Construct Drift in the MMPI-2 Restructured Clinical Scales: Further Evidence and a Possible Historic Example

P. Kevin Bolinskey¹ and David S. Nichols²

¹Indiana State University
²Pacific University

Construct drift (Nichols, 2006) describes the possibility that the MMPI-2 Reconstructed Clinical (RC) scales are less accurate measures of the primary traits than the original Clinical scales. The authors review the development of RC scales RC4, RC7, and RC9 and provide further evidence that these scales measure traits and behaviors that are not necessarily equivalent to those assessed by the original scales. Further, the development of the original Clinical Scale 7 is reviewed and evidence is provided that suggests that construct drift is not a new phenomenon, but may have been present in the development of the original scales. Suggestions for future research as well as cautions with regard to the clinical interpretation of the RC scales are provided. © 2011 Wiley Periodicals, Inc. J Clin Psychol 67:907–917, 2011.

Keywords: MMPI-2; restructured clinical scales; RC scales

Nichols (2006) invoked the concept of “construct drift” to describe the consequence of one of the procedures involved in the development of the Minnesota Multiphasic Personality Inventory-2 (MMPI-2) Restructured Clinical (RC) Scales (Tellegen, Ben-Porath, McNulty, Arbisi, Graham, & Kaemmer, 2003). Nichols asserted that the procedure (i.e., their Step 4) of correlating “seed” scales, which are purported to embody the “core” constructs of their parent Clinical Scales, with the remaining items of the inventory risked the drift of the selected core construct in the direction of substantive content areas at variance with this core, as off-scale items were added to the seeds to arrive at the final RC scales. This investigation seeks to explore construct drift in relation to three of the RC scales: RC7 (Dysfunctional Negative Emotions), RC4 (Antisocial Behavior), and RC9 (Hypomanic Activation).

In fairness, it is worth noting that if the label, construct drift (Nichols, 2006), is recent, then the phenomenon to which it refers is not. Indeed, within the MMPI itself, the phenomenon goes back to its very beginnings, specifically to the construction of Scale 7, Psychasthenia. Hathaway’s original criterion group comprised only 20 individuals who were “carefully selected as probable psychasthenia cases” (McKinley & Hathaway, 1942, p. 617) based upon such symptoms as obsessive thinking, unreasonable fears, doubts, and compulsions. On the basis of contrasts in the item endorsement rates between the criterion group and each of 339 normal married men and women and 265 college students, a preliminary scale was derived. Regrettably, McKinley and Hathaway failed to identify the items comprising their preliminary scale, or even the number of items it contained.

Whatever this number, it was apparently deemed insufficient to constitute a final psychasthenia scale. It was therefore decided to augment the preliminary scale with items “that had been doubtful in the statistics of the comparison of the criterion and normal groups” (p. 618), according to the following procedure:

Tetrachoric correlations were obtained for every preliminary item and all the doubtful items against the total scores on the preliminary scale for a sample of one hundred normal persons and for a sample of one hundred randomly selected psychiatric patients. These data combined with the original comparison data of criterion and normal cases permitted us to select a final scale of forty-eight items.

Correspondence concerning this article should be addressed to: P. Kevin Bolinskey, B-214 Root Hall, Terre Haute Indiana 47807; e-mail: Kevin.Bolinskey@indstate.edu
Because we do not know how many preliminary items there were, we do not know how many items were added by this procedure, only that the sum of the two sets of items comes to 48. We do know, however, two additional things about Scale 7:

First, McKinley and Hathaway noted that the preliminary scale was “unusually homogeneous” (p. 618). This impression is confirmed in internal consistency values of between .81 and .93 across six samples presented in Dahlstrom, Welsh, and Dahlstrom (1975, p. 260).

Second, it is found to overlap heavily with various measures of the broad, nonspecific general maladjustment or subjective distress dimension commonly referred to as the First Factor in MMPI research. This dimension is understood to be well, if not overly, represented in the MMPI/MMPI-2 item pools and to inflate the correlations among many, if not most, of the scales, thereby compromising the discriminant validity of these scales. The first factor has been given various labels, including Anxiety (Welsh, 1956) and Demoralization (Tellegen et al., 2003). For the purpose of the current manuscript, however, it will be simply referred to as the First Factor or general maladjustment. With regard to item overlap, 27% of the Scale 7 items overlap Welsh’s A (Anxiety; 27%), as do 56% and 79% of first factor items found in the item-level factor analyses of Johnson, Butcher, Null, and Johnson (1984) and Waller (1999), respectively (see Table 1).

