JOURNAL OF PERSONALITY ASSESSMENT, *87*(2), 139–147
Copyright © 2006, Lawrence Erlbaum Associates, Inc.

The MMPI–2 Restructured Clinical Scales: A Paradigmatic Shift in Scale Development

Richard Rogers, Kenneth W. Sewell, Kimberly S. Harrison, and Mandy J. Jordan

Department of Psychology
University of North Texas

Tellegen et al. (2003) proposed fundamental changes in MMPI–2 (Butcher, Dahlstrom, Graham, Tellegen, & Kaemmer, 1989) scale development by discarding empirical scale development in favor of construct validation via Jackson's (1970) sequential system of scale development. As a result of their efforts, a general distress factor (Demoralization) was identified and 8 Restructured Clinical (RC) Scales were developed. Using 7,330 clinical cases from Caldwell's (1997) data set, in this study, we sought to cross-validate the MMPI–2 RC Scales. Scale homogeneity was confirmed with high alpha coefficients and interitem correlations in the expected range. We also achieved a major objective of reducing interscale correlations. In replicating Tellegen et al.'s principal components analysis, we achieved a high concordance for 6 of the 8 RC Scales. We critically examine these results in light of Jackson's construct validation. We discuss the clinical usefulness of the MMPI–2 RC Scales within the context of current and future research.

The original development of the Minnesota Multiphasic Personality Inventory (MMPI; Hathaway & McKinley, 1940) constituted a "radical departure from preexisting personality questionnaires" (Dahlstrom, Welsh, & Dahlstrom, 1972, p. 4) in its scale development. Dissatisfied with the transparency of face-valid measures, Hathaway and McKinley (1940) relied on empirical scale development for their development of MMPI scales. Irrespective of content and theory, scale items were selected solely based on their discriminability between criterion groups. More important, no assumptions were made about the veridicality of the clients' self-ascriptions regarding either their personality characteristics or past experiences.

Tellegen et al. (2003) pioneered the second "radical departure" in MMPI history when they discarded the time-honored empirical scale development for Clinical Scales. Although content analysis has long played a role in MMPI and MMPI–2 (Butcher, Dahlstrom, Graham, Tellegen, & Kaemmer, 1989) interpretation, we consider this change to be radical because Tellegen et al. (2003; see chap. 5) deemphasized decades of clinical interpretation based on empirically keyed scales.[1] With the RC Scales, Tellegen et al. completely revamped MMPI–2 interpretation, rebuilding the Clinical Scales via construct validation that combined factor-analytic methods with Jackson's (1970) construct-oriented scale development. To provide a contextual understanding of Tellegen et al., in this introduction, we briefly outline the original MMPI conceptualization of scale development and subsequent changes introduced with the MMPI–2 (Butcher, Williams, Graham, Tellegen, & Kaemmer, 1989). In the introduction, we then examine the paradigmatic shift spurred by Tellegen et al.'s contributions.

ORIGINAL SCALES AND MMPI–2 REFINEMENTS

Dahlstrom et al. (1972) described the development of MMPI items with its intentional emphasis on ambiguity. Ambiguity was essential to "permit differential interpretations" by clients that would facilitate "dependable personological interpretations" (Dahlstrom et al., 1972, p. 6). However, the MMPI items could not be entirely irrelevant in their content because clients might begin to question the test's purpose and validity.

[1]A more categorical version of this statement was criticized during the review process as being too strong. It is true that Tellegen et al. (2003) "recommend that the RC Scales be used to help clarify the interpretation of Clinical Scale findings" (p. 59). Rather than rely on any specific quote, our "strong" conclusion is drawn from Tellegen et al.'s extensive discussion of the MMPI–2 Clinical Scales weaknesses and comparatively, the psychometric strengths of RC Scales.

With their pool of ambiguous items, Hathaway and McKinley (1940) applied a rudimentary form of empirical keying in the MMPI scale development. Scale items were selected that merely differentiated persons with specific clinical conditions from presumably unimpaired community controls. As a direct result of this methodology, items common to general distress or nonspecific to clinical conditions were often chosen. For instance, only 9 of 33 Scale 1 items (27.3%) were not overlapping with other Clinical Scales. Consequently, the MMPI scales have substantial item overlap, which limits their discriminability between clinical conditions.

The temptation to interpret the content of MMPI items was increasingly difficult to resist (Graham, 1987), despite item ambiguity and extensive research on empirically derived scales. Among the early researchers, Marks, Seeman, and Haller (1974) strenuously objected to any effort that associated MMPI item content with clinical reality. Marks et al. (1974) concluded that MMPI items "could not be taken at face value nor [sic] as substitutes for observed behavior" (p. 6). Marks et al. reasoned that item content could not be believed because (a) many clients lack insight about their functioning and behavior and (b) many others will go to great lengths to conceal their true attitudes and experiences.

Publication of the MMPI–2 (Butcher, Williams, et al., 1989) represented an appreciable shift away from empirically derived scales as the sole method of clinical interpretation. Although early MMPI manuals (e.g., Hathaway & McKinley, 1967) avoided any reference to Content scales, the MMPI–2 manual featured them prominently. As stated by Butcher, Williams, et al. (1989), "Content interpretation of the MMPI has been widely accepted as an adjunct to traditional empirical scale interpretation" (p. 41). Through nomination by three experts, scale refinement, and rational review (see Butcher, Graham, Williams, & Ben-Porath, 1989), 15 content scales were included in the MMPI–2 test manual. Subsequent research (e.g., Archer, Aiduk, Griffin, & Elkins, 1996; Ben-Porath, Butcher, & Graham, 1991; Munley, Busby, & Jaynes, 1997; but see also Jackson, Fraboni, & Helmes, 1997) has generally supported the use of content validity to augment traditional Clinical Scales. However, the level of incremental validity is sometimes very modest. Archer et al. (1996) found the average change in R^2 was 2.8% for male inpatients and 2.9% for female inpatients when using the Symptom Checklist–90–Revised (Derogatis, 1977) as a criterion. How such minor increments translate into more accurate interpretations is unclear. A potential confound in combining these scales is their nonindependence with substantial overlap between MMPI–2 Clinical and Content scales.

