JOURNAL OF PERSONALITY ASSESSMENT, *87*(2), 121–138

The Trials of Separating Bath Water From Baby: A Review and Critique of the MMPI–2 Restructured Clinical Scales

David S. Nichols
School of Professional Psychology
Pacific University

I review the development of the new set of MMPI–2 scales, the Restructured Clinical (RC) Scales (Tellegen et al., 2003). I identify several conceptual and methodological flaws in the construction of these scales, and I discuss the influence of a central shortcoming, the use of an atypical and depressively biased marker for unwanted ("first-factor") variance and its consequences for the RC Scales. I criticize the monograph introducing and describing the development of these scales for multiple important omissions. I provide examples of RC Scales in which relevant variances were overextracted or underextracted in the process of their construction. I introduce and apply the concept of "construct drift" to the RC Scales corresponding to the MMPI–2 (Butcher, Dahlstrom, Graham, Tellegen, & Kaemmer, 1989) Clinical Scales, Psychasthenia and Hypomania. I conclude that the RC Scales are highly redundant with and function as routinely scored Content scales and that their designation as Clinical Scales is at best tenuous. I present two more appropriate, unbiased markers for the first factor along with the description of several nonintrusive strategies for increasing the independence of the Clinical Scales without compromising their syndromal fidelity.

The MMPI–2 Restructured Clinical (RC; Tellegen et al., 2003) Scales were devised, at least in part, to correct a long-standing problem with the basic Minnesota Multiphasic Personality Inventory (MMPI; Hathaway& McKinley, 1940) and MMPI–2 (Butcher, Dahlstrom, Graham, Tellegen, & Kaemmer, 1989) Clinical Scales: their extensive covariation. In some samples, a pair of scales such as Scale 7, Psychasthenia, a putative "neurotic" scale, and 8, Schizophrenia, a putative "psychotic" scale, might have 75% or more of their variance in common. Although not entirely surprising considering that less than 20% of Scale 8 items describe psychotic phenomena, and a similar proportion of its items overlap with Scale 7, the extent of their covariation routinely confounds the interpretation of these two scales.

The method of contrasting normal with psychiatric criterion groups that Hathaway and McKinley (1940) employed to develop the basic MMPI Clinical Scales (i.e., Scales 1—Hs/Hypochondriasis, 2—D/Depression, 3—Hy/Hysteria, 4—Pd/Psychopathic deviate, 6—Pa/Paranoia, 7—Pt/Psychasthenia, 8—Sc/Schizophrenia, and 9—Ma/Hypomania) did not prohibit items from overlapping on scales measuring different diagnostic constructs. Although such overlap might be desirable in that the constructs measured may, by their nature, contain overlapping elements (e.g., distress, worry, impaired cognition), it exacts a price in reduced discriminant validity. The overlapping items are disproportionately represented among those with the highest face validity (i.e., "obvious" items) that refer to symptoms and personality characteristics generally recognized as problematic or "abnormal." As a consequence, these items possess a high sensitivity to psychopathology when it is present, that is, they function well to detect that "something is wrong," whereas functioning relatively poorly to indicate what that specific "something" is. Symptoms such as anxiety, dysphoria, self-consciousness, difficulty concentrating, and mild social alienation occur in differing strengths and combinations in numerous clinical syndromes and thus are unique to none. The presence of such symptoms in test findings will therefore not, as a rule, assist the clinician in discriminating, for example, an anxiety disorder from a mood disorder.

The RC monograph (Tellegen et al., 2003; a supplement to the MMPI–2 *Manual for Administration and Scoring* [Butcher et al., 2001]) describes the development, validation, and interpretation of the RC Scales. It reports Tellegen et al.'s proposed solution to the covariation problem and provides a description of the procedures followed in constructing a set of

RC scales. Because this monograph (hereafter, *Manual*) remains the major source of published research on these new scales, I focus much of the discussion to follow on it.

It is not entirely clear how the aims of the RC project evolved from the admirable goal of solving the clinical scale covariation problem to creating an entirely new set of scales. It is important to note, however, that a number of other methods (that I describe following) are already at hand for modifying the existing Clinical Scales or adjusting their scores that at once substantially reduce their covariation while preserving their multivariate structure. These methods could improve the discriminant performance of the Clinical Scales while maintaining their syndromal fidelity. Instead, the RC project took a direction that went well beyond the reduction of covariation among the Clinical Scales to fashioning a set of scales that aimed to "preserve the important descriptive properties of the existing MMPI–2 Clinical Scales while enhancing their distinctiveness" (Tellegen et al., 2003, p. 1). In this review, I describe the means by which these aims were pursued and then evaluate the methods and results presented in the RC *Manual*.

THE NATURE OF THE CLINICAL SCALES

To place the RC project in context, it is necessary to appreciate the nature of the MMPI/MMPI–2 Clinical Scales and the features that set these scales apart from scales devised using alternate procedures such as the Harris and Lingoes subscales, the Content and Content Component scales, and the Personality Psychology Five (PSY–5; Harkness, McNulty, Ben-Porath, & Graham, 2002) scales (Butcher et al., 2001; hereafter, simply Content scales unless indicated more specifically).

The essential feature of most of the basic Clinical Scales is their syndromal character. As diagnostic constructs, these scales are not merely personality or trait scales as commonly conceived but represent or model the clinical syndromes they were developed to measure by incorporating the diverse aspects of emotion, thinking, and behavior that in combination define and constitute them. The construct of clinically significant depression, for example, consists not merely of dysphoric mood but also of cognitive and vegetative signs and symptoms such as slowed or effortful thinking; impaired concentration and memory; and difficulties with sleeping, eating, elimination, and weight gain/loss. Not surprisingly, given its method of construction, Scale 2 contains items from all of these different content domains, and again, as such, it models or maps onto the clinical syndrome it purports to measure.

As multivariate models of their respective clinical syndromes, the Clinical Scales may be seen to function in ways that are analogous to sets of diagnostic criteria such as those published in the *Diagnostic and Statistical Manual of Mental Disorders* (4th ed. [*DSM–IV*]; American Psychiatric Association, 1994) wherein patients who meet the criteria for the same diagnostic category (or the same Clinical Scale score) are not required to do so in the same way. In both cases, differences in symptoms, traits, and history are recognized and codified into a polythetic system of classification that strives for a reasonable balance between the goals of adequate intragroup homogeneity to support diagnostic reliability and adequate diversity of coverage such that patients deemed appropriate for classification within a given category can be so classified, despite their differences.

Like the *DSM–IV* criteria lists, the MMPI–2 Clinical Scales allow the diverse elements of mood, cognition, and behavior to combine with each other in ways that reflect, more or less uniquely, the syndromes that they characterize, even as they allow for such aspects to manifest themselves in differing forms and strengths within syndromes and for similar aspects to be recognized across syndromes. In the case of the MMPI–2, of course, this latter feature is manifested in shared variances and overlapping items. Thus, to take only one example, packets of item content reflecting dysphoria are found in Scales 2 (D1/Subjective Depression, D5/Brooding), 3 (Hy3/Lassitude-Malaise), 4 (Pd5/Self-Alienation), 6 (Pa2/Poignancy), and 8 (Sc2/Emotional Alienation; Sc4/Lack of Ego Mastery, Conative).

Content scales, by contrast, are devised in ways that emphasize face validity and internal consistency. Such scales are especially well suited to measuring important thematic dimensions of MMPI–2 performance. By enabling the aggregation and summary of the examinee's responses to specific content domains, Content scales preserve his or her responses to areas of item content that might otherwise be overlooked as influences within the scores of the complex Clinical Scales.

It has been known for over a half of a century that the diagnostic efficiency of the Clinical Scales is far from optimal, owing in large part to their limited discriminant validity. The quest for empirical correlates, the proliferation of subscales and content-based scales, and the shift from single-scale to code type interpretation may all be seen as attempts to compensate for the deficiencies of the Clinical Scales.

MAJOR SOURCES OF CLINICAL SCALE COVARIATION

The covariation among the Clinical Scales may be traced to two major, nonmutually exclusive sources: construct overlap and item overlap. Construct overlap occurs in the example of Scales 7 (Psychasthenia) and 8 (Schizophrenia) I gave previously. Despite their differences, both scales and the syndromes they represent share elements of impaired memory and concentration, heightened fear, agitation, and apathetic dysphoria. Similarly, the Scales 1 (Hypochondriasis) and 3 (Hysteria) constructs share a focus on somatic symptoms/complaints without which either would lose a defining characteristic.

The justification for these overlapping constructs has been given by Dahlstrom (1969) who emphasized the diagnostic yield realized by exploiting the configural relationship of Scales 1 and 3. Goldberg (1965) provided data supporting the value of the configural relationship of Scales 7 and 8 in discriminating between neurotic and psychotic MMPI profiles and showed that a simple index based on the difference between these two scales yielded a rate of correct classification for these profile groups that exceeded that of groups of experienced and trainee psychologists.

The method of contrasted groups all but assured that the overlap among diagnostic constructs would result in overlapping items. For example, Scales 1 and 3 share 20 items, and Scales 7 and 8 share 17. Although it may be argued that patterns of item overlap such as these contribute to the integrity of their respective constructs, it can be readily demonstrated that such overlap exacts a price in the independence of the Clinical Scales and hence their discriminant validity.

Just how extensive is the item overlap among the Clinical Scales? Considering only Scales 1 through 4 and 6 through 9, there are a total of 259 items that are scored on one or more of these eight scales, of which 101 (39%) overlap at least one other scale. Of these 101 items, 66 items overlap only one other scale, 29 items are scored on three scales, 4 items are scored on four scales, and 2 items are scored on five scales. If one counts the actual number of overlaps between any two of the Clinical Scales, as distinct from overlapping items, the total is 197.[1] In the Caldwell (1997) clinical data set of 26,118 male and 26,425 female patients,[2] the average correlation among the clinical scales was .59. The effect of deleting the 101 overlapping items was dramatic. The average correlation amongst the scales thus modified dropped to .29, an increase in their independence of more than 26% ($.59^2 - .29^2 = .26$). Unfortunately, this modification exacts a substantial price in the fidelity of the modified scales to their unaltered parent scale constructs, as the average correlation of each with its parent scale fell to .74.

The history of the MMPI/MMPI–2 includes a number of efforts to strengthen the Clinical Scales by reducing their covariation due to item overlap to increase their discriminative efficiency (Adams & Horn, 1965; Welsh, 1952; see also Welsh, 1956) or by normative adjustments that would favor the development of scales with improved discriminant properties (e.g., Rosen, 1962). The RC project falls within this tradition.

Another way to look at the problems of construct and item overlap among the MMPI–2 scales is from the factor analytic perspective of shared variance. The first factor has been repeatedly identified in scale-level and item-level factor analytic investigations of the test and designates the first and major source of variation among its scales and items. This factor is typically identified as a general maladjustment or subjective distress dimension and is marked by a variety of item content including anxiety, tension, depression, and worry; reduced self-confidence/self-esteem; submissiveness or yielding in the face of obstacles; oversensitivity and irritability; and problems in concentration, memory, and initiative. Also, just as with construct and item overlap, the influence of the first factor is such as to compromise scale independence and therefore discriminant validity. So pervasive is this dimension that it has been shown to be present, in greater or lesser degree, in the vast majority of MMPI/MMPI–2 scales including the basic Clinical Scales. In clinical samples, for example, a common marker for this factor, Welsh's (1956) A, routinely achieves correlations of .85 or higher with both Scales 7 and 8, suggesting that there is considerable nonspecific variance in both scales. With these considerations as background, I now turn to an analysis of the project leading to the RC Scales.

