These items refer to a variety of familial issues, such as dislike of one’s family members, feeling unappreciated by one’s family, and feeling that one cannot count on one’s family. High scores are associated with familial discord, family resentment, blaming one’s family for one’s difficulties, and feeling as if one’s family lacks love, in both men and women.

## INTERPRETATION

Low ( $T < 39$ ) scores are associated with individuals who report that their past and present relationships with their family are relatively conflict-free. Individuals with scores in this range do not tend to blame their families for any difficulties they might be experiencing. Average ( $T$ -40 to 63) scores are not interpreted. These individuals are reporting an unremarkable number of family conflicts. Elevated scores ( $T \geq 65$ ) are associated with reports of family discord. These individuals tend to feel that their family does not provide them the type of support and understanding that they deserve; they may blame their family for their current and past difficulties. As scores elevate above  $T$ -80, the amount of discord within the family, as well as the amount of resentment and blame on the part of the respondent, are likely to increase.

*Interpersonal Passivity (IPP)*

The *IPP* scale contains items that broadly refer to unassertive or submissive behavior in interpersonal interactions. The scale contains 10 items, of which 9 are keyed False. Four items overlap with *R*, two with *Si*. Tellegen and Ben-Porath (2011) report a one-week test–retest correlation of .78. Internal consistency estimates ranged from .68 to .77 across a variety of samples.

High scores are associated with passivity in both men and women and are negatively related to extroversion. Elevated *IPP* scores are associated with poor sexual adjustment, a low sex drive, self-doubt, perfectionism, and pessimism in men. High scores in women are associated with introversion, social awkwardness, and submissiveness.

#### INTERPRETATION

Low scores ( $T < 39$ ) are seen in individuals who describe themselves as assertive and someone who will stand up for themselves. They often see themselves as leaders, although others may see them as domineering or self-centered. Average elevations ( $T$ -43 to 62) are not interpreted. Individuals with scores in this range are reporting an average balance of assertiveness and passivity. Elevated ( $T \geq 65$ ) scores on *IPP* are associated with individuals who describe themselves as unassertive. They often lack confidence in themselves and report a dislike for leadership roles. As scores increase above  $T$ -80, this unassertiveness becomes more pronounced. Individuals with scores in this range do not like being in social situations and may be viewed as awkward or shy by others. They are often submissive in interpersonal relationships.

#### *Social Avoidance (SAV)*

The SAV scale contains 10 items, with 9 keyed False. All items overlap *Si* (seven on *Si1*, three on *Si2*) and *SOD* (nine on *SOD1*, one on *SOD2*), and six items overlap *INTR*. Tellegen and Ben-Porath (2011) report a one-week test–retest correlation of .84. Internal consistency (Cronbach's alpha) estimates ranged from .77 to .86 across a variety of samples.

The content of the SAV items reflects the respondent's report of not enjoying social interactions and actively avoiding them. High scores in both men and women are associated with feelings of hopelessness, sadness, and depression, as well as with introversion. Elevated scores in men are also associated with sleep disturbance, feeling like a failure, self-doubt, and feeling that life is a strain. Elevated scores in women are associated with low energy and aspirations, shyness, and social awkwardness. Ben-Porath and Tellegen (2011) suggest that elevations on SAV in the presence of a non-elevated Shyness (*SHY*) score may suggest the presence of avoidant personality, rather than social anxiety, particularly if Self-Doubt (*SFD*) and Inefficacy (*NFC*) are elevated.

#### INTERPRETATION

Low scores ( $T < 39$ ) are seen in individuals who describe themselves as enjoying social interactions. They may engage in a variety of social events. Average elevations ( $T$ -44 to 59) are not interpreted. Elevated ( $T \geq 65$ ) scores on SAV are associated with individuals who describe not enjoying social activities and interactions. They may be described as introverted. As with individuals scoring high on *IPP*, they may suffer from a lack of self-confidence and avoid positions of leadership. They are often passive in interpersonal relationships. As scores increase above  $T$ -80, their difficulties increase in magnitude. They may report feelings of sadness and depression. They may feel hopeless and that they have little power to effect positive change.

