

Cultural Validity of the Minnesota Multiphasic Personality Inventory–2 Empirical Correlates: Is This the Best We Can Do?

Jill S. Hill, Rockey R. Robbins, and Terry M. Pace

This article critically reviews empirical correlates of the Minnesota Multiphasic Personality Inventory–2 (MMPI-2; Butcher, Dahlstrom, Graham, Tellegen, & Kaemmer, 1989), based on several validation studies conducted with different racial, ethnic, and cultural groups, and argues for the urgent development of culturally derived measures that would illuminate culturally relevant empirical correlates for use in validating the MMPI-2.

Este artículo examina críticamente correlatos empíricos del Test de Personalidad Multifásico de Minnesota-2 (MMPI-2, por sus siglas en inglés), basándose en varios estudios de validación conducidos con distintos grupos raciales, étnicos y culturales, y razona a favor del desarrollo urgente de medidas derivadas culturalmente que puedan ilustrar correlatos empíricos culturalmente relevantes para su uso en la validación del MMPI-2.

It is a widely held view that for a personality inventory to be accepted as useful, empirical validation of the measure is necessary (Butcher, 2000; Meehl, 1945). Citing Meehl's (1945) "empiricist manifesto," Butcher (2000) stated that the role of personality test items is to serve as "a sample of behavior that has correlates requiring further empirical validation" (p. 376). From this view, simply scoring a personality test item does not ensure a valid self-rating. According to Butcher (2000), this is "the theoretical basis of the empirical approach to scale construction" (p. 376) used by Hathaway and McKinley (1940) in developing the Minnesota Multiphasic Personality Inventory (MMPI). The process of identifying empirical correlates validates the power of a personality scale to predict external behavior. In short, this process verifies the test's credibility (Butcher, 2000).

Hathaway and McKinley (1940) selected criterion groups based on the clinical diagnosis and used an empirical approach in selecting items for a specific scale. As Greene (2000) stated, the items had to be answered differently by the criterion group (e.g., patients who were depressed) as compared with nonclinical groups. Items were selected solely on the basis that the criterion group answered the items differently than did other groups.

Jill S. Hill, Department of Counseling and Clinical Psychology, Teachers College, Columbia University; Rockey R. Robbins and Terry M. Pace, Department of Educational Psychology, University of Oklahoma. Correspondence concerning this article should be addressed to Jill S. Hill, Department of Counseling and Clinical Psychology, Teachers College, Columbia University, Box 102, 525 West 120th Street, New York, NY 10027-6696 (e-mail: jh2937@columbia.edu).

© 2012 American Counseling Association. All rights reserved.

The test's developers were guided by normative references for differential diagnosis. It is interesting that users of the test observed that it was not an effective measure for differential diagnosis, although MMPI scale scores were associated with clinically relevant symptoms, traits, and behaviors. Handel and Ben-Porath (2000) noted that "within a decade of the test's original publication, MMPI interpretation was based on the empirically established correlates of the instrument" (p. 230). Large-scale studies aimed at forming a foundation for actuarial application of the test were conducted and subsequently provided clinicians a method by which they could base their interpretations on the empirical correlates of the scales. This approach has held true for the restandardized MMPI, the MMPI-2 (Butcher, Dahlstrom, Graham, Tellegen, & Kaemmer, 1989).

In what Butcher (1999) described as the "definitive MMPI-2 empirical study" (p. xv), Graham, Ben-Porath, and McNulty (1999) examined the empirical validity of the test's clinical scales. One purpose of this study was to empirically demonstrate that the MMPI-2 reliably assessed the same constructs and clinical dimensions addressed by the original MMPI (Butcher, 1999). Butcher and Graham et al. concluded this objective was achieved.

Graham et al. (1999) conducted the study within an urban community mental health setting in the state of Ohio over 21 months. The main purpose of the study was to identify empirical correlates of MMPI-2 scales and code types (Graham et al., 1999). The study's sample consisted of 410 men and 610 women (clients who completed all standardized procedures and the MMPI-2). With regard to race, the sample consisted of 814 Caucasians (332 men, 482 women), 192 African Americans (70 men, 122 women), and 14 identified as "other." Graham et al. claimed the sample's demographic composition appropriately represents a large segment of the U.S. population that would receive mental health services and to whom the MMPI-2 would be administered in community mental health settings.