EARLIER CRITICISMS AND ALTERNATIVE APPROACHES

The first two chapters of the RC *Manual* present an overview and rationale for the development of the RC Scales. The second chapter includes a useful review of some earlier criticisms of the Clinical Scales and their empirical keying and two alternative approaches to scale development: the exploratory factor analytic and Jackson's (1970) sequential construct-oriented method. This chapter provides some helpful perspectives on the RC project, but there is no discussion of important prior critiques and alternative approaches that would have usefully enlarged the nonexpert reader's grasp of the relevant issues. For example, there is no reference to the most recent and probably best known of the MMPI/MMPI–2 critiques (Helmes & Reddon, 1993) nor to the earlier and long-neglected analysis by Norman (1972). Nor is there any discussion of alternative methods that have been proposed and have demonstrated to increase the independence of the Clinical Scales (Adams & Horn, 1965; Finney, 1968; Jackson & Reddon, 1987; Welsh, 1952, 1956). In Finney's (1968) procedure, to take one, the usual T scores of the Clinical Scales are recalculated using a formula that adjusts them to remove the nonspecific covariance of the first factor as determined by the individual's score on a marker, say Welsh's (1956) A, for this source of covariance. The chief virtue of this method is that it leaves the scales themselves, their item composition and keying, and therefore their syndromal complexity intact. One of the automated MMPI/MMPI–2 scoring and interpretation services, Behaviordata (formerly Behaviordyne) has used such a procedure for decades.

[1] An item scored on two scales equals 1 overlap (Scales VW); an item scored on three scales equals 3 overlaps (Scales VW, VX, WX); on four scales, 6 overlaps (Scales VW, VX, VY, WX, WY, XY); and on five scales, 10 overlaps (Scales VW, VX, VY, VZ, WX, WY, WZ, XY, XZ, YZ).

[2] These data were generously supplied to D. S. Nichols and R. L. Greene for unrestricted research use by A. B. Caldwell.

STEP 1: CAPTURING DEMORALIZATION

The construction of the RC Scales proceeded in four steps, of which the first was to identify and embody a general dimension, Demoralization, which Tellegen et al. (2003) considered a relabled version of A (p. 12). These steps can only be outlined here; the reader is referred to chapter 3 of the *Manual* for more adequate detail. Although Tellegen et al. described their methods as "pervasively empirical" (p. 11), the approach taken to the construction of their marker for the first factor relied heavily on theoretical considerations. Specifically, Tellegen et al. conjectured that the "broad affectively colored dimension represented to some degree in each of the Clinical Scales" (p. 1), Demoralization, corresponds to the pleasantness-unpleasantness (PU) dimension in Watson and Tellegen's (1985) model of affect.

In accordance with this conjecture, Tellegen et al. combined the items from Scales 2 (Depression) and 7 (Psychasthenia), the scales they judged to be most saturated with the PU dimension. Tellegen et al. then factored these items using principal components analysis with varimax rotation (PCA/V) twice: once to identify items with high (>|.49|) loadings on the first factor in each of their data sets and again to identify items achieving high loadings on two other factors identified in the same data sets—Positive Emotionality (PEM) and Negative Emotionality (NEM). Ten items survived in both analyses. Items drawn from the remainder of the MMPI–2 item pool (less the items on Scales 2 and 7) were added to these 10 items on the basis of their correlations with the PEM and NEM measures, which yielded a preliminary Demoralization (Dem) scale of 23 items, 14 of which overlap with A and 9 with the Depression (DEP) Content scale (Butcher, Graham, Williams, & Ben-Porath, 1990).

It would have been useful to have more details in the description of Step 1 such as how the defining PEM-like and NEM-like factors were identified and their composition. In addition, the origins of the full set of 23 Dem items and a scoring key for this preliminary scale is not provided. The *Manual* also does not indicate how five items (31T, 180T, 400T, 505T, 554T) came to be on Dem's 24-item successor scale, RCd (Demoralization), and it is likewise unclear as to how and at what point a final 24th item was added. Furthermore, on page 21 of the *Manual*, four items (299, 364, 394, 509) are reported as having been dropped from the final 24-item Demoralization scale (RCd), although previous discussion (p. 14) did not identify these items as being on either Dem or RCd in the first place. These omissions leave the fundamental definition of the Dem and RCd constructs unclear.

More broadly, the appropriateness and advantages of the decision to embrace a theoretically rather than an empirically driven strategy for constructing Dem and RCd are doubtful. There is, of course, no reason to question a scale constructed to measure a theoretically derived construct. However, the *Manual* provides no discussion of the rationale for why the construct Dem should be preferred as a means for identifying and extracting the problematic covariance from the basic Clinical Scales as described in the next stages of the RC project (Steps 2, 3, and 4) over other readily available, empirically derived first-factor markers, which I describe following.

In terms of Tellegen et al.'s (2003) goal of restructuring the Clinical Scales so as to identify and measure "the distinctive substantive core" (p. 15) of each, the selection of a less than optimal marker for the first factor would appear to court twin risks. To the degree that a first factor marker is biased, it may inadvertently extract those elements that comprise a scale's distinctive core. Alternatively, a biased or poorly focused marker may only partially, inadequately, extract the scale elements that are most responsible for its unwanted covariance with other scales. In either case, the consequence would be a restructured scale of compromised distinctiveness, that is, a scale that fails to adequately represent the scale's true core dimension(s) either because the residual dimensions following the extraction of unwanted variance are impoverished with respect to the scale's core or because the identified core continues to retain unwanted variance. A third risk, in either case, is that the identified core will contain entirely unanticipated variances that become incorporated into the selected core constructs at Steps 3 and 4, a phenomenon I refer to as "construct drift." I illustrate each of these consequences by examples to follow.

STEPS 2 TO 4: RESTRUCTURING THE CLINICAL SCALES

In Step 2, Tellegen et al. attempted to remove the first-factor covariance from each of the clinical scales. First, the 23 Dem items were appended to each[3] of the Clinical Scales in turn, and the combined item set for each scale was factored and rotated to yield from two to five factors, one of which was chosen to be defined by the Dem items. Those items from each scale gravitating toward the Dem factor were then eliminated from the scale. Tellegen et al. then selected from their exploratory factor solutions a dimension judged "the distinctive substantive core" (p. 15) for each scale.

Next, in Step 3, from the set of 321 items scored on 1 or more of the 10 clinical scales,[4,5] 158 items were selected as candidates for membership in a "seed scale" (Tellegen et al.,

[3]In the case of Scale 2, the eight correction items, those that Hathaway and McKinley (1942) found to discriminate nondepressed patients who obtained high scores on a preliminary version of Scale 2 from their criterion depressives, were omitted from this analysis as "not … relevant to our concerns" (Tellegen et al., 2003, p. 15).

[4]This does not include the correction items unique to Scale 2.

[5]Although all 10 of the standard criterion referenced scales were included in this analysis, Scales 5, Mf/Masculinity–Femininity and 0, Si/Social Introversion, were excluded from the focus of the *Manual* (Tellegen et al., 2003). Restructured versions of these scales were contemplated for the future, however (Tellegen et al., 2003, p. 15).

2003, p. 17), those items thought to comprise the distinctive core component of each Clinical Scale by meeting two conditions: (a) correlating more highly with their parent Clinical Scale than with any other and (b) failing to achieve "salient" loadings on the Dem marker in Step 2. These seed scales were then refined by various means to reduce overlap and increase internal consistency. From the remaining items, a second set of seed scales was derived from which items were eliminated or added to increase the distinctiveness of the core component of each scale. The 73 items surviving these measures were then sorted into a final set of seed scales for RC1 through RC9 and an RCd core. Table 1 presents the RC Scales and their corresponding Clinical Scales.

In Step 4, the seed scales were augmented by items drawn from the entire MMPI–2 item pool. Roughly, an item was added to a seed scale if it correlated above a minimum value with that seed (the "convergence criterion"; Tellegen et al., 2003, p. 19) and below a maximum value with all of the remaining seed scales (the "discrimination criterion"; Tellegen et al., 2003, p. 19); these values were allowed to differ for each scale. A final series of ad hoc adjustments were then made to these augmented seed scales (including, in some cases, the relaxation of these criteria) to "optimize" scale content, increase internal consistency, or to increase desired relationships with selected external criteria. These procedures culminated in the final RC Scales.

Table 2 enables a comparison between the Clinical, RC, and seed scales in terms of length and the extent to which items from the Clinical Scales persist in their RC versions and overlap with MMPI–2 first factor or content-based scales. It is noteworthy that less than one half (48%) of the items on RC1 through RC9 originate from the Clinical Scales. The 36 items that are scored on a single Clinical Scale (averaging 4.5 items per RC Scale) are outnumbered by the 45 items that overlap at least one other Clinical Scale (averaging 5.6 items per RC Scale). Hence, it is apparent that the RC Scales did not eliminate the problem of item overlap that afflicts the Clinical Scales.

TABLE 1
MMPI–2 RC Scales and Corresponding Clinical Scales

RC Scale	*Clinical Scale*
RCd—Demoralization	
RC1—Somatic Complaints	Scale 1—Hypochondriasis
RC2—Low Positive Emotions	Scale 2—Depression
RC3—Cynicism	Scale 3—Hysteria
RC4—Antisocial Behavior	Scale 4—Psychopathic Deviate
RC6—Ideas of Persecution	Scale 6—Paranoia
RC7—Dysfunctional Negative Emotions	Scale 7—Psychasthenia
RC8—Aberrant Experiences	Scale 8—Schizophrenia
RC9—Hypomanic Activation	Scale 9—Hypomania

Note. MMPI = Minnesota Multiphasic Personality Inventory–2; RC = Restructured Clinical.

Although the role of item overlap as a contributor to high covariation among the Clinical Scales is downplayed in the *Manual* (Tellegen et al., 2003, p. 6), the dramatic increase in the independence of the Clinical Scales following the removal of their 101 overlapping items, as I illustrated previously, suggests that this source of Clinical Scale covariance is hardly inconsequential. The threat to the independence of the RC Scales posed by overlapping items was at least implicitly recognized by Tellegen et al. in their prohibition of item overlap among the final versions of RC1 through RC9. However, there was apparently no similar prohibition of overlapping Clinical Scale items within the RC Scales despite the obvious conflict such items pose vis-à-vis the major goal of producing a set of scales to represent each Clinical Scale's "distinctive substantive core" (Tellegen et al., 2003, p. 15). Even among the items on the seed scales, those RC Scale precursors that were intended to embody the most definitive of each Clinical Scale's distinctive core, about 20% of the items are keyed for two or more Clinical Scales. More than one third of these overlapping items occur on RC1. Overlap in this instance may have been unavoidable given the limited supply of somatic content in the item pool. Even with this exception held aside, however, at least one half of the items that derive from their parent Clinical Scales on each of RC2, RC3, RC6, and RC8 overlap with one or more other Clinical Scales.