#### *Shyness (SHY)*

The items of the Shyness (*SHY*) scale refer to various ways in which social anxiety may be manifested, such as embarrassment and discomfort in social situations. The scale

contains seven items, six of which are keyed True. All items overlap *Si1*, five overlap *SOD* (four on *SOD2*, one on *SOD1*), and three overlap *Pt*. Tellegen and Ben-Porath (2011) report a test–retest correlation of .88. Internal consistency estimates ranged from .74 to .80 across a variety of samples.

High scores in men are associated with anxiousness, depression, worry, reports of sleep disturbance, passivity, feeling overwhelmed, self-doubt, and discomfort around women. High scores in women are associated with introversion, low energy, passivity, shyness, poor sexual adjustment, and social awkwardness. It is impossible to score above *T*-75 on this scale.

#### INTERPRETATION

Low ( $T < 39$ ) scores are associated with reports of little or no social anxiety. These individuals may be described as feeling very comfortable in social situations. Average ( $T$ -44 to 57) elevations are not interpreted. These individuals are reporting an unremarkable balance of comfort and anxiety in social situations. Elevated ( $T \geq 65$ ) scores are associated with reports of shyness. These individuals are uncomfortable being around others in social situations, especially members of the opposite sex. They may be described as being socially awkward or introverted.

#### *Disaffiliativeness* (DSF)

The *DSF* scale contains only six items, five keyed True. Scores below *T*-44 cannot be obtained on this scale. *DSF* overlaps *SOD1* by three items, and *FB* and *Sc1* by two items each. Tellegen and Ben-Porath (2011) report a one-week test–retest correlation of .60 within a subset of the normative sample, which may suggest a state-like quality to the concerns measured by this scale, or may be an artifact of the scale having so few items. Internal consistency estimates ranged from .43 to .65 across a variety of samples, which may, again, reflect the small number of items on the scale.

The *DSF* items refer to a dislike of people, having never had a close relationship, and a preference for solitude. Elevated scores reflect an asocial individual, although extremely high scores may be associated with schizoid personality. High scores in men have been associated with complaints of sleep disturbance, including nightmares; feelings of failure; hopelessness; depression; and a preoccupation with health problems. No empirical correlates have been reported for women; thus, we recommend that elevated scores in women be interpreted only in the context of self-report.

#### INTERPRETATION

Average ( $T$ -44 to 58) elevations are not interpreted. These individuals are reporting neither a preference for solitude nor a dislike of others. Elevated ( $T \geq 65$ ) scores are seen in individuals who report that they dislike being around others. As the *T*-score increases, this dislike for social involvement becomes more pronounced and individuals tend to prefer solitude. Men may be experiencing symptoms of depression, including sleep disturbance and hopelessness. Extremely elevated ( $T > 100$ ) scores may be associated with individuals who have never had a close relationship. When



*T* is elevated to this level, the clinician should consider evaluating for schizoid personality disorder.

### **Interest Scales**

As noted in the previous chapter on the RC scales, seed items for two scales were derived from clinical Scale 5 (Masculinity-Femininity) during the course of the derivation of the RC scales. These seed items formed the basis of the interest scales, which consist of the Aesthetic-Literary Interests (*AES*) and Mechanical-Physical Interests (*MEC*) scales. The scales are essentially uncorrelated with one another, which means that an individual can score high on both, low on both, or high on one and low on the other. Ben-Porath and Tellegen (2011) suggest that low scores on both scales may reflect a lack of outside interests; in some cases, low scores on both scales may indicate psychological disengagement from the environment.

#### ***Aesthetic-Literary Interests (AES)***

The *AES* scale contains seven items, with each keyed True. All overlap *Mf*, with five appearing on *Mf4* and two on *Mf3*. Tellegen and Ben-Porath (2011) report a test-retest correlation of .86 within a subset of the normative sample. Cronbach's alpha estimates ranged from .49 to .66 across a variety of samples.

The items of *AES* reflect an interest in occupations or activities of an aesthetic or literary nature, such as working in a library or with flowers. Elevated scores are associated with stereotypic feminine behavior and a rejection of traditional gender roles in men. No empirical correlates have been found for high scores among women. For this reason, we suggest that elevated scores in women be interpreted only in terms of self-report of item content.

#### ***Interpretation***

Low scores ( $T < 39$ ) indicate a lack of interest in occupations or activities associated with the arts or of an aesthetic nature. Average ( $T$ -39 to 62) elevations are not interpreted. Elevated ( $T \geq 65$ ) scores are associated with a reported interest in aesthetic or literary activities or occupations. Individuals scoring in this range are often described as being empathic. Males are described as not having stereotypical gender interests. It is impossible to score above *T*-73 on this scale.

