JOURNAL OF PERSONALITY ASSESSMENT, *87*(2), 148–171

Further Evidence on the Validity of the MMPI–2 Restructured Clinical (RC) Scales: Addressing Questions Raised by Rogers, Sewell, Harrison, and Jordan and Nichols

Auke Tellegen
Department of Psychology
University of Minnesota

Yossef S. Ben-Porath and Martin Sellbom
Department of Psychology
Kent State University

Paul A. Arbisi
Minneapolis VA Medical Center
Minneapolis, Minnesota

John L. McNulty
Department of Psychology
University of Tulsa

John R. Graham
Department of Psychology
Kent State University

The reviews by Rogers, Sewell, Harrison, and Jordan (2006/this issue), and by Nichols (2006/this issue) offer markedly contrasting appraisals of the MMPI–2 Restructured Clinical (RC) Scales introduced by Tellegen et al. (2003). The one common feature is that both reviews draw on the same atypical MMPI–2 (Butcher, Dahlstrom, Graham, Tellegen, & Kaemmer, 1989) data set for their empirical analyses, with results warranting critical scrutiny. Rogers et al.'s critique provides an evaluation of the RC Scales from the perspective of Jackson's (1970) method of test development. One significant issue in Rogers et al.'s review concerns social desirability, prompting us to clarify our own views on this topic. We also highlight and discuss problems associated with Rogers et al.'s use of the unrepresentative data set. Nichols's polemical critique neglects empirical and theoretical support for demoralization as a central construct and misconstrues as "construct drift" the purposeful process of developing the RC scales. Nichols's criticisms and proposals overlook requirements for assessing syndromes and for construct validation and even rudiments of scale development. Our reply incorporates evidence, including new findings, refuting his criticisms and confirming that demoralization is a pervasive MMPI dimension, that the RC Scales capture the major distinctive features of the original Clinical Scales, and that they generate correspondingly meaningful validity patterns.

The MMPI–2 Restructured Clinical (RC) Scales (Tellegen et al., 2003) are the outcome of a comprehensive effort to modernize the basic sources of information on the test. In our monograph introducing the RC Scales, we[1] noted that an abundance of existing Minnesota Multiphasic Personality Inventory–2 (MMPI–2; Butcher, Dahlstrom, Graham, Tellegen, & Kaemmer, 1989) data sets was available to explore the validity of these scales. Rogers, Sewell, Harrison, and Jordan (2006/this issue) and Nichols (2006/this issue) use subsets of one existing data set to explore internal, structural features of the RC Scales. Although less informative than studies that include extratest measures, internal analyses can shed light on important characteristics of psychometric instruments. It is unfortunate, as we discuss and illustrate in this article, that these efforts fall short in several important respects and are hampered by their reliance on a flawed data set. We begin with the Rogers et al. effort to replicate some of the steps we took in developing the RC Scales and to examine elevation patterns of the scales.

ROGERS ET AL.'S REPLICATION STUDY

Although generally supportive of the effort to revise the Clinical Scales, Rogers et al. raise a number of questions and concerns about the RC Scales and caution against their use in applied settings. We address each of these concerns.

Paradigm Shift

Rogers et al. describe our approach to constructing and interpreting the RC Scales as a paradigm shift. We agree that our scale construction method departed from previous MMPI–2 scale development efforts. However, contrary to Rogers et al.'s suggestion, it was not our intent simply to adopt Jackson's (1970) sequential scale construction method. Rogers et al. may have inferred this from our review of Jackson's (1970) test construction approach; and on one important point, we did follow Jackson's (1970) example: In our selections of items, we were guided by the convergent/discriminant patterns of item correlations with a set of provisional construct measures. On the other hand, in our discussion of Jackson's (1970) approach, we noted its requirement that the test developer start with an already fully formulated and corroborated trait model, a prerequisite we did not find appropriate for the Clinical Scale restructuring project. Referring to our own approach, we wrote "methodologically we did not adhere to a particular standard test construction system or recipe, and theoretically we were guided by more or less open constructs, by our own judgment, and by data" (Tellegen et al., 2003, p. 12).

Rogers et al.'s implicit characterization of the RC Scales as representing an attempted "Jacksonian" paradigm shift also implies that Tellegen et al. abandoned the empirical keying approach (as Jackson, [1970, 1974], had). In fact, we adopted the empirically keyed Clinical Scales and the constructs they measure as our crucial starting points: "Our approach to restructuring the Clinical scales was based on the assumption that these scales represent conceptually meaningful and clinically important constructs" (Tellegen et al., 2003, p. 11). We believed that this approach represented a vital next step in a bootstrapping process in which the Clinical Scales played a central role. Also, as we (Tellegen et al., 2003, p. 12) reported in our monograph, part of the final step in the development of the RC Scales was to refine the scales on the basis of relevant external correlates insofar as it was possible to do so.

In the early days of the MMPI (Hathaway & McKinley, 1943), it had become evident that although the Clinical Scales did not adequately predict the psychiatric diagnoses they were designed to assess, they did have meaningful empirical correlates. This led MMPI researchers to focus on the scales themselves and to explore the range and clarify the content of their correlates. A natural, if long delayed, next step in the bootstrapping process was to treat the Clinical Scales themselves as fallible criteria for the development of a set of new scales designed to better assess the major distinctive components of the Clinical Scales. From this perspective, we view the RC Scales as an attempt to fulfill Hathaway's (1972a) long-unrealized expectation that the Clinical Scales should and would eventually be improved on.

Insofar as the RC Scales represent a paradigm shift, it is not quite the one Rogers et al. articulate. The scales do represent a move away from the "blind empiricism" that has long been associated with the MMPI–2 toward a more theoretically grounded, construct-validity-guided approach. As we discuss and illustrate later, the RC Scales are linked in meaningful ways to contemporary models of personality and psychopathology. This allows MMPI–2 users to draw for their interpretations on not only the empirical correlates of the RC Scales but also the nomological frameworks embedding these correlates.

We now turn to some of the specific concerns Rogers et al. raise regarding construction of the RC Scales, beginning with considerations of social desirability.

Failure to Control for "Social Desirability"?

As we just discussed, Rogers et al. assume that we relied on Jackson's (1970) sequential approach to scale construction in developing the RC Scales and fault us for failing to implement the important second step of minimizing the influence of social desirability. Rogers et al. observe that in light of our "complete omission of Jackson's (1970) second principle in item selection, purists may question whether the

[1] In the following, "we," "us," and "our" refer to the authors of either this article or the monograph by Telegen et al. (2003). In addition, the text provides a specific reference to either document if the context does not clearly imply one.

RC Scales should be described as based on Jackson's scale validation" (Tellegen et al., 2003, p. 15). Even if Rogers et al. were to concede that it had not been our aim to follow Jackson's (1970) scale construction methodology, they and others might still question our apparent decision to ignore the issue of social desirability in constructing the RC Scales. To address this criticism properly requires an examination of Jackson's (1970) own implementation of his second principle.

The suppression of social desirability, that is, the maximum unconfounding of substantive content and social desirability response style, was unquestionably one of Jackson's (1974) major objectives in developing his own personality inventories. For the Personality Research Form (PRF; Jackson, 1974), Jackson's (1974) first personality questionnaire, he assembled a separate Desirability scale to create substantive scales that were minimally correlated with it. Jackson (1974) characterized it as "largely independent of psychopathology" (p. 12), stressing the importance of avoiding "the predominantly psychopathological content" (p. 17) of other desirability scales. Inspection of the PRF Desirability measure reveals that this is not an accurate description. Many of its items, when answered in the undesirable direction, are clear statements of clinically relevant difficulties. Although the use of this type of Desirability scale did not seem to hamper the development of a normal-range inventory such as the PRF, could the same approach be followed to construct a clinically oriented inventory? Jackson's (1989) Basic Personality Inventory (BPI) is such an inventory.

As was true for the PRF, the BPI was reportedly developed in a manner intended to ensure "that irrelevant sources of response variance, such as those due to desirability responding … be suppressed" (Jackson, 1989, p. 1). However, in striking contrast to the PRF, the BPI does not include a separate index of desirability. Instead, the BPI manual characterizes several substantive scales (Self-Depreciation, Denial, and Deviation) as also being indexes of (un)desirable responding. The BPI manual itself (Jackson, 1989, p. 35) recognizes these confounds of clinical content and "desirability" when it acknowledges that the assessment of some forms of psychopathology "requires that their assessment must address the presence or absence of deviant and socially disapproved behavior" (Jackson, 1989, p. 22).

In sum, the initial construct of a general "evaluative bias" dimension of self-report appears to have fragmented into alternative versions of desirability represented by substantive scales that double as response-style measures. Psychometrically, the completion of the BPI signified a significant change in the approach to response styles. The most important and sophisticated program of inventory development to have been predicated on the idea of a general and separately measurable social desirability dimension ended up abandoning that construct.

Our own view of social desirability can be summed up as follows. The major MMPI factor Jackson and Messick (1962) interpreted as a desirability dimension primarily represents substantive and clinically descriptive variance rather than invalidating "stylistic" response variance. Block (1965), in particular, marshalled convincing evidence and arguments in support of this view. The principal substantive features of this major MMPI dimension have been interpreted as reflecting psychological malaise or subjective discomfort and, in the terminology of the RC Scales, it is highly saturated with Demoralization.

Items representing this content in self-report measures of psychopathology often do describe manifestly (un)desirable attributes. Test takers who encode any of these attributes in this way may deliberately or automatically bias their responses to relevant items in a desirable or undesirable direction given certain personal characteristics and/or conducive circumstances. Assessing these distortions requires the assessment of sometimes elusive latent state and trait variables. MMPI researchers and users have long been aware of these different sources of test invalidity. The use of special validity scales to aid in identifying defensiveness or invalidating efforts to fake good or bad continues to be integral to MMPI–2 interpretive practices.

We revisit the issue of social desirability as it applies to the RC Scales later in this article.

Cross-Validation of the RC Scales

Under this heading, Rogers et al. discuss their efforts to replicate our scale construction findings. In Table 3 of Rogers et al.s' article, they present results based on separate analyses of the items in each of the eight numbered RC Scales (RC1, RC2, etc.), augmented with the Demoralization items. For each item set, the table presents two kinds of information about three item subsets (the seed items of the specified RC Scale, the "added" items in that scale, and the accompanying Demoralization items): (a) the average loading on the target factor and (b) the proportions of replicated items, that is, the items loading not only strongest but at least |.40| on the target factor. From Rogers et al.'s Table 3, it is clear that the average factor loadings are uniformly strong but that the replication proportions vary markedly, ranging, in the authors' words, "from modest to exceptionally high" (p. 143).

However, these proportions present a misleading picture. Although one would not have expected a step-by-step duplication of our procedure, the target loading criterion of at least |.40| is too great a departure to qualify as a replication effort. Adoption of this criterion for the restructuring project itself would have aborted it at Step 3, the derivation of seed scales. In our monograph (Tellegen et al., 2003, p. 18), we reported choosing for that step a decidedly more lenient minimum, a loading of |.27| in at least two of our four development samples, to define a factor marker. This criterion allowed us to recruit a large enough number of items for each seed scale. A more veridical attempt to replicate our findings would have adopted a similar requirement.

How much of a difference does the choice between these two minimum loadings in fact make? We applied Rogers et al.'s procedure in two clinical samples[2] (not the four development samples we used to construct the RC Scales) to examine the Demoralization items and the seed and nonseed items of the numbered RC Scales (instead of the very small subsets of "added" items) in each of the same eight item sets. We obtained the following results. When the |.40| marker item criterion was applied to the numbered RC Scales, the percentages of confirmed markers across the 16 (scale/setting) combinations ranged from 37% to 100% ($M = 78\%$) for the seed items and from 40% to 100% ($M = 77\%$) for the nonseed items, results similar to the wide ranges reported by Rogers et al. However, with the |.27| criterion, the corresponding percentages ranged from 71% to 100% ($M = 95\%$) for the seed items and from 77% to 100% ($M = 93\%$) for the nonseed items. In other words, under the |.27| criterion, we found that for both sets of items, the proportions of replicated markers clearly exceeded 70% and were "high" in Rogers et al.'s definition (and "very high," i.e., larger than 90%, in three fourths of the cases). For the Demoralization scale (RCd), the proportion of confirmed markers, already high under the |.40| criterion, did not change under the |.27| criterion, in both cases ranging from 87% to 100%, with a mean of 99%. In summary, an amended version of Rogers et al.'s replication method that is identical to theirs, except for more faithfully paralleling ours on the crucial issue of marker item definition, yielded uniformly strong replication statistics.

Infrequent RC Scale Elevations?

Rogers et al. report that in their data set of 7,330 cases from the Caldwell (1997) MMPI–2 collection, 44.8% of the men and 40.4% of the women had within normal limits (WNL) RC Scale profiles compared with 36.1% and 30.8% WNL Clinical Scale profiles for men and women, respectively. Based on these numbers, Rogers et al. (2006/this issue) express concern that "almost half of clinically referred cases assessed [with the RC Scales] will have WNL profiles for which any clinical interpretations would likely be minimal" (p. 143). Rogers et al.'s conclusions are explicitly predicated on the assumption that they were analyzing a set of "clinically referred cases." In the Method section of Rogers et al.'s article, they state that their sample is "entirely composed of MMPI–2 raw data from clinical settings" (p. 141) and that "excluded from this database were child custody cases; these referrals markedly underreported their psychopathology … and could skew these findings" (p. 141).

The Caldwell (1997) MMPI–2 collection was not only the source of Rogers et al.'s study sample but also of Nichols's larger data set. Contrary to Rogers et al.'s assumption, R. Greene (personal communication to Y. S. Ben-Porath, February 15, 2006), who assembled this data set, informed us that it included an unknown but presumably large number of cases that had been sent to Caldwell's service for scoring only, without accompanying information regarding the settings and circumstances in which these MMPI–2s had been obtained. There is, in fact, no reason to believe that this sample did not include individuals tested for child custody evaluations and preemployment assessments and no way to exclude such cases from the data set. As a result, Rogers et al.'s concern that marked underreporting could skew their results was justified. R. Greene (personal communication to Y. S. Ben-Porath, December 5, 2005) informed us that nearly one-third of the cases in this data set produced a T score of at least 65 on validity Scales L or K. For a clinical setting, this is an implausibly high proportion of defensive profiles. By contrast, in the clinical samples used in our validation analyses, less than 20% of the participants produced similarly elevated scores on these validity scales.