MMPI-2 RESTRUCTURED SCALES

Tellegen et al. proposed the de-emphasis of empirically derived MMPI–2 scales and interpretations. Like previous commentators (e.g., Friedman, Webb, & Lewak, 1989), Tellegen et al. correctly observed that MMPI criterion groups were often small, single-site samples based on imprecise inclusion criteria. For the Clinical Scales, they also noted that the high interscale correlations and lack of theoretically based scales may limit the clinical usefulness of the MMPI–2.

We agree with Tellegen et al.'s criticisms of early MMPI research that used empirical keying. However, these criticisms apply to the early studies and not necessarily to empirically keyed validation per se. In contrast to early studies, refinements in empirical keying had focused on the statistically based methods of discrimination (Golden, Sawicki, & Franzen, 1984). In addition, the problems with item overlap would be largely resolved by "near-neighbor" comparisons (Frances et al., 1991) that distinguish between closely related diagnoses. For instance, a near-neighbor comparison on the MMPI–2 would select items that differentiated paranoid (Scale 6) from other psychotic characteristics (Scale 8). These fully refined scales would likely have more homogeneity than the traditional MMPI–2 Clinical Scales. Their theoretical relevance would be dependent on the coherence of the clinical constructs underlying each of the individual MMPI–2 scales.

Rather than refining MMPI–2 Clinical Scales via empirical keying, Tellegen et al. opted to develop new scales by using exploratory factor analysis supplemented by Jackson's (1970) sequential approach to construct validity. Although acknowledging Jackson's sequential procedure involved a formal analysis of social desirability, Tellegen et al. omitted this step in their scale development. Tellegen et al. chose principal component analysis (PCA) over other exploratory factor analyses. In treating the each MMPI scale as a separate and independent measure, the PCA maximizes the variance accounted for and makes latent constructs more easily identifiable. Because Tellegen et al. were building a model rather than estimating the amount of variance for each scale, the PCA was a reasonable choice over common factor methods (Gorsuch, 1990; Velicer & Jackson, 1990).

Tellegen et al. postulated a "general distress" factor labeled Demoralization, which is composed primarily of negative affect. Relying on Watson and Tellegen (1985), negative affect is a second-order dimension that includes sadness, anger, disgust, fear, shame, guilt, and contempt. Tellegen et al. hypothesized that Scales 2 and 7 best captured Demoralization; the first factor of a PCA on these scales was used to operationally define Demoralization. Tellegen et al.'s objective was to remove this nonspecific Demoralization from each of the Clinical Scales. To accomplish this goal, PCAs were performed separately for each of the eight Clinical Scales augmented with the Demoralization items.

The overall goal was to establish Restructured Clinical (RC) Scales that were (a) nonoverlapping, (b) distinct from nonspecific distress (Demoralization), and (c) relevant to the clinical construct measured by each scale. Although complicated, a general understanding of Tellegen et al.'s statistical

methods is critical to evaluating their results and the contributions of this cross-validation effort. First, each Clinical Scale was examined by one or more PCAs with the goal of identifying and removing the Demoralization factor. The remaining items were further refined by removing the overlapping items assessing more than one clinical construct and those whose item–scale correlations were less than .20. These 109 core items were used to form the "seed" scales. These items were further tested with four separate samples; an additional 10 items were excluded because they failed to obtain consistently the highest loading on their identified component (making special accommodations for Scale 9, which had insufficient seed items). This process yielded 99 items on the seed scales.

The seed scales became the nucleus of the RC Scales. Additional items were added to individual RC Scales if they met three criteria: the "general criterion" (i.e., correlated most highly with the corresponding seed scale), the "convergent criterion" (i.e., correlated with the corresponding seed scale above a specified value), and the "discrimination criterion" (i.e., correlated weakly with noncorresponding seed scales). The specified values for convergent and discriminant criteria were scale specific (see Tellegen et al., 2003). Several additional modifications improved internal consistency of the RC Scales, increased external validity, and maximized the differences between RC2 and RC7. The final result was a 24-item Demoralization scale and 168 items assigned to one of the eight RC Scales corresponding to MMPI–2 Scales 1, 2, 3, 4, 6, 7, 8, and 9.

THE CURRENT STUDY

In this investigation, we sought to replicate the Tellegen et al. RC Scales with a very large clinical sample of MMPI–2 protocols. In establishing their reliability analysis, Tellegen et al. evaluated alpha coefficients for males and females across individual clinical samples (see Tellegen et al., 2003, Table 4.4, p. 35). To test the homogeneity of the MMPI–2 RC Scales, we examined alpha coefficients and interitem correlations separately by gender. In addition, our confidence in the reliability of individual scales (Anastasi, 1988) relies on standard errors of measurement (*SEM*). For comparative purposes, we evaluated *SEM* by gender for the RC Scales.

The primary purpose of this study was the cross-validation of the Demoralization, seed, and added items for each of the RC Scales. We were particularly interested in whether we could replicate Tellegen et al.'s findings for seed items, which form the nucleus of the RC Scales. In addition, the RC Scales sought to improve clinical interpretation through the reduction of interscale correlations. We used two methods of comparison for RC interscale correlations: (a) general (i.e., the average correlation with all other Clinical Scales) and (b) near neighbor (i.e., where applicable, average correlations for scales within the neurotic triad or psychotic tetrad).

