The Deconstruction of the Hy Scale of MMPI–2: Failure of RC3 in Measuring Somatic Symptom Expression

James N. Butcher

Department of Psychology
University of Minnesota

Cassia K. Hamilton and Steven V. Rouse

Department of Psychology
Pepperdine University

Edward J. Cumella

Remuda Programs for Eating Disorders
Wickenburg, Arizona

The MMPI–2 (Butcher, Dahlstrom, Graham, Tellegen, & Kaemmer, 1989) Clinical Scales have a long history in psychological assessment. Recently, Tellegen et al. (2003) conducted a series of analyses to restructure the scales to reduce what they considered to be problems that limit scale functioning. In a critique of the Restructured Clinical (RC) Scales published in this issue, Nichols (2006/this issue) questions a number of aspects of the approach Tellegen et al. took including their theoretical assumptions, methods of analysis, and failures to report important information needed for scale evaluation such as relationships with existing scales. We concur with many points raised by Nichols. In our analysis of the performance of the RC3 scale, we found that it has “drifted” so far from the original Hy scale as to be a completely different measure—a scale of cynical attitudes that is already well represented in existing MMPI–2 measures. In this article, we take these concerns a step further and examine the history and construct validity of the Hy scale in evaluating the somatic expression of problems that the original authors (McKinley & Hathaway, 1944) intended. We also include new information from a medical setting, an application not represented in Tellegen et al.’s RC Scale monograph. In agreement with Rogers et al. (2006/this issue), it is our conclusion that some RC Scales do not represent the measurement domain of the original scales and should not be relied on for or used to refine traditional interpretation, particularly in medical or forensic situations (such as personal injury cases) because of their confusing and conflicting results.

Hathaway and McKinley (1940) developed the Minnesota Multiphasic Personality Inventory (MMPI) to assess mental health problems in psychiatric and medical settings through patients’ true–false responses to a large number of symptoms and attitudes. By the 1960s, the MMPI became the most widely used personality assessment measure and a standard for self-report assessment of psychopathology in a broad variety of settings. The inventory underwent a major revision and expansion in the 1980s, culminating, after nearly 10 years of research, in the MMPI–2 (Butcher, Dahlstrom, Graham, Tellegen, & Kaemmer, 1989) and MMPI for Adolescents (Butcher et al., 1992) with new, more representative norms and extensive validation across a number of clinical applications. The MMPI revision maintained the integrity of the traditional Clinical Scales (Tellegen et al., 2003) to ensure that the test would continue to address the original constructs (Hathaway & McKinley, 1940) and provided new content-based measures to provide more comprehensive assessment (Butcher, 2000; Butcher et al., 2001).

Although change may be inevitable, it does not always signal progress. The recent effort to restructure the original MMPI Clinical Scales is an example of developing revised
measures that do not offer uniform improvement. For example, one scale, Hy, has been altered so radically as to drastically change the constructs underlying the original scale, the measured correlates, and the scale interpretations. Nichols’s (2006/this issue) article in this issue provides an extensive critique of the Restructured Clinical (RC) Scales. Nichols’s analysis of the strategy used to create the RC Scales and his conclusions about the adverse impact of the psychometric decisions provide an important contribution to understanding the RC Scales. The article by Rogers, Sewell, Harrison, and Jordan, (2006/this issue) also in this issue, places the RC Scales in historical perspective and notes both prospects and problems. Rogers et al. note that there is a pressing need for research to support their clinical use, especially now that the RC Scales are being profiled in Pearson Assessments’ extended scoring service. The wide availability of the RC Scales and stated theoretical promise have produced unfounded expectations about their utility. The RC Scales have thus far been insufficiently researched to support their use in many settings.

In this article, we contribute to the emerging critique of the RC Scales by offering an analysis of one of the new scales, RC3: its relationship to the original MMPI Hy scale, how far the construct has drifted (to use a characterization provided by Nichols) away from the original scale and its personality correlates, and what attributes RC3 actually addresses.