In addition to the MMPI-2, the researchers used the following extratest measures in the study: An intake form specifically designed for this project, the Symptom Checklist-90-Revised (SCL-90-R; Derogatis, 1983; a self-report instrument developed that reflects patterns of psychological symptoms), Symptom Checklist-90 Analogue (SCL-90 Analogue; Derogatis, 1983; a visual analogue scale rated by clinicians), and a patient description form (PDF) developed specifically for this project (Graham et al., 1999). In the data analysis, the researchers (i.e., Graham et al., 1999) calculated correlations between MMPI-2 scale and subscale scores and measures from the extratest instruments; the correlation coefficients provided a measure of effect size for the empirical correlates. The researchers dichotomized MMPI-2 scale scores ($T \geq 65$, clinically significant vs. $T \leq 65$, nonclinically significant; Graham et al., 1999) and reported an immense number of significant correlations between extratest characteristics and the MMPI-2 clinical scales for both men and women (see Graham et al., 1999).

Identifying significant empirical correlates, although highly praised within the field of personality assessment, is but one method of test validation. Extratest measures, especially those designed specifically for research purposes and based on clinician ratings can be problematic when used to determine cultural bias within a test. Extratest measures, like the tests they are correlated with, are developed from and reflect a specific worldview, typically the dominant culture's worldview. The practice of using measures developed from the same worldview to validate each other repeatedly fails to take into consideration cultural context, cultural factors, or local cultural norms of behavior specific to the racial and ethnic minority persons to whom results of research studies will be generalized.

Although the information gained from the Graham et al.'s (1999) study is valuable, it remains questionable whether this study was truly definitive given only Caucasians and African Americans (with strikingly disparate sample sizes between these groups) were included in the study and the assumed universality of the extratest measures used in the study. Making sweeping generalizations from conclusions based on observed differences or similarities between only two racial groups is dangerous. Assumptions regarding the universality of extratest measures and the generalizations made from Graham et al. (1999) need to be examined much more closely before zealously claiming the study's definitiveness.

multicultural issues and the mmpi-2

When investigators began researching multicultural issues with the MMPI, initially, they attended to primarily normative questions in a manner that was inconsistent with the empirical approach to MMPI interpretation (Handel & Ben-Porath, 2000). Because the original norms of the MMPI were developed on an exclusively White sample from Minnesota, concern existed about the applicability of the inventory with members of various ethnic minority groups (Graham, 1993). Unfortunately, multicultural factors in personality assessment using an empirical perspective were not investigated until after the restandardization of the MMPI in 1989.

One of the major goals of the MMPI Restandardization Project was to obtain a contemporary normative sample, one that was consistent with the 1980 U.S. Census (Butcher, Graham, Williams, & Ben-Porath, 1990). However, the achievement of that goal was questionable. One of the criticisms of the restandardization was its shortcomings in representativeness, for example, inclusion of few Hispanic/Latino(a) Americans, Asian Americans, and older women and inclusion of unrepresented American Indians (Nichols, 1992; Nichols, Padilla, & Gomez-Maqueo, 2000). In fact, Nichols et al. (2000) argued the representation of various ethnic minorities in the restandardization sample "is insufficient to guarantee freedom from ethnic biases, particularly among

persons without competence in English” (p. 262). Other criticisms included continued untested assumptions regarding item content and phrasing of the items that reflect a purely Western perspective (Hill, Pace, & Robbins, 2010; Pace et al., 2006). These criticisms remain unanswered, unfortunately, considering the extent to which the MMPI-2 has been translated and adapted internationally.