To help us further ascertain characteristic features of the Caldwell data set that would be atypical for a clinical sample, we compared a random subset of 10,000 cases from this data set provided by R. Greene (personal communication, December 9, 2005), which we have given the shorthand label "Caldwell subsample," with several of our own data sets. For our comparisons, we chose the MMPI–2 normative sample as well as our outpatient and inpatient samples. As potentially informative measures, we selected two global emotional discomfort indicators, RCd and Waller's first-factor scale (discussed by Nichols), labeled here WF1 as well as the five standard MMPI–2 validity indicators L, K, F, F_P, and F_B. The descriptive statistics for the four data sets on these seven measures are presented in Tables 1 and 2.

Table 1 shows lower mean discomfort scores, closer to the normative means, in the Caldwell subsample (and by the same token in representative subsets such as Rogers et al.'s, which in the following will not always be specifically mentioned) than in our clinical samples. In addition, Table 2 reveals that the distributions of these two measures are *positively* skewed in the Caldwell subsample as they are (to an even greater extent) in the normative sample, whereas the skew is consistently *negative* in our two clinical samples. In other words, the Caldwell scale distributions tend to identify *high* levels of measured discomfort as deviant (as "out of the ordinary"), which, as expected, is also true for the normative sample. On the other hand, in our clinical samples, deviancy was on balance associated more with *low* observed levels of discomfort. It is evident that with respect to both the location and shape of the "first-factor" distributions, the Caldwell MMPI–2 data set is more similar to the MMPI–2 normative sample than to our clinical samples.

The five validity indicators present a picture that is compatible with that provided by the subjective discomfort measures. The overall pattern suggests that in the Caldwell subsample, the tendency to "accentuate the positive" is more prominent

[2] A description of all the samples used in the data analyses we report in this article appears in Appendix A.

TABLE 1
Comparisons of the Caldwell Sample With the Normative, Outpatient, and Inpatient Samples: Means, Standard Deviations, and Effect Sizes

	Means							*Standard Deviations*							*Cohen's d*						
	First Factor		*Standard Validity Indicators*					*First Factor*		*Standard Validity Indicators*					*First Factor*		*Standard Validity Indicators*				
Sample	*RCd*	*WF1*	*L*	*K*	*F*	F_B	F_P	*RCd*	*WF1*	*L*	*K*	*F*	F_B	F_P	*RCd*	*WF1*	*L*	*K*	*F*	F_B	F_P
Normative	4.82	33.55	3.55	15.15	4.04	1.91	1.09	4.86	19.39	2.17	4.66	3.09	2.52	1.32	.46	.27	.54	.26	.49	.44	.24
Caldwell	7.94	40.85	4.91	16.48	6.43	3.89	1.46	7.33	29.05	2.60	5.30	5.27	4.93	1.62	—	—	—	—	—	—	—
Outpatient	13.51	64.14	4.28	12.63	9.85	7.50	2.11	7.08	28.15	2.45	4.76	5.72	5.90	1.95	.76	.80	.24	.73	.64	.72	.39
Inpatient	13.65	64.26	4.23	13.32	11.31	9.35	2.40	7.45	29.96	2.58	5.23	6.30	6.29	2.08	.78	.80	.26	.60	.89	1.04	.54

Note. All of the differences are statistically significant at the .001 level. Normative sample $N = 2{,}600$; Caldwell sample $N = 9{,}012$; outpatient sample $N = 1{,}020$; inpatient sample $N = 2{,}378$. First Factor = first-factor scales; RCd = RC Demoralization scale; WF1 = Waller's (1999) first-factor scale; L = Lie scale; K = Correction scale; F = Infrequency scale; F_B = Back Infrequency scale; F_P = Infrequency-Psychopathology scale; Cohen's d = effect size comparing the means of the Caldwell sample with the normative, outpatient, and inpatient means in Rows 1, 2, and 4, respectively.

TABLE 2
Comparisons of the Caldwell Sample With the Normative, Outpatient, and Inpatient Samples: Skewnesses and Kurtoses

	Skewness							*Kurtosis*						
	First Factor		*Standard Validity Indicators*					*First Factor*		*Standard Validity Indicators*				
Sample	*RCd*	*WF1*	*L*	*K*	*F*	F_B	F_P	*RCd*	*WF1*	*L*	*K*	*F*	F_B	F_P
Normative	1.34	.85	.94	–.02	1.37	2.04	1.69	1.37	.40	1.46	–.44	2.40	4.79	4.11
Caldwell	0.60	.60	.52	–.11	1.32	1.62	1.70	–1.00	–.68	0.01	–.70	1.63	2.16	4.42
Outpatient	–0.42	–.20	.68	.43	0.61	0.73	1.33	–1.04	–.75	0.40	–.25	–0.26	–0.39	2.20
Inpatient	–0.37	–.21	.75	.41	0.53	0.40	1.17	–1.15	–.94	0.51	–.42	–0.47	–0.95	1.69

Note. First Factor = first-factor scales; RCd = RC Demoralization scale; WF1 = Waller's (1999) first-factor scale; L = Lie scale; K = Correction scale; F = Infrequency scale; F_B = Back Infrequency scale; F_P = Infrequency-Psychopathology scale.

than in the other three samples. To start with Scales L and K, Table 1 shows that the Caldwell subsample yields higher mean scores on these two favorable self-presentation indicators than do the other three samples. The findings for the negative self-presentation indicators F, F_P, and F_B are basically consistent with those obtained for RCd and WF1. That is, in the Caldwell subsample, the F, F_P, and F_B means are at intermediate levels: above the normative means and below the two clinical sample means. Furthermore, the skewness values are more strongly positive in the normative sample and the Caldwell subsample than they are in the two clinical samples.

Together, these findings on the subjective discomfort and validity scales strongly suggest that the Caldwell data set is not representative of any specific clinical or other meaningfully defined population and is more appropriately described as a composite sample, an amalgam, than as a well-defined clinical sample.

As a corrective, and to allow a more realistic appraisal, we determined the percentages of WNL[3] RC and Clinical Scale profiles in representative female and male samples collected in a variety of identified settings. Table 3 shows the results for 10 such groups: two male Veterans' Administration (VA) samples, one including psychiatric inpatients, the other veterans in substance abuse treatment, and separate samples of men and women from each of the following four settings: inpatients in a community hospital, outpatients at a community mental health center, individuals assessed at admission to a state prison system, and college students from a general psychology participant pool. Comparisons within Table 3 show, in contrast to Rogers et al.'s findings, similar WNL percentages for the RC Scales and the Clinical Scales: for example, median percentages of 14% and 13%, respectively, in the six patient samples (represented in the first four rows of Table 3).[4] Furthermore, in each of these six samples and for both

[3]We calculated WNL percentages for the nine RC Scales including RCd and the eight original Clinical Scales. The nine RC Scales were specifically designed to represent the eight original Clinical Scales and include RCd as a psychologically meaningful and interpretable indicator. Therefore, if one wants to know how many "uninterpretable" RC Scale profiles will be generated in a setting, then RCd needs to be included in the analysis. Because there are no restructured versions of Scales 5 and 0 available at this time, an interpreter seeking this information must rely on these scales themselves; therefore, whether they are elevated is irrelevant to the Clinical–RC Scale WNL comparison.

[4]Readers may wonder how to reconcile the observation that percentages of WNL profiles are similar for the two sets of scales with

sets of scales, the frequencies of WNL profiles do not exceed 20%, well below the values reported by Rogers et al. Even our college samples generated WNL percentages lower than their findings.

By highlighting the atypicality of their "clinical sample," these results clearly contradict Rogers et al.'s (2006/this issue) conclusion that "almost half of clinically referred cases assessed [with the RC Scales] will have WNL profiles" (p. 143). They are consistent with our preceding assessment of the findings reported in Tables 1 and 2, indicating the strong likelihood that the Caldwell MMPI–2 collection includes a substantial number of protocols of individuals motivated to underreport (contrary to the assurances Rogers et al. had been given that such cases had been excluded from the data set they were provided). The unfortunate result is that their findings, particularly those relating to the scale and profile elevations presented in their tables, are largely uninterpretable.

A Lack of Validation Research on the RC Scales?

Rogers et al. assert that more research is called for in certain areas before the RC Scales can be used in applied settings. Rogers et al. (2006/this issue) focus primarily on the susceptibility of the RC Scales to social desirability effects and malingering, and on "the need for programmatic research on the clinical characteristics uniquely associated with each RC Scale elevation" (p. 145). Before addressing these issues specifically, we note surprisingly that Rogers et al. did not consider any of the research on the RC Scales that has appeared since our monograph was published, some of which bears directly on the research needs they identify.

TABLE 3
Percentage of Individuals Scoring Within Normal Limits on RC, Clinical, and Non-K-Corrected Clinical Scales Across Different Samples

	RC Scales		*Clinical Scales*		*Clinical Scales (Non-K)*	
Sample	*Men*	*Women*	*Men*	*Women*	*Men*	*Women*
MVA[a]	11.3	NA	7.2	NA	8.4	NA
HCMC[b]	17.2	14.0	13.8	9.1	16.4	11.6
Portage Path[c]	19.8	14.6	16.3	11.8	16.8	12.3
VARC[d]	14.2	NA	14.2	NA	15.0	NA
Ohio DOC[e]	28.9	24.3	32.7	28.9	37.0	32.2
College[f]	34.0	32.2	41.3	38.7	44.9	41.3

Note. MVA = Minneapolis Veterans Administration (VA); NA = not applicable; HCMC = Hennepin County Medical Center; VARC = Cleveland VA Substance Abuse Recovery Unit; DOC = Department of Rehabilitation and Correction.
[a]*N* = 1,128. [b]*N* = 1,182 (709 men, 473 women). [c]*N* = 1,020 (410 men, 610 women). [d]*N* = 1,235. [e]*N* = 43,095 (35,982 men, 7,113 women). [f]*N* = 758 (276 men, 482 women).

Malingering and Social Desirability

On the topic of malingering, Sellbom, Ben-Porath, Graham, Arbisi, and Bagby (2005) compared the susceptibility of the MMPI–2 RC, Clinical, and Content Scales to overreporting and underreporting using archival analog simulation data sets. Sellbom, Ben-Porath, Graham, et al. concluded that the RC and Content Scales were comparably susceptible to misleading responding and that the seeming advantage of the Clinical scales in this regard was attributable to "subtle" items, which are less susceptible to manipulation but cannot generally be considered valid markers of the constructs targeted by their home scales. As we noted earlier in our discussion of social desirability, MMPI–2 researchers and users have long recognized that it has not been possible to construct scales that are immune to misleading test-taking approaches and that the solution to this challenge is to rely on a variety of well-validated, test-taking measures. In the case of the MMPI–2, a wealth of empirical data—most recently meta-analyzed by Rogers, Sewell, Martin, and Vitaco (2003) and Baer and Miller (2002)—is available to help guide effective, empirically grounded assessment of misleading responding. Such assessments should always precede interpretation of any of the MMPI–2 substantive scales. Their

findings that in clinical samples, mean scores on the Clinical Scales were generally higher than on the RC Scales. The following idealized illustration clarifies how this can happen. Consider the following two results for four hypothetical clinical cases. Assume there are no clinical cases without an elevation on at least one of the Clinical Scales and on at least one of the RC Scales. Also, for simplicity, let the four cases be clinically different from one another. The first data set consists of the T-score profiles on four Clinical Scales that are highly overlapping and correlated, probably confounded with demoralization:

	80	70	70	80
	80	80	70	70
	70	80	80	70
	70	70	80	80
Mean	75	75	75	75

The second set includes the scores of the same cases on four near-perfectly distinctive RC Scales (for simplicity, not including RCd):

	80	60	50	50
	50	80	60	50
	50	50	80	60
	60	50	50	80
Mean	60	60	60	60

These two sets of results show that in both samples, the percentage of WNL profiles equals zero but that the mean elevation is far lower for the RC Scales. The reason is that the RC Scales are more discriminantly valid such that each profile shows a substantial elevation on only one scale, whereas the Clinical Scale profiles show multiple elevations (in this case, elevations of all scales). This example is obviously an idealization, but it highlights the following point: If the RC Scales as a set are not only convergently valid but also more discriminant than the Clinical Scales, then (a) the percentage of profiles with at least one elevated scale could be roughly the same for the Clinical and RC Scales, but (b) mean scores would be substantially lower for the RC Scales.

inherent susceptibility to misleading responding is no more an impediment to using the RC Scales than it is to using any other self-report measures of psychopathology.

We already noted that the primary proponent of eliminating social desirability from self-report measures, Jackson, abandoned this objective in constructing his own measure of psychopathology. Other developers of self-report measures have reached similar conclusions. In describing the development of the Personality Assessment Inventory (PAI), Morey (1991) noted

> The idea of eliminating all stylistic variance from the test was neither desirable nor practical because there is no reason to suspect that certain response styles will be orthogonal to certain syndromes of personality or to certain personality traits. The psychological phenomena experienced by the schizophrenic will never be seen as socially desirable, while the depressed individual usually manages to see the cloud surrounding every silver lining. (p. 64)

In practice, Morey required that items assigned to a given scale be more highly correlated with that scale than with three response-style indicators, the PAI Positive and Negative Impression scales (PIM and NIM, respectively) and the Marlowe–Crowne Social Desirability Scale (Crowne & Marlowe, 1960). The resulting correlations, reported in the PAI manual, between PIM and NIM scales and some of the substantive PAI scales reached the .60s and .70s for some scales.

Rogers et al. nonetheless assert that the effects of response styles such as social desirability must be investigated prior to the clinical use of the RC Scales. Although for reasons just discussed, we believe that this expectation holds the RC Scales to a standard not applied to other self-report measures of psychopathology, existing data can be used to attend to Rogers et al.'s concerns. Using the two large clinical samples we examined in our monograph, we computed correlations between the RC and Clinical Scales and two commonly used measures of social desirability and present these in Table 4. The two measures are MMPI–2 versions of the Edwards (1957) Social Desirability (Esd) and Wiggins (1959) Social Desirability (Wsd) scales.

The Esd was developed through rational item selection and figured prominently in much of Edwards's (1957) and Jackson and Messick's (1962) early research on social desirability in the original MMPI. Subsequently, critics such as Block (1965) and later Jackson (1974) himself faulted the scale for being greatly confounded with actual psychopathology. As we discussed previously, when Jackson (1989) developed his own psychopathology inventory, he abandoned efforts to suppress this type of variance and instead recognized that it is inherently confounded with genuine pathology. The Wsd was constructed empirically by Wiggins (1959) to differentiate between college students instructed to take the MMPI in a socially desirable manner and those responding to the test following standard instructions.