#### ***Mechanical-Physical Interests (MEC)***

The items of the *MEC* scale refer broadly to an interest in activities or occupations of a mechanical or physical nature. Such activities include building things, sports, and other outdoor activities. High scores on *MEC* in men are associated with stereotypical masculine interests, few concerns about homosexuality, low self-doubt, and little difficulty making decisions. As with *AES*, no empirical correlates have been found for high scores among women. For this reason, we suggest that elevated scores in women be interpreted only in terms of self-report of item content.

MEC contains nine items, all keyed True. All overlap *MfI*, and two items each overlap *MAC-R* and *DISC*. Tellegen and Ben-Porath (2011) report a one-week test-retest correlation of .92. Internal consistency (Cronbach's alpha) estimates ranged from .55 to .64 across a variety of samples.

### *Interpretation*

Low scores ( $T < 39$ ) are associated with a lack of interest in activities or occupations of a mechanical or physical nature. Average ( $T=39$  to  $62$ ) elevations are not interpreted. Elevated ( $T \geq 65$ ) scores are associated with an above-average interest in physical or mechanical activities or occupations. Individuals scoring in this range may be described as having stereotypically masculine interests. They may be high in sensation-seeking or adventurousness.

### **Personality Psychopathology Five (PSY-5) Scales**

The MMPI-2 PSY-5 scales (Harkness, McNulty, & Ben-Porath, 1995) reflect a dimensional five-factor trait model developed specifically for application to personality pathology. The PSY-5 constructs were originally developed by Harkness and McNulty (1994) from normal personality terms and from descriptors of abnormal personality taken from the *Diagnostic and Statistical Manual of Mental Disorders III—Revised* (DSM-III-R; American Psychiatric Association, 1987), as an aid for the description of normal personality and to provide a dimensional complement to the diagnosis of personality disorders. The MMPI-2 scales were developed using a combination of rational and statistical procedures, termed *replicated rational selection*, to select MMPI-2 items that measured each of the PSY-5 constructs. The resultant MMPI-2 PSY-5 scales were named Aggressiveness (*AGGR*), Psychoticism (*PSYC*), Disconstraint (*DISC*), Negative Emotionality/Neuroticism (*NEGE*), and Introversion/Low Positive Emotionality (*INTR*). The reader is encouraged to read the recent review of these scales by Harkness, Finn, McNulty, and Shields (2012).

Although there are some similarities between the constructs of the Five-Factor Model (FFM), as measured by instruments such as the NEO-PI-R (Costa & McCrae, 1992a) and the PSY-5 scales, there are also important differences, and these differences have important implications for psychological assessment. For example, as the PSY-5 scales are intended to measure the domains of disordered personality, some scales, such as *PSYC*, have no direct correspondent among the NEO-PI-R scales, just as the NEO-PI-R *Openness* scale has no direct PSY-5 correspondent. Although there is some overlap among the remaining scales in the two instruments, it is important to note that even for the PSY-5 scales that show conceptual and empirical overlap with the FFM constructs, the PSY-5 scales tend to have a higher “ceiling” for maladaptive levels of the personality traits. Commenting specifically on the PSY-5's inclusion of a psychoticism factor, Krueger et al. (2011, p. 182) have suggested that the “... PSY-5 is a highly prescient model in this regard and aligns closely with a model suitable for DSM-5.”

The reader is encouraged to review the earlier description of the development of the MMPI-2 PSY-5 scales and their clinical correlates for additional information regarding

their clinical utility. This section will focus on the creation of a revised group of PSY-5 scales for MMPI-2-RF.

To adapt the PSY-5 scales for MMPI-2, Harkness and McNulty (2007) first began with the 96 items (of the original 138 MMPI-2 PSY-5 items) that remained in the 338-item MMPI-2-RF pool. Incorporating a dual-method of item-scale and item-criterion analyses, Harkness and McNulty dropped 22 of the surviving items from the revised scale and added 30 items that had not previously appeared on any of the PSY-5 scales (although they were available in the MMPI-2 item pool). The resultant scales consist of 104 items; there is no item overlap among the revised PSY-5 scales, although there is substantial overlap with other MMPI-2-RF scales. The revised scales are known as Aggressiveness-Revised (*AGGR-r*), Psychoticism-Revised (*PSYC-r*), Disconstraint-Revised (*DISC-r*), Negative Emotionality/Neuroticism-Revised (*NEGE-r*), and Introversion/Low Positive Emotionality-Revised (*INTR-r*). Each will be described individually.