**RC SCALES**

The purpose of the altered or restructured Clinical Scales was to preserve the core constructs of the scales and improve their effectiveness by removing some items thought to assess general maladjustment, reducing item overlap, lowering scale intercorrelation, eliminating the so-called subtle items (i.e., items without content validity), and improving discriminant validity. The RC Scales were developed through several steps. The initial step involved the development of a “Demoralization” measure that because it was viewed as common to most of the Clinical Scales resulted in concept overlap. The Demoralization factor was used to remove items from the eight Clinical Scales that were considered to be influenced by this tendency toward general maladjustment, leaving a set of “seed” scales for the eight Clinical Scales. (Both Nichols and Rogers et al. express reservations about the elimination of general distress from the Clinical Scales in that this resulted in the alteration of important facets of some Clinical Scales.) Then, the seeds were expanded by including correlated MMPI–2 items from the remainder of MMPI–2 item pool. Finally, Tellegen et al. conducted both internal and external validity analyses to examine the operation of the new scales. Tellegen et al. conducted several analyses of the internal and predictive validity of the RC Scales using patients from three samples: the Portage Path outpatient sample (Graham, Ben-Porath & McNulty, 1999) and two inpatient samples from prior studies (Arbisi, Ben-Porath, & McNulty, 2003). The results described in Tellegen et al.’s RC monograph were thought to show that the RC Scales have equal or greater association with external behavioral correlates than the traditional Clinical Scales. Graham (2006) noted that the correlational data indicate that most RC Scales are measuring characteristics that are “similar but not identical to their clinical and content scale counterparts” (p. 158). The scale construction strategies used to develop the RC Scales assured that the resulting measures would be unidimensional in scope and homogeneous in content. However, very little information has yet been provided about the operation of the RC Scales relative to other widely used measures such as the MMPI–2 Content scales and Personality Psychopathology Five (PSY–5; Harkness, McNulty, & Ben-Porath, 1995) scales. Moreover, the samples employed to validate the RC Scales included only psychiatric patients and inpatient substance abusers. No information has been provided about how the RC Scales operate in medical, forensic, or personnel settings.

Although most of the RC Scales bear some content resemblance to the original Clinical Scales, some, such as RC3, do not. The RC3 scale underwent an extreme makeover, and the original focus was not preserved. The RC3 factors were Demoralization, Somatization, and a small number of items representing Cynical Attitudes. Tellegen et al. (2003) chose the smallest of the factors—the five Cynical Attitudes items—as the seed scale to represent the Hy scale and eliminated the most important elements (defensive somatization) that Hathaway and McKinley (1940) strove to maintain. As a result, RC3 is actually negatively correlated with the original Hy scale; across various samples reported in the Tellegen et al. monograph, correlations between RC3 and Hy were –.24 for women and –.42 for men. Thus, RC3 functions very differently from what Hathaway and McKinley had in mind when they created Hy. Clearly, before RC3 finds a distinct place in personality assessment, more needs to be known about its operation. It appears as though RC3 is a redundant measure of cynicism and anger already measured by two content scales: CYN and ANG (Butcher, Graham, Williams, & Ben-Porath, 1990). In light of the development of RC3 and its proposed use in clarifying the interpretation of Hy, it is valuable to review the history, construct validity, and interpretive guidelines for the Hy scale; assessment psychologists must weigh its strengths and weaknesses for measuring several important personality dimensions and to consider whether RC3 assesses important constructs measured by Hy.

**ORIGINAL DEVELOPMENT OF THE Hy SCALE OF THE MMPI**

McKinley and Hathaway (1944) developed the Hy scale to assess hysteria, a psychological pattern associated with the development of physical symptoms for which there is no clear organic basis. McKinley and Hathaway followed the empirical
keying approach in developing Hy. McKinley and Hathaway made efforts to eliminate items overlapping with other measures, such as the Hypochondriasis scale (Hs), but eliminating these items from Hy lowered statistical discrimination between patients and normal persons. Thus, most items were restored to maintain the power of empirical separation.

McKinley and Hathaway (1944) concluded that

All clinicians who used both scales were emphatic that there was indubitably a valid clinical difference between two persons having both scores on Hy and Hs but differing in that one score was higher. There was a different prognosis and treatment indicated for the two. (p. 159)

McKinley and Hathaway (1944) further pointed out that

The scale appears to measure a rather variable trait which is closely allied to and likely includes the earlier scale of hypochondriasis. The person who is especially characterized by Hy tends to be less obviously neurotic and to have, during disabled periods, a more specific set of physical symptoms. (p. 162)

Although Hathaway and McKinley (1940) recognized the item-overlap and construct similarity, they made compelling arguments that the subtle distinctions between these scales provide meaningful clinical information. Early studies on MMPI Clinical scale content had established further the internal makeup of Hy. Three aspects of physical symptom expression were found in Hy’s content (Little & Fisher, 1958): somatic complaints, social facility or participation, and denial of any kind of problem.