METHOD

In this investigation, we used a large subset of approximately 8,000 MMPI–2 protocols from the Caldwell (1997) database to replicate and evaluate the Tellegen et al. RC Scales. Although a larger data set was previously used by Greene (2000) to provide critical percentages and prototypic code types, these analyses of RC and traditional scales are entirely original. The University of North Texas Institutional Review Board exempted this use of an anonymous archival data set from further review.

Sample

The sample was entirely composed of MMPI–2 raw data from clinical settings that used the Caldwell report for MMPI–2 scoring and interpretations. Excluding most adolescents and invalid protocols (see following), the final sample was composed of 3,601 (49.1%) male clients and 3,729 (50.9%) female clients with an overall average age of 38.81 years (*SD* = 11.72; range = 17 to 92). Regarding marital status, data were available on 6,822. Of these, a total of 2,754 (37.6%) of the clients were married at the time of assessment, with 2,248 (30.7%) single, 1,711 (23.3%) divorced/separated, and 109 (1.5%) widowed. Consistent with MMPI–2 answer sheets, demographic data on ethnicity are not available.

Measure

The MMPI–2 is an established and well-validated multiscale inventory consisting of 567 true–false items. Psychometric data on its reliability, validity, and interpretation are readily available (Butcher, Graham, Ben-Porath, Tellegen, & Dahlstrom, 2001; Graham, 2000; Greene, 2000).

Procedure

R. Greene (personal communication, January 20, 2004) randomly selected 8,000 clinical cases from his computerized version of the Caldwell database. These data were completely anonymous without names or personal identifiers. Excluded from this database were child custody cases; these referrals markedly underreported their psychopathology (Strong, Greene, Hoppe, Johnston, & Olesen, 1999) and could skew these findings.

We refined the Caldwell data by age parameters and missing values. In selecting age parameters, our goal was to make the sample as representative as possible of the MMPI–2 clinical applications. Although recent textbooks (e.g., Greene, 2000) have warned about applying the MMPI–2 to any persons below age 18, 17-year-olds were included in core research on the MMPI–2's validation. These studies, including Butcher and other leading MMPI–2 researchers, have tested 17-year-olds in research

on college (Ben Porath & Butcher, 1989; Butcher, Graham, Dahlstrom, & Bowman, 1990; Roper, Ben-Porath, & Butcher, 1995), military (Butcher, Jeffrey, et al., 1990), outpatient (Crossman, Casey, & Reilley, 1994), and forensic (Humphrey & Dahlstrom, 1995) settings. These studies had raised no concern about the validity of the MMPI–2 with persons aged 17 and older and have appeared confident in their clinical conclusions. Because we were not using MMPI–2 norms, we included 17-year-olds even though they represented a small proportion of the entire data set (n = 65 or 0.8%). Similarly, we included a small number of clients who were slightly older than the upper range (i.e., 84) of the normative sample; six clients ranged in age from 86 to 92. We excluded younger adolescents (n = 138) and those with missing or likely incorrect ages (n = 127). We also refined the sample by eliminating profiles with missing or misscored items. Given the item-based analyses required for the study, we adopted a stringent criterion for missing values: We excluded 405 cases with 10 or more MMPI–2 items unscored or double scored. Following the Tellegen et al. procedure, we did not attempt to eliminate "invalid" profiles (e.g., for inconsistencies or overreporting). As a result of these refinements, a total of 7,330 cases were retained for analysis.

RESULTS

Reliability and *SEM*

Tellegen et al. reported excellent alpha coefficients for the RC Scales in comparison to the Clinical Scales; unweighted averages across clinical samples consistently exceed .80. As summarized in Table 1, in our sample, RC Scales had excellent alpha coefficients that were comparable between males (M = .87) and females (M = .85) and almost identical to the Tellegen et al. results. The interitem correlations were mostly in the optimal range between .15 to .50 (Clark & Watson, 1995). The only exception was RC9, which fell slightly below this range (see Table 1).

The American Psychological Association, in coauthoring the official test standards (American Education Research Association, American Psychological Association, & National Council on Measurement in Education, 1999), require that *SEM* be used to evaluate to the confidence and reliability for individual scales (see also Anastasi, 1988). Because these data were omitted from the Tellegen et al. test manual, we calculated *SEM* estimates for Tellegen's data by combining clinical and normative data (see Table 2). The results on Tellegen's data indicate generally low *SEM* estimates for raw scores with the exception of RC9. In our cross-validation of Tellegen et al., these *SEM* estimates were comparable for males (M = 1.39) and for females (M = 1.37). When computed as T scores, the *SEM* for RC Scales averaged close to 5 points.

Cross-Validation of the RC Scales

Following the Tellegen et al. methods, we performed PCAs for each of the RC Scales. We replicated the Demoralization factors for each PCA with substantial and unique loadings (i.e., items defined the target factor and were not cross-loaded; *M*s ranged from .52 to .66). As summarized in Table 3, the proportion of loadings consistent from Tellegen et al. to this study was high (> 70.0%) for five RC Scales.