other mmpi-2 empirical validation studies with diverse groups

Addressing the question of test bias, Timbrook and Graham (1994) used scales developed by Long (1993) as extratest criteria and compared the validity of only the five following standard clinical scales: 2 (Depression), 4 (Psychopathic Deviate), 7 (Psychasthenia), 9 (Hypomania), and 0 (Social Introversion) across Caucasian and African American individuals (Handel & Ben-Porath, 2000). Using a combination of rational and statistical methods, Long (1993) developed scales for a partner rating form (PRF). This form was based on a modified version of the Katz Adjustment Scales (Katz & Lyster, as cited in Long, Graham, & Timbrook, 1994). Data for the development of scales for this measure were obtained using the MMPI-2 normative sample. The 110 items on the PRF were categorized for content by four judges. From these categories, preliminary scales were developed and then refined using internal consistency procedures, resulting in eight scales that represented the content dimensions of the PRF items. The scales were labeled as follows: Antisocial, Anger–Hostility/Irritability, Psychological Maladjustment, Mania–Hyperactivity, Social Discomfort, Depression, Low Self-Esteem, and Anxiety (Long et al., 1994). Long reported a mean internal consistency coefficient alpha of .78 for the scales.

Timbrook and Graham (1994) reported no significant differences in the accuracy of prediction of extratest criteria for African American versus Caucasian individuals, with the exception of Scale 7 (Psychasthenia) for African American women, which underpredicted ratings of anxiety. However, Timbrook and Graham examined only five of the standard clinical scales and reported that they were unsure whether similar results would be found with the remaining clinical scales (Timbrook & Graham, 1994). Additionally, potential sampling bias may have occurred because of the inclusion of such a measure post hoc.

It is essential to place the extratest measures used in this study within context. The scales developed by Long (1993) were based on a modified version of the Katz Adjustment Scales (Katz & Lyster, as cited in Long et al., 1994). It seems very likely that this instrument and any measures derived from it would be systematically biased in the sense of a Eurocentric monocultural perspective of adjustment. In effect, using such an approach to evaluate the validity of the MMPI-2 is tautological.

No new information is gained, and the status quo of addressing (or, in this case, not addressing) test bias is maintained. Studies that use such approaches typically result in conclusions that favor the MMPI-2, and these conclusions are unsound.

Reed, Walker, Williams, McLeod, and Jones (1996) administered the MMPI-2 along with other measures of personality, achievement, and coping style to 78 female African American college students. Reed et al. found that 76% of their sample had elevated scale scores. Additionally, the authors concluded that such elevations might be more related to adaptive coping style rather than psychopathology and emphasized the importance of careful interpretation of MMPI-2 clinical profiles.

Using an inpatient sample, Arbisi, Ben-Porath, and McNulty (1998) compared the validity of MMPI-2 scores across African American and Caucasian individuals. Although some significant mean differences were revealed in the results of the study, there were no significant differences in the ability of MMPI-2 scores to account for variance in extratest measures incorporated in the study (Arbisi et al., as cited in Handel & Ben-Porath, 2000, p. 239). Arbisi et al. concluded that there is an absence of bias in the ability of the MMPI-2 scales to predict variance in supposedly relevant extratest criteria in African Americans. McNulty, Graham, Ben-Porath, and Stein (1997) examined the comparative validity of MMPI-2 scores for Caucasian and African American community mental health center clients. For males, African American individuals scored significantly higher than did Caucasian individuals on Scale L (Lie), whereas for females, African American individuals scored higher on Scale 9 (Hypomania). McNulty et al. included therapists' ratings of clients' symptoms and personality characteristics and examined the correlations of those ratings with MMPI-2 scale scores across both groups. They reported that correlations between MMPI-2 scale scores and extratest criteria were not significantly different between the groups. In their conclusions, McNulty et al. maintained that there was a lack of evidence for MMPI-2 test bias against African American individuals.

The studies conducted by Arbisi et al. (1998) and McNulty et al. (1997) may be biased, however, because the extratest measures and therapist ratings these authors used did not appear to exert any effort to address cultural differences in perspectives across their samples. Thus, if the criteria used were biased, any meaningful results could not be obtained or interpreted. Again, the authors' conclusions appear to be suspect based on the epistemological perspective of the extratest measures used. This simply underscores the need for the application of culturally and contextually informed extratest measures to meaningfully address the validity of using the MMPI-2 with diverse populations.