Research has established a rich network of correlates for Hy ranging from physical symptoms to managing aggression. Hy’s complex content, with items from both social and somatic domains, is clearly shown by its multifaceted construct validity. Space limitations prevent us from completely reviewing the hundreds of articles on Hy. Hy’s correlates have been widely studied and summarized in textbooks (e.g., Graham, 2006; Greene, 2000). Graham’s (2006) popular text lists descriptors for high Hy scorers such as often feeling overwhelmed; reacting to stress and avoiding responsibility by developing physical symptoms; headaches, stomach discomfort, chest pains, weakness; symptoms that appear and disappear suddenly; lack of acute emotional turmoil; feeling sadness, depression, anxiety at times; lack of energy, feeling worn-out, sleep disturbances; and frequent diagnoses of conversion disorder, somatoform disorder, or pain disorder. In short, Hy has an important place in assessment, and its functions cannot be fulfilled by RC3 whose assessment is limited to cynicism.

Hy and Assessment of Chronic Pain

Numerous researchers have found Hy associated with somatization (Keller & Butcher, 1991; Sellbom, Graham, & Schenk, 2005), particularly chronic pain. Summaries of this research are found in Arbisi (2006), Arbisi and Butcher (2004), and Keller and Butcher (1991). Arbisi and Butcher (2004) and Arbisi and Seime (2006) concluded from their review of personality factors associated with chronic pain that somatic preoccupation and naïve denial of emotional and interpersonal difficulties—constructs measured by Hy—render individuals vulnerable to develop chronic pain conditions and become disabled.

Prediction of Future Disability

The MMPI Clinical Scales have been used in prospective studies of psychological factors in the recovery from physical illness or injury. Fordyce, Bigos, Battie, and Fisher (1992) tested aircraft manufacturing employees at employment initiation and then contacted them several years later. Fordyce et al. found that persons with high Hy scores at the time of hiring were more likely to develop disabling chronic pain later. Hy also predicts failure to return to work after an alleged injury (Gatchel, Polatin, & Kinney, 1995).

Response to Treatment

Hy has been found in multiple studies to be related to poor response to treatment. McCrery (1985) and Strassberg, Reimherr, Ward, Russell, and Cole (1981) reported that Hy (and Hs) elevations are associated with poor response to treatment among chronic pain patients. Wiltse and Rocchio (1975) found among chronic pain patients with extremely high elevations on both Hs and Hy (T > 85) only a 10% chance of good to excellent recovery.

Litigation or Work Compensation

To understand the impact of Hy in assessing compensation and its role in forensic evaluations, readers can consult Arbisi (2006) and Shaffer, Nussbaum, and Little (1972). For example, in a study by Long, Rouse, Nelson, and Butcher (2004), Hy was prominently elevated among persons involved in sexual harassment personal injury claims.

DECONSTRUCTING HY: DEVELOPMENT OF RC3

Why does one need to restructure a scale that has strong validity, reliability, and utility in a range of assessments in medical, personnel, and forensic settings? The influence of a factor that Tellegen et al. referred to as Demoralization, or the tendency for psychiatric patients to approach personality test items with an extreme set to endorse severe psychopathology, was a central theme in the removal of items from the original scales. In his critique, Nichols points out a number of problems with the demoralization construct. We concur with his viewpoint and how this ap-
The Tellegen et al. RC monograph did not provide information on the relationships between the RC Scales and the MMPI–2 Content scales, PSY–5 scales, or any other supplementary measures on the test. For example, there is an 80% overlap between items on RC3 and the CYN scale. The cynicism factor is not a new concept; it was described in earlier factor analytic studies by Comrey (1957) who found a small cluster of cynicism items in his factor analysis of Hy, exactly the same five items in the Tellegen et al. core for RC3. Later, in an item factor analysis of the entire item pool, Johnson, Butcher, Null, and Johnson (1984) found a 20-item cynicism cluster that contains all of the Tellegen et al. core items. There is an over 73% overlap between RC3 and the Johnson et al. Cynicism factor. The RC3 also has a 73% overlap with the supplemental scale Hostility (Cook & Medley, 1954; Han, Weed, Calhoun, & Butcher, 1995), indicating further that the content has drifted away from the original Hy scale and toward an aggressive, hostile, cynical personality cluster.