In most respects, chapter 3 is written clearly enough to allow the reader to follow Tellegen et al.'s thinking and methods. However, despite controversy regarding the suitability of PCA/V for the MMPI/MMPI–2 (Waller, 1999; Waller, Tellegen, McDonald, & Lykken, 1996; more generally, see Gorsuch, 2003; Reise, Waller, & Comrey, 2000), no rationale is provided for the selection of PCA/V as the primary method for all of the RC factor analyses nor for their choice of orthogonal rather than oblique rotation. Indeed, this choice is less discussed than announced and is disposed of in a single sentence (Tellegen et al., 2003, p. 14). More puzzling, the *Manual* contains no explanation for why the preliminary Dem scale was used in the factor analyses of the Clinical Scales in Step 4 rather than its final and presumably improved version, RCd.

The details of the ad hoc analyses and decisions in Step 4 are generally clear and well rationalized, but here too, there is incomplete information regarding some important issues. For example, Tellegen et al. (2003) reported that "We examined the correlations of the items in RC1, RC2, RC4, RC6, RC7, and RC8 with appropriate external criterion measures that were available for these six scales" (p. 21) without further comment on the source or nature of the criterion measures used. Absent such comment, one assumes these measures were among those used later (Tellegen et al., 2003, chap. 4) for the validation of the RC Scales, raising the question of sample-optimized validity coefficients, at least for those scales altered by this procedure (RC2, RC6, and RC8).

In one respect, Step 4 partially cancels out the effect of Step 2. That is, the decision to go outside the clinical scales

TABLE 2
Item Composition and Overlap for the Clinical, Seed, RC, and Selected First Factor and Content Based Scales

		Item Overlap and % of the Seed and Full RC Scales With the MMPI–2 Clinical and Selected First-Factor and Content-Based Scales								
	Length	*Seed Items*		*Clinical Scale Items (Unique/Overlapping)*		*Off-Scale Items*		*Items From Factor and Content-Based Scales*		
Scale	*Clinical/RC*	*No.*	*%*	*No.*	*%*	*No.*	*%*	*Scale*	*No.*	*%*
Scale 1/RC1	32/27	15	56	20 (3/17)	74	7	26	HEA	20	74
Scale 2/RC2	57/17	4	24	8 (4/4)	47	9	53	INTR	9	53
								DEP	2	12
Scale 3/RC3	60/15	5	33	5 (2/3)	33	10	67	CYN	12	80
								HEA	0	0
Scale 4/RC4	50/22	5	23	9 (5/4)	41	13	59	DISC	8	36
								ASP	6	31
								AAS	7	32
Scale 6/RC6	40/17	6	37	13 (6/7)	76	4	24	BIZ	10	59
								PSYC	10	59
Scale 7/RC7	48/24	7	29	8 (6/2)	33	16	67	JBW	72	14
								A	10	42
								JB1	13	54
								W1	20	83
								ANG	4	17
								OBS	3	13
								ANX	2	8
Scale 8/RC8	78/18	6	33	10 (4/6)	56	8	44	BIZ	12	67
								PSYC	8	44
Scale 9/RC9	46/28	8	29	8 (6/2)	29	20	71	AGGR	7	25
								ANG	4	14
								TPA	4	14
M	51/21	7	33	10 (4.5/5.6)	48	11	52			
RCd	—/24	17	71	13	54	11	46	DEP	11	46
								NEGE	1	4

Note. RC = Reconstructed Clinical; Scale 1 = Hs—Hypochondriasis; RC1 = Somatic Complaints; HEA = Health Concerns; Scale 2 = D—Depression; RC2 = Low Positive Emotions; INTR = Introversion/Low Positive Emotionality; DEP = Depression; Scale 3 = Hy—Hysteria; RC3 = Cynicism; CYN = Cynicism; Scale 4 = Pd—Psychopathic Deviate; RC4 = Antisocial Behavior; DISC = Disconstraint; ASP = Antisocial Practices; AAS = Addiction Admission Scale; Scale 6 = Pa—Paranoia; RC6 = Ideas of Persecution; BIZ = Bizarre Mentation; PSYC = Psychoticism; Scale 7 = Pt—Psychasthenia; RC7 = Dysfunctional Negative Emotions; JBW72 = items overlapping both the Johnson–Butcher (Johnson et al., 1984) and Waller (1999) first factors; JB1 = Johnson–Butcher first factor; W1 = Waller (1999) first factor; ANG = Anger; OBS = Obsessiveness; ANX = Anxiety; Scale 8 = Sc—Schizophrenia; RC8 = Aberrant Experiences; Scale 9 = Ma—Hypomania; RC9 = Hypomanic Activation; AGGR = Aggressiveness; TPA = Type A; RCd = Final Demoralization; NEGE = Negative Emotionality/Neuroticism.

to embody the "distinctive substantive core" (Tellegen et al., 2003, p. 15) of each undermined the independence from the first-factor variance of the seed scales achieved in Steps 2 and 3. In the process of augmenting the seed scales, some of the first-factor variance that was removed from the Clinical Scales in Step 2 was reintroduced to the seeds in Step 4. This effect may be seen in Table 3, which allows the comparison between the seed and nonseed portions of each RC Scale in terms of their respective correlations with the seven first-factor markers. With one exception (RC1), the nonseed correlations with the first-factor markers average about 10% higher than for their seed counterparts. This effect is particularly strong for RC2, RC4, and RC9. As a consequence of Step 4, the RC Scales contain about 90% of the first-factor variance as do the unaltered Clinical Scales. That is, only about 10% of the first-factor covariance was removed from the RC Scales.

INTERNAL VALIDATION AND MISSING ANALYSES

Chapter 4 presents a description of the samples used to develop and validate the RC Scales: the MMPI–2 restandardization sample (Butcher et al., 2001), the Portage Path outpatients (Graham, Ben-Porath & McNulty, 1999), and two inpatient samples (Arbisi, Ben-Porath, & McNulty, 2003). It also provides basic psychometric data for the final RC Scales and a set of tables displaying their internal and external correlates. It is this chapter that is potentially the most problematic of the *Manual* as to the character of the RC Scales.

The procedures I described in Steps 2, 3, and 4 previously—including the selection of which constructs were to be represented in the RC Scales, the seed items that defined them, and the augmentation of these seeds with other items in the pool, always with an eye to increasing internal con-

sistency—all ensured scales that would resemble scales developed by rational/statistical rather than empirical means. Not surprisingly, then, the internal analyses described in Tables 4–4 to 4–12 (Tellegen et al., 2003, pp. 35–42) show that (a) most of the RC Scales correlate with their parent Clinical Scales; (b) these correlations are more or less distinctive in the sense that the RC versions are usually correlated more highly with their parent scales than with the other Clinical Scales; and (c) the correlations among the RC Scales are, on the whole, lower than the correlations among the Clinical Scales. This result is due, at least in part, to the removal of the Dem variance from and the avoidance of overlapping items across the RC Scales. The results of these analyses show startling exceptions at times. For example, the correlation between RC3 and Scale 3 reported for the normal men is actually negative at –.42 (Tellegen et al., 2003, p. 36), and RC3 shows an average correlation of .54 with Scale 8 versus an average correlation of –.18 with Scale 3 across samples. This result appears to be a consequence of the reverse keying of the RC3 items to reflect cynicism rather than the denial thereof, which is the thrust of the items on Hy2/Need for Affection, from which all of the RC3 seed items originate.

The internal analyses reported in the *Manual* are essential to showing the extent to which the goals for the RC Scales vis-à-vis the Clinical Scales were met. However, given the strategy used in their construction, one wonders why the explorations of the internal validity of the RC Scales reported in the *Manual* was limited to the Clinical Scales alone when such explorations could so easily have been extended to content-based scales, that is, scales constructed using comparable strategies. Providing correlations for the RC Scales with content-based scales would have met the important goal of comparing like with like.

Table 4 presents several correlations drawn from the Caldwell (1997) clinical data set between the RC Scales and the Clinical Scales, first-factor scales, and some of the content-dominated scales to illustrate the trend of such analyses had they been undertaken. These correlations—.85 or greater for most, .90 or greater for RC1, RC3, RC7, and RC8—exhibit extremely high levels of redundancy between the RC Scales and content-based scales that are already in wide use and have substantial empirical correlates. In some cases, as with RC1 and RC3, the correlational redundancy is so high as to suggest that the latter scales are mere congeners of their corresponding Content scales. The reasons for the high correlations between the RC and Content scales are not hard to find: They simply follow the patterns of item overlap between the former and the latter.

EXTERNAL VALIDITY: A NEED FOR MORE APPROPRIATE ANALYSES

It is the next series of analyses, those that purported to demonstrate the external validity of the RC Scales, that is most likely to give the reader an incomplete picture as to the value of the comparisons presented. In these analyses, we compared the ability of the RC Scales to predict molecular ratings (e.g., somatic symptoms, anxiousness, delusions) with that of the Clinical Scales (Tables 4–13 to 4–20; Tellegen et al., 2003). These comparisons overlook fundamental differences in the nature of the scales compared. Just as the multidimensional Clinical Scales may possess advantages in the prediction of complex criteria such as psychiatric diagnoses, unidimensional scales will almost uniformly have an advantage over complex, syndromal scales in predicting discrete, unidimensional ratings variables.

In other words, the external correlates reported in the *Manual* were relatively "soft targets" for the RC Scales, and their probative value for establishing the superiority of the RC over the Clinical Scales is quite limited. These comparisons are on a par with demonstrations that the Block Design subtest of the Wechsler (Wechsler, 1958) better predicts performance on visual-spatial tasks than does Full Scale Intelligence Quotient. For complex criteria, a far more informative demonstration of the comparative validities of the RC versus the Clinical scales would involve more appropriate but more challenging multivariate correlates such as the Structured Clinical Interview Axis I *DSM–IV* (First, Spitzer, Gibbon, & Williams, 1997) diagnoses. For more unidimensional ratings

TABLE 3
Mean Correlations Among the Seed and Nonseed Portions of Each of the RC Scales and Various First-Factor Markers in the Caldwell Clinical Data Set

Scale	*RC1S*	*RC1N*	*RC2S*	*RC2N*	*RC3S*	*RC3N*	*RC4S*	*RC4N*	*RC6S*	*RC6N*	*RC7S*	*RC7N*	*RC8S*	*RC8N*	*RC9S*	*RC9N*
Dem	66	63	56	76	44	47	27	42	47	51	77	82	51	55	22	35
RCd	65	63	56	76	45	49	28	44	48	52	77	82	52	56	23	37
A	65	63	51	70	52	57	31	47	51	56	84	90	58	62	30	47
JB1	66	64	49	68	53	58	33	49	51	55	85	92	58	62	35	51
W1	67	66	53	71	54	59	34	50	56	61	85	91	60	66	32	50
AJBW	64	62	50	69	52	56	32	48	51	55	82	89	57	61	31	47
JBW72	67	65	50	69	53	57	32	48	52	56	84	91	58	62	33	50
Seed *M*	66		52		50		31		51		82		56		29	
Nonseed *M*	64		71		55		47		55		88		61		45	

Note. Decimals omitted. RC = Restructured Clinical; S = seed items; N = nonseed items; JB1 = Johnson–Butcher (Johnson et al., 1984) first factor; W1 = Waller (1999) first factor; JBW72 = items overlapping both the Johnson–Butcher and Waller first factors; AJBW = items overlapping the A scale and JBW72.

criteria, the more appropriate comparison would contrast the predictive ability of the RC scales with those of similarly derived content-based scales such as the MMPI–2 Content scales or the PSY–5 scales. Such a comparison would have provided immeasurably clearer and more cogent evidence for the advantages (or deficiencies) of the RC scales. Even among the comparisons presented in the *Manual*, the ability of the RC scales to predict ratings are not always reassuring. For example, in Table 4–13 (Tellegen et al., 2003, p. 43), one finds that the best predictor of the Patient Description Form (PDF) Psychotic Symptoms scale for the Portage Path (Graham et al., 1999) men ($n = 410$) was RC4 (Antisocial Behavior) at .21, higher than both RC6 (Ideas of Persecution) at .16 and RC8 (Aberrant Experiences) at .15.