The replication of seed items is crucial to construct validation of the RC Scales. Three scales (RC1, RC3, and RC6) vir-

TABLE 1
Reliability Analysis for Tellegen et al. Clinical Samples and the Current Data

	Tellegen et al. Samples		*Current Data*			
			Male		*Female*	
Scale	*Male* α	*Female* α	α	*M Interitem r*	α	*M Interitem r*
RCd	.94	.94	.95	.45	.95	.42
RC1	.87	.89	.92	.30	.92	.29
RC2	.84	.85	.85	.26	.86	.26
RC3	.84	.82	.85	.28	.83	.25
RC4	.83	.80	.82	.18	.79	.16
RC6	.84	.82	.85	.26	.83	.23
RC7	.89	.89	.91	.30	.90	.28
RC8	.84	.83	.86	.26	.83	.23
RC9	.83	.81	.81	.14	.78	.12
Average	.86	.85	.87	.27	.85	.25

Note. Alphas for Tellegen et al. (2003) Clinical Samples are nonweighted averages across three clinical samples for males and two clinical samples for females. RC = Restructured Clinical.

TABLE 2
Standard Errors of Measurement for MMPI–2 RC Scales

	Tellegen et al. Samples		*Current Data*			
			Male		*Female*	
Scale	*Male Raw*	*Female Raw*	*Raw*	*T*	*Raw*	*T*
RCd	1.25	1.42	1.37	3.49	1.43	3.30
RC1	1.19	1.41	0.74	5.33	0.78	4.77
RC2	1.22	1.21	1.58	5.77	1.59	6.48
RC3	1.60	1.50	1.55	4.32	1.56	4.46
RC4	1.59	1.43	1.13	4.93	0.94	5.18
RC6	0.67	0.64	0.99	5.44	0.97	5.71
RC7	1.50	1.66	1.76	4.34	1.84	4.28
RC8	0.92	1.10	1.20	5.24	1.19	5.33
RC9	2.17	2.13	2.17	4.27	2.05	4.51
Average	1.35	1.39	1.39	4.79	1.37	4.89

Note. Standard error of measurement values for the Tellegen et al. (2003) samples are estimates based on (a) alphas averaged across three different clinical samples plus a normative sample for males (two clinical samples plus a normative sample for women) and (b) scale standard deviations that were reported for the normative sample (Tellegen et al., 2003, p. 101). MMPI–2 = Minnesota Multiphasic Personality Inventory–2; RC = Restructured Clinical.

TABLE 3
Cross-Validation of the MMPI–2 Restructured Clinical Scales

		Demoralization		*Seed Items*		*Added Items*	
Scale	*Factors*	*M Load*	*Proportion*	*M Load*	*Proportion*	*M Load*	*Proportion*
RC1	2	.61	0.79	.53	0.93	.52	0.80
RC2	2	.52	0.50	.47	0.75	.55	0.75
RC3	3	.54	0.75	.54	1.00	—[a]	—[a]
RC4	3	.54	0.58	.48	0.60	.40	0.50
RC6	3	.64	1.00	.52	1.00	.53	1.00
RC7	2	.52	0.46	.45	0.71	.39	0.00[b]
RC8	2	.59	0.92	.43	0.67	.44	0.50
RC9[c]	2	.66	0.96	.40	0.50	.40	0.50
M	2.38	.58	0.75	.48	0.77	.46	0.58

Note. MMPI–2 = Minnesota Multiphasic Personality Inventory–2; Factors = the number of factors established by Tellegen et al. (2003) and cross-validated in the current study; Proportion = the proportion of unique and replicated loadings (greater than or equal to .40); Added Items = those items from the original clinical scales that were added back onto to RC Scale in "Step 4" (Tellegen et al., 2003, p. 19).
[a]No items from the original Scale 3 were added back onto RC3. [b]This value is zero because the single added item loaded at .39 (just below the criterion of .40). [c]RC9 analyses included items 189 and 226 as Seed Items (see Tellegen et al., 2003, p. 18) even though these items did not appear on the original Scale 9.

tually replicated Tellegen et al.'s loadings for seed items with a very high (> 90%) proportion of correspondence. Two scales (RC2 and RC7) had a high (> 70%) proportion of unique and replicated findings. Of the remaining three scales, RC9 produced relatively modest results, with 50% of its items being replicated. We also used the PCAs to cross-validate Tellegen's added items. The proportion of unique and replicated loadings were much more variable; correspondence ranged from modest to exceptionally high.[2]

A major objective of Tellegen's RC Scales was to improve their discriminant validity by reducing the high intercorrelations reported with the traditional MMPI–2 scales. We generated interscale correlations for both RC and traditional scales (see Table 4). Using .70 as the benchmark for highly correlated scales (i.e., sharing close to 50% or more of the variance), we examined RC and traditional scales. Almost no RC Scales were highly correlated; only one of 28 comparisons (3.6%) met this benchmark. In marked contrast, nearly half (46.4%) of the traditional scales were highly correlated. In Tellegen et al.'s samples, high RC intercorrelations were rarely observed and varied from 0.0% to 4.6% (see Tellegen et al., 2003, Tables 4.8 to 4.12). Clearly, Tellegen et al. achieved the general goal of substantially reducing the intercorrelations of MMPI–2 Clinical Scales.

As a further analysis, we examined near-neighbor comparisons for RC and traditional scales. On the neurotic triad (see Table 4), traditional scales lacked discriminability because of their very high intercorrelations ($M\ r$ = .81). In marked contrast, the RC Scales for the neurotic triad were much less correlated ($M\ r$ = .47) and allowed for discrimination between scales. A similar though less obvious pattern emerged for near-neighbor comparisons on the psychotic tetrad between traditional ($M\ r$ = .62) and RC ($M\ r$ = .56) scales.