TABLE 4
Correlations Between Clinical and RC Scales, Edward's Desirability Scale, and Wiggin's Desirability Scale With Corrections for Item Overlap

	Esd								*Wsd*							
	Outpatient Sample Scale				*Inpatient Sample Scale*				*Outpatient Sample Scale*				*Inpatient Sample Scale*			
	Corrected			*Corrected*		*Corrected*			*Corrected*			*Corrected*		*Corrected*		
Clinical	*r*	r^a	*RC*	*r*	*Clinical*	*r*	r^a	*RC*	*r*	*Clinical*	*r*	r^a	*RC*	*r*	*Clinical*	*r*	r^a	*RC*	*r*	r^a
			RCd	–.87				RCd	–.87	–.86			RCd	–.46				RCd	–.50	
Hs	–.69	–.68	RC1	–.66	Hs	–.69	–.68	RC1	–.64	–.63	Hs	–.33	RC1	–.27	Hs	–.31		RC1	–.23	
D	–.76	–.73	RC2	–.69	D	–.71	–.68	RC2	–.66		D	–.44	RC2	–.55	D	–.53		RC2	–.66	–.64
Hy	–.45	–.43	RC3	–.51	Hy	–.37	–.35	RC3	–.53		Hy	–.32	RC3	–.05	Hy	–.35	–.35	RC3	–.02	
Pd	–.66	–.65	RC4	–.34	Pd	–.65	–.64	RC4	–.39		Pd	–.37	RC4	–.25	Pd	–.44		RC4	–.28	
Pa	–.60	–.59	RC6	–.48	Pa	–.60	–.59	RC6	–.46		Pa	–.24	RC6	–.03	Pa	–.26		RC6	–.01	
Pt	–.91	–.90	RC7	–.86	Pt	–.92	–.91	RC7	–.87	–.85	Pt	–.43	RC7	–.34	Pt	–.45		RC7	–.28	
Sc	–.86	–.85	RC8	–.83	Sc	–.89	–.88	RC8	–.57		Sc	–.35	RC8	–.07	Sc	–.36		RC8	–.03	
Ma	–.24	–.24	RC9	–.28	Ma	–.36	–.36	RC9	–.37		Ma	.13	RC9	.06	Ma	.14	.13	RC9	.11	.09

Note. Outpatient sample, $N = 1{,}020$; inpatient sample, $N = 2{,}378$. RC = Restructured Clinical; Esd = Edward's (1957) Social Desirability Scale; Wsd = Wiggin's (1959) Social Desirability Scale.
[a]Correlations were corrected for item overlap.

Wiggins developed the scale to address concerns that the Esd and similar measures were overly confounded with psychopathology and therefore did not differentiate effectively between misleading responding and reporting of genuine psychological problems. Wiggins demonstrated that his empirically derived scale did indeed substantially better than the Esd in identifying misleading responding, and this finding has been replicated repeatedly with both the original MMPI (Baer, Wetter, & Berry, 1992) and MMPI–2 (Baer, Wetter, Nichols, Greene, & Berry, 1995).

Examination of the correlations reported in Table 4 indicates, as expected, that Esd is generally more highly correlated with both the Clinical and the RC Scales than is Wsd, consistent with past findings that Esd is more highly confounded with genuine psychopathology. The high correlations of RCd and RC7 (and the even higher correlations of Clinical Scale 7) with Esd are consistent with Block's (1965) observation that social desirability as operationalized by Edwards (1957) and Jackson and Messick (1962) was confounded with general maladjustment and particularly with anxiety. Both Esd and Wsd are generally more highly correlated with the Clinical Scales than with the RC Scales. With Esd, this difference is most dramatic for RC4, RC6, and RC8, and the only reversal occurs for RC3. With Wsd, substantially lower correlations are found for all of the RC Scales but RC2, likely reflecting the large number of Wsd items that describe positive engagement with one's environment.

In sum, empirical research indicates that the RC Scales are no more susceptible to misleading responding than are the Clinical Scales. In fact, whether one adopts Edwards's (1957) and Jackson and Messick's (1962) broad concept of social desirability or Wiggins's (1959) more narrowly focused (and better validated) definition, the RC Scales, despite their greater transparency, are overall less highly correlated with socially desirable responding than are the Clinical Scales.

Empirical Correlates

Rogers et al. suggest that more information is needed on the empirical correlates of the RC Scales and emphasize in particular the need for information on what is *uniquely* associated with each scale. They assert that "for the MMPI–2 RC Scales to demonstrate discriminant validity, clinical characteristics singular to each scale must be established" (Rogers et al., 2006/this issue, p. 145). We are not aware of any definition of discriminant validity that requires each scale in a multiscale instrument to have strictly unique correlates. Nor are we aware of any instrument that satisfies this requirement. This is not surprising, given the well-known problem of "comorbidity," the frequent co-occurrence of psychiatric diagnoses, especially, we assume, the "near-neighbor" diagnoses that are the focus of Rogers et al.'s. attention. Without a set of fully distinctive clinical criteria, it is futile to insist on scales boasting fully distinctive clinical correlates. Discriminant validity is better examined by looking at the relative strengths of the extratest correlates of the scales of an instrument and then evaluating the scales in light of conceptually and empirically informed expectations.

The external correlate analyses reported in our RC Scale monograph were conducted with data on thousands of clinical participants who were tested at different facilities and on whom various types of collateral data were collected. These analyses demonstrate comparable to markedly improved convergent validity for the RC Scales in comparison with the Clinical Scales and markedly improved discriminant validity. These findings have to date been replicated with different samples and collateral measures by Forbey and Ben-Porath (in press); Simms, Casillas, Clark, Watson, and Doebbeling (2005); Sellbom and Ben-Porath (2005); Sellbom, Ben-Porath, and Graham (2006); and Sellbom, Graham, and Schenk (2006). Studies of the RC Scales in medical, forensic, correctional, and other mental health facilities have been presented by our research group and others at scientific conferences and have yielded a growing body of peer-reviewed studies showing that the RC Scales have markedly improved criterion validity in comparison with the Clinical Scales, with findings that generalize nicely across settings and criteria. Very few, if any, MMPI–2 scales or other scales and instruments that have been in clinical use for years have firmer empirical foundations to support their interpretive use in clinical and other settings.

An Important "Omission" From the RC Scale Monograph?

Noting that the Standards for Educational and Psychological Testing (American Educational Research Association, American Psychological Association, & National Council on Measurement in Education, 1999) require that standard error of measurement (*SEM*) be used to evaluate the reliability of individual scales, Rogers et al. derived *SEM*s from the subset of the Caldwell database that they had at their disposal, and Rogers et al. report their results "because these data were omitted from the Tellegen et al. test manual" (p. 142). This observation is not entirely correct because *SEM*s of the RC Scales are, in fact, reported in Tables 4–4 and 4–5 of our RC scale monograph in a format comparable to what is reported in the MMPI–2 manual (i.e., based on test–retest correlations for a subset of the normative sample). This information was omitted from an initial printing of our monograph, and this may explain Rogers et al.'s observation that no *SEM* information was provided for the RC Scales. In any event, the *SEM*s provided by Rogers et al. are based on internal consistency analyses of the Caldwell data set, a sample of questionable usefulness for reasons we discussed earlier.

Concluding Remarks Regarding Rogers et al.

We appreciate Rogers et al.'s interest in the RC Scales and their effort to independently replicate some of the factor ana-

lytic findings from the developmental phase of the RC scale project, but we demonstrate that closer adherence to our methodology yields uniformly strong replications, unlike the more equivocal Rogers et al. results. We also note their reservations about using the RC Scales in applied settings. We show that, first of all, Rogers et al.'s recommendation is in large part based on analyses of a highly problematic collection of cases that does not represent a defined population and that analyses of several representative data sets yield results that show their concern about a dearth of elevated RC Scale scores in clinical settings to be unfounded. We also review relevant literature showing that their perspective on social desirability is not consistent with current knowledge and insights, particularly those bearing on clinically oriented self-report measures. Finally, we point out that Rogers et al.'s concerns about the empirical correlates of the RC Scales not only reflect discriminant validity standards that cannot be imposed but also overlook a growing body of research-based empirical RC Scale correlates. We believe that Rogers et al.'s reservations about the use of the RC Scales are shown to be unfounded in view of these methodological and empirical considerations.

NICHOLS'S OBJECTIONS

Nichols raises four primary criticisms of the RC Scales. First, the construct Demoralization and RCd, the scale we constructed to measure it, do not adequately capture the "first factor" of the MMPI–2; second, our scale construction methodology resulted in "construct drift," and as a result, some of the RC Scales do not adequately measure the intended constructs; third, the RC Scales do not maintain "syndromal fidelity"; and fourth, we omitted important information from the RC Scale monograph. On this last point, and before addressing these criticisms, three important omissions from Nichols's own article need to be pointed out.

To begin with, Nichols ignores all of the research published and presented on the scales since the release of the monograph in 2003. When this was pointed out in response to Nichols's request for feedback on an earlier version of his critique, he replied "I am acting as a reviewer of the monograph in question, not of every public forum in which the RC Scales may be discussed, including SPA [Society for Personality Assessment] meetings" (D. Nichols, personal communication to Y. S. Ben-Porath, April 25, 2004). Nichols's decision to ignore the literature on the RC Scales (with the lone exception of a chapter by Butcher & Miller [2006], that is critical of the RC Scales) explains his failure to take into account findings that address points raised in his critique (we mentioned earlier that Rogers et al. likewise did not consider such studies in their critique).

A second omission from Nichols's article is any detailed information about the "clinical" data set used in his analyses, a larger subset of the Caldwell collection than the one used by Rogers et al. This omission is not trivial. The scant information Nichols does provide—his characterization of it as a "clinical sample"—is inaccurate. We showed earlier that it is an ill-defined but clearly idiosyncratic collection of MMPI–2 protocols sent for processing to a particular scoring service.

The final problematic omission from Nichols's critique is the absence of extratest data. Nichols reports dozens of correlations between a large array of MMPI–2 scales, many unpublished and unknown. As we just noted, these analyses were conducted on a collection of MMPI–2 protocols that is not representative of any clinical or nonclinical population. Furthermore, although internal structural analyses can be informative, they are not sufficient to address Nichols's questions and challenges. In our monograph introducing the RC Scales, we observed that many existing data sets can be harvested for further exploration of the scales in samples of populations other than those reported in the monograph and, an important point, with additional external criteria. A serious effort to critically examine the RC Scales would surely include at least some collateral data.

We turn now to Nichols's four bones of contention with the RC Scales, beginning with the centerpiece of his critique that the construct and scale labeled Demoralization do not adequately represent the first factor of the MMPI–2.

Does Demoralization Define the First MMPI–2 Factor?

Throughout the Nichols review, one encounters questions about and criticism of our identification of Demoralization as the source of overly pervasive common variance in the Clinical Scales.

What Is Demoralization?

Introduction. In our RC Scale monograph, we noted that Tellegen (1985) described Demoralization as a general factor that inflates correlations in clinical inventories such as the MMPI between attributes that would be expected to be relatively independent. Tellegen's observations regarding the role of Demoralization in inflating correlations between measures of psychopathology were based on his study of the structure of mood. As often happens in scientific inquiry, other authors have reached similar conclusions from very different perspectives. Frank (1974a, 1974b) postulated that if disparate therapeutic modalities had common results, this might be because they address a problem common across patient types and presenting complaints. Frank (1974b) labeled this common factor Demoralization and observed that

> Only a small proportion of persons with psychopathology come to therapy; apparently something else must be added that interacts with their symptoms. This state of mind, which may be termed "demoralization," results from persistent failure to cope with internally or externally induced stresses. …

> Its characteristic features, not all of which need to be present in any one person, are feelings of impotence, isolation, and despair. (p. 271)

Thus, according to Frank (1974b), demoralization is common across various forms of psychopathology and unique to none.

Dohrenwend, Shrout, Egri, and Mendelson (1980) linked Frank's (1974a, 1974b) concept of demoralization to a common general distress dimension in psychiatric screening scales administered in community studies. Dohrenwend et al. observed that there was considerable phenotypic overlap between the common nonspecific distress dimension they attributed to demoralization and depression. However, Dohrenwend et al. also noted important differences between the constructs. Demoralization appeared to be closer to minor depressive disorder as described by Spitzer, Endicott, and Robbins (1978) in the research diagnostic criteria that preceded the *Diagnostic and Statistical Manual of Mental Disorders* (3rd ed. [*DSM–III*]; American Psychological Association, 1980). Other authors have likewise commented on the similarities and differences between demoralization and depression. de Figueiredo (1993) noted that whereas subjective distress and dysphoric affect may be manifest in both conditions, vegetative symptoms of depression such as poor sleep and appetite are less likely to accompany demoralization. de Figueiredo also observed that the dysphoric affect often found in individuals with medical disorders is more likely a product of demoralization than of depression, and Clarke and Kissane (2002) suggested that depression is characterized uniquely by anhedonia and that suicidal ideation is more likely associated with demoralization than it is with anhedonic depression.

In the psychological assessment literature, Joiner, Walker, Pettit, Perez, and Cukrowicz (2005) also noted the distinction between depression, which they associated uniquely with anhedonia, and the broader experience of depressed mood, which they related to general distress and demoralization. Joiner et al. concluded that "Depression is clearly more than just distress, demoralization, or depressed mood … depressed mood, although very common among those experiencing major depression, is not very specific to the syndrome; anhedonia, by contrast, is more unique to major depression" (p. 230).

To summarize, research conducted in a broad array of settings and with a variety of patients presenting with physical and psychological problems has identified demoralization as a ubiquitous, affect-laden dimension bearing phenotypic similarity to depression but distinguished from it because of the unique association of depression with anhedonia. Demoralization is characterized by unhappy, dysphoric mood, a sense of helplessness and inability to cope with one's current circumstances, and general dissatisfaction with one's condition. It is also associated with suicidality. Ample, converging evidence in the area of mood has confirmed that demoralization operates as a higher order dimension marked by sadness and unhappiness at its negatively valenced pole and that it is a common factor contributing to excessive correlations between measures of seemingly independent manifestations of psychopathology.

Having provided this background, we address Nichols's questions about Demoralization.