Such details should not distract, however, from the larger issue: The choice of the Clinical Scales as contrasts for the RC Scales in the prediction of molecular ratings violates the spirit of comparing like with like.

Next, chapter 5 of the RC *Manual* presents interpretive guidelines for the RC Scales and their use in nine case examples. The data provided for these examples include scores for the Consistency and Response Style scales, the Clinical Scales, and the RC Scales. Here again, however, information about the Harris–Lingoes subscales, Content and Component scales, and PSY–5 scales—the kinds of scales most likely to make the redundancy of the RC Scales conspicuous—is not provided for any of the case examples. A final chapter offers conclusions and future directions.

Taking the *Manual* as a whole, one is struck by the unevenness in the coverage of the RC project. On one hand, matters relating to judgments about the placement of individual items on the various scales are usually described in considerable detail. On the other hand, although the *Manual* provides a review of two alternatives to the empirical or contrasted groups approach to scale construction, discussion of previous research in solving or illuminating the problem of the covariation among the clinical scales was omitted completely, including the highly relevant findings of the previous item factor analyses of the MMPI item pool by Johnson, Butcher, Null, and Johnson (1984) and Waller (1999); nor is any rationale offered for adopting PCA/V as the primary factor analytic procedure.

The strategy for the construction of the Dem and RCd scales was reasonably well described but, again, there is no discussion to inform the reader of the already existing methods for controlling the covariation among the Clinical Scales or about scales that are already available for its measurement. Nor is there any explanation to assist the reader's understanding of why the theoretically driven strategy adopted for developing Dem was considered preferable to an empirical approach. Many other concerns have been noted here about the Tellegen et al.'s choices of what to include/exclude in the *Manual.* Suffice it to say that those choices, like the ones just mentioned, make it difficult for the reader to understand and in particular, to evaluate, the steps followed in developing the RC Scales.

At this point, in Tellegen et al.'s defense, it must be acknowledged that test manuals or monographs for that matter are not required to live up to the balance, depth, and rigor required, or at least expected, of scholarly reviews. Nor should they be. On the other hand, no intellectual product that aspires to scientific status is entitled to fully set the terms of its own evaluation. As we proceed, the reader should bear in mind that the RC *Manual* is not a scholarly review and makes no claim to this effect. It is ultimately for the reader of this critique, and of the *Manual*, to decide the merits and shortcomings of both.

TABLE 4
Correlates of the RC Scales From the Caldwell Clinical Data Set

RCd:	A = 95, DEP = 95, Mt = 94, Pt[a] = 93, PS = 93, WRK = 92, PK = 92, D5 = 90
RC1:	Hs[a] = 96, HEA = 95, Hy4 = 94, HEA2 = 88
RC2:	INTR = 88, Anhedonia = 87, D1 = 85, D = 84
RC3:	CYN1 = 93, CYN =93 , Ho = 86, Pa3 = –81, ASP1 = 81, Hy2 = –81, S = –76, ASP = 76, Hy = 01
RC4:	PSP = 87, AAS = 82, ASP2 = 78, Re = –73, Pd[a] = 66
RC6:	Pa1 = 88, PSYC = 82, BIZ = 80, BIZ2 = 72, Sc1 = 71, BIZ1 = 71, F = 70, Pa[a] = 70
RC7:	JBW72 = 94, W1 = 94, JB1 = 93, A = 92, NEGE = 90, Pt[a] = 89, PS = 88, Sc[a] = 86, OBS = 86, Si3 = 85, WRK = 85, Rcd = 84, Mt = 84
RC8:	BIZ = 92, BIZ2 = 86, PSYC = 84, Sc6 = 83, Sc[a] = 80, BIZ1 = 77, Sc5 = 77, F = 74
RC9:	HOS = 75, Ma[a] = 74, Ma2 = 71, TPA= 69, Ho = 69, AGGR = 67, TPA2 = 67, ANG = 66

Note. Decimals omitted. RC = Restructured Clinical; RCd = Final Demoralization; A = Anxiety (Welsh, 1956); DEP = Depression; Mt = Maladjustment; Pt = Psychasthenia; PS = Posttraumatic stress disorder (Schlenger & Kulka, 1987); WRK = Work Interference; PK = Posttraumatic stress disorder–Keane; D5 = Brooding; RC1 = Somatic Complaints; Hs = Hypochondriasis; HEA = Health Concerns; Hy4 = Somatic Complaints; HEA2 = Neurological Symptoms; RC2 = Low Positive Emotions; INTR = Introversion/Low Positive Emotionality; Anhedonia (Watson, Klett, & Lorei, 1970); D1 = Subjective Depression; CYN1 = Misanthropic Beliefs; CYN = Cynicism; Ho = Hostility; Pa3 = Moral Virtue; ASP1 = Antisocial Attitudes; Hy2 = Need for Affection; S =Superlative Self-Presentation; ASP = Antisocial Practices; RC4 = Antisocial Behavior; PSP = Psychopathy (Nichols, 1989); AAS = Addiction Admission Scale; ASP2 = Antisocial Behavior; Re = Responsibility; RC6 = Ideas of Persecution; Pa1 = Persecutory Ideas; PSYC = Psychoticism; BIZ = Bizarre Mentation; BIZ2 = Schizotypal Characteristics; Sc1 = Social Alienation; BIZ1 = Psychotic Symptomatology; F = Infrequency; Pa = Paranoia; RC7 = Dysfunctional Negative Emotions; JBW72 = items overlapping both the Johnson–Butcher (JB; Johnson et al., 1984) and Waller (W; 1999) first factors; JB1 = JB first factor; W1 = W first factor; NEGE = Negative Emotionality/Neuroticism; Pt = Psychasthenia; OBS = Obsessiveness; Si3 = Alienation—Self and Others; RC8 = Aberrant Experiences; Sc6 = Bizarre Sensory Experiences; Sc = Schizophrenia; Sc5 = Lack of Ago Mastery, Defective Inhibition; RC9 = Hypomanic Activation; HOS = Manifest Hostility (Wiggins, 1966); Ma = Hypomania; Ma2 = Psychomotor Acceleration; TPA = Type A; AGGR = Aggressiveness; TPA2 = Competitiveness; ANG = Anger.

[a]Non-K corrected.

CRITIQUE OF DEMORALIZATION

Turning from the *Manual* to the RC project itself, a number of questions arise about both the general strategy and the particular procedures followed and about the new set of scales that resulted from the application of these procedures. With respect to overall strategy, it seems doubtful that the goal of developing a measure of the "broad, emotionally colored variable that underlies much of the variance common to the MMPI–2 Clinical Scales" (Tellegen et al., 2003, p. 11) was more likely to be advanced by following the guidance afforded by a model of affect of uncertain applicability to the MMPI–2 than by exploiting existing and multiply replicated empirical findings of large scale investigations into the structure of the test.

Although A scale has given relatively good service as a marker for the first factor over the years, there are reasons to consider it obsolete. Developed long before computer capacity could accommodate the factor analysis of MMPI items for very large samples, the A scale was derived through analyses applied to only 34 items that have shown high overlap among the Clinical Scales in relatively small, mostly male Veterans Administration samples. By the 1980s, the growth in computational capacity made possible the factor analytic investigation of the structure of the full MMPI item pool in samples of many thousands of clinical subjects, for example, Johnson et al. (1984) and Waller (1999).

Johnson et al. (1984) adopted a cross-validation design with an initial sample of 5,506 and a replication sample of 5,632 inpatients and outpatients from the Missouri Department of Mental Health. Johnson et al. used PCA/V and phi coefficients. The first factor (JB1) consisted of 87 items, of which 83 survive in the MMPI–2 item pool. Waller (1999) selected 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. Waller used binary-item factor analysis with tetrachoric correlations. Waller's first factor (W1) consisted of 143 items, of which 135 survive in the MMPI–2 item pool. It is notable that despite major differences in the samples, the factor methods employed and the type of correlation coefficient selected, 72 (87%) of the 83 MMPI–2 items that appeared on Johnson et al.'s replicated first factor were again replicated by Waller. Of these 72 items, 37 were scored on one or more of the Clinical Scales. Were one to seek an ideal factor analytically derived marker for the major source of covariation among the Clinical Scales or within the entire MMPI–2 item pool, it would seem difficult to find a better starting point than that provided by these sets of 37 and 72 items (hereafter JBW37 and JBW72), respectively.

It can certainly be argued that the goal of identifying and extracting the most relevant and problematic source of clinical Scale covariance would be approached most directly by limiting the search for this source to only those items that are actually scored on one or more of the Clinical Scales rather than involving the entire MMPI–2 item pool and risking the reintroduction of first-factor or other undesirable variances from off-scale items (i.e., items that are scored on none of the eight basic Clinical Scales). One might, for example, select the 37 items (JBW37) scored on one or more of the Clinical Scales that have been independently found to have high loadings on the first factor across the large item factor analyses of Johnson et al. (1984) and Waller (1999). A logical alternative approach might be to focus on the 35 items that overlap at least three of the eight clinical scales. Or, an even more productive approach might be to combine these two sets of items to benefit from both sources of covariation. A fourth possibility would be to focus on the primary factor following the factor analysis of the items of JBW37 when combined with all of the 101 items that are scored on at least two of the Clinical Scales. These are, of course, only a few of the most obvious possibilities.

It is an empirical question whether any of these alternative markers would be superior to Dem in drawing away from the Clinical Scales those items most responsible for their covariation from those that best reflect the residual core dimension(s) of each. However, the case for items scored on the Clinical Scales, especially those scored on multiple Clinical Scales, as those most clearly responsible for their covariation is prima facie. Rather than addressing this issue, the *Manual* is silent on these relatively simple, empirical approaches and fails to articulate the reason for turning instead to a theoretical model of affect of far less certain relevance to the task at hand.

In any event, as can be seen in Table 5, comparing the correlations of Dem, RCd, and the empirically derived first-factor markers A, JB1, and W1 with several scales that measure one or another facet of general maladjustment/subjective distress, there is reason to believe that Dem and RCd are somewhat unrepresentative markers for the major source of covariation within the MMPI–2 item pool. Whereas the empirically derived markers for the first factor, A, JB1, and W1 show similar and relatively balanced relationships with content related to depression, anxiety, tension, obsessiveness, and low self-esteem, Dem and RCd disproportionately favor depressive content.