Maintaining the established trend of strikingly disparate cell sizes, Arbisi, Ben-Porath, and McNulty (2002) examined ethnic differences between African American ($n = 229$) and Caucasian ($n = 1,558$) psychiatric inpatients on the MMPI-2. The purpose of the study was to test for racial bias within the MMPI-2. Although the authors had no actual contact with the participants

in their study, they devised a record review form that included extratest information, such as presenting problems at intake, medication history, and *Diagnostic and Statistical Manual of Mental Disorders* (DSM, 3rd ed., rev.; DSM-III-R; 4th ed., DSM-IV; American Psychiatric Association [APA], 1987, 1994, respectively) multi-axial diagnoses. This form also included 87 items addressing range of affect, mood, cognition, and organic indicators that were developed by the researchers. It is important to note that research assistants reviewed patients' charts to provide ratings on the record review form. The ratings for the 87 items were based on rater reviews of the psychiatrist's interviews recorded in the charts (Arbisi et al., 2002). The authors used data from this form to create conceptually relevant criterion variables with which they correlated particular MMPI-2 scales.

Results of the analyses revealed clinically and statistically significant differences between the African American and Caucasian groups. African American male inpatients scored significantly higher than did Caucasian male patients on Scales F, 4, 6, 8, and 9 (Arbisi et al., 2002). Conversely, Caucasian male inpatients scored significantly higher on Scale K than did African American male inpatients. All differences were considered clinically significant with the exception of Scale K. Significant differences between the female subgroups were also found on several scales, such that African American women scored significantly lower on Scale K and significantly higher on Scales 6 and 9 (Arbisi et al., 2002). These differences were clinically significant as well.

The authors used a moderated multiple regression procedure to assess for racial bias. Results indicated evidence for overprediction of psychopathology for African American men on three basic scale criterion predictions, Scales 2, 8, and 9. For African American women, evidence for overprediction of psychopathology was found on two basic scales, Scales 4 and 9. Ironically, the authors did not elaborate on this evidence. However, they detailed the findings of slight underprediction of psychopathology in their African American subgroups and concluded that the MMPI-2 underpredicts psychopathology more frequently than it overpredicts psychopathology (Arbisi et al., 2002).

Once again, a major limitation of this study relates to the extratest measures the researchers used. The chart reviews conducted by research assistants relied on information recorded on intake by mental health workers. Systematic bias seems more than possible because no effort was exerted to account for cultural differences in perspectives across their groups. Simply dividing the participants by race and gender and then imposing extratest measures or ratings on participants' charts as rated by research assistants is not enough to adequately address cultural differences. The implicit assumption is that the extratest variables are the same and mean the same thing across the racial/ethnic groups. That is a dangerous leap for two reasons. First, the criterion measures used in this study did not take into consideration any contextual factors. Second, there was a complete lack of

any type of measure or even a single question having to do with the cultural perspective of the participants.

Thus, the question remains: If a study of racial bias within the MMPI-2 uses systematically biased criterion measures, how is one able to properly address the research problem? Unexamined assumptions about the objectivity of extratest measures do nothing to address the issue of bias within the MMPI-2. In fact, use of these supposedly objective extratest measures appear to simply provide blind support to the validity of the MMPI-2 with ethnic minority groups. If anything, these measures simply validate their own epistemological framework from which they and the MMPI-2 were derived, evidencing a dangerous yet consistently used tautological approach in these types of investigations.

Concerning studies of the MMPI-2 with Asian Americans, the research is limited at best. The studies that have been published used primarily convenient college student samples. Sue, Keefe, Enomoto, Durvasula, and Chao (1996) compared MMPI-2 profiles of Asian American and Caucasian college students. The investigators categorized the Asian American group into lower and higher acculturation subgroups. They found that low-acculturation Asian American college students scored significantly higher than did Caucasian students on several MMPI-2 scales. Sue et al. (1996) concluded that their results might indicate cultural bias in the MMPI-2 scores of Asian Americans, especially those of lower acculturative status. To further examine the MMPI-2's generalizability, Handel and Ben-Porath (2000) suggested the importance of future research to investigate within-group differences among Asian Americans on the MMPI-2 because there are several cultural subgroups under the umbrella term *Asian American*.

In contrast to MMPI-2 research conducted with other ethnic minority groups, American Indian individuals have been drastically ignored. Few published studies exist that directly examine the use of the MMPI-2 with nonclinical samples of American Indians. Included in these studies are comparisons between MMPI-2 scale scores of American Indian individuals and those of Caucasian individuals. With barely a handful of published studies, clearly there is a substantial deficit in the research literature.