Why was a theoretical rather than an empirical approach taken to the development of the Demoralization construct and scale? In Nichols's (2006/this issue) words, "the appropriateness and advantages of the decision to embrace a theoretically rather than an empirically driven strategy for constructing Dem and RCd are doubtful" (p. 124). This is a false dichotomy. We readily acknowledged that the identification of Demoralization as the source of common variance in the Clinical Scales and our decisions about how to construct a measure of this attribute were guided by theoretical considerations. However, as described in detail in the monograph, our theoretical expectations yielded five explicit hypotheses about Demoralization, which were tested empirically, yielding an empirically supported and theoretically informed construct and scale.

Reliance on blind empiricism, as Nichols advocates and exemplifies in his subsequent recommendations of a variety of alternative first-factor markers, has yielded unsatisfactory results in MMPI and MMPI–2 research and interpretation. If interpretive guidelines focus almost exclusively on lists of empirical correlates and do not provide organizing conceptual principles, the MMPI–2 will be left out of the mainstream of current thinking about personality and psychopathology and be increasingly disconnected from important conceptual developments in the field. A recent addition to the MMPI–2, the Personality Psychopathology Five (PSY–5; Harkness, McNulty, & Ben-Porath, 1995) scales, represents an important effort to tie the instrument to current developments in the field of personality and psychopathology. Likewise, the RC Scales provide a link between the measures at the center of the test, the Clinical Scales, and current models of personality and psychopathology.

Sellbom and Ben-Porath (2005) compared correlational connections of the Clinical Scales and the RC Scales with the scales of the Multidimensional Personality Questionnaire (MPQ; Tellegen, in press), a broad-spectrum personality inventory that includes higher order scales assessing positive and negative emotionality. Consistent with the hierarchical findings of Tellegen, Watson, and Clark (1999a, 1999b) in the mood domain, Sellbom and Ben-Porath (2005) found that Demoralization mapped as expected on both Positive (reversed) and negative Emotionality, whereas RC2 and RC7 correlated more distinctively with Positive (reversed) and negative Emotionality, respectively. Demoralization (RCd), along with the scales Low Positive Emotions (RC2) and Dysfunctional Negative Emotions (RC7), link the MMPI–2 directly to contemporary conceptions of psychopathology such

as Watson's (2005) proposed hierarchical model of mood and anxiety disorders for the *DSM* (5th ed. [*DSM–V*]) and Joiner et al.'s (2005) recommendations for evidenced-based assessment of depression, discussed later.

Why were alternative first-factor markers not used? Nichols questions our choice of Demoralization as the common variance source we aimed to control as part of our restructuring effort and recommends a host of alternative variables derived primarily from two factor analyses of the entire MMPI or MMPI–2 item pool. Nichols (2006/this issue) states

> The *Manual* [sic] provides no discussion of the rationale for why the construct Dem should be preferred as a means for identifying and extracting problematic covariance from the basic Clinical Scales … over other readily available, empirically derived first-factor markers. (p. 124)

Two considerations led to our decision to explore, and ultimately to develop, a new measure. First, we needed to focus our effort on the Clinical Scales, which include just over half the MMPI–2 items, unlike the factor analyses cited by Nichols, which involved the entire MMPI/MMPI–2 item pool. Second, as we noted earlier, we preferred a theoretically guided process for identifying and developing a measure of this source of common variance (and focused as a first step on Clinical Scales 2 and 7) over blindly empirical factor analyses of the entire MMPI/MMPI–2 item pool such as those Nichols cites. In our RC Scale monograph, we made clear that we hypothesized Demoralization as a major source of common variance on the basis of arguments and evidence advanced by Tellegen (1985) that this, in itself, meaningful affective dimension underlies excessive overlap between measures of seemingly disparate forms of psychopathology.

We conceived of Demoralization as the equivalent to the pleasant-versus-unpleasant axis of Watson and Tellegen's (1985) mood model, a dimension of hedonic valence marked by adjectives such as *happy* and *content* on one pole and *sad* or *blue* on the other. Informed by this model and the empirical research it generated, we hypothesized that the items of Clinical Scales 2 and 7 would be the most promising starting point of our search for Demoralization markers. Through the procedures reported in the monograph, we tested and confirmed a series of hypotheses about Demoralization and arrived at a scale with item content clearly corresponding to the targeted dimension.

We did not require, nor did our procedures guarantee, that any of the scales they derived would be markedly different from any existing scales, including those that measure the first factor. For us, this was an empirical question. We fully expected our set of RC Scales to show substantial resemblances to existing scales (the similarity of RC1 and Hs is an obvious example from another content domain). However, we did not expect our new scales to be interchangeable with existing measures. Data reported by Nichols in Tables 4 and 5 of his critique show that RCd is indeed different from the various alternative markers he prefers. However, we note that Nichols does not indicate a preference for one of these alternatives. Although according to Nichols (2006/this issue), "it is an empirical question whether any of these alternative markers would be superior to Dem" (p. 129), he makes no effort to answer the question. Following are results of two of our own analyses, both providing additional empirical support for our perspective on RCd.

In one analysis, we set up a competition between sets of alternative first-factor markers. For rivals, we drew on three item sources: the 24 items in RCd, the "D" set, and two item sets Nichols favors, namely, the 83 first-factor markers reported by Johnson, Butcher, Null, and Johnson (1974); the "J" set; and the 135 markers reported by Waller (1999), the "W" set. Together, the three sources comprise 152 different items, with overlap patterns that allowed us to make an informative, near-full use of the total item pool (leaving out only 8 items). We created 18 nonoverlapping, 8-item scales representing five of the seven possible source combinations or subsets with either a single scale or several parallel scales—namely, (D), (DJW), (J) (JW), and (W)—where, for example, (JW) designates the subset of items in the Johnson et al. and Waller sets but not in the Demoralization set. In the case of parallel scales, we assigned items randomly to the scales under the constraint that in each scale the items be distributed throughout the booklet. We thus assembled the following 18 scales for the five subsets: D1, DJW1–2, J1, JW1–7, and W1–7 in which, for example, JW1–7 designates the seven 8-item scales assembled from the (J,W) subset of items. The one "impurity" among the 18 scales is the inclusion in D1 of two items in the (D,W) subset, which was done so D1 would also contain eight items.

We used principal factor analysis (PFA) here to capture shared variance only. If, as claimed by Nichols, RCd is a relatively poor marker of the MMPI–2 first factor compared to other first-factor scales, then the scales constructed from items contained in RCd will have a weaker association with the first extracted factor than scales derived from Nichols's preferred first-factor sources. Table 5 shows the loadings on the first unrotated common factor that were obtained from the inpatient and outpatient samples. These two similar loading patterns show the following features: substantial loadings for all configurations but overall the highest loadings on scales made up of items belonging to the Demoralization subsets, (D) and (D, J, W). In other words, RCd emerges here as the item source that best defines the first-factor domain.

In a second analysis, we calculated, separately for inpatient and outpatient samples, the overlap-corrected correlations between 16 MMPI–2 scales consisting of RCd and the 15 MMPI–2 Content scales. We used PFA again, and on the basis of the slope of the eigenvalues, we extracted and varimax rotated the first 3 factors. The results, reported in Table 6, show that within the broad spectrum of the 15 Con-

TABLE 5
Factor Loadings for 18 Independent First-Factor Markers in Outpatient and Inpatient Samples

	Factor Loadings	
Marker	*Outpatient*	*Inpatient*
D1	.86	.87
DJW1	.85	.87
DJW2	.84	.83
JW1	.78	.80
JW2	.76	.80
JW3	.84	.87
JW4	.85	.87
JW5	.84	.83
JW6	.83	.85
JW7	.82	.84
J1	.72	.75
W1	.71	.77
W2	.77	.80
W3	.76	.77
W4	.77	.79
W5	.79	.80
W6	.74	.77
W7	.75	.77

tent scales, RCd emerged in both samples as the strongest first-factor marker followed, as expected, by the various indicators of "internalizing" tendencies: the Depression, Work Interference, Low Self-Esteem, Negative Treatment Indicators, and Anxiety scales. These outcomes should help answer Nichols's question. In addition, we now consider more closely Nichols's misgivings about defining Demoralization as the core of the first factor, given its depressive content.

Why does Demoralization disproportionally favor "depressive" content? The centerpiece of Nichols's case against the RC Scales is his contention that Demoralization is too heavily laden with depressive content and not sufficiently balanced with markers of anxiety, tension, obsessiveness, and low self-esteem. In the following, we show why Nichols's criticism, based on MMPI–2 files devoid of collateral information and uninformed by relevant literature, is incorrect.

In Table 5 of his review, Nichols presents correlations of RCd with other scales, which supposedly show that Demoralization is "biased" toward depression. However, internal MMPI–2 analyses are complicated by item overlap, and it cannot be assumed that scale names accurately reflect the content of a scale or the underlying disposition. To adequately assess the alleged depressive bias, extratest criteria are needed. Sellbom, Ben-Porath, and Graham (2006); Simms et al. (2005); and Tellegen et al. (2003) all reported correlations showing that RCd is consistently more highly correlated with criteria related to depression than with those related to anxiety. However, do they suggest (as Nichols infers) that RCd is *too* closely related to depression?

In our earlier discussion of the demoralization construct, we referred to the converging evidence that the latent variable it refers to, which is responsible for excessive correlations between self-report measures of distinct forms of psychopathology, bears considerable phenotypic similarity to depression, more so than it does to anxiety-related difficulties, although it would be expected to be correlated with both these emotional disturbances. This conclusion is supported by studies in the areas of mood disorder, trauma, and behavioral medicine. The existing literature leads one to expect that a good measure of this source of common variance should, in fact, be associated with indications of depression and suicidality and to a lesser, but still substantial extent with markers of anxiety.

In Tables 7 and 8, we compare in our outpatient and inpatient samples the relevant external correlates (just mentioned) of RCd and the alternative first-factor markers proposed by Nichols. As expected, the first-factor markers were more highly correlated with depression and suicidality than with anxiety, but all were substantially correlated with the criteria included in these tables. In both samples, RCd showed stronger predictive validity than the alternatives. Coupled with our own analyses, including those we reported in the preceding section, these findings indicate that from a structural and an external validity perspective, RCd is a particularly strong marker of the major common source of variance of the MMPI–2 Clinical Scales and that none of Nichols's proposed alternatives performs better.

TABLE 6
Factor Loadings for Demoralization and the Content Scales in Outpatient and Inpatient Samples

	Factor Loadings	
Scale	*Outpatient*	*Inpatient*
Demoralization	.90	.90
Anxiety	.74	.78
Fears	.26	.19
Obsessiveness	.66	.65
Depression	.85	.89
Health Concerns	.47	.41
Bizarre Mentations	.24	.22
Cynicism	.40	.42
Anger	.24	.25
Antisocial Practices	.11	.20
Type A Behavior	.28	.24
Low Self-Esteem	.81	.79
Social Discomfort	.57	.63
Family Problems	.46	.44
Work Interference	.81	.82
Negative Treatment Indicator	.78	.80

TABLE 7
Correlations in the Outpatient Mental Health Sample of First-Factor Scales Discussed by Nichols With PDF Scales

	RCd	*A*	*JBF1*	*WF1*	*JBW72*	*AJBW*	*JBW37*
Depressed	.45	.39	.40	.41	.40	.39	.43
Pessimistic	.24	.18	.18	.21	.18	.19	.20
Insecure	.33	.29	.30	.31	.30	.29	.31
Anxious	.30	.28	.30	.31	.30	.28	.31
Obsessive–Compulsive	.16	.14	.15	.16	.15	.15	.16

Note. $N = 560$ to 669 for whom these five criteria were available. PDF = Patient Description Form; RCd = Demoralization; A = Anxiety; JBF1 = Johnson–Butcher (Johnson et al., 1984) first factor; WF1 = 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 8
Correlations Between First-Factor Markers Reported by Nichols and RRF Factors in the Psychiatric Inpatient Sample

RRF Scale	*RCd*	*A*	*JBF1*	*WF1*	*JBW72*	*AJBW*	*JBW37*
Depression	.28	.21	.20	.21	.21	.22	.22
Intrusive Ideation	.18	.18	.20	.22	.20	.18	.22
Suicidality	.32	.23	.23	.26	.24	.23	.24

Note. $N = 2{,}378$. RRF = Record Review Form; RCd = Demoralization; A = Anxiety; JBF1 = Johnson–Butcher (Johnson et al., 1984) first factor; WF1 = 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.

Does removal of Demoralization content make RC2 a less effective measure of depression than Clinical Scale 2? To support this contention, Nichols again uses internal correlational analyses (Table 7 of his article). However, as we noted earlier, there is ample evidence (cf. discussions by Clarke & Kissane, 2002; de Figuiredo, 1993; Joiner et al., 2005; Tellegen, 1985) that what is distinctive about depression is anhedonia or the absence of positive emotional experiences. Meehl (1975) previously also identified one form of depression as a manifestation of low hedonic capacity. Therefore, contrary to Nichols's complaint, removal of demoralization variance (as well as other extraneous sources) from the Clinical Scale 2 item pool achieved exactly what was intended, namely, a measure of the valid and distinctive affective component of that scale.

Referring to four RC2 seed items, Nichols (2006/this issue) writes "the heterogeneity of these four items creates reason for doubt that they could stand as a suitable core for anything" (p. 131), noting that each item marks a different dimension in Waller's (1999) factor analysis of the original MMPI item pool. We do not dwell on Nichols's inconsistency in criticizing the alleged heterogeneity of an RC Scale while defending the factorial complexity of the Clinical Scales and disregarding the fact that all four items are reverse-keyed members of Clinical Scale 2. Instead, we note that the correlations between the four seed items were all positive (with median phi coefficients of .19 and .27 and median tetrachoric correlations of .38 and .51 in our outpatient and inpatient samples, respectively). We also note that scale items can be factorially heterogeneous at a lower order level but cohere structurally at a higher level. In this case, the content of the four RC2 seed items—describing the self as pleasurably engaged, energetic, competent, and sociable—corresponds to the Wellbeing, Achievement, Social Potency, and Social Closeness primary traits that make up the higher order Positive Emotionality dimension (Tellegen & Waller, 1992). Sellbom and Ben-Porath (2005) found that RC2 as a whole is actually more highly correlated than Scale 2 with MPQ measures of these constructs. Similarly, Simms et al. (2005) showed RC2 to be more highly (negatively) correlated than Scale 2 with Positive Temperament in a community sample of veterans (but not psychological clinic patients).

Nichols claims elsewhere that Demoralization is actually too strongly represented in many of the RC Scale items. Specifically, Nichols suggests that the nonseed portions of the RC Scale items are too highly saturated with "first-factor variance," singling out RC2 as one of the more conspicuous offenders. Nichols's Table 3 compares the correlations of two kinds of RC subscales with Demoralization and other first-factor markers: subscales consisting of seed items and those consisting of nonseed items. The table shows consistently higher correlations with the nonseed scales.