### ***Aggressiveness-Revised (AGGR-r)***

The *AGGR-r* scale contains 18 items referring to “aggressively assertive behavior” (Ben-Porath & Tellegen, 2011), with 16 of these keyed True. Fourteen items are common to the MMPI-2 *AGGR* scale, whereas four are unique to *AGGR-r*. Of the four items appearing on *AGGR*, but not on *AGGR-r*, two were lost from the MMPI-2-RF item pool and two were dropped from the revised scale.

Tellegen and Ben-Porath (2011) report very strong negative correlations, ranging from  $-.87$  to  $-.89$ , between *AGGR-r* and Interpersonal Passivity (*IPP*) in both men and women, across a variety of treatment settings and within the normative sample. Moderate positive correlations, ranging from  $.56$  to  $.68$ , were reported with *RC9*. Internal consistency estimates (Cronbach’s  $\alpha$ ) ranged from  $.71$  to  $.75$  within these same samples. The test–retest coefficient within a subset of the normative sample was  $.84$ .

High scores are associated with antisocial behavior, aggressiveness, extraversion, and assertiveness. Low scores are associated with passivity, submissiveness, and proneness to guilt.

### ***Interpretation***

Low ( $T < 39$ ) scores are associated with a lack of assertiveness. These individuals may be described as being submissive in interpersonal relationships. They prefer to let others take the lead and rarely stand up for themselves. They tend to blame themselves for their shortcomings.

Average ( $T$ -41 to 60) elevations are not interpreted. Individuals scoring in this range are reporting a balance between passivity and assertiveness.

Elevated ( $T \geq 65$ ) scores may reflect assertiveness and self-confidence at the lower end of the elevated range. They may be extraverted and view themselves as having leadership qualities. As  $T$  elevates above 70, however, the likelihood of aggressive and domineering behavior increases. These individuals may have a history of physical or instrumental aggression and they may try to intimidate others through the threat of aggression. At this level, the positive self-impression seen in lower elevations takes on a narcissistic quality. High scores are associated with the absences of feelings of guilt or remorse.

***Psychoticism-Revised (PSYC-r)***

The 26 items of the *PSYC-r* scale refer to experiences that are associated with thought disturbance; 25 of these items are keyed True. *PSYC-r* shares 17 items with the MMPI-2 *PSYC* scale, which contains 25 items; 9 items appearing on *PSYC-r* were not included on *PSYC*. Of the eight items appearing on *PSYC*, but not on *PSYC-r*, five do not appear in the MMPI-2-RF item pool, whereas three were not included on the revised scale. As noted in the discussion of the higher-order scales, *PSYC-r* shares 22 items with *THD*, which essentially makes it more similar to another scale appearing on MMPI-2-RF than the scale upon which it was based. Whereas one would expect a correlation of .85 between *PSYC-r* and *THD* on the basis of item overlap, the correlation between *PSYC-r* and *PSYC* would be only .67. For this reason, we reiterate our earlier warning that *PSYC-R* and *THD* are redundant; thus, the clinician should interpret only one of these scales and should never use an elevated score on one of these scales as independent confirmation of traits or symptoms suggested by an elevated score on the other.

As noted in the discussion of *THD*, Tellegen and Ben-Porath (2011) report very strong positive correlations, ranging from .95 to .98 between *PSYC-r* and *THD* in both men and women, across a variety of treatment settings and within the normative sample. Additionally, *PSYC-r* scores were highly related to *RC8* scores, with correlations ranging from .87 to .90, in these same samples. Cronbach's alpha estimates ranged from .69 to .88 within these samples, with nominally higher estimates seen among psychiatric patients than within non-patient samples. The test-retest coefficient within a subset of the normative sample was .76.

High scores are associated with a variety of unusual thoughts and perceptual experiences, along with a feeling of alienation from others. Low scores are associated with an absence of these experiences and feelings.