Robin, Greene, Albaugh, Caldwell, and Goldman (2003) compared a Plains nonclinical sample ($n = 297$) and a Southwest nonclinical sample ($n = 535$) with the MMPI-2 normative group and found clinically significant score elevations on the following scales: L (Lie), F (Infrequency), 1 (Hypochondriasis), 4 (Psychopathic Deviate), 8 (Schizophrenia), and 9 (Hypomania). Conducting within-group analyses, Robin et al. (2003) found no clinically significant differences on any of the MMPI-2 scales between the American Indian samples.

Matching for gender, age, and education, results indicated a reduction in the magnitude of scale differences observed between the American Indian samples and the MMPI-2 normative group (Robin et al., 2003). Although the

magnitude of the differences was reduced, the scale differences still exceeded 5 T points on each of the scales. These results suggest differences between American Indian and White individuals do not dissipate completely when groups are matched on socioeconomic status variables. The researchers suggested that such substantive cross-cultural differences warrant further investigation of empirical correlates because their study could not directly address the issue of test bias through comparison of mean group profiles alone.

Robin et al. (2003) also included a modified version of the Schedule for Affective Disorders and Schizophrenia–Lifetime version (SADS-L; Spitzer, Endicott, & Robins, as cited in Robin et al., 2003). The SADS-L interview, administered to all participants in the nonclinical American Indian samples, was used to diagnose psychiatric disorders based on *DSM-III-R* (APA, 1987) criteria. The researchers used descriptive phrases from this interview to assess empirical correlates for the MMPI-2 scales. Such phrases generally reflect negative affective/general distress, symptoms of alcohol/drug use, and antisocial behaviors (Greene, Robin, Albaugh, Caldwell, & Goldman, 2003). Analysis of the interviews and empirical correlates were the subject of a companion follow-up study conducted by Greene et al. (2003).

In their investigation of empirical correlates of MMPI-2 scales with the same nonclinical American Indian samples, Greene et al. (2003) suggested that the differences reported by Robin et al. (2003) may actually reflect behaviors and symptoms that are the result of the American Indian participants' adverse backgrounds and current conditions. To test this hypothesis, the researchers used data gathered from SADS-L interviews and independently assigned clinical (*DSM-III-R*; APA, 1987) diagnoses derived from SADS-L narrative interview documents. Greene et al.'s rationale for using the SADS-L is that it has been found to be reliable when administered to American Indians by clinicians experienced in providing psychiatric assessment to American Indian individuals. The authors listed several studies in support of this statement.

Results of their investigation revealed significant correlations between several MMPI-2 scales related to the descriptive phrases of the SADS-L interview, which reflected negative or depressive affect, antisocial behaviors, and symptoms of alcohol or drug use (Greene et al., 2003). Specifically, significant correlates were found with antisocial symptoms on Scales 4 and 9 and negative affect and generalized distress on Scales 7 and 8. It is interesting that Scale 2 was found to have few significant correlates with negative affect and generalized distress in men; however, none were found with women (Greene et al., 2003).

Examining the significant correlates, Greene et al. (2003) concluded that considering content of MMPI-2 scales, American Indian participants reported symptoms and behaviors that would be anticipated by observed scale score elevations. Therefore, Greene et al. cautioned that such elevations should

not be simply dismissed as being due to test bias but should be considered as accurate reflections of behavioral and symptomatic experiences.

Several critiques of the Greene et al. (2003) follow-up study are warranted. First, empirical correlates account for only one aspect of determining cultural bias in testing. For reasons stated previously, it should not be considered the only or even the best manner in which to determine cultural bias.

Second, although Greene et al. (2003) supported their rationale for using the SADS-L with their nonclinical American Indian samples with previous research, it is very questionable whether diagnoses of nonclinical participants that were derived from the SADS-L would be endorsed by traditional healers or other members of the respective tribes. Nor was there any attempt to validate this data from a cultural framework by simply checking with the participants themselves, members of their families, or other community leaders or elders as to the accuracy of such ratings or diagnoses.