For example, in the case of RC2, the correlations with Demoralization are .56 for the seed item subscale and .76 for the nonseed item counterpart. In our outpatient and inpatient samples, we found somewhat similar correlations of .50 versus .71 and .49 versus .73, respectively (resulting in an average ratio of the squared seed/nonseed correlations with RCd of .5). The direction of these correlational differences (resulting in a ratio below 1.0) is in itself unsurprising because the addition of nonseed items (needed to achieve sufficiently reliable measures) required flexible inclusion criteria. However, the difference of interest here is between seed and nonseed items, whereas the correlational differences shown

in Nichols's Table 3 are between seed and nonseed subscales. These reported subscale differences are therefore misleading in this context because the seed subscales are invariably shorter than their nonseed counterparts and therefore are less reliable, with weaker correlates, than if the two kinds of subscales had been of the same length. To remove this confound between item type and scale length, one could, for example, estimate the correlations with RCd for a seed-item subscale of the same length as its nonseed-item counterpart (using the Spearman–Brown formula). This adjustment increased the two just reported correlations of the RC2 seed-item subscale with RCd to .59 (increasing the earlier mentioned average ratio of squared correlations to a value closer to 1.0, namely, from .5 to .7).

Does removal of Demoralization from RC7 make it a less effective measure than Clinical Scale 7? Nichols asserts that the removal of demoralization from RC7 has resulted in a scale that although less confounded with depression, is also less sensitive to "psychasthenia" and more highly correlated with psychoticism and anger. Here too, Nichols bases his conclusions on analyses of internal correlations. To properly assess the success of the restructuring of Scale 7, a conceptual framework and appropriate collateral measures are needed.

Watson (2005) recently outlined a hierarchical model for rethinking mood and anxiety disorders for the *DSM–V*. This model builds on the tripartite model (Clark & Watson, 1991), which posits a general distress factor that includes both anxious and depressed mood and corresponds to Demoralization in the RC Scale framework. According to Watson's (2005) updated model, several mood and anxiety disorders fall under the general category of distress disorders (specifically, major depression; dysthymic, generalized anxiety; and posttraumatic stress disorders [PTSDs]), whereas other anxiety disorders (panic, agoraphobia, social phobia, specific phobias, and possibly obsessive–compulsive) are primarily fear related. This model yields empirically testable hypotheses for the RC Scales, namely, that measures associated primarily with the first group of disorders should be most highly correlated with RCd, whereas the second group should be associated with RC7.

Because the data on the clinical samples examined in our monograph do not include criteria allowing a test of these hypotheses, we report in Table 9 data from a college student sample (described in Appendix A) for which relevant collateral measures were available. As expected, general measures of depression and anxiety (the Beck Depression Inventory [Beck, Ward, Mendelson, Mock, & Erbaugh, 1961], the Internal States Scale [Bauer et al., 1991] Depression and Well-Being scales, and the State–Trait Personality Inventory Trait Anxiety scale [Spielberger, 1979]) were most strongly correlated with Demoralization. On the other hand, measures of fear-related disorders (the Fears Questionnaire [Marks & Mathews, 1979] and its subscales and the Obsessive–Compulsive Scale [Gibb, Bailey, Best, & Lambirth, 1983]) were consistently most highly correlated with RC7. Moreover, a measure of the Behavioral Inhibition System (Carver & White, 1994), associated in the literature with negative emotionality, was also considerably more strongly correlated with RC7 than with RC2. For both the fear and negative emotionality related criteria, RC2 and RC7 show a more differential pattern of correlations than Clinical Scales 2 and 7. These findings lend considerable support to the construct validity of three major affect-related RC Scales (RCd, RC2, and RC7) and provide another illustration of how the restructuring ties the MMPI–2 to contemporary models of psychopathology.

TABLE 9
Correlations Between Affective Clinical and RC Scales and Measures of Depression, Anxiety, Fear, and Psychoticism in a College Student Sample

Collateral Measures	*RCd*	*2*	*7*	*RC2*	*RC7*
Demoralization/distress related					
Beck Depression Inventory	.72	.63	.68	.54	.58
STPI trait anxiety	.78	.60	.75	.55	.65
ISS Depression	.55	.43	.51	.40	.42
ISS Well-Being	–.42	–.38	–.37	–.39	–.29
Negative emotionality related					
Behavioral Inhibition System	.42	.40	.47	.28	.51
STPI trait anger	.45	.18	.48	.16	.56
Fear related					
FQ total score	.29	.30	.38	.24	.41
FQ Agoraphobia	.21	.24	.30	.18	.32
FQ Blood/Injury Phobia	.15	.15	.21	.09	.24
FQ Social Phobia	.39	.37	.46	.33	.47
Obsessive–Compulsive Scale	.27	.21	.36	.11	.39
Thought disorder-related					
Perceptual Aberration Scale	.33	.15	.36	.16	.35
Magical Ideation Scale	.34	.12	.43	.08	.43

Correlations with two other criteria are reported in Table 9, measures of anger (the State–Trait Personality Inventory Trait Anger scale) and unusual thinking (the Magical Ideation Scale [Eckblad & Chapman, 1983] and the Perceptual Aberration Scale [Chapman, Chapman, & Raulin, 1978]). We included these variables to address Nichols's concern that RC7 is overly correlated with anger and psychoticism. With respect to the former, our data show that RC7 was indeed more highly correlated with anger than was Clinical Scale 7. However, rather than revealing a problem, this finding lends further support to the construct validity of RC7, a measure of a variety of dysfunctional negative emotions including anger. As to the measures of unusual thinking, our findings show that RC7 is not more highly correlated with the Magical Ideation and Perceptual Aberration Scales than is Clinical Scale 7 and that the modest correlations of both the Clinical and RC Scales with these thought disorder indicators are consistent with the correlations between these

indicators and Demoralization. Overall, the data in Table 9 show that RC7 is less highly correlated than is Clinical Scale 7 with measures of general distress and comparably or more highly correlated with indicators of fear-related symptoms and negative emotionality, thus providing evidence of the improved discriminant and convergent validity of this restructured version of Clinical Scale 7.

To summarize, Nichols's concerns about Demoralization reflect a neglect of the literature on this topic coupled with inadequate analyses of a problematic data set. Contrary to Nichols's assertions, there is ample evidence that Demoralization best represents the common factor inflating correlations between self-report measures of psychopathology; that it is appropriately associated with depressed mood; and that reduction of Demoralization variance has resulted in restructured versions of Clinical Scales 2 and 7 that better tap the distinctive manifestations of, respectively, depression and anxiety, and similarly of anhedonia and fears.

"Construct Drift"?

The focus of Nichols's second general concern is a problem he labels *construct drift.* Nichols (2006/this issue) appears to be concerned that in the final stage of constructing the RC Scales (described in detail in Tellegen et al.'s monograph), nonseed items were added that inadvertently contained similar "occult variances" (p. 133) so that "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" (p. 133). Nichols's concept of occult and alien scale variances is understandably opaque. However, Nichols clearly believes that items not originally scored on a Clinical Scale are not likely to be good measures of its distinctive core component. Nichols goes on to speculate about a number of possible areas of drift and faults us for failing to consider these. We acknowledge that we did not anticipate Nichols's drift scenarios, and space limitations preclude responding here to all the ones woven into his critique. We have noted that contrary to Nichols's concern, a stronger association of RC7 (than of Clinical Scale 7) with measures of anger reflects evidence of the improved convergent validity of this measure of dysfunctional negative emotions. We turn next to several other "drift"-related concerns.

RC9 Has Drifted in the Direction of Aggression

According to Nichols, the introduction of items connoting hostility, vindictiveness, and intimidation in RC9 is an example of drift. Although we did not have relevant criteria for RC9 in our initial samples, subsequent studies (e.g., Sellbom & Ben-Porath, 2005; Simms et al., 2005) have shown that RC9 is more highly correlated with measures of aggression than is Scale 9. One need not invoke construct drift to understand why this occurred. McKinley and Hathaway's (1944) description of the 24 patients who made up their criterion group for the development of Scale 9 reveals that they were characterized by elated mood, excitation, and flight of ideas but not irritability or more extreme agitation, which are among the current defining criteria for mania and hypomania in the *DSM* (4th ed. [*DSM–IV*]; American Psychiatric Association, 1994). Consequently, it is not surprising that the seed items for RC9 (which were limited to Scale 9 items) more closely resemble the euphoric mood and cognitive symptoms characteristic of the original criterion group. However, it is an empirical fact that items describing aggressive behavior were uniquely correlated with this seed scale and were therefore added to RC9. Moreover, their addition is consistent with current understanding of behavioral manifestations of hypomanic activation.

Nichols speculates that the addition of "hostile" item content to RC9 makes it a less effective predictor of hypomanic symptomatology. Examination of the existing literature indicates that he is mistaken. As noted, Simms et al. (2005) reported that RC9 was more highly correlated with aggression than is Clinical Scale 9. However, Simms et al. also reported that RC9 was more highly correlated than Scale 9 with disinhibition, impulsivity, and exhibitionism—other manifestations of hypomanic activation. In a study comparing the prediction of self-reported symptoms of psychopathology by the Clinical and RC Scales, Sellbom, Graham, et al. (2006) found that RC9 was more highly correlated with symptoms of bipolar disorder and mania than is Scale 9. These findings indicate that the restructuring of Scale 9 has resulted in improved predictive validity likely resulting from the addition of aggression-related items.

What About Scale 3?

Although Nichols includes a discussion of Scale 3 under the topic "construct drift," he acknowledges that the substantial changes in this scale were not accidental. Nichols asserts that Scale 3 was not so much restructured as it was replaced by RC3, which he finds highly redundant with the Content scale Cynicism (CYN) because they have 80% item overlap. On this last point, Nichols neglects to consider that item overlap is a bidirectional concept. Although 80% of the items of RC3 appear on CYN, only 52% of the CYN scale items appear on RC3. All RC3 items, including the 12 that also appear on CYN, describe a non-self-referential belief that others look out only for their own interests and are generally untrustworthy. By contrast, the remaining CYN items deal mainly with interpersonal suspiciousness, adding an element of self-referential concern that was left entirely out of RC3 and assigned instead to RC6. This is one of several ways in which the RC Scales differentiate phenomena that are confounded in the Content scales.[5]

[5]Other examples include Bizarre Mentation (BIZ), which confounds the self-referential persecutory ideation of RC6 and other

Nichols cites Butcher and Miller (2006) to support his conclusion that the change in Scale 3 would impose significant limits on the application of the RC Scales to medical, chronic pain, and personal injury assessments. This opinion appears predicated on the assumption that Scale 3 measures something uniquely relevant to these assessments, presumably in relation to symptoms associated with conversion disorder. However, although there is abundant lore surrounding Scale 3, there are virtually no empirical data to support its purported validity as a measure of somatic symptomatology beyond what is assessed by Scale 1. For example, in their study of chronic pain patients, Keller and Butcher (1991) found no unique correlates for Scale 3 in men or women.

Studies have shown that scores on Scale 3 are associated with increased somatic complaining and somatization, a prediction of correlates entirely redundant with Scale 1 predictions. For example, in a study of community mental health outpatients, Graham, Ben-Porath, and McNulty (1999) found no somatic correlates of Scale 3 that were not also associated to the same degree or more strongly with Scale 1. Consistent with these correlational patterns, Scale 1 and Scale 3 scores are often both elevated in individuals who present with somatic complaints. This is the origin of the so-called Conversion V profile or the 13/31 code type. Some MMPI–2 sources (e.g., Greene, 2000) have recommended different interpretations of the 1-3/3-1 code type depending on the relative elevation of the two scales (i.e., more somatization if 1 is higher than 3 and more hysterical features if 3 is higher than 1). Most interpretive guides recommend emphasizing hysterical or conversion symptoms if the T score on Scale 3 exceeds 79. Both of these highly inferential interpretive strategies are necessitated by the heterogeneity of Scale 3.

Scale 1 is relatively homogeneous, its item content focusing primarily on somatic complaints. Scale 3 has four distinct content areas: somatic complaints, demoralization, disavowal of cynicism, and denial of social discomfort. In the recursive restructuring process of identifying distinctive features (reviewed shortly in more detail), the content domain of somatic complaints had to be assigned to the restructured version of Scale 1 being its one major non-Demoralization component. And because introversion-related content did not emerge as a dominant feature (it is being attended to in work currently in progress), disavowal of cynicism was identified as the major distinctive component of the scale. Because responses to this item content are negatively correlated with somatic complaining, demoralization, and psychopathology in general, the scoring key was reversed, resulting in a restructured scale labeled Cynicism.

abberant experiences assessed by RC8, and Antisocial Practices, which confounds antisocial behavior as assessed by RC4 and cynicism, which is measured by RC3.

Like others (e.g., Butcher, 2006), Nichols appears to be concerned about what Tellegen et al. (2003) called the "dissolution" of Scale 3 qua scale. In this context, Butcher (2006) stated

> For a patient who has a high Hy score on the Hy scale [sic] (or a code type with Hy as a member) the RC3 is not likely to be prominent and is not likely to be a factor in the interpretive process. (p. 13)

These critics are particularly alarmed over the negligible and even slightly negative correlations between this scale and RC3. For example, in the outpatient mental health validation sample included in the Tellegen et al. RC scale monograph, the correlation between Scale 3 and RC3 is –.05. In the inpatient mental health sample, the correlation is –.13. However, although there are compelling reasons (discussed shortly in the context of syndromal assessment) not to mix the discordant manifest content of Scale 3 into one measure, major components of Scale 3 are well represented in the RC Scales. As we mentioned earlier, the main source of variance in Scale 3 is avowal of somatic concerns—the RC scale for examining this type of problem being RC1 and not RC3. Correlations between Scale 3 and RC1 in the two samples just mentioned are .74 and .66, respectively, magnitudes consistent with correlations between most of the Clinical Scales and their restructured counterparts. To the extent that the combination of somatic concerns and disavowal of cynicism, associated with elevated scores on Scale 3, is meaningful, it can be identified by considering together scores on RC1 and RC3. Individuals who produce elevated scores on RC1 and below-average scores on RC3 present with that combination of self-reports. Indeed, a composite formed by subtracting the score on RC3 from the RC1 score yields an even higher correlation (than RC1 alone) with Scale 3: .79 and .76 in the outpatient and inpatient samples, respectively. Thus, reports of the demise of Scale 3 as a composite have been greatly exaggerated. Its largest components are well represented in the RC Scales.