Although the magnitude of the differences in the correlations between the various first-factor markers and the selected domains of item content in Table 5 tends to be small, much larger differences become apparent when the various facets of content were represented as proportions of each first-factor scale's total item content in Table 6. When represented in this way, Dem and RCd are seen to contain a preponderance of depressive content relative to the other facets of content comprising the first factor and that these other facets are distributed more equally in the empirically derived markers, A, JB1, and W1. For example, whereas items with depressive content comprise roughly 40% of the total content of Dem and Rcd (Dem = approximately 39%, Rcd = approximately 42%, respectively), such content comprises only about 20% (A = approximately 23%, JB1 = approximately 17%; W1 =

TABLE 5
Correlations Among Dem, RCd and Selected Scales Reflecting Various First-Factor-Related Areas of Item Content in the Caldwell Clinical Data Set

	Depressive Content		*Anxiety, Tension, Obsessive, Low Self-Esteem Content*					
Scale	*D*	*DEP*	*TSCD*	*Pt*	*ANX*	*TSCT*	*OBS*	*LSE*
Dem	81	94	96	93	90	87	84	86
RCd	81	95	97	94	89	86	83	87
A	75	92	94	95	90	94	89	87
JB1	73	90	93	96	92	93	90	87
W1	75	93	94	97	91	93	89	89
JBW72	74	91	94	96	92	94	89	87
AJBW	74	91	94	94	90	90	88	86
JBW37	79	91	93	97	93	93	86	84

Note. Decimals omitted. D = Depression; DEP = Depression; TSCD = T–S–C Depression & Apathy (Stein, 1968); Pt = Psychasthenia; ANX = Anxiety; TSCT = T–S–C Tension, Worry, & Fear (Stein, 1968); OBS = Obsessiveness; LSE = Low Self-Esteem; Dem = Preliminary Demoralization; RCd = Final Demoralization; A = Anxiety; JB1 = Johnson–Butcher (Johnson et al., 1984) first factor; W1 = Waller (1999) first factor; JBW72 = items overlapping both the Johnson–Butcher and Waller first factors; AJBW = items overlapping A and JBW72; JBW37 = items overlapping both the Johnson–Butcher and Waller first factors and scored on one or more Clinical Scales.

TABLE 6
Proportion of Selected Domains of Item Content Closely Related to the First Factor Found Within First-Factor Markers

	Depressive Content		*Anxiety, Tension, Obsessive, Low Self-Esteem Content*					
Scale	*D*	*DEP*	*TSCD*	*Pt*[a]	*ANX*	*TSCT*	*OBS*	*LSE*
Dem	30	39	48	39	22	9	13	22
RCd	30	46	50	38	21	8	8	17
A	15	23	31	33	18	10	13	5
JB1	17	14	20	33	19	25	8	8
W1	15	25	20	28	19	21	10	13
JBW72	18	15	22	35	24	29	8	10
AJBW	21	28	38	31	24	14	7	7

Note. Decimals omitted. D = Depression; DEP = Depression; TSCD = T–S–C Depression & Apathy (Stein, 1968); Pt = Psychasthenia; ANX = Anxiety; TSCT = T–S–C Tension, Worry, & Fear (Stein, 1968); OBS = Obsessiveness; LSE = Low Self-Esteem; Dem = Preliminary Demoralization; RCd = Final Demoralization; A = Anxiety; JB1 = Johnson–Butcher (Johnson et al., 1984) first factor; W1 = Waller (1999) first factor; JBW72 = items overlapping both the Johnson–Butcher and Waller first factors; AJBW = items overlapping A and JBW72.
[a]Non-K corrected.

approximately 20%, respectively) of A, JB1, and W1. By contrast, content related to anxiety, tension, obsessiveness, and low self-esteem can be seen to be more equitably distributed across these markers, particularly in the balance between depression on one hand and anxiety/tension on the other. Thus, the data presented in Table 6 suggest a rather strong depressive bias in the item composition of Dem and RCd.

It might be supposed that the depressive composition of Dem and RCd occurred more by accident than by design, but this would be incorrect. Other than a brief and incidental mention of the A scale (Tellegen et al., 2003, p. 12), the *Manual* contains no discussion of alternative markers for the major source of Clinical Scale covariance or the first factor. However, such a comparison is fundamental here. There are at least two related issues that present themselves: The first is the nature of the Demoralization construct (Dem and RCd) in terms of its PU parent construct. Second, if Demoralization is deemed to satisfy the aim of encapsulating and segregating the "*broad* [italics added], affectively colored dimension" (Tellegen et al., 2003, p. 1) underlying the variance common to all of the Clinical scales, it is fair to ask whether the demoralization construct is sufficiently broad to adequately embody this variance.

With respect to the first issue, it is clear that the PU dimension, or at least its unpleasantness pole, is predominantly depressive in character. According to Watson and Tellegen (1985), mood at this pole is described as blue, grouchy, lonely, sad, sorry, and unhappy. Six of the RCd items (25%) explicitly reference such moods (56, 65, 82, 95, 277, 388F), and all appear on the DEP Content scale. Judged by the number of overlapping items between DEP and both Dem (9/23 = 39%) and RCd (11/24 = 46%), a far greater proportion of

such items than appear on Scale 2 (16%), the Demoralization measures contain much more depressive content than their empirically derived first-factor markers A (9/39 = 23%), JB1 (11/83 = 14%), W1 (20/135 = 15%), and the replicated items from both JB1 and W1, JBW72 (11/72 = 15%). By contrast, the proportion of Anxiety (ANX) content contained in Dem, RCd, A, JB1, W1, and JBW72 varies much less, amounting to between 18% (A) and 24% (JBW72) across all of these measures.

Modeling Dem after an explicitly depressive theoretical dimension, unpleasantness, all but assured that when Dem was added to the items of each of the Clinical Scales and the resulting combinations factored at Step 2, the depressive facet of First-factor variance would be preferentially extracted relative to other facets of the first factor such as anxiety. In the case of Scale 2, to take the most egregious example, factoring the items of Scale 2 (Depression) when combined with those of Dem guaranteed that some of the "distinctive substantive core" (Tellegen et al., 2003, p. 15) variance of Scale 2 would itself be extracted as nuisance variance while leaving the noncore components of Scale 2 to be concentrated within the residual factor(s). That is, using one depressively biased construct to extract the unwanted variance from another depression construct like Scale 2 could not help but withdraw from the latter scale elements of the distinctive core variance that the procedures applied in Step 2 were explicitly intended to preserve.

The pattern of correlations presented in Table 7 suggest that this outcome was indeed the result of the method of construction of RC2. When measured against a dozen depression scales and their components, Dem and RCd are, overall, only slightly less saturated with depression than Scale 2. Moreover, the removal of Dem variance from Scale 2 resulted in an overall decline in the depressive variance within RC2, at least as measured by these scales. This impression is confirmed in the correlations with external ratings of depression presented in Tables 4 to 13 to 4 to 20 of the *Manual*. These correlations show RCd to have a small but consistent advantage over RC2 (but not Scale 2) in predicting depression in all of the samples for which such ratings are provided.

The loss of depressive variance from Scale 2 in Step 2 left a set of items described as having "positive emotional content" (Tellegen et al., 2003, p. 15), four of which—49, 109, 188, and 330—when keyed False, were adopted as the seed items for its "distinctive core component" (Tellegen et al., 2003, p. 15). Yet the heterogeneity of these four items creates reason for doubt that they could stand as a suitable core for anything. In the Waller (1999) factor analysis, for example, these four items obtained their primary loadings on no fewer than four different factors: Social Inhibition, General Maladjustment, Extroversion, and Hypomania, respectively.

In summary, based uon considerations of item content, scale intercorrelations, and external ratings criteria, the overall goal of the theoretically driven Demoralization construct—to encapsulate the "broad affectively colored dimension" (Tellegen et al., 2003, p. 1) underlying the variance common to all of the Clinical Scales—appears not to have been realized. Rather, Dem and RCd are disproportionately weighted with depressive content relative to other facets of content within the first factor such as anxiety, tension, obsessiveness, and low self-esteem. Furthermore, the consequence of removing Dem variance from Scale 2 was an overall decline in the depressive variance in RC2.

MORE ANOMALOUS RESULTS: THE EXAMPLE OF RC7

In the previous section, I suggested that the depression bias of Dem resulted in the overextraction of depressive variance from Scale 2, leading to a restructured scale, RC2, that is missing substantial core variance for depression. However, it is important to consider another, more comprehensive, consequence of employing the depressively biased Dem marker: the underextraction of a sufficiently broad range of unwanted variance, the nondepressive facets of the first factor, when Dem was appended to the other seven Clinical Scales in Step 2. That is, it can be anticipated that in Step 2, a broad range of nondepressive nuisance variance (i.e., anxiety, etc.) would have remained in the residual factor(s), possibly in some cases to inhabit the Clinical Scales' core dimensions.

TABLE 7
Correlations Among Dem, RCd, D, and RC2 and Clinical and Content Depression Scales and Their Components in the Caldwell Clinical Data Set

	D	*D1*	*D2*	*D3*	*D4*	*D5*	*DEP*	*DEP1*	*DEP2*	*DEP3*	*DEP4*	*TSCD*
Dem	81	90	43	62	88	91	94	86	88	85	64	96
RCd	81	90	42	61	87	92	95	88	89	86	66	97
D	—	95	68	80	90	82	80	77	79	64	55	81
RC2	83	86	60	62	82	74	76	75	70	63	56	77

Note. Decimals omitted. D = Depression; D1 = Subjective Depression; D2 = Psychomotor Retardation; D3 = Physical Malfunctioning; D4 = Mental Dullness; D5 = Brooding; DEP = Depression; DEP1 = Lack of Drive; DEP2 = Dysphoria; DEP3 = Self-Depreciation; DEP4 = Suicidal Ideation; TSCD = T–S–C Depression & Apathy (Stein, 1968); Dem = Preliminary Demoralization; RCd = Final Demoralization; RC2 = Low Positive Emotions.

The risk of such underextraction would appear to be especially great in the case of Scale 7 (Psychasthenia), a scale known to be especially highly saturated with first-factor variance (Scale 7 × A: $r = .95$). Indeed, of the 24 items appearing on RC7, the proportion of items retained from Scale 7 (8; 33%) is actually exceeded by the proportion of items overlapping with the empirically derived first-factor markers A (10; 42%), JB1 (13; 54%), W1 (20; 83%), and JBW72 (14; 58%), presumably just the kind of variance that Step 2 was designed to eliminate from Scale 7. Perhaps even more surprisingly, as compared with RC7, the original Scale 7 actually has, if anything, a somewhat lower proportion of overlap with A (13; 27%), JB1 (26; 54%), W1 (37; 77%), and JBW72 (25; 52%). In other words, RC7 tends to have a greater overlap with the empirically derived first-factor markers than with Scale 7, from which first-factor covariance was to be extracted. Moreover, if Dem is a suitable marker for the first factor, its use to extract such variance from Scale 7 should not leave an RC7 seed scale with a greater affinity for the items of other first-factor markers than for items from scales more closely related to the original Scale 7 construct such as Anxiety (ANX) or Obsessiveness (OBS). It is further noteworthy that of the four Scale 7 items that overlap with neither any of the other seven basic Clinical Scales nor with any of the empirically derived first-factor markers A, JB1, W1, or JBW72 (313, 174F, 293F, 321F), presumably those items that are most distinctive of the core Scale-7 construct, none appear on RC7.