Interpretations

A major limitation of MMPI–2 interpretations is the number of clinical profiles with no clinical elevations (i.e., within-normal-limits [WNL] profiles). According to Greene (2000), this code type is the most common and accounts for nearly one third (30.1%) of clinical profiles. We examined the RC profiles to assess the frequency of WNL profiles. They occurred with 44.8% of male and 40.4% of female clients. In contrast, the clinical profiles produced lower percentages of WNL profiles, with 36.1% for male and 30.8% for female clients. These results for the RC profiles indicate that almost half of clinically referred cases assessed will have WNL profiles for which any clinical interpretations would likely be minimal.

MMPI–2 interpretations are often guided by code types that are buttressed by extensive clinical research. Tellegen et al. suggested that the RC profiles would not be congruent with traditional profiles. Moreover, Tellegen et al.'s eight case studies revealed markedly different profiles and code types. To investigate this issue, we examined code types for RC and traditional profiles and found 20 code types that were represented by at least 1% of the cases for which code types could be derived.[3] As shown in Table 5, RC code types were predominated by elevations on RC1, RC2, and RC6; they accounted for 17 of 20 (85.0%) code types. Concordance of RC and traditional code types was generally modest (M = 14.6%) and varied from none (3 to 6) to approximately 50% (6 to 8).

[2]The result for the RC7 scale should not be interpreted because only one added item (.39) was too close to the benchmark of .40.

[3]For purposes of clarity, we eliminated cases with triple ties or ties on the second elevation.

TABLE 4
Intercorrelations of MMPI–2 RC Scales (Above the Diagonal) and MMPI–2 Clinical Scales (Below the Diagonal)

	Scale								*RC Scale Correlations*		*Traditional Scale Correlations*	
Scale	*1*	*2*	*3*	*4*	*6*	*7*	*8*	*9*	*Near*	*Distant*	*Near*	*Distant*
1	—	.64	.40	.15	.52	.65	.60	.19	.52	.42	.83	.61
2	.84	—	.29	.19	.38	.61	.40	–.03	.47	.31	.80	.58
3	.82	.76	—	.38	.52	.61	.52	.54	.35	.51	.79	.41
4	.55	.59	.41	—	.30	.41	.37	.53	—	—	—	—
6	.61	.61	.51	.66	—	.61	.69	.37	.56	.43	.60	.60
7	.79	.82	.55	.72	.70	—	.70	.48	.60	.57	.69	.72
8	.79	.75	.52	.75	.75	.93	—	.48	.62	.47	.74	.70
9	.32	.13	.08	.43	.36	.43	.53	—	.44	.31	.44	.24

Note. MMPI–2 = Minnesota Multiphasic Personality Inventory–2; RC = Restructured Clinical; For correlations: near = the average intercorrelations within the neurotic triad (Scales 1, 2, and 3) and within the psychotic tetrad (Scales 6, 7, 8, and 9); distant = average of the remaining intercorrelations.

TABLE 5
Common MMPI–2 RC Codetypes (≥ 1.0%) With and Without Demoralization and Their Concordance With Traditional MMPI–2 Code Types

		Demoralization		
Code Type	*Total %*	*% Elevated*	*% Unelevated*	*Concordance*
1–2	19.4	14.2	5.2	9.4
1	11.0	1.6	9.4	4.1
1–6	9.8	4.6	5.2	6.6
6	8.3	0.4	7.9	16.0
2	7.6	2.9	4.7	5.0
1–8	5.9	4.5	1.4	7.7
4	5.1	0.6	4.5	17.8
2–6	3.7	2.6	1.1	32.2
1–7	3.5	2.9	0.5	3.6
6–8	3.3	2.1	1.2	51.9
2–7	3.0	2.9	0.1	33.9
2–4	2.0	1.5	0.6	16.3
4–6	1.7	0.7	1.0	27.3
3–6	1.5	0.5	1.1	0.0
3	1.5	0.1	1.3	1.7
1–4	1.4	0.8	0.6	1.8
1–3	1.3	0.8	0.6	18.9
6–7	1.2	1.0	0.2	18.4
2–8	1.2	1.1	0.1	17.4
8	1.0	0.1	0.9	2.4

Note: Code type percentages are based on the 3,995 cases that had codifiable elevations; concordance values refer to the percentage of RC Scale code types that had identical corresponding code types on traditional MMPI–2 scales. MMPI–2 = Minnesota Multiphasic Personality Inventory–2; RC = Restructured Clinical.

DISCUSSION

The discussion is organized into three major sections, beginning with a detailed examination of the MMPI–2 RC Scales that uses Jackson's (1970) principles of scale construction as its conceptual framework. We selectively raise several professional issues regarding the RC test interpretation. We end with a brief section summarizing the major conclusions and offering directions for subsequent research.

MMPI–2 RC Scale Construction

Jackson's (1970) groundbreaking work on scale construction takes strong issue with empirically derived scales. On this point, he asserted "No longer is it necessary, or even desirable, to take refuge from our psychological ignorance by relying on an external criterion" (p. 63). Instead, Jackson proposed that test development be guided by four interrelated principles: (a) preeminence of psychological theory, (b) suppression of response-style variance, (c) scale homogeneity and generalizability, and (d) convergent and discriminant validity. The Tellegen et al. development of MMPI–2 RC Scales clearly embraced three of these principles.

From a theoretical perspective, Tellegen et al. posited a nonspecific distress (Demoralization) factor as common to clinical populations. For specific clinical constructs, Tellegen et al. remained faithful to the MMPI–2 traditional scales but sought to refine these constructs via factor analytic work. Tellegen et al. elected to eliminate somatic issues from Scale 3 and retain only those items assessing cynicism. In addition, items related to cynical distrust were dropped from Scale 6, which became much more focused on ideas of persecution. The revised Scale 8 removed content areas related to poor familial relationships, impulse control, issues of self-worth and self-identity. Tellegen et al. concentrated RC8 on aberrant experiences. Overall, the general theoretical perspective of removing general distress and narrowing the clinical constructs appears sound and logical. More open to question are the individual decisions to eliminate important facets of MMPI–2 Clinical Scales. Should problems with identity and alienation, dropped from Scale 8, be constituted as separate scales? In light of the fundamental changes in MMPI–2 Clinical Scales, we would submit that the theoretical contributions of these facets be closely evaluated before they are summarily removed.