These patterns of overlap indicate that Hathaway’s procedure for augmenting his preliminary scale would be at least as likely to recruit items with First Factor variances to his preliminary version of Scale 7, as it would to draw items with the desired specific psychasthenia variances. Remembe that these added “doubtful” items had already failed to pass the contrasted groups criterion. Thus, their addition to the preliminary scale appears to constitute an early case of construct drift away from the core construct embodied in the preliminary scale.

Although there is no way of determining the nature of this core construct with any substantial confidence at this late date, it might be possible to identify an approximation of this core by a process of “backward engineering,” one in which a credible empirically derived marker for the first factor would be appended to Scale 7 and the combination factored, precisely the method followed by Tellegen et al. (2003) in Step 2 of their construction of the RC scales. In this way, the least distinctive items of Scale 7 are drawn away from that scale by the marker, with the most distinctive psychasthenia items remaining on a residual factor.

Note that the argument is not that these residual items constitute Scale 7’s true “core,” though this may be the case. It is rather that these items can be demonstrated to show the greatest resistance to being drawn to the marker and, on those grounds, may fairly be considered the most distinctive and therefore core-eligible. The results of this analysis could then be compared with the 7 seed items Tellegen et al. (2003) selected to embody the core construct of Scale 7, and with RC7.

RC4 and RC9

In response to Nichols’ (2006) criticism of the inclusion of 7 items from the Addiction Admission Scale (AAS; that is, 32% of the RC4 items) risked false-positive inferences of broad antisocial dispositions and behavior based on substance abuse alone (p. 135), the RC authors have been largely silent. This may be because the data from their own samples, for which

<table>
<thead>
<tr>
<th></th>
<th>Seed</th>
<th>Pt</th>
<th>A</th>
<th>JB1</th>
<th>W1</th>
<th>JBW72</th>
<th>AJBW</th>
</tr>
</thead>
<tbody>
<tr>
<td>RC7</td>
<td>7 (29)</td>
<td>8 (33)</td>
<td>10 (42)</td>
<td>16 (67)</td>
<td>19 (79)</td>
<td>14 (58)</td>
<td>8 (33)</td>
</tr>
<tr>
<td>Seed</td>
<td>7 (100)</td>
<td>4 (57)</td>
<td>5 (71)</td>
<td>6 (86)</td>
<td>4 (57)</td>
<td>2 (29)</td>
<td></td>
</tr>
<tr>
<td>Pt</td>
<td>13 (27)</td>
<td>27 (56)</td>
<td>38 (79)</td>
<td>25 (52)</td>
<td>15 (31)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>A</td>
<td>32 (82)</td>
<td>72 (87)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>JB1</td>
<td>72 (87)</td>
<td>72 (87)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>W1</td>
<td></td>
<td>72 (87)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>JBW72</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>29 (40)</td>
<td></td>
</tr>
</tbody>
</table>
suitable external correlates are available, clearly indicate a more substantial association between RC4 and substance abuse than with criminal justice involvement (Tellegen et al., pp. 45–47; 51–52), an association that has been thrice replicated in subsequent research (Arbisi, Sellbom, & Ben-Porath, 2008; Sellbom, Ben-Porath, Baum, Erez, & Gregory, 2008; Sellbom, Ben-Porath, & Stafford, 2007). Although Nichols did not raise the question of construct drift for RC4 at the time, these findings raise the question of a drift from antisocial behavior and toward substance abuse, as five of the seven items from AAS imported into the final version of RC4 do not overlap with Scale 4 (see Table 2).

Nichols (2006) also raised the question of construct drift in RC9 with regard to the amount of hostile, vindictive, and aggressive content present in the final scale. He pointed out a correlation between RC9 and Wiggins’ (1966) Manifest Hostility (HOS) scale, with which it shares five items, of .75 within the Caldwell Clinical Data Set (Caldwell, 1997a); in contrast, HOS and Scale 9 correlate at .51 in the MMPI-2 Normative Sample. Scale 9 contains only one item (50; 2% of the scale items) that overlaps with the PSY-5 Aggressiveness (AGGR) scale. To this, the procedures followed in Step 4 added 6 additional AGGR items, for a total of 7 items, 25% of RC9 scale length. This impression is confirmed in a recent article by Forbey and Ben-Porath (2007), which found RC9 to correlate more highly than Scale 9 with self-reported physical aggression, verbal aggression, anger, and hostility in a sample of 1,238 men in a VA substance abuse treatment program.