***Interpretation***

Low ( $T < 39$ ) scores are associated with denial of any type of perceptual disturbance, unusual thoughts, or feelings of alienation.

Average ( $T$ -47 to 63) elevations are not interpreted. Individuals scoring in this range are reporting an unremarkable number of unusual perceptual experiences and thoughts.

Elevated ( $T \geq 65$ ) scores are associated with unrealistic thinking. They are reporting some unusual thoughts and perceptual disturbances. These individuals may feel that life is a strain and have difficulty coping with the demands of their daily lives. As the score elevates above  $T$ -75, the magnitude of the disturbance increases. These individuals are reporting a greater likelihood of psychotic symptoms, such as delusional beliefs or hallucinations. They may exhibit impaired reality testing. They are often described as anxious or depressed. They are likely to have difficulties with interpersonal functioning. The clinician should evaluate for the presence of a psychotic disorder or an associated personality disorder.

***Disconstraint-Revised (DISC-r)***

The *DISC-r* scale contains 20 items, with 17 keyed True, which refer broadly to impulsiveness and risk-taking. Of the 20 items on *DISC-r*, 13 are common to the MMPI-

2 *DISC* scale, which contains 29 items, whereas 7 items are unique to *DISC-r*. Of the 16 items that appear on *DISC*, but not on *DISC-r*, 11 were not included in the MMPI-2-RF item pool and 5 were dropped from the revised scale. As noted previously in the discussion of the higher-order scales, *DISC-r* shares 15 items with *BXD*, which means that it has more overlap with another MMPI-2-RF scale than with its namesake MMPI-2 scale. On the basis of item overlap alone, one would expect a correlation of .70 between *DISC-r* and *BXD*, as opposed to a correlation of .54 between *DISC-r* and *DISC*.

As might be expected, given the substantial item overlap, Tellegen and Ben-Porath (2011) report correlations, ranging from .89 to .92, between *DISC-r* and *BXD* across a variety of treatment settings and within the normative sample. Further, correlations ranging from .76 to .83 were reported in these samples for *DISC-r* and *RC4*, although only three items are shared between the two scales. Moderate positive correlations, ranging from .57 to .62, were reported with *RC9*, with which *DISC-r* shares five items. *DISC-r* and *AGGR-r* had positive correlations ranging from .22 to .50. Cronbach's alpha estimates ranged from .69 to .75 within these same samples. The test-retest coefficient within a subset of the normative sample was .93. Given the substantial item overlap and high correlations between *DISC-r* and *BXD*, it is not necessary to interpret both scales. High scores on *DISC-r* have been associated with antisocial behavior, lack of impulse control, narcissism, deception, and superficial relationships.

### *Interpretation*

Low ( $T < 39$ ) scores on *DISC-r* are associated with reports of overly constrained behavior. These individuals may be described as being inhibited and over-conventional.

Average ( $T$ -41 to 63) elevations are not interpreted. Individuals scoring in this range are reporting an unremarkable balance between behavioral constraint and disinconstraint.

Elevated ( $T \geq 65$ ) scores are associated with behavioral disinconstraint. Individuals at the lower end of the elevated range may be described as impulsive, nonconformist, or sensation-seeking. As scores elevate above  $T \geq 75$ , however, this lack of behavioral constraint becomes more severe, as do the problems and consequences associated with it. Individuals scoring in this range show a decreasing ability to use good judgment. They are often described as rule-breakers and may have significant histories of legal difficulties and behavioral problems in school. They have narcissistic tendencies and may manipulate or deceive others in order to get what they want. The clinician should evaluate for narcissistic or antisocial personality disorder. Individuals with elevated scores are not often motivated to engage in treatment and are unlikely to initiate treatment of their own accord. Once enrolled in treatment, they tend to be non-compliant and/or to drop out of treatment early.

### ***Negative Emotionality/Neuroticism-Revised (NEGE-r)***

The items of the *NEGE-r* scale reflect a variety of negative emotional experiences and are associated with the Neuroticism dimension of the five-factor model of personality. *NEGE-r* contains 20 items, with 15 keyed True. *NEGE-r* shares 14 items with the 33-item MMPI-2 *NEGE* scale; 6 of the *NEGE-r* items are not included on *NEGE*. Of the 19 items appearing on *NEGE*, but not on *NEGE-r*, 11 are not in the MMPI-2-RF item pool,

whereas 8 items were dropped from the revised scale. *NEGE-r* shares five items with the *EID* Higher-Order scale and six with *RC7*.