That the restructuring procedures applied to Scale 7 resulted in the underextraction of first-factor variance can be readily seen by comparing the empirically driven first-factor correlates of Scale 7 given in Table 5 (JB1, W1, and JBW72) with the RC7 correlations for the same markers presented in Table 4. Based on these values, the reduction of first-factor variance in RC7 as compared with Scale 7 works out to less than 1%. The problems with RC7 do not end here, however.

Table 8 presents a comparison between Scale 7 and RC7 across four categories of constructs: depression, psychasthenia, psychoticism, and anger/hostility. Overall, the pattern of correlations among these four construct categories suggests, predictably, that the primary consequence of restructuring Scale 7 was the extraction of depressive variance, with the first set of correlations showing a decline in such variance in RC7 relative to Scales 7, RCd, and A. In the next set of correlations, both RC7 and Scale 7 demonstrate moderate and presumably desirable relationships with scales related to the psychasthenia construct that Scale 7 was intended to measure, although RC7 appears to be slightly weaker than Scale 7 in this respect.

However, the third set of correlations in Table 8 show that the relationships of Scale 7 and RC7 with scales measuring psychotic phenomena, presumably much less desirable, are also relatively strong, with RC7 achieving marginally higher correlations with the psychoticism scales than Scale 7. That is, if anything, the RC7 scale is slightly more psychotic and slightly less psychasthenic in character than its Clinical Scale precursor. Thus, it appears that the restructuring procedures applied to Scale 7 achieved neither an increase in desirable substantive variances (psychasthenia) nor a decrease in undesirable ones (psychoticism), at least as measured by these scale correlates.

The example of RC7 provides an illustration of the central problem with the RC Scales: the use of a first-factor marker that fails to optimally extract unwanted covariance (vs. overly extract desirable variance) in subsequent restructuring efforts. In this case, there is reason to believe that the narrowness, that is, depression dominance, of Dem acted on Scale 7 to preferentially extract depressive variance and underextract psychoticism while failing to leave an enhanced residual core of variance(s) related to the original psychasthenia construct. In a sense, this illustration is the obverse of the case of Scale 2/RC2 in which the restructuring procedures had the opposite effect, namely, the overextraction of desirable construct variance, leaving the core substantive variance(s) impoverished.

CONSTRUCT DRIFT

The influence of a depression-biased marker like Dem, as in the Scale 2/RC2 and Scale 7/RC7 examples, was not the only source of potential distortions in the RC project. Distortions introduced in Step 2, when the clinical scales were factored with Dem appended, could then be exaggerated or ameliorated by the procedures applied in Step 4 to augment the seed scales. That is, the potential exists for the procedures applied in Step 4 to shift the essential character of the selected RC core constructs as new content was added to each of the scale seeds, resulting in construct drift.

Items contain multiple variances. Thus, in the Step 4 process of selecting items from elsewhere in the item pool to augment the seed scales, some of these variances were selected for (the "convergence criterion"), and others were selected against (the "discrimination criterion"). However, what of those variances that were neither explicitly selected for nor against? These remain occult through the process of seed augmentation as each item meets the criteria set for its retention. Provided that these occult variances differ more or less randomly from item to item, as each is added to its seed, the risk of construct drift is negligible.

However, when a sufficient number of the items selected by the convergence and discrimination criteria contain similar variances, these may accumulate to the extent of creating new scale variances that were not anticipated in the seed construct and may be undesirable. In some cases, a new, alien, scale variance may attain sufficient strength to dominate the variances that were intended for inclusion in the final RC Scale, thereby adversely influencing scale performance. Also, of course, the more items that are added, especially relative to the number of the seed items, the greater the risk that the content of the augmented RC Scale will have drifted

TABLE 8
Comparison of Construct-Relevant Scales With Scale 7 and RC7, and RCd and A, From the Caldwell Clinical Data Set

Constructs Relevant Scale	*Scale 7*	*RC7*	*RCd*	*A*
Depression				
DEP	91	82	95	92
DEP1	85	76	88	85
DEP2	83	71	89	83
DEP3	81	77	86	84
TSCD	93	83	97	94
M	83	85	87	83
Psychasthenia				
ANX	91	83	89	90
TSCT	93	88	86	90
FRS	57	59	48	53
OBS	86	86	83	89
LSE	85	83	87	87
MOR	90	88	90	97
M	84	81	81	84
Psychoticism				
RC6	59	61	54	58
RC8	69	70	60	66
BIZ	67	69	59	65
BIZ1	52	55	45	50
BIZ2	70	73	64	70
PSYC	71	73	65	71
PSY	77	79	70	76
TSCA	89	85	81	87
M	69	71	62	68
Anger-Hostility				
ANG	72	79	68	73
ANG1	58	64	55	58
ANG2	73	79	70	74
TPA	64	72	58	66
TPA1	68	75	63	69
TPA2	51	59	47	55
HOS	69	77	64	72
JB7	47	54	44	50
W16	61	68	57	63
M	61	68	57	63

Note. Decimals omitted. RC = Restructured Clinical; Scale 7 = Pt—Psychasthenia; RC7 = Dysfunctional Negative Emotions; RCd = Final Demoralization; A = Anxiety; DEP = Depression; DEP1 = Lack of Drive; DEP2 = Dysphoria; DEP3 = Self-Depreciation; DEP4 = Suicidal Ideation; TSCD = T–S–C Depression & Apathy (Stein, 1968); ANX = Anxiety; TSCT = T–S–C Tension, Worry, & Fear (Stein, 1968); FRS = Fears; OBS = Obsessiveness; LSE = Low Self-Esteem; MOR = Poor Morale (Wiggins, 1966); RC6 = Ideas of Persecution; RC8 = Aberrant Experiences; BIZ = Bizarre Mentation; BIZ1 = Psychotic Symptomatology; BIZ2 = Schizotypal Characteristics; PSYC = Psychoticism; PSY = Psychoticism (Wiggins, 1966); TSCA = T–S–C Autism & Disruptive Thoughts (Stein, 1968); ANG = Anger; ANG1 = Explosive Behavior; ANG2 = Irritability; TPA = Type A; TPA1 = Impatience; TPA2 = Competitive Drive; HOS = Manifest Hostility (Wiggins, 1966); JB7 = Johnson–Butcher Item Factor 7: Aggressive Hostility; W16 = Waller (1999) Item Factor 16: Hostility.

away from its core construct, perhaps to the extent of measuring a substantially different construct.

Referring to Table 2, one can see that the nonseed, mostly off-scale items dominate the seed items in the final RC Scales in an average ratio of 3:1. Thus, the final RC Scales (average length = 21 items) are balanced on a rather small base of seed items (average length = 7). Among the scales appearing most at risk for construct drift in the process of augmenting the seed scales with other items in Step 4 are RC7 and RC9 (71% nonseed items each). In the case of Scale7/RC7, as we have shown, whatever essential psychasthenic variance existed in Scale 7 was not increased in the transition to its RC version (Table 8). Rather, such variance was, if anything, diminished as the original Scale 7 seed items become dominated by items drawn from elsewhere in the item pool to complete RC7. Not surprisingly then, the correlations between RC7 and two relevant external ratings items from the PDF, Anxious and Obsessive–Compulsive, are smaller than their correlations with RCd, which, in turn, are smaller than their correlations with Scale 7 for both men and women in the Portage Path outpatients (Tellegen et al., 2003, pp. 43–44). This finding suggests, again, that, if anything, the core construct-relevant attributes of Scale 7 were decreased rather than increased in RC7.

Evidence for construct drift in the case of RC7 is suggested by the number of its items that overlap the Anger Content scale. RC7 has twice as many overlapping items with Anger (four) as it does with the presumably more construct relevant Anxiety Content scale (two). As can be seen in Table 8, RC7 shows higher correlations than Scale 7 across all of nine measures of Anger/Hostility, suggesting that anger/hostility variance was imported into RC7 during the Step 4 restructuring procedures. It appears that RC7 not only fails to enhance the distinctive and desirable core variances of Scale 7, it also fails to reduce the least desirable psychotic variance of Scale 7, the variance that accounts in part for the high correlation between Scales 7 and 8, and augments RC7 variance related to anger/hostility.

Scale 9 (Hypomania)/RC9 provides a clearer example of construct drift in that the variance that was imported into RC9 in Step 4 appears to be more specific and influential. The items forming the seed scale for RC9 reflect familiar aspects of the hypomanic syndrome including excitement and euphoria, racing thoughts, grandiosity, social disinhibition, stimulation seeking, tension, overreactivity, drivenness, and irritability. However, as many as one fourth of the items on the final RC9 scale connote hostility, vindictiveness, and intimidation. Of these, seven overlap PSY–5 Aggressiveness (AGGR), a scale measuring offensive and instrumental aggression (Harkness, McNulty, Ben-Porath, & Graham, 2002, p. 3). Six items overlap Wiggins (1966) Hostility (HOS) Content scale, five of which overlap AGGR. Yet the original Scale 9 contains only one item that appears on either AGGR or HOS. This suggests that considerable variance of an aggressive, hostile nature was imported into RC9 at Step 4. The influence of this new source of variance on RC9 versus Scale 9 can be judged from their correlations with a variety of scales referencing anger, aggression, and hostility presented in Table 9. Although manic/hypomanic patients are certainly known to become hostile on a reactive basis, as when thwarted, the expansion of hostile content in RC9 appears sufficient to dominate its variances,

leading to the underprediction of manic symptoms in nonhostile manic conditions and their overprediction in hostile nonmanic conditions.

Another instance of construct drift may appear to have occurred in the case of Scale 3 (Hysteria)/RC3. In this case, however, the shift away from the original Scale 3 construct appears to have occurred more as a result of design than drift. RC3 contains only five original items, all from Hy2 (Need for Affection); the somatization component of Scale 3 was dropped entirely. In addition, three items were imported into RC3 from Scale 6 (Paranoia). In a move that further distanced RC3 from the original Scale 3 construct, the RC3 items were keyed in the opposite direction from Scale 3, thereby making the RC and clinical profiles incongruent.

Thus, Scale 3 was not so much restructured as replaced by RC3, a scale highly redundant (80% item overlap) with the Cynicism (CYN) Content scale. The lack of continuity between Scale 3 and RC3 would appear to impose significant limits on the application of the RC Scales to medical, chronic pain, and personal injury cases (Butcher & Miller, 2006). Scale 3, which is of particular importance in such contexts, simply has no counterpart among the RC Scales.

Given the risk of construct drift, it is surprising that there is nothing in the *Manual* to indicate that an attempt was made to ascertain whether the goal to embody and concentrate the variance of the distinctive core construct selected for each of the Clinical Scales was achieved. This would involve a fairly straightforward follow-up task: The final RC Scales would be factored again to assure that the selected scale seeds/core constructs survived within the dominant variance component of each. Such an analysis would have been highly informative for establishing the appropriateness and success of Tellegen et al.'s procedures and would serve to identify such sources of variance as may have been imported into the scales inadvertently as the steps in restructuring proceeded.