In response to Nichols initial criticism, Tellegen et al. (2009) acknowledged that Simms, Casillas, Clark, Watson, and Doebeleing (2005) had “reported that RC9 was more highly correlated with aggression than is Clinical Scale 9” (p. 162), but claim this as a virtue. For example, they noted that “Simms et al. also reported that RC9 was more highly correlated than Scale 9 with disinhibition, impulsivity, and exhibitionism—other manifestations of hypomanic activation” (p. 162). These authors neglected, however, to report that RC9 was also more highly correlated than Scale 9 in Simms et al.’s data with a wide variety of attributes with a presumably more distant relationship to hypomanic activation, including negative temperament, mistrust, manipulativeness, workaholism, and even propriety and dependency. This pattern is repeated in the correlations Simms et al. report for their personality disorder scales. In these data, RC9 is shown to bear a closer relationship than Scale 9 to no fewer than 9 of these 10 scales, including histrionic, avoidant, dependent, and obsessive-compulsive.

With this outline of the issues as we understand them, we can move on to describe three studies that look into the question of construct drift. First, we examined Scale 7 to clarify its relationship to the First Factor of the MMPI-2 and to RC7. Then, we examined scales RC4 and RC9 for evidence of construct drift.

**Method**

**Participants**

Data from 29,983 participants were drawn from 24 research samples collected by Rouse, Greene, Butcher, Nichols, and Williams (2008), reflecting the wide variety of settings in which the MMPI-2 is used. Excluded from these analyses was the Caldwell Clinical Data Set (1997; N> 50,000), about which Tellegen et al. (2006) raised objections. These objections have been addressed elsewhere (Ranson, Nichols, Rouse, & Harrington, 2009). The remaining samples upon which the analyses used in this report are based are described in Table 4. For each sample, the exclusion criteria proposed by Butcher, Graham, and Ben-Porath (1995) were applied, that is, respondents were eliminated from the research samples if they had 30 or more omitted items,

<table>
<thead>
<tr>
<th>Seed</th>
<th>Pd</th>
<th>Pd1</th>
<th>FAM</th>
<th>FAM1</th>
<th>FAM2</th>
<th>Pd2</th>
<th>ASP1</th>
<th>ASP2</th>
<th>DISC</th>
<th>AAS</th>
<th>MAC-R</th>
</tr>
</thead>
<tbody>
<tr>
<td>RC4</td>
<td>5</td>
<td>9</td>
<td>3</td>
<td>4</td>
<td>2</td>
<td>1</td>
<td>5</td>
<td>0</td>
<td>5</td>
<td>8</td>
<td>7</td>
</tr>
<tr>
<td>Seed</td>
<td>5</td>
<td>2</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>3</td>
<td>0</td>
<td>3</td>
<td>3</td>
<td>1</td>
<td>3</td>
</tr>
<tr>
<td>Pd</td>
<td>9</td>
<td>6</td>
<td>3</td>
<td>2</td>
<td>8</td>
<td>0</td>
<td>3</td>
<td>5</td>
<td>2</td>
<td>6</td>
<td></td>
</tr>
</tbody>
</table>
raw scores less than 6 or greater than 12 on TRIN (True Response Inconsistency), raw scores greater than 30 on \( F \) (Infrequency), or T-scores greater than 80 on \( VRIN \) (Variable Response Inconsistency), \( L \) (Lie), \( K \) (Correction), or \( S \) (Superlative) (Tables 3 and 4).

**Measures**

The five markers we incorporate for the first factor have been previously described (c.f., Nichols, 2006). They are, briefly, as follows: (a) Welsh’s \( A \); (b) \( JB1 \), an 83-item scale comprising the replicated first component from the Johnson, Butcher, Null, and Johnson (1984) item-level principal components analysis of MMPI data from 11,138 Missouri psychiatric patients incorporating phi coefficients and Varimax rotation (PCA/V); (c) \( W1 \), the