Empirical Evidence of No Drift

Nichols concludes his discussion of "construct drift" by observing that we present no evidence in our monograph that the RC Scales did not drift. Although claiming that the task of providing such evidence is "straightforward," Nichols fails to undertake it himself. In the following, we report the results of analyses designed to answer the question: Do the RC Scales measure the distinctive core components of the Clinical Scales?

The goal of the restructuring project was to identify the major non-Demoralization component of each Clinical Scale, which in the ideal case would be clearly distinctive from all the others. However, success was by no means assured. It required, first of all, that each Clinical scale indeed did contain a major and distinctive non-Demoralization core

component (mdc). As we reported in our monograph, for some of the Clinical Scales the identification of such a component was straightforward, but for other cases it was less so. The factorial complexity of Clinical Scales 3 and 6, in particular, required a more indirect and recursive approach. First, Somatic Complaints was designated the mdc of Scale 1, being that scale's only major non-Demoralization component. This constrained us to designate Cynicism as the mdc of Scale 3 (because Scale 3's other major non-Demoralization component was Somatic Complaints, already designated as the mdc of Scale 1). This in turn left us no choice but to identify Persecutory Ideas as the mdc of Scale 6 (because Scale 6's other major component was Cynicism and had been designated the mdc of Scale 3).

This sometimes indirect approach did result in eight distinctive and meaningful Clinical scale dimensions. However, were these dimensions in fact representative of the major distinctive core features of the original Clinical Scales? A negative answer might set the stage for a variety of reactions including suggestions of "drift." On the other hand, a positive answer would mean not only that each Clinical Scale did in fact contain a major distinctive core but also that the RC Scales had captured these core components.

To arrive at an empirical answer, we carried out a variant of an analysis pioneered by Jackson (1989) and conducted the same analyses for two clinical data sets: one inpatient, one outpatient. We began with separate principal components analyses (PCA) of the Restructured Scales (excluding RCd) and of the eight corresponding Clinical Scales, which enabled us to extract as many dimensions as there were scales. Varimax rotations of the extracted eight components to simple structure generated for each set of scales a solution in which each scale was strongly represented by its own component.[6]

Next, we included the two sets of eight rotated components in a single 16-variable analysis in both samples, extracted eight dimensions, and again rotated to simple structure. In this case, we elected to extract common factors to avoid the inflated loadings to be expected here with PCA because of the small number of markers available for each dimension. Because our initial choice, PFA, did not converge, we used image factor analysis (IFA) to extract common factors, which generated factor loadings that corresponded well to the magnitudes of the observed correlations between the two scale sets.

Table 10 depicts the varimax-rotated IFA solutions in the two data sets and reveals very similar convergent/discriminant structures. In both solutions, each factor is anchored to one and only one Clinical scale component and its RC Scale counterpart such that each pair of components defines its own independent dimension. This pattern corroborates the idea that each RC Scale represents the major distinctive component of its parent Clinical Scale and disfavors the idea of "construct drift."

Syndromal Fidelity

Nichols expects the more homogeneous RC Scales to be found ill-suited for assessing the inherently heterogeneous syndromes targeted by the original Clinical Scales. Setting aside the long recognized limitations of the Clinical Scales themselves as diagnostic measures, we address in the following section basic methodological considerations bearing on the assessment of complex variables, which are neglected in Nichols's critique. We then review recent and new empirical findings disconfirming his expectations.

Methodological Considerations

In published writings and informal communications, Hathaway (1956, 1972b; McKinley & Hathaway, 1944) himself did not hesitate to discuss and criticize his own work. Especially significant in this context was Hathaway's evident concern over the overlap and high correlations between some of the Clinical Scales. This phenomenon did not come as a surprise to Hathaway. Hathaway considered the syndromes targeted by his scales to be multifaceted and to share facets and was not startled to find this symptomatic overlap reflected in overlapping scales.

However, Hathaway (1956, p. 104) did deem it desirable to try to "hold down" certain correlations, which confronted him with a dilemma. By allowing substantial overlap, Hathaway might achieve what Nichols calls the "syndromal fidelity" of the overlapping scales (i.e., increase their convergent validity) but would reduce their capacity for making reliable differentiations (decrease their discriminant validity). Conversely, by reducing overlap, Hathaway could decrease their convergent validity but would increase their discriminant validity. It is evident that Hathaway tried to optimize the simultaneous attainment of both these validity goals, relying on his subjective judgment and on sometimes making an admittedly arbitrary decision.

At the heart of Hathaway's dilemma was his commitment to what we may call a "one-scale, one-syndrome" strategy, that is, his determination to develop one scale for each of his nine targeted clinical syndromes. Application of this strategy im-

[6] Analyses of the set of eight RC Scales produced in both data sets the desired pattern of each scale being saliently linked to and represented by its own rotated component. A one-step rotation did not produce this result for the Clinical Scales, undoubtedly because of the very high correlations between several of the scales. We therefore took a still exploratory but stepwise approach. First, we only analyzed Clinical Scales 7 and 8 (the two most highly correlated scales) and rotated and saved the two principal components. Next, we analyzed these two components together with Clinical Scale 2, and now rotated and retained the three principal components. We then jointly analyzed these three and Clinical Scale 1, and retained the resulting four rotated components. Finally, we included these four components and the remaining four Clinical Scales in one analysis. On rotation of the obtained eight principal components, we found in both samples that each Clinical Scale clearly marked a different component.

TABLE 10
Factor Analysis of Orthogonalized Clinical Scales and RC Scales in Outpatient and Inpatient Samples

	Factor							
Scale	*1*	*2*	*3*	*4*	*5*	*6*	*7*	*8*
Hs_{or}	.91/.90	–.01/–.01	.05/.07	–.01/.00	–.03/–.03	.02/.01	.01/–.03	.00/.00
$RC1_{or}$	.92/.91	.06/.05	–.11/–.14	.01/.00	.04/.04	–.01/–.02	.03/.05	.04/.06
D_{or}	.06/.05	.69/.71	.03/.02	–.11/–.12	.03/–.02	–.05/–.02	–.04/–.04	–.12/–.08
$RC2_{or}$	–.02/–.03	.77/.80	–.07/–.09	.14/.18	.04/.04	.13/.13	.08/.08	–.09/–.09
Hy_{or}	10/.14	.08/.10	–.53/–.58	.03/.05	–.07/–.06	–.03/–.04	.10/.11	–.01/–.01
$RC3_{or}$	.08/.10	.07/.07	.62/.67	.09/.12	–.16/–.14	.04/.03	.15/.15	.06/.08
Pd_{or}	.02/.01	.11/.14	.07/.10	.68/.70	.06/.04	.02/–.04	–.07/–.08	–.01/–.01
$RC4_{or}$	–.02/–.01	–.08/–.11	–.02/–.05	.61/.66	–.01/.00	–.02/.05	.08/.04	.04/.05
Pa_{or}	.03/.04	.06/.05	–.19/–.17	–.02/.00	.70/.76	.08/.09	–.01/–.03	.02/.01
$RC6_{or}$	–.03/–.03	.01/–.03	.12/.11	.07/.06	.70/.76	–.08/–.10	.09/.14	.05/.01
Pt_{or}	–.01/–.02	.13/.13	.03/.01	–.01/.08	–.10/–.11	.77/.79	–.01/.00	.12/.13
$RC7_{or}$	.01/.01	–.05/–.04	.05/.06	.01/–.05	.08/.09	.75/.73	.06/.07	–.05/–.04
Sc_{or}	.02/.04	.08/.08	.18/.18	.11/.08	.10/.12	.05/.04	.78/.81	.01/.00
$RC8_{or}$	.01/–.02	–.03/–.03	–.13/–.13	–.07/–.09	.00/.01	.01/.03	.62/.59	.10/.12
Ma_{or}	.04/.05	–.09/–.10	.04/.05	.00/.00	.05/.02	–.06/–.07	.05/.05	.64/.72
$RC9_{or}$	–.01/–.01	–.05/–.07	.02/.02	.03/.05	.01/.00	.11/.13	.05/.07	.62/.66

Note. Factor loadings to the left of the slash are from an analysis of the psychiatric outpatient sample (N = 1,020); those to the right are from an analysis of the psychiatric inpatient sample (N = 2,378). Extraction method was image factoring; rotation method was varimax. Loadings > |.20| are underlined. RC = Restructured Clinical; or = orthogonalized.

plied the adoption of a univariate measurement model for assessing multidimensional syndromal entities, a basic model–target mismatch. Hathaway's selection of the one-scale, one-syndrome strategy is from a historical perspective entirely understandable. The field of applied measurement was still underdeveloped, and the success of the Strong Vocational Interest scales and pragmatic appeal of empirical keying must have been compelling. However, this was almost 70 years ago.

In light of contemporary measurement principles, the application of a univariate measurement model to a multivariate target is readily recognized as a mistake. Almost 40 years ago, Nunnally (1967) called attention to this model–target mismatch when he rejected as a "fallacy" the notion that a heterogeneous variable could "be predicted with *one* test, which to be effective must be heterogeneous in content" (italics in original, p. 248). Instead, Nunnally strongly advocated the use of relatively homogeneous scales in multiple regressions to achieve optimal prediction (as illustrated in our monograph).

Multiple-linear regression may go a long way toward optimally predicting complex clinical variables. However, the assessment of true syndromes requires going beyond the additive (or linear) measurement model. This is true even for additive measures that are dimensionally complex, that is, heterogeneous. Heterogeneous additive measures can be characterized as "disjunctive" or "compensatory," that is, as measures that allow different configurations of observations to add up to the same total score so that a high score on one component can in principle compensate for a low score on a different component. Examples of heterogeneous additive measures include linear multiple-regression estimates made up of a mix of heterogeneous components, which they often are. Among single scales, the MMPI–2 Clinical Scales 3 and 6 are particularly striking examples of heterogeneous additive measures.

Whether homogeneous or heterogeneous, univariate or multivariate, additive measurements cannot fully accommodate true syndromes. To infer the presence of a true syndrome requires the simultaneous presence, or conjunction, of distinctive attributes, which therefore need to be ascertained separately. To avoid another model–target mismatch, the assessment of syndromes would have to conform to a "conjunctive" or "multiple-hurdles" model, which is nonadditive, rather than to the disjunctive/compensatory additive model.

The traditional MMPI–2 code types can be said to have syndromal features. However, the long-standing use of the largely heterogenous Clinical Scales as the components of these code types is not optimal for syndromal assessments. Profile configurations defined by well-chosen more homogeneous scales should in principle achieve a more accurate recognition of distinctive syndromal response patterns. We believe the RC Scales will make it easier to explore true syndromal assessment. The first requirement of course is to recognize the special character of this type of assessment, a recognition lacking in Nichols's treatment of the topic despite numerous references to syndromes including "syndromal fidelity," "syndromal complexity," and the modeling of syndromes. By contrast, Ben-Porath (2006) recently noted the potential of broadening the scope of multivariate prediction beyond the additive model to include true syndromal models.

The one-scale, one-syndrome strategy resulted not only in a set of overly heterogeneous and overinclusive Clinical Scales but virtually ensured the familiar side effects of overlap and redundancy. Nichols's agreement that reducing overlap between the Clinical Scales is desirable is inconsistent

with his endorsement of its cause: the one-scale, one-syndrome strategy. It amounts to being against both the symptoms and the cure.

Empirical Findings

Based on his assumptions regarding "syndromal fidelity," Nichols clearly expects the Clinical Scales to outperform the RC Scales in the prediction of psychiatric diagnoses. Empirical data addressing this question have already been published. Simms et al. (2005) reported correlations between scores on the Clinical and RC Scales and Structured Clinical Interview Axis I *DSM–IV* disorders diagnoses (First, Spitzer, Gibbon, & Williams, 1997) in a sample of 564 community-dwelling military veterans. Simms et al.'s results show that the correlations of Clinical Scales 2, 7, and 8 with depressive disorders were indistinguishable (i.e., nondiscriminant), whereas RCd achieved the strongest convergent correlation with all of the scales analyzed, followed by RC2, which was more highly correlated with depressive disorder than were RC7 and RC8. The high correlation with RCd is expected because the depressive group included individuals diagnosed with dysthymia and depressive disorder not otherwise specified, which, as we discussed earlier, are more likely to be associated with Demoralization than with anhedonia.

The correlations of Clinical Scales 2, 7, and 8 with anxiety disorders were again indistinguishable in Simms et al. (2005). RCd was correlated at about the same level with this diagnosis followed by RC7 and RC2. Here too the correlation between anxiety disorder and RCd was expected because this disorder group included individuals with generalized anxiety disorder and PTSD. These conditions are in Watson's (2005) conception more similar to depressive disorders because they are associated with psychological distress, which in the RC Scale scheme is conceived of as Demoralization and is measured by RCd. Although RCd is more highly correlated than is RC7 with anxiety disorders, inclusion of the two diagnostic groups just mentioned and the finding that the demoralization-saturated Clinical Scales 7 and 8 are equally correlated with anxiety disorders suggests that this disorder group is characterized by a high degree of demoralization. Finally, Simms et al. (2005) reported that the RC Scales clearly outperformed the Clinical Scales in predicting substance-use diagnoses, and the two scale sets were comparable in predicting somatoform disorders.

In the mental health outpatient and inpatient data sets used in the analyses reported in our monograph, clinical diagnoses made independently of the MMPI–2 at intake were available for comparisons between the RC Scales and the Clinical Scales, which have not been reported before. The diagnoses were recorded by clinicians who interviewed the patients on admission to their respective facilities. For the outpatients, diagnoses were generated by intake workers (mainly master's-level counselors) at the community mental health center where these data were collected. For the inpatients, diagnoses were entered by the admitting psychiatrist following an intake interview. Figure 1 allows visual pairwise comparisons of the zero order correlations of the Clinical Scales and the corresponding RC Scales with these clinician-generated diagnoses. The display format of the figure provides an overall picture of the comparative performance of the two sets of scales in the two samples.

Points in Figure 1 that fall along the diagonal represent comparable validity coefficients. For example, Clinical Scale 7 and RC7 produced near-identical correlations with anxiety disorders and PTSD (.18 and .17, respectively) in the inpatient samples. Points that fall below the diagonal represent better predictive performance by the Clinical Scales; points above the diagonal are evidence that the RC Scales did better. Most of the comparisons clearly fall in this area, indicating that the overall validity picture favors the RC Scales.