Finally, the correlates derived from the SADS-L interview, as conceptualized by the researchers, may not be conceptualized in the same way by the participants. Therefore, it is also questionable and most unlikely that Greene et al. (2003) and tribal members shared the same epistemological framework with regard to psychological adjustment. As Gone (2001) stated,

With regard to the cross-cultural validation of “established” diagnostic categories . . . it is not yet clear that any such category as currently conceptualized within conventional neo-Kraepelinian psychiatry rests upon sufficient scientific evidence to withstand the difficult questions posed by the local and variable cultural construction of affect, distress, and disorder. Again, the absence of compelling empirical anchors renders the field an unstable “science” shaped more by theoretical assumptions and methodological commitments than by *consensually* [emphasis added] identified validators for purported disorders. In this state of flux, it may simply be that one’s position on the existence “in nature” of various identifiable psychiatric disorders is ultimately determined by a priori conceptual assumptions (as opposed to, say, compelling scientific evidence that speaks beyond its own self-contained language game). (p. 122)

Responding to these significant limitations, Pace et al. (2006) investigated the normative validity of the MMPI-2 using two distinct nonclinical American Indian samples, comparing these samples to the normative standard. Although the focus of the study was not an examination of empirical correlates, Pace et al. suggested that using extratest measures derived from a majority culture perspective in MMPI-2 research disables such work from fully validating the instrument with American Indian groups. Other researchers (Dana, 2000; Hoffman, Dana, & Bolton, 1985) have made similar conclusions.

Hill et al. (2010) conducted a companion follow-up study of the cultural validity of the MMPI-2 that used a mixed methods approach privileging an indigenous perspective within the research. This study examined item-level differences between one of the American Indian samples from Pace et al. (2006) and the MMPI-2 normative group. Hill et al. used an indigenous

framework for the study that emphasized the involvement of participants from the very beginning of the project through the end, including manuscript development and submission. Reflecting on this process, Hill et al. emphasized that not only is it imperative to use culturally valid measures in this type of research but also methods deemed as culturally appropriate and valid by the participating/partnering community. Furthermore, the authors cautioned that when participating Native Nations receive no direct benefit from assessment research or any other type of psychological research, the research should not be conducted (Hill et al., 2010).

discussion

The central issue surrounds researchers' use and application of extratest measures that masquerade as etics in formulating diagnoses or conceptualizations of pathology and behavior. In fact, it is quite apparent that such measures are actually European American emics forcefully imposed on persons from cultural, racial, and ethnic minority groups within research contexts. Results from these contexts are then generalized to psychological practice with sometimes devastating consequences. For example, given the use of the MMPI-2 across a broad range of settings and for a variety of purposes (e.g., personnel selection, child custody evaluations, parole review, and treatment planning), its lack of cultural validity can lead to devastating consequences when the test is applied in a decontextualized manner, devoid of cultural variables and other influences.

In his extensive review of MMPI/MMPI-2 research, Greene (1987) concluded that the best way to determine whether the consistently observed scale differences between various racial/ethnic groups and the overall normative group on the MMPI-2 are the result of cultural bias is through investigation of the scales' empirical correlates. We have attempted to repeatedly address the problematic issue of empirical correlates through this brief review of MMPI-2 studies with ethnic minorities.

In research that addresses the issue of cultural bias within the MMPI-2, leading MMPI-2 researchers have strenuously advocated for investigations of empirical correlates (Arbisi et al., 1998; Butcher & Williams, 2000; Graham et al., 1999; Greene, 2000; Greene et al., 2003; Robin et al., 2003). Although these researchers consider this to be a strong method for determining cultural bias within the MMPI-2, this approach is quite flawed because of the glaringly unexamined ethnocentric assumption inherent within it: Extratest measures used to determine empirical correlates are "universally" applicable. Unfortunately, the assumption of universality is fundamental within Western psychology. Most theories within the field are based on this assumption. The problem is that Western psychology and its derivatives are cultural products that reflect and reproduce a specific (Western) cultural context (Cushman, 1993). The history of Western psychology demonstrates its power to define

supposed “universal” constructs, dismiss conflicting conceptualizations, and maintain its hegemonic status.