Even with the fairly low number of shared items, Tellegen and Ben-Porath note large positive correlations of *NEGE-r* scores with both *EID* and *RC7*, ranging from .73 to .81 and from .82 to .87, respectively, in both men and women across a variety of samples. Cronbach's alpha estimates ranged from .76 to .84 within these samples. The test-retest coefficient within a subset of the normative sample was .85.

High scores are associated with anxiety, depression, insecurity, and worry. High scores are also associated with the tendency to feel hopeless, to feel as if one's family lacked love, and to develop physical responses to stress. Low scores are associated with a lack of negative emotions, as well as feeling energetic.

### *Interpretation*

Low ( $T < 39$ ) scores reflect the absence of negative emotionality. Individuals scoring in this range can be described as being essentially free from worry. They report having good energy and having a positive outlook on life. They report a good capability to cope with stress.

Average ( $T$ -40 to 62) elevations are not interpreted. Individuals scoring in this range are reporting an unremarkable number of negative emotional experiences.

Elevated ( $T \geq 65$ ) scores are associated with reports of emotional distress. These individuals may be described as "worriers." They have a tendency to expect the worst and may blame themselves if it comes to pass. As scores elevate above  $T$ -75, this anxiousness becomes more predominant. Individuals scoring in this range may find themselves behaviorally inhibited due to excessive anxiety regarding possible negative consequences. They may develop physical complaints in response to stressors; thus, the clinician is urged to examine *RC1* and the somatic scales for possible elevations.

### ***Introversion/Low Positive Emotionality (INTR-r)***

*INTR-r* contains 20 items, all keyed False. The items reflect the broad category of a dearth of positive emotional experiences, as well as avoidance of social interaction. *INTR-r* shares 16 items with the MMPI-2 *INTR* scale, which contains 34 items; 4 items appear on *INTR-r*, but are not included on *INTR*. Of the 18 *INTR* items not included in *INTR-r*, 13 do not appear in the MMPI-2-RF item pool and 5 were not included in the revised scale. *INTR-r* shares 10 items with *RC2*, 8 with *EID*, and 5 (reverse keyed) with *RC9*.

Given the number of shared items between *INTR-r* and the scales noted above, one would expect to find significant relationships. The pattern of relationships reported by Tellegen and Ben-Porath is interesting, and bears examination. Within the non-clinical normative sample, moderate sized correlations of  $-.52$  for men and  $-.46$  for women were reported with *RC9*. However, within clinical samples, these correlations were nominally smaller, ranging from  $-.35$  to  $-.38$ . An opposite pattern is observed, however, with regard to *EID* and *INTR-r*. In the normative sample, *EID* correlated with *INTR-r* at only .34 for men and .37 for women; in the clinical samples, however, this correlation was substantially larger, ranging from .63 to .65. The effect was similar, although not as dramatic for *RC2*, as the normative sample evidenced a correlation of .74 for men and .69



for women, but the clinical samples evidenced correlations ranging from .84 to .86. That similarities increase as a function of clinical distress is not surprising, and may indicate that *EID* and *INTR-r*, in particular, measure separate facets of the same construct.

High scores on *INTR-r* have been associated with feeling depressed, anxious, hopeless, and like a failure. Low scores are associated with optimism, extraversion, and feeling energetic.

### *Interpretation*

Low ( $T \leq 39$ ) scores are associated with individuals who report more positive emotional experiences than average. Individuals scoring in this range can be described as knowing what they want and having the energy to go after it. They enjoy social interactions and have confidence in their social skills. They report having a good ability to cope with stress.

Average ( $T$ -42 to 64) elevations are generally not interpreted. These scores are obtained by individuals who endorse an unremarkable number of positive emotional experiences.

Elevated ( $T \geq 65$ ) scores are associated with individuals who report fewer positive emotional experiences than average. They are likely to present as socially awkward or introverted; they have little confidence in their social skills. As scores elevate into the range of  $T$ -75, the lack of positive emotional experiences becomes more prominent. Such individuals may appear anhedonic and lacking in energy. They may experience significant depression and anxiety. These individuals are likely to feel overwhelmed by the difficulties they are facing. They tend to feel like a failure and have little faith in their ability to effect positive change in their lives. Given their pessimism and poor self-concept, it may be difficult to engage them in psychotherapy.