<table>
<thead>
<tr>
<th>Table 3</th>
<th>Number of Overlapping Items for RC9, RC9 Seeds, and Ma With Other Scales</th>
</tr>
</thead>
<tbody>
<tr>
<td>Seed Ma</td>
<td>Ma2</td>
</tr>
<tr>
<td>RC9</td>
<td>8</td>
</tr>
<tr>
<td>Seed</td>
<td>8</td>
</tr>
<tr>
<td>Ma</td>
<td>11</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Table 4</th>
<th>Description of Samples</th>
</tr>
</thead>
<tbody>
<tr>
<td>Samples</td>
<td>Men</td>
</tr>
<tr>
<td>Clinical</td>
<td></td>
</tr>
<tr>
<td>Alcohol/Drug 1</td>
<td>510</td>
</tr>
<tr>
<td>Alcohol/Drug 2</td>
<td>799</td>
</tr>
<tr>
<td>Psychotherapy</td>
<td>176</td>
</tr>
<tr>
<td>State Hospital</td>
<td>234</td>
</tr>
<tr>
<td>VA Inpatient 1</td>
<td>1,161</td>
</tr>
<tr>
<td>College Students</td>
<td></td>
</tr>
<tr>
<td>Sample 1</td>
<td>57</td>
</tr>
<tr>
<td>Sample 2</td>
<td>510</td>
</tr>
<tr>
<td>Forensic</td>
<td></td>
</tr>
<tr>
<td>Child Custody 1</td>
<td>844</td>
</tr>
<tr>
<td>Child Custody 2</td>
<td>952</td>
</tr>
<tr>
<td>Child Custody 3</td>
<td>534</td>
</tr>
<tr>
<td>Corrections</td>
<td>263</td>
</tr>
<tr>
<td>Criminal Psych.</td>
<td>780</td>
</tr>
<tr>
<td>Death Row</td>
<td>74</td>
</tr>
<tr>
<td>Personal Injury 1</td>
<td>49</td>
</tr>
<tr>
<td>Personal Injury 2</td>
<td>74</td>
</tr>
<tr>
<td>Medical</td>
<td></td>
</tr>
<tr>
<td>Chronic Pain</td>
<td>104</td>
</tr>
<tr>
<td>Neuropsych. 1</td>
<td>97</td>
</tr>
<tr>
<td>Neuropsych. 2</td>
<td>41</td>
</tr>
<tr>
<td>Personnel Screening</td>
<td></td>
</tr>
<tr>
<td>Clergy</td>
<td>234</td>
</tr>
<tr>
<td>General</td>
<td>3,630</td>
</tr>
<tr>
<td>Military, Active Duty</td>
<td>1,354</td>
</tr>
<tr>
<td>MMPI-2 Normative</td>
<td>1,108</td>
</tr>
<tr>
<td>Native American</td>
<td>264</td>
</tr>
</tbody>
</table>
135-item first factor of Waller’s (1999) item factor analysis of 28,390 MMPI protocols from the Hathaway Data Bank of medical and psychiatric patients treated at the University of Minnesota Hospitals between 1940 and 1976, using tetrachoric correlations with binary-item factor analysis; (d) JBW72, a scale comprising the 72 items common to JBI and W1, i.e., twice-replicated first factor items; (e) AJBW, a 29-item scale containing items common to A, JBI, and W1 (see Table 3).

Of these scales, it appears that the use of JBW72 for the current investigation is advantageous in several ways: (a) this scale is based on the original MMPI item pool (i.e., the item pool of the original competition for membership on Scale 7); (b) this scale was derived from a large sample of individuals (N = 39,528); (c) the sample size comprised both psychiatric and medical patients; (d) the participants resided in different states; (e) although different methods of extraction were incorporated for the parent scales, they still show remarkable overlap.

Procedures

Study 1. Following the procedure of Tellegen et al. (2003), the items of JBW72 were appended to Scale 7 and the items overlapping the two scales deleted. This aggregate scale was then subjected to PCA/V to a two-factor solution in the combined samples. This method was then repeated for each of our 24 samples, individually. It is worth noting from the outset that more than one half (58%) of the RC7 items overlap with JBW72, while only one-third overlapped with Scale 7.

Studies two and three. The items of RC4 and RC9, respectively, were subjected to PCA/V to a two-factor solution for the aggregate samples. These analyses were then repeated for each of our 24 samples, individually. The failure of the RC4 and RC9 seed items to load preferentially on the primary, i.e., first, factor, and/or to find these items significantly displaced on this factor by AAS or AGGR items, respectively, would be considered evidence for construct drift.