Many of the correlations reported in Figure 1 are of rather low magnitude, probably reflecting two constraints. First, these correlations represent a comparison between patients with a given clinical diagnosis and the remaining individuals in the sample, many of whom likely had symptoms associated with these disorders without actually satisfying their diagnostic criteria. Second, they may reflect the limitations just discussed of using univariate measurements to predict complex (and possibly nonadditively defined syndromal) variables. As a partial corrective, we portray in Figure 2 the comparative performances of the two sets of scales when we let (strictly linear) multiple regression estimates optimally combine three scale scores to predict each of the diagnostic variables. As expected, the multiple correlations shown in Figure 2 show a general increase in the proportions of variance accounted for by the two sets of scales and especially, because they are decidedly less redundant, by the RC Scales. These empirical findings clearly do not bear out the advantage of the Clinical Scales over the RC Scales that Nichols expected in the syndromal area.

The Question of Redundancy

As discussed earlier, our goal was to develop a set of RC Scales that would represent the major distinctive core components of the Clinical scales. This is why in our monograph, we focused exclusively on the connections between the RC Scales and the Clinical Scales from which each was derived. Given our methods, it would have been entirely possible for one or more of the resulting scales to be essentially redundant with existing measures. This would not have compromised the restructuring project or detracted from its value. Although not relevant to the goals of the restructuring project, questions about possible redundancies are certainly legitimate. Nichols's way of addressing the issue is displayed in his Table 4 in which he presents correlations between the RC Scales and no fewer than 51 other MMPI–2 measures. Nichols (2006/this issue) concludes from these correlations that there are "extremely high levels of redundancy between

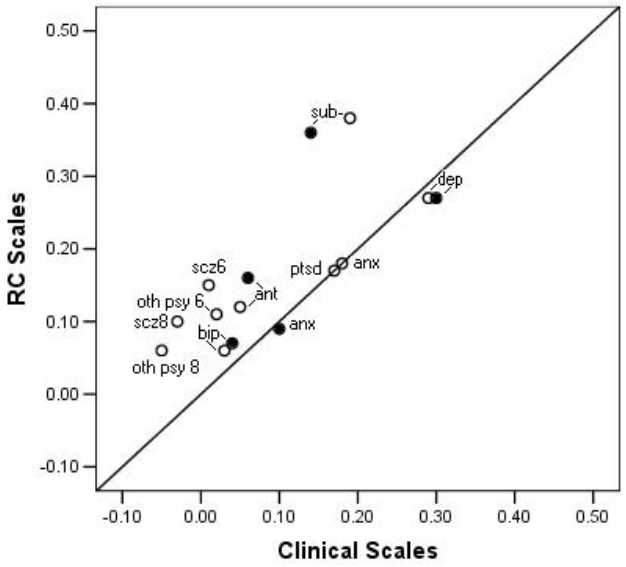

FIGURE 1 A comparison of bivariate correlations of Restructured Clinical (RC) Scales versus Clinical Scales in predicting admission diagnosis. Correlations to the left of the diagonal indicate more variance accounted for by the RC Scales; correlations to the right of the diagonal indicate more variance accounted for by the Clinical Scales. Blackened circles indicate outpatient sample correlates, whereas open circles indicate inpatient sample correlates. ant = antisocial personality disorder correlated with RC 4 and Scale 4 (*r*s = .12 [RC] and .05 [Clinical] for inpatient sample, and *r*s = .16 [RC] and .06 [Clinical] for outpatient sample); anx = anxiety disorders correlated with RC7 and Scale 7 (*r*s = .18 [RC] and .18 [Clinical] for inpatient sample, and *r*s = .09 [RC] and .10 [Clinical] for outpatient sample); bip = bipolar disorders correlated with RC9 and Scale 9 (*r*s = .06 [RC] and .03 [Clinical] for inpatient sample, and *r*s = .07 [RC] and .04 [Clinical] for outpatient sample); dep = depressive disorders correlated with RC2 and Scale 2 (*r*s = .27 [RC] and .29 [Clinical] for inpatient sample, and *r*s = .27 [RC] and .30 [Clinical] for outpatient sample); oth psy 6 = other psychotic disorders when correlated with RC6 and Scale 6 (*r*s = .11 [RC] and .02 [Clinical]); oth psy 8 = other psychotic disorders when correlated with RC8 and Scale 8 (*r*s = .06 [RC] and –.05 [Clinical]); ptsd = posttraumatic stress disorder correlated with RC7 and Scale 7 (*r*s = .17 [RC] and .17 [Clinical]); scz6 = schizophrenia when correlated with RC6 and Scale 6 (*r*s = .15 [RC] and .01 [Clinical]); scz8 = schizophrenia when correlated with RC8 and Scale 8 (*r*s = .10 [RC] and –.03 [Clinical]); and sub = substance-related disorders correlated with RC4 and Scale 4 (*r*s = .38 [RC] and .19 [Clinical] for inpatient sample, and *r*s = .36 [RC] and .14 [Clinical] for outpatient sample).

the RC Scales and [these] content-based scales that are already in wide use and have substantial empirical correlates" (p. 127).

First, even if this assertion were accurate, one could conclude that the nine RC Scales can efficiently replace 51 scales and would do so on the grounds of parsimony. However, many of the scales included in Nichols's Table 4 hardly qualify as being "in wide use and hav[ing] substantial empirical correlates." Nichols's redundancy argument led the action editor handling a recent validation study published in this journal (Sellbom, Ben-Porath, & Graham, 2006) to suggest that Sellbom, Ben-Porath, et al. compare the predictive validity of the RC Scales with that of a plausible alternative set of existing scales he had selected from Nichols's list. Complying with this request, Sellbom, Ben-Porath, et al. found consistently larger validity coefficients for the RC Scales than for the suggested proxies. The seeming inconsistency with Nichols's correlations may be attributed to his reliance on a problematic data set and his failure to correct for the spurious effects of correlated measurement error owing to item overlap, which resulted in inflated correlations (some of which reach or exceed the limit set by the reliabilities of the scales in question).

In summary, although a number of RC Scales resemble a number of existing scales, the individual RC Scales are not redundant with other MMPI–2 measures; and jointly, unlike the ad hoc collection of scales in Nichols's Table 4 or even carefully selected proxies, the nine RC Scales form a set of nonoverlapping and conceptually distinctive indicators.

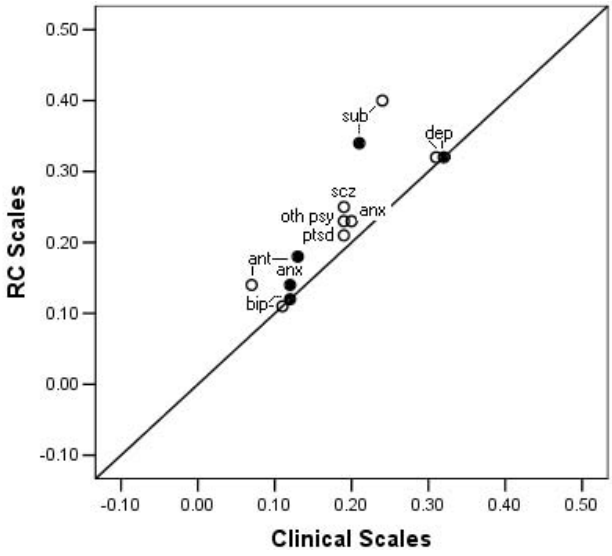

FIGURE 2 A comparison of Multiple *R*s for top three predictors of Restructured Clinical (RC) Scales versus Clinical Scales in predicting admission diagnosis. Correlations to the left of the diagonal indicate more variance accounted for by the RC Scales; correlations to the right of the diagonal indicate more variance accounted for by the Clinical Scales. Blackened circles indicate outpatient sample *R*s, whereas open circles indicate inpatient sample *R*s. ant = antisocial personality disorder (*R*s = .14 [RC] and .07 [Clinical] for inpatient sample, and *R*s = .18 [RC] and .13 [Clinical] for outpatient sample); anx = anxiety disorders (*R*s = .23 [RC] and .20 [Clinical] for inpatient sample, and *R*s = .14 [RC] and .12 [Clinical] for outpatient sample); bip = bipolar disorders (*R*s = .11 [RC] and .11 [Clinical] for inpatient sample, and *R*s = .12 [RC] and .12 [Clinical] for outpatient sample); dep = depressive disorders (*R*s = .32 [RC] and .31 [Clinical] for inpatient sample, and *R*s = .32 [RC] and .32 [Clinical] for outpatient sample); oth psy = other psychotic disorders (*R*s = .23 [RC] and .19 [Clinical]); ptsd = posttraumatic stress disorder (*R*s = .21 [RC] and .19 [Clinical]); scz = schizophrenia (*R*s = .25 [RC] and .19 [Clinical]); and sub = substance-related disorders (*R*s = .40 [RC] and .24 [Clinical] for inpatient sample, and *R*s = .34 [RC] and .21 [Clinical] for outpatient sample).

Omissions

Throughout his critique, Nichols expresses concerns about information he believes to be missing from the Tellegen et al. monograph that introduced the RC Scales. Nichols (2006/this issue) faults us for failing to refer to the "long-neglected analysis by Norman (1972)" (p. 123) criticizing the MMPI Clinical Scales and to Helmes and Reddon's (1993) critique from the perspective of Jackson's (1970, 1974) methodology. Nichols does so despite the direct relevance of our own partially critical reviews of the Clinical Scales and of Jackson's approach presented in our monograph and without telling how these omitted references would have informed readers of the monograph or even whether and how they have informed his own views.

Nichols's (2006/this issue, pp. 124–125) claims that there are gaps in our account of how the seed scale for RCd was derived reflect inattention to its sequential development: selection of 23 provisional Demoralization markers in Step 1, their use in the factor analyses in Step 2, the reported construction (which Nichols clearly missed) of a 19-item Demoralization seed scale in Step 3, and the development of the final 24-item RCd in Step 4. Each of these steps was reported in our monograph.

Nichols also alleges that we failed to address prior efforts to increase the independence of the Clinical Scales and singles out one developed by Finney (1968). Finney's procedure involved estimating a corrected T score following removal of unwanted variance from a scale. According to Nichols (2006/this issue), "The chief virtue of this method is that it leaves the scales themselves, their item composition and keying, and therefore their syndromal complexity intact" (pp. 136–137). However, Nichols neglects to include Finney's (1968) own appraisal of this approach:

> A drawback of all correction formulas is the relative or absolute increase in random error. If the given scores are highly correlated with the dimension to be removed, the formula will greatly multiply the random error. It is clear that having uncontaminated measures to begin with is far superior to removing contamination by any correction formula. (p. 1235)

We concur with Finney and note that the corrected T scores he described are in effect difference scores, known to be unreliable, which is Finney's point. We suggest that the RC Scales represent precisely the kind of solution Finney called for. In addition, they measure examined substantive dimensions instead of the unexamined residuals favored by Nichols.

Nichols (2006/this issue) deplores our failure to consider "that the RC Scales did not eliminate the problem of item overlap that afflicts the Clinical Scales" (p. 125). If we understand Nichols correctly, restructured scales containing items belonging to a Clinical Scale other than their parent scale are unacceptable. For example, informative somatic items included in both Clinical Scales 1 and 3 or cynicism items included in both Clinical Scales 3 and 6 would have to be dropped. Instead, the restructuring effort would aim for truncated Clinical Scales that contain only items unique to the scale in question, in other words, "pure" MMPI scales such as those Welsh (1952) created more than 50 years ago. Today, a mechanical approach along these lines would be recognized as unpromising: arbitrarily restrictive and substantively and structurally uninformed.

Nichols criticizes us for not explaining why in our factor analyses, we chose PCA over other extraction methods, phi coefficients over tetrachoric correlations, and orthogonal rotations over oblique rotations. Issues relevant to these choices have been extensively discussed but have not been resolved in the methodological literature. Space limitations preclude a detailed discussion of these issues here, but we offer the following brief observations. PCA can generate spuriously high loadings when the number of factor markers is very small. In our own comparisons of PCA with other extraction methods, this problem did not occur in analyses representative of our restructuring studies. For example, whenever enough marker variables were available for each factor, as has invariably been the case in our restructuring analyses, the magnitudes of differences between loadings obtained with PCA and with PFA have been negligible.

A problem sometimes encountered when phi coefficients are used in multifactor MMPI item factor analyses is the emergence of "difficulty factors." An example is the fragmentation of a bipolar MMPI somatic complaint dimension into a number of spurious separate factors, which occurred when phi coefficients were used but not when tetrachorics were analyzed (Waller, 1999). However, in our own analyses that relied on phi coefficients, but in which few dimensions were extracted, difficulty factors did not emerge as a problem and have regularly yielded bipolar factors including the clearly bipolar somatic complaint dimension represented by RC1. In addition, our direct factor analytic comparisons between phi coefficients and tetrachoric correlations have yielded consonant findings.

Finally, we have difficulty seeing why oblique rotations would be a serious option in a project that aims to pinpoint strong markers of maximally distinctive dimensions. Nichols again omits specifics.

A FINAL CONCERN

Rogers et al. and Nichols raise a number of concerns about the RC Scales. Unfortunately, Rogers et al.'s and Nichols's data analyses were conducted with a problematic collection of MMPI–2 protocols that does not represent a meaningful population and, in the case of Rogers et al.'s elevation frequency analyses, yields results that led Rogers et al. to arrive at incorrect conclusions about the RC Scales. We hope that the publication of these results and conclusions will not re-

sult in a perpetuation of errors in the literature. We also hope that readers of Nichols's critique of the RC Scales will not be influenced or distracted by the sheer number of objections he voices but rather will attend to their substance, which does not take into account basic psychometric principles, the existing literature on the RC Scales, and current knowledge of personality and psychopathology.

Our defense of the RC Scales against these criticisms should not be construed as reflecting the belief that they are beyond improvements. Like any other psychometric devices, the RC Scales are imperfect, and we fully expect that in due course, a growing body of conceptually informed empirical research will identify ways to improve them.

ACKNOWLEDGMENTS

We thank Roger Greene for making available a random subset of the Caldwell MMPI–2 collection and providing clarifying information about this data set and Beverly Kaemmer for her unstinting support and helpful editorial suggestions.

REFERENCES

American Educational Research Association, American Psychological Association, & National Council on Measurement in Education (AERA/APA/NCME). (1999). *Standards for educational and psychological testing.* Washington, DC: American Educational Research Association.

American Psychiatric Association. (1980). *Diagnostic and statistical manual of mental disorders* (3rd ed.). Washington, DC: Author.