### **Perspective**

As noted earlier, the MMPI-2-RF *Manual* notes that the instrument is not being marketed as a replacement for MMPI-2, but as an alternate version of the form (Ben-Porath & Tellegen, 2011), although its creators tout its purported psychometric superiority over the earlier form. The MMPI-2-RF shares some similarities to the earlier form, especially with regard to the validity scales and the RC scales, but also contains many unique features that reflect the particular goals of its authors, among which were improved psychometric characteristics and the creation of non-overlapping scales within content areas.

Given the overwhelming and enduring popularity of the MMPI-2 and its forerunner, the MMPI, it should be expected that the introduction of the MMPI-2-RF has been met with its share of detractors. Greene (2011, p. 22), in fact, has argued that the “MMPI-2-RF should *not* be conceptualized as a revised or restructured form of the MMPI-2, but as a *new* self-report inventory that chose to select its items from the MMPI-2 item pool and to use its normative group” (emphasis in original). This chapter is not the place for a thorough review of all of the criticisms and defenses of the MMPI-2-RF; the reader is encouraged to review Butcher and Williams’ (2012) criticism, as well as Ben-Porath and Flens’ (2012) response.

It has been observed that a great strength can also be a weakness. Among the advantages of the MMPI-2-RF is its substantially shorter length (338 items) in comparison to the

standard MMPI-2 (567 items). As noted by Graham (2012), many test takers may find the abbreviated length more manageable, especially if they are being asked to complete a battery of multiple assessments. However, with decreased length comes decreased coverage, although the creators of MMPI-2-RF claim that they adequately covered all substantive content areas of the MMPI-2 item pool.

Another advantage of the MMPI-2-RF is that it is laid out in somewhat of a top-down approach with higher-order scales and specific problems scales. As both Graham (2012) and Greene (2011) note, this arrangement can make interpretation much simpler and less time-consuming than with the standard MMPI-2. It can also greatly decrease the learning curve for students.

One issue that, from our experience, appears to be a common misunderstanding with regard to MMPI-2-RF is the belief that the MMPI-2-RF is composed of completely non-overlapping scales. It should be noted that Tellegen and Ben-Porath (2011) are quite clear with respect to this issue; the fault for this misunderstanding does not lie with them. Although the scales within each area (i.e. higher-order, RC, specific problem, PSY-5 scales) do not share items, there is *substantial* item overlap across areas. As noted above, the *PSYC-r* PSY-5 scale shares more items with *THD*, an MMPI-2-RF higher-order scale, than with the original *PSYC* scale as it appears on MMPI-2. This issue, of course, is not unique to MMPI-2-RF, as the scales of MMPI-2 also show substantial overlap. There are few scales, however, that overlap to the extent of *PSYC-r* and *THD*. The clinician has a responsibility to be aware of the issue of item overlap and scale redundancy when interpreting the MMPI-2-RF.

One of the primary concerns with regard to the use of MMPI-2-RF concerns the lack of empirical support for its use relative to the amount of empirical support available for use of the MMPI-2 (Graham, 2012). The MMPI-2-RF was introduced in 2008, with a minor update in 2011. The RC scales, which comprise the core of the RF, were introduced in 2003. Given the relatively brief time that the RC scales and form RF have been available in comparison to MMPI /MMPI-2, it is not surprising that they do not have the wealth of empirical literature to support their use that is enjoyed by the earlier versions of the MMPI. A PsycINFO search performed on September 15, 2013, revealed 122 hits for a search of “MMPI-2” and “RC,” 101 hits for a search of “MMPI-2” and “RF,” and 176 hits for a search of “MMPI-2” and “Restructured;” note that there is undoubtedly some overlap among the obtained results. In comparison, a search of “MMPI-2” revealed 2,503 hits. Again, there is undoubtedly some overlap with the prior searches; nevertheless, it is clear that there remains substantially more empirical support for the standard MMPI-2 than for MMPI-2-RF.