Results

Study 1

In the analysis of the aggregate sample, 42 of the JBW72 items, 21 of the Scale 7 items, and 13 of 24 RC7 items, including 4 of the 7 seed items, received their highest loadings on the general maladjustment factor. The specific Pt factor showed higher loadings for 30 of the JBW72 items, 27 Scale 7 items, and 5 RC7 items, including three seed items. More than twice as many RC7 items, including a majority of the RC7 seed items, were found on the general maladjustment factor, rather than the specific Pt factor in this analysis.

These findings set the pattern found for the analyses of our individual samples. In one sample, identified in Table 4 as Personal Injury 2, the number of JBW72 items receiving their primary loadings on a given factor was equal; thus, this sample was eliminated from further analysis. The more noteworthy results found in our analyses of the remaining 23 samples are as follows:

- The probability of any Scale 7 item appearing on the specific Pt factor was .51 across the 23 samples.
- The probability of any Scale 7 item on RC7 appearing on the specific Pt factor was .35 across 23 samples.
- The probability of any RC7 item overlapping JBW72 appearing on the marker factor was .70 across 23 samples. Hence, an RC7 item overlapping JBW72 is about twice as likely to appear on the marker factor as an RC7 item overlapping Scale 7 is to appear on the specific Pt factor.
- The probability of any of the original Scale 7 item appearing on the Pt factor, as distinct from the general maladjustment, factor in any of our 23 samples is .51, whereas the probability of any of the RC7 items overlapping with Scale 7 appearing on the Pt factor is
Thus, the odds of any item of Scale 7 appearing on the specific Pt factor, are 46% greater than any RC7 item overlapping Scale 7 to appear on the specific Pt factor.

Alternately, the odds of any given RC7 item overlapping JBW72 and appearing on the general maladjustment factor in the same samples is .70. Thus, an RC7 item overlapping JBW72 is about twice as likely to appear on the general maladjustment factor than is an RC7 item overlapping Scale 7 to appear on the substantive factor in our analyses.

Twenty-five items, slightly more than half of the Scale 7 items, replicated on the specific Pt factor in at least half of our research samples: True: 11, 31, 38, 65, 73, 94, 130, 147, 170, 175, 273, 277, 301, 308, 317, 325, and 329; False: 3, 9, 109, 140, 165, 174, and 321. Of these, only two are found on RC7 (301 and 329), and only one of these is a seed item (301). In contrast, 11 of the RC7 items (127, 251, 274, 302, 320, 328, 390, 421, 430, 442, and 451) received their primary loadings on the general maladjustment factor in at least half of our research samples, including three seed items (302, 320, & 328).

Of the 25 Scale 7 items replicating on the Pt factor in at least half of our research samples, 15 of those (items 11, 31, 38, 65, 73, 94, 147, 170, 175, 273, 277, 301, 109, 140, and 165) replicate in at least two-thirds, and six (items 11, 31, 65, 175, 273, and 9) in at least 80% of our samples. None of the latter items appear on RC7. These six items would seem to be among the reasonable candidates to represent the true core construct for Scale 7.

Study 2

In the analysis of the aggregate sample, five of the seven AAS items (264, 487, 489, 511, and 429F) loaded on the first factor, while only two of the five RC4 seed (202 and 83F) items loaded on the same factor. Again, the pattern found for this analysis is substantially evident in the research samples when these are considered individually.

In 15 of our samples, including both alcohol/drug samples, the criminal psychiatric and the death row samples, accounting for 18,434, or 61% of the total number of participants, the proportion of AAS items exceeded the proportion of RC4 seed items on the primary factor. In contrast, the proportion of RC4 seed items exceeded the proportion of AAS items on the first factor in nine samples, including the corrections sample, accounting for 11,549 or 39% of total participants.

Across all of our samples, the probability of any given AAS item appearing on the first factor was .61, while the probability of any given RC4 seed item appearing on the same factor was .54. Thus, the odds of the first factor being dominated by AAS was about 13% more likely than this factor being dominated by RC4 seed items.

Study 3

The analysis of the aggregate sample found that all seven of the AGGR items loaded on the first factor, while the eight seed items for RC9 were evenly divided between the first and second factors.

In 15 of our samples, accounting for 20,750, or 69% of the total number of participants, the proportion of AGGR items exceeded the proportion of RC9 seed items on the primary factor. By contrast, the proportion of RC9 seed items exceeded the proportion of AGGR items on the first factor in nine samples, accounting for 9,233 or 31% of total participants.