**Relationship with other scales.** The RC3 scale is negatively correlated with the Hy scale. This supports Nichols’s view that the RC Scales suffer from “construct drift” and are not measuring the constructs represented in the parent scales. The correlation noted between the RC3 scale and the Hy scale was –.42 for the men in the normative sample and –.24 for the women in the normative sample; for samples taken from clinical populations, however, all correlations between RC3 and Hy reported by Tellegen et al. fell below an absolute value of .20. This indicates that the scale is not measuring the same constructs defined by Hathaway and McKinley (1940). In our subsequent analysis of the MMPI–2 normative sample, we found that the scores for the men on the RC3 were correlated with several other measures of cynical hostility: CYN (r = .90), Ho (r = .84), and AGGR (r = .38). For the women in the MMPI–2 normative sample, RC3 was correlated with CYN (r = .91), Ho (r = .82), and AGGR (r = .34). The correlations between RC3 and CYN were actually higher than the alpha coefficient for RC3 (.80 for men and .79 for women). In other words, RC3 is best considered a psychometrically parallel form of CYN.

**Preliminary research correlates.** Research on the RC Scales in a nonclinical sample of college undergraduates showed that the RC3 scale was moderately correlated with the Machiavellian Scale (Forbey & Ben-Porath, 2005). RC3 was modestly associated with externalization of blame in a study of men in a corrections sample (Sellbom, Ben-Porath, Barum, et al., 2005). These studies support the interpretation that RC3 is measuring an aggressive-cynical personality cluster rather than the somatic expression as measured by the original Hy scale. The only study to date in which the RC Scales have been examined in a medical psychological application involves eating disorder inpatients (Hamilton, Rouse, Miller-Perrin, & Cumella, 2006). The value of the MMPI–2 in assessing individuals with eating disorders has been well established. Clinical elevations have been commonly reported on several MMPI–2 scales including Hy (Cumella, Wall, & Kerr-Almeida, 2000). Research has shown small Clinical Scale elevation differences between anorexia and bulimia groups but no statistical difference between subtypes of these disorders. Although Hy has been shown to be a valuable mea-
In this article, we addressed the development of the new RC Scales, reviewed and evaluated the critiques provided by Nichols and Rogers et al., and reported on the failure of the RC3 scale to assess somatic aspects of a large sample of clients being treated in a health care setting. Our conclusions are as follows. The data reported in the RC manual pertaining to the RC3 scale and RC3’s preliminary research results, especially its performance in the study of patients in a medical setting we reported in this article, lead to a clear conclusion. RC Scale development has not captured the essence of the Hy scale intended by Hathaway and McKinley (1940). The RC3 scale is negatively and weakly correlated with Scale Hy from which it was derived: −.24 in women in the RC monograph and −.05 among female eating disorder patients. The RC3 scale is a different measure than the Hy scale and should be recognized as such. With respect to the RC3, we concur with Nichols’s view that it is a redundant measure of other widely used MMPI–2 scales such as the CYN scale and to some extent the Hostility scale, Anger Content scale, and PSY–5 Aggressiveness in use today.

In developing new MMPI–2 scales, it is important to properly align them with the existing and widely accepted measures already serving as standards, clearly describing similarities and differences (see discussions by Butcher, Graham, & Ben-Porath, 1995). The RC Scales have not been sufficiently compared with existing similar measures such as the Content scales and other supplemental scales such as Ho and the PSY–5 scales. The initial publication produced confusion as to what new and old constructs were actually being addressed. Rogers et al. conclude that there is a major limitation in the RC Scale interpretative approach in that about half of the clinical cases in their study had normal limits profiles. Moreover, Rogers et al. point out that the concordance between the RC Scales and traditional scales is too low to extrapolate any meaningful interpretations to the RC Scales. The methodology followed in the RC development project resulted in drift for many of the Clinical Scales because the theoretical model of removing Demoralization from the scales also eliminated impor-
tant items for describing personality in situations in which demoralization is not pertinent.

We concur with Rogers et al.’s caution against using the RC Scales in professional settings at this time. The need for establishing a network of empirical meanings for RC3 is apparent in such settings as forensic and medical applications. In the Rogers et al. study, they eliminated a forensic sample (custody cases) from their data set because these individuals were thought to underreport psychopathology—a test-taking strategy quite different from the clinical populations on which the scales were developed. The RC Scales have been insufficiently validated to meet legal criteria for admissibility in court. Just because a measure is derived from MMPI–2 items is insufficient to assure validity in forensic assessments. More information about the performance of the RC Scales in forensic and medical settings is needed to support their use.

REFERENCES