The point is significant. The RC Scales (RC1 to RC9) contain 168 items, of which only 81 (48%) derive from the original Clinical Scales. Of the 87 off-scale items that were selected to augment the seed scales in Step 4, 21 are among the 107 items that were added at the time of the MMPI–2 restandardization and were therefore unavailable as candidate items for discriminating Hathaway's original criterion groups (Hathaway & McKinley, 1940) from his normals. The remaining 66 off-scale items (39% of the total number of items on the RC Scales) were included in the original MMPI item pool and thus were available to discriminate any of the original criterion groups but failed to do so. The failure of these items to qualify for inclusion in any of the basic Clinical Scales at the time that they were constructed inevitably raises doubts as to their distinctive substantive status vis-à-vis these Clinical Scale constructs, whether core or otherwise. In fact, the inclusion of these items as representative of the "distinctive substantive core" (Tellegen et al., 2003, p. 15) of one or another of the Clinical Scales is ironic: Far from inhabiting the supposed core of any of the Clinical Scales, they were specifically excluded from them at the time of the construction of these scales.

As empirical correlates were gathered from the Portage Path and Veterans Administration Medical Center samples to provide initial validity anchors for the new RC Scales, the authors of the *Manual* as well as their readers could be expected to have a keen interest in knowing the precise nature of the variances most likely responsible for whatever correlates were found. Leaving aside the matter of which scales, Clinical or Content, would be the more appropriate and informative as contrasts for the RC Scales, the question of the extent to which these scales actually retain the core Clinical Scale variances selected for them remains. Did the final versions of the RC Scales build on their seed scales in a way that left these intended variances dominant? Or did the procedures used in their construction allow these core variances to be removed (e.g., Scale 2/RC2), contaminated (e.g., Scale 7/Rc7, Scale 9/RC9), or overtaken (e.g., Scale 9/RC9) by unintended variances introduced by items from elsewhere in the MMPI–2 item pool as these came to be assigned to the seeds in Step 4? The *Manual* does not address these questions.

SOME MORE FORTUNATE OUTCOMES

Not all of the transformations of the Clinical Scales into their RC versions appear undesirable. For example, RC8 possesses features that make it unique among MMPI–2 scales. The best of the previous MMPI–2 psychoticism scales, BIZ and PSYC, include considerable paranoid content. In the construction of the RC Scales, this content was segregated to

TABLE 9
Comparison of Correlations for Construct-Relevant Scales With Scale 9 and RC9 From the Caldwell Clinical Data Set

	ANG	*ANG1*	*ANG2*	*TPA*	*TPA1*	*TPA2*	*AGGR*	*TSCR*	*HOS*	*JB7Hos*	*W16Hos*
Scale 9	53	52	45	53	38	54	53	53	62	51	33
RC9	66	63	56	70	50	68	66	62	75	70	50

Note. Decimals omitted. ANG = Anger; ANG1 = Explosive Behavior; ANG2 = Irritability; TPA = Type A; TPA1 = Impatience; TPA2 = Competitive Drive; AGGR = Aggressiveness; TSCR = T–S–C Resentment & Aggression (Stein, 1968); HOS = Manifest Hostility (Wiggins, 1966); JB7 = Johnson–Butcher Item Factor 7: Aggressive Hostility; W16 = Waller (1999) Item Factor 16: Hostility; Scale 9 = Ma-Hypomania; RC9 = Hypomanic Activation.

RC6. RC8 reflects an unusually good balance of content reflecting anomalous experience such as acute dissociation and derealization; auditory, visual, and olfactory hallucinations; and so-called first rank (Schneider, 1959) symptoms such as thought broadcasting. There is no other MMPI–2 Clinical or Content scale in which this content is better represented and concentrated. Relative to Scale 8, therefore, RC8 may be quite sensitive to the positive—particularly the disordered perceptual—symptoms of schizophrenia. Conversely, it may be less specific in the sense of generating false positive predictions for schizophrenia in patients with other disorders that may also manifest through such symptoms (e.g., mania, depression with psychotic features, some Cluster A personality disorders, toxic states, etc.).

The psychoticism content of Scale 8 is limited to perhaps one fifth of its items, the remainder of which reflect severe alienation, disrupted cognition, amotivation, anergia, bewildering emotional experience, and sensorimotor symptoms. Scale 8 is therefore probably more sensitive to the deficit or negative aspects of the schizophrenia syndrome and thus vulnerable to false negative predictions for schizophrenia manifesting through positive symptoms. Although RC8 bears an unknown relationship to schizophrenia at present, it seems likely that it would lack sensitivity (i.e., risk false negatives) for negative symptom schizophrenia, at least relative to Scale 8. Given their complementary emphases, however—negative symptoms in Scale 8 and positive symptoms in RC8—the use of these two scales in conjunction may well improve predictions related to this disorder.

RC4 is another scale that may possess some advantages over comparable scales such as Pd2 (Authority Problems), Antisocial Practices (ASP), and Disconstraint (DISC). Antisocial attitudes and behavior are, if anything, underrepresented in the content of Scale 4 (Psychopathic Deviate), especially relative to its substantial content related to dysphoria and guilt (e.g., six items overlap DEP) and alienation. RC4 largely constitutes an extension of Pd2 and ASP2 (Antisocial Behavior) with some Pd1 (Familial Discord). In other words, RC4 is a collection of items reflecting antisocial conduct, dispositions thereto, and attendant family conflict. Seven items (22%) overlap the Addiction Admission Scale (AAS), a proportion sufficient to risk false positive inferences of broad antisocial dispositions and behavior based on substance abuse alone. Similar to ASP2 and Pd2, RC4 retains a significant bias toward conduct problems in childhood and adolescence, but this bias appears to be less severe in RC4 than in the former scales.

RESTRUCTURED? CLINICAL?

Despite their title, "Restructured Clinical Scales," the RC Scales manifest neither the gross appearance (i.e., heterogeneous content, syndromal complexity) nor a pattern of internal correlates typical of the Clinical Scales. Indeed, they are not clinical scales in any traditional sense of this usage. Although the RCs do have their origins in the Clinical Scales, less than half of the RC items, on the average, derive from their parent scales (range 29%, RC9; 77%, RC6; Table 2). For half of the RC Scales (RC1, RC2, RC3, RC8—with RC4 and RC9 coming very close), the number of items retained from the parent Clinical Scales is equaled or exceeded by items that overlap with content driven scales (Table 2). These patterns of overlap render most of the RC Scales highly redundant with these other face valid, unithematic scales.

If having only a minority of items that appear on one of the Clinical Scales is a fair standard against which to evaluate the RC project's success in identifying and embodying each Clinical Scale's "distinctive substantive core" (Tellegen et al., 2003, p. 15), this goal has obviously not been achieved. The RC Scales will therefore not serve as proxies for the original Clinical Scales (RC1 is a possible exception). They will not even function in ways that are analogous to the Clinical Scales; hence, the designation of these scales as clinical is overstated. At best, the RC Scales are hybrids: Content scales with clinical roots.

Moreover, attempts to use the RC Scales as analogues of their parent scales may occasion substantial interpretive blunders. RC3 is the most obvious case in point, but a case can be made that elevations on RC4, RC8, and RC9 may substantially misdirect diagnostic inferences even in the presence of concurrent elevations on their parent scales.

THE ROAD NOT TAKEN

Although not new, continuing efforts to strengthen the Clinical Scales in ways that stand to improve their discriminant functioning, of which the RC project is an example, remain important. Tellegen et al., as authors of the RC Scales, are to be credited for their attempt to tackle this problem. However, in contemplating Tellegen et al.'s project, nonintrusive adjustments were overlooked.

At least two sorts of such adjustments come to mind: first, those items from the Clinical Scales that appear on existing, well-replicated markers for first-factor covariance could simply be deleted from each Clinical Scale. For example, in the Caldwell (1997) clinical data set, such an adjustment was completed based on the removal of the 37 Clinical Scale items overlapping both the Johnson et al. (1984) and the Waller (1999) first factors (JBW37). This adjustment produced an average reduction in the level of Clinical Scale covariance from .59 to .47, whereas the scales thus modified maintained an average correlation with their unaltered parent scales of .97, a negligible loss in fidelity to the original Clinical Scales and a reduction of 55 item overlaps.[6]

[6]As a point of reference, the RC Scales do demonstrate a desirable gain in independence, with an average level of intercorrelation of

A second effective strategy could focus on the 35 items that occur on three or more of the Clinical Scales. When these items were deleted from all of the Clinical Scales, the average correlation among the scales thus adjusted dropped from .59 to .39 in the same sample while maintaining an average correlation with their unaltered parent scales of .94. This strategy resulted in a robust 20% gain in scale independence, with an average loss of only .06 in fidelity to the original Clinical Scale constructs. This significant drop in the average correlation among the Clinical Scales occurs because the 35 deleted items account for nearly two thirds (131/197) of the total item overlaps among the scales.

These two strategies could be combined. For example, there are a total of 62 individual scale items that either overlap with three or more clinical scales or appear on JBW37. Deleting these 62 items from the Clinical Scales on which they appear results in an average reduction of the intercorrelation among these scales from .59 to .32 in the same sample, a 25% increase in scale independence, whereas the scales thus modified maintain an average correlation with their parent scales of .88, a still-respectable level of construct fidelity. There are other, similarly straightforward ways of increasing scale independence, all of which can be calibrated to whatever level of construct fidelity is desired. These kinds of preliminary scale modifications would provide strong and defensible starting points for scale restructuring.

Regardless of the need or desirability one may attach to the pursuit of core dimensions of the Clinical Scales, the broader problem of their covariation and the routine challenges this problem presents to clinical interpretation with the MMPI–2 remain unresolved. Yet the availability of at least one adequate and unbiased marker for the first factor and various uncomplicated means for applying such a marker to the MMPI–2 in ways that reduce unwanted variance from the Clinical (or other) Scales, thereby increasing their independence and discriminant performance, should be a cause for optimism.

The availability of such means does not, of course, obviate the need for the careful consideration of context: What level of control or reduction of first-factor covariance provides the greatest benefit in practice? Is it preferable to think in terms of optimizing removal of this variance or of maximizing it? It might be, for example, that different syndromes may involve somewhat different characteristic styles of distress and suffering that the overextraction of the first factor would obscure. It may be advantageous for the clinician to retain some access to such variance to support better personality description and clinical prediction.

The quest for the distinctive core features of the Clinical Scales need not be pursued in an inflexible, one-size-fits-all fashion. Some scale constructs, Scale 3, for example, might be restructured to better advantage if features of both its somatic focus and pattern of social dispositions were allowed to persist in a restructured scale. At the very least, however, future scale restructuring efforts should be able to demonstrate clear advantages over the kinds of relatively simple and minimally intrusive fixes like those I described previously, something that the RC project did not do.

.46, but their correlations with their parent scales averages only .71 in the Caldwell (1997) sample.

SUMMARY

The history of efforts to strengthen the MMPI Clinical Scales in ways that improve their discriminant functioning is lengthy but of continuing importance, and the RC project falls within this tradition. However, the RC *Manual* is uneven, notable at least as much for what it excludes as for what it contains. Previous approaches to the Clinical Scale covariance problem, the methods used and the solutions proposed, are treated perfunctorily or, more often, not at all. Similarly, considering the level of detail provided to explain and justify the ultimate assignment of scale items, the rationale given for the factor-analytical approach on which Tellegen et al.'s primary results rest is terse, and considerations of alternative approaches such as the use of tetrachoric correlations or an oblique rotation strategy are nowhere to be found.