Tellegen et al. tested Demoralization and specific RC Scales via a series of PCAs on four separate samples. As a very positive finding, RC1, RC3, and RC6 were nearly replicated in this cross-validation. With one exception, the remaining scales had moderate to high correspondence from the original validation to this cross-validation. Only RC9 evidenced a modest level of correspondence (50%). This scale will likely require further validation and refinement. In general, the Tellegen et al. model of general distress and refined clinical constructs were cross-validated.

The reconceptualization of the MMPI–2 failed to consider adequately Jackson's (1970) second principle involving the suppression of response variance. As observed by Jackson (1970), "response styles often will determine major portions of the variance" (p. 64), thereby frustrating efforts to make reliable and accurate diagnostic distinctions. To address this important matter, Jackson proposed that items be systematically screened for social desirability before their inclusion on any personality scale. Only those items that correlated appreciably higher with the clinical construct than desirability should be retained. Although Tellegen et al. briefly mentioned the problems with response styles, they did not follow Jackson's principle or his specific procedures for the suppression of response variance.

Test items are more easily manipulated when their content is obvious and transparent to clients (Rogers & Bender, 2003). Tellegen et al. openly acknowledged the "relative transparent content of RC Scale items" (p. 54) and their greater susceptibility to response distortion. Given the complete omission of Jackson's (1970) second principle in item selection, purists may question whether RC Scales should be described as based on Jackson's scale validation. Pragmatists will likely want to test the desirability of RC items and assess the potential consequences of omitting Jackson's principle and concomitant methods from this restructuring of the MMPI–2.

Tellegen et al.'s efforts for improving scale homogeneity were clearly fulfilled. Alpha coefficients were high to exceptionally high in both the original validation and this cross-validation. More important, these alphas were consistent across gender. Interitem correlations are used to examine whether individual items are contributing to scale homogeneity. Again, the results were very promising, with little item redundancy (i.e., all *r*s ≤.45). Only RC9 evidenced lower than expected correlations in this cross-validation (i.e., *M r*s of .12 and .14), but the alpha coefficients remained satisfactory.

The Tellegen et al. data on convergent and discriminant validity are complex and difficult to interpret. Based on a systematic record review for inpatients (i.e., Hennepin County Medical Center), depression had the highest correlation with RCd for inpatients (.39 for women and .40 for men) followed by RC2 (.36 and .34, respectively). Reassuringly, hallucinations had the highest correlations (.29 and .38, respectively) with RC8. Criminal justice involvement evidenced low correlations on RC4, which varied by gender: .06 for women and .16 for men. With both genders, however, substance abuse had much higher correlations (.55 and .47, respectively).

A formal analysis of convergent-discriminant validity using Campbell and Fiske (1959) multitrait-multimethod approach would be very helpful in compiling and integrating the Tellegen et al. correlate research. One concern is the magnitude of the convergent correlations; Fiske and Campbell (1992) noted that modest correlations in the .30 to .50 range could be successful. Although many of the Tellegen et al. convergent correlations were below the threshold of .30, they could likely be improved by the implementation of better validated external criteria (Rogers, 2001).

In this cross-validation, we addressed one important component of convergent-discriminant validity, namely, heterotrait-monomethod comparisons. An examination of RC interscale correlations revealed moderate correlations (*M r* = .45) that were substantially improved over the traditional MMPI–2 Clinical Scales (*M r* = .60). These correlations are similar to those found by Tellegen et al. with outpatient samples (see Tellegen et al., 2003, Tables 4.8 to 4.11 on pp. 38–41). In establishing convergent-discriminant validity, the objective is to have convergent correlations be appreciably higher than discriminant correlations. The goal is to have very few comparison violations in which discriminant correlations exceed convergent correlations (Byrne & Goffin, 1993). This goal has yet to be achieved with the MMPI–2 RC Scales, with most convergent correlations below the .45 average for heterotrait-monomethod comparisons.

Interpretation of RC Scales and Profiles

This cross-validation raises practical issues regarding the clinical interpretation of RC Scales, which are presently provided in the Extended Score Report by Pearson Assessments. The most pressing issue is the need for programmatic research on the clinical characteristics uniquely associated with each RC Scale elevation. Clinical correlate research is insufficient to address Tellegen et al.'s predominant goal of removing Demoralization from profile interpretation because many characteristics of general impairment and distress will be included. Cashel, Rogers, Sewell, and Holliman's (1998) earlier work on the MMPI–Adolescent suggested that most correlates lacked discriminability if they were positively associated with three or more Clinical Scales. Items associated with general distress should be removed from both the measure and the external criteria. For the MMPI–2 RC Scales to demonstrate discriminant validity, clinical characteristics singular to each scale must be established.

Scale interpretations, based largely on convention, have relied principally on clinical elevations. In providing interpretive guidelines, Tellegen et al. (2003) proposed that elevations should be used as "heuristic guidelines … rather than as fixed points demarcating qualitative change" (p. 54).

Tellegen et al. proposed a fundamental change in MMPI–2 interpretation based on a dimensional rather than categorical approach. Our data on *SEM*s (see Table 2) indicated variable accuracy of individual scores (Anastasi, 1988); at 2 *SEM*s, the range of RC Scales is from 13.20 to 25.92 T points. These ranges support Tellegen et al.'s proposal about elevations as heuristic guidelines rather than discrete categories.