There is no doubt that with the passage of time the research base will increase with regard to empirical correlates for MMPI-2-RF scales. At the present time, however, there appears to be limited specificity with regard to empirical content for several of the MMPI-2-RF scales, even within the *Technical Manual* (Tellegen & Ben-Porath, 2011). An example of this phenomenon can be seen in the somatic/cognitive scales in which nearly every scale has correlates of multiple somatic complaints and a preoccupation with physical symptoms. Although the item content of the scales may address different phenomena, the correlates for each of the scales are substantively similar and not especially clinically informative. Thus, these scales can best be interpreted only as a respondent's self-report. A similar phenomenon is observed with regard to the *ACT*

scale, for which no empirical correlates have been reported, as well as *DSF*, *AES*, and *MEC*, for which no empirical correlates have been reported for women. The reader is encouraged to keep this in mind when interpreting these scales until such time as more research is available with regard to empirically supported correlates.

Graham (2012, p. 415), who, it should be noted, is one of the authors of the MMPI-2 RC scales, which enjoy pride of place on the MMPI-2-RF, suggested that interpretations based on the MMPI-2 “can yield a more in-depth analysis of personality and psychopathology” than interpretation based on MMPI-2-RF. He suggested that the MMPI-2-RF is preferable when brevity is a primary concern or when a screening instrument is desired. At the present time, the authors of this text partially concur with that statement; we agree that the MMPI-2-RF is best incorporated as a screening instrument or situations wherein administration of the 567-item MMPI-2 is impractical; however, we note that the MMPI-2-RF contains only 32 fewer items than the MMPI-2 370 form, whereas the latter enjoys the advantage of the extensively researched clinical scales and the literature on clinical interpretation of codetypes.

We are reminded of Alexander Pope’s words from *An Essay on Criticism*, in 1711: “Be not the first by whom the new are tried, nor yet the last to lay the old aside.” For now, however, we would encourage clinicians to administer the full MMPI-2 item pool, as doing so allows for scoring of both the MMPI-2 and MMPI-2-RF profiles. Tellegen and Ben-Porath (2011), as well as van der Heijden, Egger, and Derksen (2010), have reported that MMPI-2-RF scale scores obtained from an MMPI-2 administration are comparable to those obtained with the MMPI-2-RF booklet. By scoring both forms, the clinician who chooses to base their interpretation on the MMPI-2-RF scales has the MMPI-2 profile available if they are faced with interpreting an MMPI-2-RF profile containing elevations only on those scales for which limited empirical support is available. Moreover, the availability of the MMPI-2 scores and profile will often enable the clinician to place noteworthy MMPI-2-RF scores within a broader and potentially more useful context. Given the comparability of MMPI-2-RF scores obtained in this manner, extant databases—those that have given us our rich empirical knowledge base for the MMPI-2—can be mined for additional correlates for MMPI-2-RF scales.

Despite the MMPI-2 designation for both the standard MMPI-2 and MMPI-2-RF versions, the RF form should be considered to be an essentially new instrument, as distinct from a mere revision or updating of the MMPI-2, as was the case in its transition from the original MMPI. To be sure, the RF does have its roots in the MMPI-2 item pool, the 1989 norms gathered for the MMPI-2, similar (and, in at least one case, improved, see *Fp-r*) validity scales, and revisions of the MMPI-2 PSY-5 scales. However, the substitution of a theory-driven methodology for one that was empirically-driven in the construction of the RF’s central set of scales, the RC scales, marks a significant departure from MMPI/MMPI-2 tradition and, in turn, a significant obstacle in applying to the RF form the vast research literature for the MMPI/MMPI-2 that has accumulated over the past 70 years. In short, the MMPI-2-RF is a new psychometric instrument and does not yet have the wealth of empirical support and interpretive data enjoyed by the MMPI-2. With time, we will surely come to develop a complete understanding of the MMPI-2-RF; for now, however, its strengths and weaknesses, and the patterns thereof, largely remain to be clarified in research efforts that the future must await.

**Note**

- 1 Using the Guilford's (1936) formula for a baseline correlation due to shared items,

$$r = \frac{N_s}{\sqrt{A_u + N_s} \cdot \sqrt{B_u + N_s}}, \text{ where } N_s = \text{the number of shared items between scales } A \text{ and } B,$$

$A_u$  = the number of unique items in scale  $A$ , and  $B_u$  = the number of unique items in scale  $B$ . Shared items equals the number of items scored in the same direction minus the number of items scored in the opposite direction. Unique items for each scale equals the number of items on that scale that do not appear on the opposite scale, regardless of scoring direction.