Across all of our samples, the probability of any given AGGR item appearing on the first factor was .75, while the probability of any given RC9 seed item appearing on the same factor was .52. Thus, the odds of the first factor being dominated by AGGR was about half-again more likely than this factor being dominated by RC9 seed items.

Discussion

When Scale 7 is analyzed using the empirically derived JBW72 marker for the First Factor, the indicators of the core construct of the scale differ substantially from those found by Tellegen et al. (2003), who used a marker fashioned to represent their theoretically derived demoralization/unpleasantness construct. Given that the item overlap of RC7 is 75% greater
Psychasthenia (Janet, 1903) was already considerably dated by the time Hathaway the 27 items comprising the specific interest, many of whom would now be said to have obsessive-compulsive disorder. To be sure, that this scale possesses sensitivity for identifying members of his primary target group of Scale (measure of obsessive compulsive symptoms such as the Yale-Brown Obsessive Compulsive investigation, and an analysis of their performance in predicting scores on a better established initial step.

Price, Rasmussen, Mazure, Delgado et al., 1989; Storch et al., 2010) could be an informative both specific Pt factor in at least eighteen of our samples. Among these, items 31 and 273 overlap 11, 31, 65, 175, 273, 277, 9, and 165, none of which are 302, 320, and 328) comprise more than half of the appeared on the specific Pt factor in six or fewer of our samples. Note that four of these (289, can be judged. For example, items 89, 242, 289, 302, 304, 309, 313, 320, 326, 328, and 331 in explaining the finding of Rouse et al. that Welsh’s A and RC7 exceeds that between Scale 7 and RC7. These findings help to explain the relatively minor reductions of .01 to .10 (Caldwell, 2006, p. 194) in the covariation among the RC scales as compared with that among the Clinical Scales, particularly as compared with the magnitude of reductions achievable for the latter scales by more straightforward means (Nichols, 2006). The enhanced concentration of First Factor variance in RC7 would certainly act to attenuate the reduction in covariation among the RC scales as a set, thereby frustrating one of the RC authors’ major stated goals of the RC project. In addition, and assuming that the authors’ construct of “dysfunctional negative emotions” can be reliably distinguished from First Factor variance, the saturation of RC7 with the latter may specifically contribute to the correlation averaging about.41 between RC2 and RC7 across the authors’ seven samples (Tellegen et al., 2003, pp. 36–41) when, according to the Watson and Tellegen (1985) model of mood, this correlation should be close to zero. When combined with the results reported here, the findings of Study 1 also go some distance in explaining the finding of Rouse et al. that Welsh’s A was the best proxy for both RCd and RC7 in their exploration of the redundancy of the RC scales. Nevertheless, given that, overall, the RC7 seed items are not appreciably less saturated with the First Factor than those items added in Step 4, the concept of construct drift appears to apply no more to RC7 than to Scale 7 itself.

Whether these analyses have brought us any closer to determining the core of Scale 7 is unclear. They do, however, provide a rough gradient by which the resistance of any given Scale 7 item to being incorporated into the First Factor, at least when measured by JBW72, can be judged. For example, items 89, 242, 289, 302, 304, 309, 313, 320, 326, 328, and 331 appeared on the specific Pt factor in six or fewer of our samples. Note that four of these (289, 302, 320, and 328) comprise more than half of the RC7 seeds items. On the other hand, items 11, 31, 65, 175, 273, 277, 9, and 165, none of which are RC7 seed items, appeared on the specific Pt factor in at least eighteen of our samples. Among these, items 31 and 273 overlap both Scale 7 and JBW72. Although our analyses have tended to emphasize those Scale 7 items that appear vulnerable to being drawn away to the marker factor, the analysis of our aggregate sample found 10 items (31, 38, 73, 94, 147, 170, 273, 277, 301, and 325) that the specific Pt factor drew away from JBW72. At this point, it may be fair to ask if there is any particular virtue in the quest for determining a core for Scale 7 or its psychasthenia construct. Psychasthenia (Janet, 1903) was already considerably dated by the time Hathaway commenced upon the construction of Scale 7, and there is little or no subsequent evidence that this scale possesses sensitivity for identifying members of his primary target group of interest, many of whom would now be said to have obsessive-compulsive disorder. To be sure, the 27 items comprising the specific Pt factor described above may be worthy of further investigation, and an analysis of their performance in predicting scores on a better established measure of obsessive compulsive symptoms such as the Yale-Brown Obsessive Compulsive Scale (Y-BOCS; Goodman, Price, Rasmussen, Mazure, Fleischmann et al., 1989; Goodman, Price, Rasmussen, Mazure, Delgado et al., 1989; Storch et al., 2010) could be an informative initial step.
The findings from the analysis of RC4 provide definite, though by no means overwhelming, evidence for construct drift. Although a majority of AAS and RC4 seed items appeared on the primary factor across our individual samples and their aggregate, the AAS items were nevertheless more reliably present thereon than were the RC4 seed items.