One challenge to the MMPI–2 RC clinical usefulness is the large proportion of WNL profiles. Although such profiles are common with traditional MMPI–2 scales, the percentage increased to more than 40% of client cases in this study, thereby limiting interpretations. RC clinical elevations also show marked variations in their prevalence rates (see Table 5). For example, RC1 was elevated in almost two thirds (i.e., 62.4%) of the client profiles with any clinical elevations.[4] Is RC1 clinically useful if its interpretations apply to the large majority of interpretable cases? In contrast, RC9 elevations rarely occurred as the highest or even second highest elevation (see Table 5). Given the absence of common code types, infrequency of clinical elevations (7.5%), and limited cross-validation, the RC9 scale might be characterized as a subsidiary scale, which has circumscribed clinical usefulness.

The RCd scale is viewed as "the starting point for RC Scale interpretation" and provides "an indication of overall emotional discomfort" (Tellegen et al., 2003, p. 54). How should unelevated RCd scales affect interpretation? As shown in Table 5, nearly all clients (7.9/8.3 or 95.2%) with a spike RC6 had unelevated RCd scales. Would it be accurate to say that most clients did not experience emotional discomfort when believing they were "targeted, controlled, or victimized by outside forces, usually malevolent others" (Tellegen et al., 2003, p. 56)? Clearly, further research is needed on differential interpretations with elevated and unelevated RCd scales.

Tellegen et al. observed that RC code types would be substantially different from traditional code types because of the substantial differences in item composition. Because no data were presented, we tested this assumption. As reported in Table 5, the concordance between RC and traditional scales was too low to extrapolate any meaningful interpretations for the RC Scales.

Concluding Remarks

Tellegen et al. noted that their work on the RC Scales might stimulate further scale development and validation for the MMPI–2. This laudable goal has been achieved by proposing fundamental changes in MMPI–2 scale development with a paradigmatic shift to refined Clinical Scales that have established discriminant validity. Despite these achievements, we would find the peripheralization of empirically derived scales to be premature. Empirical keying is not moribund (Butcher, 2000) but requires further refinements such as near-neighbor comparisons.

The MMPI–2 RC Scales face several important challenges prior to their use in professional practice. Current interpretations rely mostly on clinical correlates rather than distinguishing clinical characteristics. To confidently establish discriminant validity, general distress must be removed from both the MMPI–2 and corresponding external measures. Because of the omission of response style issues during the scale development phase, the effects of response styles, such as social desirability and malingering, must be investigated prior to the clinical use of the MMPI–2 RC Scales. We caution against the use of the MMPI–2 RC Scales in professional settings until these and other issues of test validation are satisfactorily addressed.

The RC9 scale appears to be the weakest of the RC Scales. Its items evidenced less homogeneity (*M* interitem *r*s < .15) than other RC Scales and replication of its components was only modest (50%) for the seed scales. From an interpretive perspective, the RC9 scale was never the highest Clinical Scale and tended only to be elevated when several other RC Scales had higher elevations. The RC9 scale may require further refinement before applied in clinical settings.

Further studies will likely build on the Tellegen et al. accomplishments in validating the MMPI–2 scales in different professional settings with diverse clinical populations. Competing validational models can be tested to evaluate the respective contributions: (a) the original empirical keying, (b) a refined empirical keying with near-neighbor comparisons, (c) Tellegen et al.'s adaptation of Jackson's (1970) construct validation, and (d) Jackson's full model of construct validation. The eventual goal is a further refinement of the MMPI–2 as a theory driven and empirically validated multiscale inventory with specific clinical applications in a broad range of professional settings. The research by Tellegen et al. is a vital impetus toward this worthy goal.

[4]The total percentage is different from Table 5 because it includes infrequent code types and the third highest clinical elevations.

ACKNOWLEDGMENTS

We appreciate Alex Caldwell's gracious generosity in sharing his data set and the unstinting efforts of Roger Greene in compiling an anonymous sample of 8,000 MMPI–2 cases.

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.

Anastasi, A. (1988). *Psychological testing* (6th ed.). New York: Macmillan.

Archer, R. P., Aiduk, R., Griffin, R., & Elkins, D. E. (1996). Incremental validity of the MMPI–2 content scales in a psychiatric sample. *Assessment, 3,* 79–90.

Ben-Porath, Y. S., & Butcher, J. N. (1989). Psychometric stability of rewritten MMPI items. *Journal of Personality Assessment, 53,* 645–653.

Ben-Porath, Y. S., Butcher, J. N., & Graham, J. R. (1991). Contribution of the MMPI–2 content scales to the differential diagnosis of schizophrenia and major depression. *Psychological Assessment, 3,* 634–640.

Butcher, J. N. (2000). Dynamics of personality test scores: The empiricist's manifesto revisited. *Journal of Clinical Psychology, 56,* 375–386.

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. (2001). *Minnesota Multiphasic Personality Inventory–2 (MMPI–2): Manual for administration, scoring and interpretation* (Rev. ed.). Minneapolis: University of Minnesota Press.

Butcher, J. N., Graham, J. R., Dahlstrom, W. G., & Bowman, E. (1990). The MMPI–2 with college students. *Journal of Personality Assessment, 54,* 1–15.

Butcher, J. N., Graham, J. R., Williams, C. L., & Ben-Porath, Y. S. (1989). *Development and use of the MMPI–2 content scales.* Minneapolis: University of Minnesota Press.