The RC Scales selectively emphasize a single content theme embodied within each Clinical Scale. As such, they stand at considerable remove from the Clinical Scales because of the loss of the syndromal complexity that characterizes their parent scales and closer to content-based scales, scales that are all but invisible in the *Manual*. In virtually all cases, the selected RC core dimensions are already adequately, if not abundantly, represented in one or more of the numerous content-based scales of the MMPI–2 as indicated by the extremely high correlations of the RC Scales with their respective content-based scales.

Certain methodological features of the RC project appear problematic. For example, the extraction of unwanted variance from the Clinical Scales to create the initial "distinctive substantive core" (Tellegen et al., 2003, p. 15) dimension of each at an early stage in the project was at least partially reversed in a later stage that made it possible for unwanted variance to reinfiltrate the core item sets. In some cases, the procedures followed in this latter stage also resulted in the unintended recruitment of item content that was substantially alien to the core RC constructs selected earlier. RC7 and RC9 manifest significant flaws due to construct drift as a result of these procedures. The failure to provide evidence for the dominance of the variances Tellegen et al. selected to be embodied in their seeds for the final RC Scales is a particularly unfortunate omission in the *Manual*. Such a simple check would almost certainly have uncovered the construct drift in RC9.

Regardless of their redundancy with other scales and of whatever particular flaws may characterize each scale individually, the RC Scales are compromised as a set by the adoption of Dem, the theoretically inspired and depressively

biased marker that played a major role in their construction. Empirical evidence I presented in this article suggests that this marker served to overextract depressive variance from each of the Clinical Scales relative to other facets of unwanted variance such as anxiety that, in combination, comprise and define the first factor. In one case, the depression-biased Dem was used to extract unwanted variance from the Depression scale (Scale 2), thereby assuring that significant core depressive variance would be lost rather than preserved in a restructured scale (RC2).

Progress toward goal of the RC project to construct a set of scales that represent the "distinctive substantive core" (Tellegen et al., 2003, p. 15) of each Clinical Scale must be judged against the fact that fewer than half of the RC Scale items derive from their parent Clinical Scales. A total of 66 or nearly 40% of the RC items were available for inclusion in the Clinical Scales at the time these scales were constructed but failed to be selected for them. This failure surely raises doubts about the ability of these items to represent the peripheral, much less the core dimensions of the Clinical Scales that the RC project claims for them. The inclusion of items within the final RC Scales that overlap with two or more Clinical Scales and that outnumber RC items that are unique to their parent scale is another cause for doubt that the goal of embodying distinctive core clinical scale constructs in the RC Scales was achieved.

As a result of these various problems, the RC Scales should be used with caution. Their application in medical, chronic pain, personnel, and forensic evaluations may be of particular concern because of their increased likelihood of forensic challenge. Even within outpatient and inpatient settings, the interpretation of the RC Scales may mislead clinical inferences because of their defects in design and composition.

Given the long history of the MMPI/MMPI–2 and the breadth of its use worldwide, it is unlikely that the RC project will be the final effort to find a sound and efficient means of establishing better control over the covariation afflicting the Clinical Scales. The problem is certainly not going away. Enough is now known about the MMPI–2, the sources of covariation among the Clinical Scales, and viable means for its control and amelioration to support a sense of optimism that future research will have a better outcome than the research that has been applied to the covariation problem thus far.

ACKNOWLEDGMENTS

I am indebted to two colleagues for their extraordinary assistance with this article. Roger L. Greene generously performed numerous and extensive statistical analyses that were used in the preparation of this review and repeatedly read and edited the manuscript in its various stages. Heather E. P. Cattell gave not hours but days of her life to an exceptionally meticulous reading of the penultimate draft, making bountiful, detailed comments on every page. Both have vastly improved the style of this article, and both have made nonnegligible substantive contributions as well. I also thank David Bradford, Jim Butcher, Alex Caldwell, Brenton Crowhurst, Lew Goldberg, Irv Gottesman, Leonard Simms, Niels Waller, and three anonymous reviewers for their readings and suggestions on earlier drafts and Bob Dean for his expert assistance with some difficult wording.

REFERENCES

Adams, D. K., & Horn, J. L. (1965). Nonoverlapping keys for the MMPI scales. *Journal of Consulting Psychology, 29,* 284.

American Psychiatric Association. (1994). *Diagnostic and statistical manual of mental disorders* (4th ed.). Washington, DC: Author.

Arbisi, P. A., Ben-Porath, Y. S., & McNulty, J. L. (2003). Empirical correlates of common MMPI–2 two-point codes in male psychiatric inpatients. *Assessment, 10,* 237–247.

Butcher, J. N., Dahlstrom, W. G., Graham, J. R., Tellegen, A., & Kaemmer, B. (1989). *MMPI–2: Minnesota Multiphasic Personality Inventory–2: Manual for administration and scoring.* Minneapolis: University of Minnesota Press.

Butcher, J. N., Graham, J. R., Williams, C. L., & Ben-Porath, Y. S. (1990). *Development and use of the MMPI–2 content scales.* Minneapolis: University of Minnesota Press.

Butcher, J. N., Graham, J. R., Ben-Porath, Y. S., Tellegen, A., Dahlstrom, W. G., & Kaemmer, B. (2001). *MMPI–2: Manual for administration and scoring* (Rev. ed.). Minneapolis: University of Minnesota Press.

Butcher, J. N. & Miller, K. (2006). Personality assessment in personal injury litigation. In I. B. Weiner & A. Hess (Eds.), *Handbook of forensic psychology* (3rd ed.). New York: Wiley.

Caldwell, A. B. (1997). [MMPI–2 data research file for clinical patients.] Unpublished raw data.

Dahlstrom, W. G. (1969). Recurrent issues in the development of the MMPI. In J. N. Butcher (Ed.), *MMPI: Research developments and clinical applications* (pp. 1–40). New York: McGraw-Hill.

Finney, J. C. (1968). Correction for unwanted variance. *Psychological Reports, 23,* 1231–1235.

First, M. B., Spitzer, R. L, Gibbon, M., & Williams, J. B. W. (1997). *Structured Clinical Interview Axis I DSM–IV Disorders–Patient edition (SCID–I/P).* New York: Biometrics Research, New York State Psychiatric Institute.

Goldberg, L. R. (1965). Diagnosticians vs. diagnostic signs: The diagnosis of psychosis vs. neurosis from the MMPI. *Psychological Monographs: General & Applied, 79,* 29.

Gorsuch, R. L. (2003). Factor analysis. In J. A. Schinka & W. F. Velicer (Eds.), *Handbook of psychology: Research methods in psychology* (Vol. 2, pp. 143–297). New York: Wiley.

Graham, J. R., Ben-Porath, Y. S., & McNulty, J. (1999). *MMPI–2 correlates for outpatient community mental health settings.* Minneapolis: University of Minnesota Press.

Harkness, A. R., McNulty, J. L., Ben-Porath, Y. S., & Graham, J. R. (2002). *MMPI–2 Personality Psychopathology Five (PSY–5) scales.* Minneapolis: University of Minnesota Press.

Hathaway, S. R., & McKinley, J. C. (1940). A multiphasic personality schedule (Minnesota): 1. Construction of the schedule. *Journal of Psychology: Interdisciplinary & Applied, 10,* 249–254.

Hathaway, S. R., & McKinley, J. C. (1942). A multiphasic personality schedule (Minnesota): 3. The measurement of symptomatic depression. *Journal of Psychology, 14,* 73–84.

Helmes, E., & Reddon, J. R. (1993). A perspective on developments in assessing psychopathology: A critical review of the MMPI and MMPI–2. *Psychological Bulletin, 113,* 453–471.

Jackson, D. N. (1970). A sequential system for personality scale development. In C. D. Spielberger (Ed.), *Current topics in clinical and community psychology* (Vol. 2, pp. 61–96). New York: Academic.

Jackson, D. N., & Reddon, J. R. (1987). Construct interpretation of Minnesota Multiphasic Personality Inventory (MMPI) clinical scales: An orthogonal transformation. *Journal of Psychopathology and Behavioral Assessment, 9,* 149–160.

Johnson, J. H., Butcher, J. N., Null, C., & Johnson, K. N. (1984). Replicated item level factor analysis of the full MMPI. *Journal of Personality & Social Psychology, 47,* 105–114.

Nichols, D. S. (1989, March). *Construction of an MMPI psychopathy scale.* Paper presented at the 24th Annual Symposium on Recent Developments in the Use of the MMPI, Honolulu, HI.

Norman, W. T. (1972). Psychometric considerations for a revision of the MMPI. In J. N. Butcher (Ed.), *Objective personality assessment: Changing perspectives* (pp. 59–83). New York: Academic.

Reise, S. P., Waller, N. G., & Comrey, A. L. (2000). Factor analysis and scale revision. *Psychological Assessment, 12,* 287–297.

Rosen, A. (1962). Development of MMPI scales based on a reference group of psychiatric patients. *Psychological Monographs, 76*(8 Whole No. 527).

Schlenger, W. E., & Kulka, R. A. (1987, August). *Performance of the Keane-Fairbank MMPI scale and other self-report measures in identifying posttraumatic stress disorder.* Paper presented at the annual meeting of the American Psychological Association, New York.

Schneider, K. (1959). *Clinical psychopathology.* New York: Grune & Stratton.

Stein, K. B. (1968). The TSC scales: The outcome of a cluster analysis of the 550 MMPI items. In P. McReynolds (Ed.), *Advances in psychological assessment* (Vol. 1, pp. 80–104). Palo Alto, CA: Science & Behavior.

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

Waller, N. G. (1999). Searching for structure in the MMPI. In S. E. Embretson & S. L. Hershberger (Eds.), *The new rules of measurement: What every psychologist and educator should know* (pp. 185–217). Mahwah, NJ: Lawrence Erlbaum Associates, Inc.

Waller, N. G., Tellegen, A., McDonald, R. P., & Lykken, D. T. (1996). Exploring nonlinear models in personality assessment: Development and preliminary validation of a negative emotionality scale. *Journal of Personality, 64,* 545–576.

Watson, C. G., Klett, W. G., & Lorei, T. W. (1970). Toward an operational definition of anhedonia. *Psychological Reports, 26,* 371–376.

Watson, D., & Tellegen, A. (1985). Toward a consensual structure of mood. *Psychological Bulletin, 98,* 219–235.

Wechsler, D. (1958). *The measurement and appraisal of adult intelligence* (4th ed.). Baltimore: Williams & Wilkins.

Welsh, G. S. (1952). A factor study of the MMPI using scales with the item overlap eliminated. *American Psychologist, 7,* 341.

Welsh, G. S. (1956). Factor dimensions A and R. In G. S. Welsh & W. G. Dahlstrom (Eds.), *Basic readings on the MMPI in psychology and medicine* (pp. 264–281). Minneapolis: University of Minnesota Press.

Wiggins, J. S. (1966). Substantive dimensions of self-report in the MMPI item pool. *Psychological Monographs: General & Applied, 80*(22, Whole No. 630).

David S. Nichols
5107 NE Couch Street
Portland, OR 97213
Email: davemult@aol.com

Received September 29, 2005