Whether considered from the standpoint of samples, numbers of participants, or items, it appears that the intended core construct that was selected for RC4, antisocial behavior, as embodied in its seed items, has been at least partially obscured by the substance abuse item content that was introduced to the final version of RC4 at Step 4.

Tellegen et al. (2006) have argued strongly against what they call “heterogeneous additive measures,” i.e., “measures that allow different configurations of observations to add up to the same total score so that a high score on one component can in principle compensate for a low score on a different component” (p. 165), yet it appears that this is precisely what occurred in their construction of RC4. By combining items of heterogeneous content (antisocial conduct [all 5 items from ASP2 Antisocial Behavior], but, oddly, none from ASP1 Antisocial Attitudes, substance abuse [7 of 13 items from AAS], family problems [3 of 25 items from FAM], disconstraint [8 items from DISC]), Tellegen et al. arrived at the kind of scale they explicitly argue against, one in which “a high score on one component,” say, antisocial conduct or substance abuse or disconstraint, “can...compensate for a low score on a different component,” say, antisocial conduct or substance abuse or disconstraint, with the attendant risks of diagnostic misdirection. To be sure, there is known to be very considerable comorbidity between antisocial behavior and substance abuse, but this would seem only to enhance the importance of accurately discriminating between the two conditions wherever possible, hence the ongoing utility of the kinds of relatively homogeneous measures Tellegen et al. recommend, like ASP (Antisocial Practices) and AAS, for this purpose. In short, from the separate standpoints of its own psychometric characteristics, including its drift from its core construct as evidenced in the data presented here, and from the argument Tellegen et al. present against “heterogeneous additive measures,” RC4 appears poorly suited to meet the claims, construct and otherwise, made for it.

If the case for construct drift in RC4 might be considered somewhat equivocal in the data presented here, the evidence for such drift in RC9 is clear and compelling. For a clear majority of our samples, including the aggregate sample, and encompassing more than twothirds of our participants, AGGR items exceeded RC9 seed items on the primary factor in a ratio of about 3:2. These findings strongly confirm Nichols’ (2006) apprehension that RC9 risks the “underprediction of manic symptoms in non-hostile manic conditions and their overprediction in hostile non-manic conditions” (p. 134).

Whether considered from the standpoint of samples, numbers of participants, or items, it appears that the intended core construct that was selected for RC9, hypomanic activation, as embodied in its seed items, has been clearly overtaken by the hostile, vindictive, and intimidating item content that was introduced to the final version of RC9 at Step 4.

Our investigation focused on the comparison of the AGGR and seed items within RC9, but there are three other RC9 items that are thematically related to AGGR content: #212, scored on TPA2, with content that might be viewed as obstructionist; #389, scored on ANG1, reflecting angry volatility; and #542, scored on ANG2, reflecting rage. With the AGGR items, all of these items appeared on the first factor of our aggregate sample. They also appeared two (#212) to three (#s 389 & 542) times as frequently on the first than on the second factor in our samples considered individually, wherein the overall probability of any one of these items appearing on the primary factor was .72. Thus, the influence of angry, hostile content on RC9 scores is not attenuated, and may even be increased, by the participation of these items. As in the case of RC4, RC9 appears to be another “heterogeneous additive measure” of the sort Tellegen et al. argue against; further, in this case, the core construct is not merely obscured but overshadowed by hostile affect to an extent that its interpretation in terms of the construct it purportedly represents may be seriously misleading.

The literature on the RC scales is at this point only about 7 years old, and much additional research on these scales, and on a newly released form of the MMPI, the MMPI-2-RF (Ben-Porath & Tellegen, 2008), which is centered on these scales and others that mostly
followed the methodology used in the development of the RC scales, may be anticipated. It is noteworthy that the bidimensionality of both RC4 and RC9 has been implicitly acknowledged in the MMPI-2-RF wherein externalizing scales (Juvenile Conduct Problems (JCP) and Substance Abuse (SUB) essentially subdivide the major content components of RC4, as do the Aggression (AGG) and Activation (ACT) scales for the major content components of RC9.