Butcher, J. N., Jeffrey, T. B., Cayton, T. G., Colligan, S., DeVore, J. R., & Minegawa, R. (1990). A study of active duty military personnel with the MMPI–2. *Military Psychology, 2,* 47–61.

Butcher, J. N., Williams, C. L., Graham, J. R., Tellegen, A., & Kaemmer, B. (1989). *MMPI–2: Manual for administration and scoring.* Minneapolis: University of Minnesota Press.

Byrne, B. M., & Goffin, R. D. (1993). Modeling MTMM data from additive and multiplicative covariance structures: An audit of construct validity concordance. *Multivariate Behavioral Research, 28,* 67–96.

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

Campbell, D. T., & Fiske, D. W. (1959). Convergent and discriminant validation by the multitrait-multimethod matrix. *Psychological Bulletin, 56,* 81–105.

Cashel, M. L., Rogers, R., Sewell, K. W., & Holliman, N. (1998). Preliminary validation of the MMPI–A for a male delinquent sample: An investigation of clinical correlates and discriminant validity. *Journal of Personality Assessment, 71,* 49–69.

Clark, L. A., & Watson, D. (1995). Constructing validity: Basic issues in objective scale development. *Psychological Assessment, 7,* 309–319.

Crossman, L. L., Casey, T. A., & Reilley, R. R. (1994). Influence of cognitive variables on MMPI–2 scale scores. *Measurement and Evaluation in Counseling and Development, 27,* 151–157.

Dahlstrom, W. G., Welsh, G. S., & Dahlstrom, L. E. (1972). *An MMPI handbook: Volume 1. Clinical interpretation* (Rev. ed.). Minneapolis: University of Minnesota Press.

Derogatis, L. R. (1977). *The SCL–90–R version manual: Scoring, administration, and procedures for the SCL–90–R.* Baltimore: John Hopkins University, School of Medicine.

Fiske, D. W., & Campbell, D. T. (1992). Citations do not solve problems. *Psychological Bulletin, 112,* 393–395.

Frances, A. J., First, M. B., Widiger, T. A., Miele, G. M., Tilly, S. M., Davis, W. W., et al. (1991). An A to Z guide to *DSM–IV* conundrums. *Journal of Abnormal Psychology, 100,* 407–412.

Friedman, A. F., Webb, J. T., & Lewak, R. (1989). *Psychological assessment with the MMPI.* Hillsdale, NJ: Lawrence Erlbaum Associates, Inc.

Golden, C. J., Sawicki, R. F., & Franzen, M. D. (1984). Test construction. In A. S. Bellack & M. Hersen (Eds.), *Research methods in clinical psychology* (pp. 233–265). New York: Pergamon.

Gorsuch, R. L. (1990). Common factor analysis versus component analysis: Some well and little known facts. *Multivariate Behavioral Research, 25,* 33–39.

Graham, J. R. (1987). *The MMPI: A practical guide* (2nd ed.). New York: Oxford University Press.

Graham, J. R. (2000). *The MMPI–2: Assessing personality and psychopathology* (3rd ed.). New York: Oxford University Press.

Greene, R. L. (2000). *The MMPI–2: An interpretive manual.* Boston: Allyn & Bacon.

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

Hathaway, S. R., & McKinley, J. C. (1967). *MMPI manual* (Rev. ed.). Minneapolis: University of Minnesota Press.

Humphrey, D. H., & Dahlstrom, W. G. (1995). The impact of changing from the MMPI to the MMPI–2 on profile configurations. *Journal of Personality Assessment, 64,* 428–439.

Jackson, D. N. (1970). A sequential system for personality scale development. In C. D. Spielberger (Ed.), *Current topics in clinical and community psychology* (Vol. 2, pp. 60–96). New York: Academic.

Jackson, D. N., Fraboni, M., & Helmes, E. (1997). MMPI–2 content scales: How much content do they measure? *Assessment, 4,* 111–117.

Marks, P. A., Seeman, W., & Haller, D. L. (1974). *The actuarial use of the MMPI with adolescents and adults.* New York: Oxford University Press.

Munley, P. H., Busby, R. M., & Jaynes, G. (1997). MMPI–2 findings in schizophrenia and depression. *Psychological Assessment, 9,* 508–511.

Rogers, R. (2001). *Handbook of diagnostic and structured interviewing.* New York: Guilford.

Rogers, R., & Bender, S. D. (2003). Evaluation of malingering and deception. In A. M. Goldstein (Ed.), *Comprehensive handbook of psychology: Forensic psychology* (Vol. 11, pp. 109–129). New York: Wiley.

Roper, B. L., Ben-Porath, Y. S., & Butcher, J. N. (1995). Comparability and validity of computerized adaptive testing with the MMPI–2. *Journal of Personality Assessment, 65,* 358–371.

Strong, D. R., Greene, R. L., Hoppe, C., Johnston, T., & Olesen, N. (1999). Taxometric analysis of impression management and self-deception on the MMPI–2 in child-custody litigants. *Journal of Personality Assessment, 73,* 1–18.

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

Velicer, W. F., & Jackson, D. N. (1990). Component analysis versus common factor analysis: Some issues in selecting an appropriate procedure. *Multivariate Behavioral Research, 25,* 1–28.

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

Richard Rogers
Department of Psychology
P.O. Box 311280
University of North Texas
Denton, TX 76203–1280
Email: rogers_1@hotmail.com

Received January 13, 2005
Revised November 9, 2005

Copyright of Journal of Personality Assessment is the property of Lawrence Erlbaum Associates and its content may not be copied or emailed to multiple sites or posted to a listserv without the copyright holder's express written permission. However, users may print, download, or email articles for individual use.