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. (2002). A comparison of MMPI–2 validity in African American and Caucasian psychiatric inpatients. *Psychological Assessment, 14,* 3–15.

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

Baer, R. A., & Miller, J. (2002). Underreporting of psychopathology on the MMPI–2: A meta-analytic review. *Psychological Assessment, 14,* 16–26.

Baer, R. A., Wetter, M. W., & Berry, D. T. (1992). Detection of underreporting of psychopathology on the MMPI: A meta-analysis. *Clinical Psychology Review, 12,* 509–525.

Baer, R. A., Wetter, M. W., Nichols, D. S., Greene, R., & Berry, D. T. R. (1995). Sensitivity of MMPI–2 validity scales to underreporting of symptoms. *Psychological Assessment, 7,* 419–423.

Bauer, M. S., Crits-Christoph, P., Ball, W. A., Dewees, E., McAllister, T., Alahi, P., et al. (1991). Independent assessment of manic and depressive symptoms by self-rating. *Archives of General Psychiatry, 48,* 807–812.

Beck, A. T., Ward, C. H., Mendelson, M., Mock, J., & Erbaugh, J. (1961). An inventory for measuring depression. *Archives of General Psychiatry, 12,* 57–62.

Ben-Porath, Y. S. (2006). Differentiating normal from abnormal personalities with the MMPI–2. In S. Strack (Ed.), *Differentiating normal from abnormal personalities* (pp. 337–381). New York: Springer.

Black, M. S., Forbey, J. D., Ben-Porath, Y. S., Graham, J. R., McNulty, J. L., Anderson, S. V., et al. (2004). Using the Minnesota Multiphasic Personality Inventory–2 (MMPI–2) to detect psychological distress and dysfunction in a state correctional setting. *Criminal Justice and Behavior, 31,* 734–751.

Block, J. (1965). *The challenge of response sets.* New York: Appleton-Century-Crofts.

Butcher, J. N. (2006). Pathways to MMPI–2 use: A practitioner's guide to test usage in diverse settings. In J. N. Butcher (Ed.), *MMPI–2: A practitioner's guide* (pp. 3–13). Washington, DC: American Psychological Association.

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., Ben-Porath, Y. S., Tellegen, A., Dahlstrom, W. G., & Kaemmer, B. (2001). *Minnesota Multiphasic Personality Inventory–2 (MMPI–2): Manual for administration, scoring and interpretation* (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., pp. 140–166). New York: Wiley.

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

Carver, C. S., & White, T. L. (1994). Behavioral inhibition, behavioral activation, and affective responses to impending reward and punishment: The BIS/BAS scales. *Journal of Personality and Social Psychology, 67,* 319–333.

Chapman, L. J., Chapman, J. P., & Raulin, M. L. (1978). Body-image aberration in schizophrenia. *Journal of Abnormal Psychology, 87,* 399–407.

Clark, L. A., & Watson, D. (1991). Tripartite model of anxiety and depression: Psychometric evidence and taxonomic implications. *Journal of Abnormal Psychology, 100,* 316–336.

Clarke, D. M., & Kissane, D. W. (2002). Demoralization: Its phenomenology and importance. *Australian and New Zealand Journal of Psychiatry, 36,* 733–742.

Crowne, D. P., & Marlowe, D. (1960). A new scale of social desirability independent of psychopathology. *Journal of Consulting Psychology, 24,* 349–354.

de Figueiredo, J. M. (1993). Depression and demoralization: Phenomenological differences and research perspectives. *Comprehensive Psychiatry, 34,* 308–311.

Dohrenwend, B. P., Shrout, P. E., Egri, G., & Mendelsohn, F. S. (1980). Nonspecific psychological distress and other dimensions of psychopathology. *Archives of General Psychiatry, 37,* 1229–1236.

Eckblad, M. L., & Chapman, L. J. (1983). Magical ideation as an indicator of schizotypy. *Journal of Consulting and Clinical Psychology, 51,* 215–225.

Edwards, A. L. (1957). Social desirability and probability of endorsement of items in the interpersonal check list. *Journal of Abnormal & Social Psychology, 55,* 394–396.

Fenig, S., & Levav, I. (1991). Demoralization and social supports among Holocaust survivors. *Journal of Nervous and Mental Disease, 179,* 167–172.

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 for Axis I DSM–IV Disorders—Patient Edition (SCID–I/P).* New York: Biometrics Research Department, New York State Psychiatric Institute.

Forbey, J. D., & Ben-Porath, Y. S. (in press). Computerized adaptive personality testing: A review and illustration with the MMPI–2 Computerized Adaptive Version. *Psychological Assessment.*

Frank, J. D. (1974a). Common features of psychotherapies and their patients. *Psychotherapy and Psychosomatics, 24,* 368–371.

Frank, J. D. (1974b). Psychotherapy: The restoration of morale. *American Journal of Psychiatry, 131,* 271–274.

Frank, J. D. (1985). Further thoughts on the anti-demoralization hypothesis of psychotherapeutic effectiveness. *Integrative Psychiatry, 3,* 17–20.

Gibb, G., Bailey, J., Best, R., & Lambirth, T. (1983). The measurement of obsessive compulsive personality. *Educational and Psychological Measurement, 43,* 1233–1237.

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

Greene, R. L. (2000). *The MMPI–2: An interpretive manual* (2nd ed.). Needham Heights, MA: Allyn & Bacon.

Harkness, A. R., McNulty, J. L., & Ben Porath, Y. S. (1995). The Personality Psychopathology Five (PSY–5): Constructs and MMPI–2 scales. *Psychological Assessment, 7,* 104–114.

Hathaway, S. R. (1956). Scales 5 (Masculinity-Femininity), 6 (Paranoia), and 8 (Schizophrenia). In G. S. Wlesh & W. G. Dahlstrom (Eds.), *Basic readings on the MMPI in psychology and medicine* (pp. 104–111). Minneapolis: University of Minnesota Press.

Hathaway, S. R. (1972a). Foreword. In W. G. Dahlstrom, G. S. Welsh, & L. E. Dahlstrom (Eds.), *An MMPI handbook: Volume 1. Clinical interpretation* (pp. xiii–xiv). Minneapolis: University of Minnesota Press.

Hathaway, S. R. (1972b). Where have we gone wrong? The mystery of the missing progress. In J. N. Butcher (Ed.), *Objective personality assessment: Changing perspectives* (pp. 21–43). New York: Academic.

Hathaway, S. R., & McKinley, J. C. (1943). *The Minnesota Multiphasic Personality Inventory.* Minneapolis: University of Minnesota Press.

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 practices in clinical and community psychology* (Vol. 2, pp. 60–96). New York: Academic.

Jackson, D. N. (1971). The dynamics of structured personality tests: 1971. *Psychological Bulletin, 78,* 229–248.

Jackson, D. N. (1974). *Personality Research Form manual.* Goshen, NY: Research Psychologists Press.

Jackson, D. N. (1989) *Basic Personality Inventory manual.* Port Huron, MI: Sigma Assessment Systems.

Jackson, D. N., & Messick, S. (1962). Response styles on the MMPI: Comparison of clinical and normal samples. *Journal of Abnormal and Social Psychology, 65,* 285–299.

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.

Joiner, T. E., Jr., Walker, R. L., Pettit, J. W., Perez, M., & Cukrowicz, K. C. (2005). Evidence-based assessment of depression in adults. *Psychological Assessment, 17,* 267–277.

Keller, L. S., & Butcher, J. N. (1991). *Use of the MMPI–2 with chronic pain patients.* Minneapolis: University of Minnesota Press.

Marks, I. M., & Mathews, A. M. (1979). Brief standard self-rating for phobic patients. *Behaviour Research and Therapy, 23,* 563–569.

McKinley, J. C., & Hathaway, S. R. (1944). A Multiphasic Personality Schedule (Minnesota): 5. Hysteria, hypomania, and psychopathic deviate. *Journal of Applied Psychology, 28,* 153–174.

Meehl, P. E. (1975). Hedonic capacity: Some conjectures. *Bulletin of the Menninger Clinic, 39,* 295–307.

Morey, L. C. (1991). *The Personality Assessment Inventory: Professional manual.* Odessa, CA: Psychological Assessment Resources.

Nichols, D. (2006/this issue). The trials of separating bath water from baby: A review and critique of the MMPI–2 Restructured Clinical Scales. *Journal of Personality Assessment, 87,* 121–138.

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.

Nunally, J. C. (1967). *Psychometric theory.* New York: Mcgraw-Hill.

Quirk, S. W., Christiansen, N. D., Wagner, S. H., & McNulty, J. L. (2003). On the usefulness of measures of normal personality for clinical assessment: Evidence of the incremental validity of the Revised NEO Personality Inventory. *Psychological Assessment, 15,* 311–325.

Rogers, R., Sewell, K. W., Harrison, K. S., & Jordan, M. J. (2006/this issue). The MMPI–2 Restructured Clinical Scales: A paradigmatic shift in scale development. *Journal of Personality Assessment, 87,* 139–147.

Rogers, R., Sewell, K. W., Martin, M. A., & Vitacco, M. J. (2003). Detection of feigned mental disorders: A meta-analysis of the MMPI–2 and malingering. *Assessment, 10,* 160–177.

Sellbom, M., & Ben-Porath, Y. S. (2005). Mapping the MMPI–2 Restructured Clinical scales onto normal personality traits: Evidence of construct validity. *Journal of Personality Assessment, 85,* 179–187.

Sellbom, M., Ben-Porath, Y. S., & Graham, J. R. (2006). Correlates of the MMPI–2 Restructured Clinical (RC) scales in a college counseling setting. *Journal of Personality Assessment, 86,* 89–99.

Sellbom, M., Ben-Porath, Y. S., Graham, J. R., Arbisi, P. A., & Bagby, R. M. (2005). Susceptibility of the MMPI–2 Clinical, Restructured Clinical (RC), and Content scales to overreporting and underreporting. *Assessment, 12,* 79–85.

Sellbom, M., Ben-Porath, Y. S., Lilienfeld, S. O., Patrick, C. J., & Graham, J. R. (2005). Assessing psychopathic personality traits with the MMPI–2. *Journal of Personality Assessment, 85,* 334–343.

Sellbom, M., Graham, J. R., & Schenk, P. W. (2006). Incremental validity of the MMPI–2 Restructured Clinical (RC) scales in a private practice sample. *Journal of Personality Assessment, 86,* 196–205.

Simms, L. J., Casillas, A., Clark, L. A., Watson, D., & Doebbeling, B. N. (2005). Psychometric evaluation of the Restructured Clinical scales of the MMPI–2. *Psychological Assessment, 17,* 345–358.

Spielberger, C. D. (1979). *Preliminary manual for the State–Trait Personality Inventory (STPI).* Tampa: University of South Florida.

Spitzer, R. L., Endicott, J., & Robins, E. (1978). Research diagnostic criteria: Rationale and reliability. *Archives of General Psychiatry, 35,* 773–782.

Tellegen, A. (1985). Structures of mood and personality and their relevance to assessing anxiety, with an emphasis on self-report. In A. Hussain Tuma, & J. D. Maser (Eds.), *Anxiety and the anxiety disorders* (pp. 681–706). Hillsdale, NJ: Lawrence Erlbaum Associates, Inc.

Tellegen, A. (in press). *Manual for the Multidimensional Personality Questionnaire.* Minneapolis: University of Minnesota Press.

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

Tellegen, A., & Waller, N. G. (1992). *Exploring personality through test construction: Development of the Multi-Dimensional Personality Questionnaire (MPQ).* Unpublished manuscript, Department of Psychology, University of Minnesota.

Tellegen, A., Watson, D., & Clark, L. A. (1999a). Further support for a hierarchical model of affect: Reply to Green and Salovey. *Psychological Science, 10,* 307–309.

Tellegen, A., Watson, D., & Clark, L. A. (1999b). On the dimensional and hierarchical structure of affect. *Psychological Science, 10,* 297–303.

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.

Watson, D. (2005). Rethinking the mood and anxiety disorders: A quantitative hierarchical model for *DSM–V. Journal of Abnormal Psychology, 114,* 522–536.

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

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

Wiggins, J. S. (1959). Interrelationships among MMPI measures of dissimulation under standard and social desirability instruction. *Journal of Consulting Psychology, 23,* 419–427.

APPENDIX A
Basic Demographics for Samples Used in This Article

Demographic	*CMHC*[a]	*Inpatient*[b]	*VARC*[c]	*ODOC*[d]	*College*[e]	*Normative*[f]	*Caldwell Subsample*[g]
Men (*N*)	410	1,837	1,235	35,982	634	1,138	4,448
Women (*N*)	610	541	0	7,113	958	1,462	4,564
Age range (years)	18 to 76	18 to 90	23 to 76	18 to 93	18 to 53	18 to 85	18 to 92
Age							
M	32.94	40.6	44.51	29.36	19.58	41.04	38.99
SD	10.27	14.1	4.82	8.81	3.32	15.29	11.55
Ethnicity (%)							
White	79	82	35	48	89	81	NA
African American	18	12	60	51	8	12	NA
Other	3	6	5	1	3	7	NA

Note. All demographics are for the final sample after participants with invalid Minnesota Mulitphasic Personality Inventory–2 protocols have been excluded. CMHC = Community Mental Health Center (outpatient); VARC = Cleveland Veterans Administration Substance Abuse Recovery Unit; ODOC = Ohio Department of Rehabilitation and Correction; NA = not available .

APPENDIX B
External Criterion Measures Used in This Article

Sample	*Criterion Measures*
CMHC	Patient Description Form (Graham, Ben-Porath, & McNulty, 1999)
Inpatient	Record Review Form (Arbisi, Ben-Porath, & McNulty, 2002)
College	Beck Depression Inventory (Beck, Ward, Mendelson, Mock, & Erbaugh, 1961), State–Trait Personality Inventory (Spielberger, 1979), Internal States Scale (Bauer et al., 1991), Behavioral Inhibition and Activation System Scales (Carver & White, 1994), Fears Questionnaire (Marks & Mathews, 1979), Obsessive–Compulsive Scale (Gibb, Bailey, Best, & Lambirth, 1983), Perceptual Aberration Scale (Chapman, Chapman, & Raulin, 1978), and Magical Ideation Scale (Eckblad & Chapman, 1983)

Note. CMHC = Community Mental Health Center.

Auke Tellegen
Department of Psychology
Elliott Hall
University of Minnesota
75 East River Road
Minneapolis, MN 55455
Email: telle001@umn.edu

Received March 21, 2006
Revised April 28, 2006