The findings reported here relative to the saturation of RC7 with first factor variance, the heterogeneity of RC4, and the hostile drift in RC9 all raise questions about the developmental strategy and the procedures used in the construction of the RC, and now, MMPI-2-RF scales. These findings add to a growing body of literature (Butcher, Hamilton, Rouse, & Cumella, 2006; Caldwell, 2006; Gordon, 2006; Greene, Rouse, Butcher, Nichols, & Williams, 2009; Nichols, 2006; Ranson et al., 2009; Rogers, Sewell, Harrison, & Jordan, 2006; Rouse, Greene, Butcher, Nichols, & Williams, 2008) critical of the methodology and results of the RC project. Specifically, both argument and data have been presented that raise doubts that the assumptions and procedures followed in constructing the RC scales met either of the RC authors’ stated goals: “The RC Scales were designed to preserve the important descriptive properties of the existing Clinical Scales while enhancing their distinctiveness” (Tellegen et al., 2003, p. 1). For RC7, our results indicate that the RC methodology led to a scale that is actually substantially less distinctive than its Scale 7 parent in that the former is even more saturated with nonspecific or First Factor variance. For RC4, the goal of enhanced distinctiveness for antisocial behavior has been compromised, indeed somewhat overshadowed, by new substance abuse content that their methodology imported. For RC9, the distinctive hypomanic activation variance that Tellegen et al. wished to concentrate has been significantly displaced by the hostile item content their procedures added. Moreover, in both RC4 and RC9, the heterogeneity that these authors argue against remains; it has only changed from the character dictated by Hathaway’s focus on clinical syndromes, and denigrated by some as the result of “blind empiricism,” to one occasioned by accidents of procedure, perhaps a blind empiricism in alternate guise.

It will therefore be essential as research on the RC and MMPI-2-RF scales progresses, that investigators expand the domain of scales recruited as comparisons in the evaluation of performance against external criteria. Where the focus of interest is on one or more of the RC scales, investigators would be well advised to include the corresponding seed scales among the scales selected for comparison. The inclusion of both the full RC scales and their associated seeds provides another valuable strategy in studies seeking to demonstrate the construct validity of the full scales, for clarifying the definition and range of the constructs in question, and for assessing construct drift, or the lack thereof. It is a matter of some concern that among the more than six dozen publications on the RC scales to date, not a single one has compared the performance of an RC scale with its own seed in predicting any external criterion. In a review of the construct validity of the RC scales (Tellegen et al., 2009), the RC seed scales go unmentioned.

As one example of such an expanded net of comparison scales, when external criteria for antisocial behavior are available, the performance of RC4 can be informatively compared with that of many other MMPI/MMPI-2 measures, including RC4 seed items, Scale 4, Pd1 (Family Discord), Pd2 (Authority Conflict), Pd3 (Social Imperturbability), Pd4 (Social Alienation), Pd5 (Self-Alienation), Ma1 (Amorality), Ma2 (Psychomotor Acceleration), Ma3 (Imperturbability), Ma4 (Ego Inflation), AAS (Addiction Admission), MAC-R (MacAndrew Alcoholism), Re (Social Responsibility), ASP, ASP1 (Antisocial Attitudes), ASP2, and DISC (Disconstraint). Similarly, in evaluating the performance of RC9 in relation to a diagnosis of mania, comparison scales should include RC9 seed items, Scale 9, Ma1, Ma2, Ma3, Ma4, MAC-R, SOC (Social Discomfort), INTR (Introversion/Low Positive Emotionality), RC2 (Low Positive Emotions), HYP (Hypomania; Wiggins, 1966). Casting a wider net for comparing the performance of an RC, or any other MMPI/MMPI-2 scale, need not be seen only in the context of establishing incremental validity, important as that may be. Expanded comparisons serve multiple purposes. Including older, often better established, scales measuring similar or related constructs stands to illuminate both old and new, and to not only speed the growth of knowledge, but also refine it along the way.
References


McKenna, T., & Butcher, J. (April, 1987). Continuity of the MMPI with alcoholics. Paper given at the 22nd Annual Symposium on Recent Developments in the Use of the MMPI, Seattle, WA.