Ethnocentric assumptions allow researchers to define correlates of psychological adjustment and maladjustment from their own (Western) epistemological framework(s) without having to ever acknowledge or even take into account epistemological frameworks different from their own. Indeed, Hathaway and Monachesi (1953) explicitly privileged the point of view of the clinician (most assuredly trained in the Western tradition) in determining characteristics of psychological abnormality. This assumption precludes any acknowledgment of differing cultural perspectives on psychological maladjustment. In fact, this assumption renders invisible local cultural knowledge regarding psychological adjustment or maladjustment. Whether consciously admitted or not, the purpose of privileging this system is to ensure constant validation and perpetuation of its own hegemony. What is very clear from this brief review is that cultural and contextually informed extratest measures are desperately needed because the MMPI-2 does not assess psychopathology as defined from non-Western worldviews or cultural contexts. The absence of these essential perspectives on psychopathology in the form of extratest measures upholds the foundational question of the MMPI-2’s validity.

In her seminal work, *Decolonizing Methodologies: Research and Indigenous Peoples*, Maori scholar Linda Smith (1999) reflected on the baggage of Western research:

From an Indigenous perspective Western research is more than just research that is located in a positivist tradition. It is research which brings to bear, on any study of Indigenous peoples, a cultural orientation, a set of values, a different conceptualization of such things as time, space and subjectivity, different and competing theories of knowledge, highly specialized forms of language, and structures of power. (p. 42)

It appears that MMPI-2 research of empirical correlates adheres strongly to this unfortunate tradition. This adherence prevents the research from effectively addressing issues of cultural bias within the instrument (Hill et al., 2010). Justifying the use of extratest measures derived from a Western framework in MMPI-2 research with tautological logic only privileges a covert ideological framework (ethnocentric research paradigm) and perpetuates the seemingly endless cycle of studies that do everything but adequately address cultural bias.

conclusions and future research

A major critique of the reviewed MMPI-2 studies was focused on the use of extratest measures with origins in a purely Western worldview. To address significant ethical issues, future studies that investigate the validity of using the MMPI-2 with persons from diverse cultural backgrounds must incorporate

extratest measures grounded in relevant cultural systems. Using culturally and contextually informed extratest measures to identify empirical correlates and validate the MMPI-2 will improve the research in this area by leaps and bounds.

references

- American Psychiatric Association. (1987). *Diagnostic and statistical manual of mental disorders* (3rd ed., rev.). 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. L. (1998, August). *Impact of ethnicity on the MMPI-2 in inpatient settings*. Paper presented at the 106th annual meeting of the American Psychological Association, San Francisco, CA.
- 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. doi:10.1037//1040-3590.14.1.3
- Butcher, J. N. (1999). Foreword. In J. R. Graham, Y. S. Ben-Porath, & J. L. McNulty, *MMPI-2 correlates for outpatient community mental health settings*. Minneapolis: University of Minnesota Press.
- Butcher, J. N. (2000). Dynamics of personality test responses: 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). *Minnesota Multiphasic Personality Inventory (MMPI-2): Manual for administration and scoring*. Minneapolis: University of Minnesota Press.
- Butcher, J. N., Graham, J. R., Williams, C. L., & Ben-Porath, Y. S. (1990). *Development and use of the MMPI-2 content scales*. Minneapolis: University of Minnesota Press.
- Butcher, J. N., & Williams, C. L. (2000). *Essentials of MMPI-2 and MMPI-A interpretation* (2nd ed.). Minneapolis: University of Minnesota Press.
- Cushman, P. (1993). Psychotherapy and moral discourse. *Theoretical and Philosophical Psychology, 13*, 103–113.
- Dana, R. H. (2000). *Handbook of cross-cultural and multicultural personality assessment*. Mahwah, NJ: Erlbaum.
- Derogatis, L. R. (1983). *SCL-90-R: Administration, scoring, and procedures manual—II for the R(revised) version*. Towson, MD: Clinical Psychometric Research.
- Gone, J. P. (2001). Affect and its disorders in a Northern Plains Indian community: Issues in cross-cultural discourse and diagnosis. *Dissertation Abstracts International: Section B. Sciences and Engineering, 62*(6-B), 2957.
- Graham, J. R. (1993). *MMPI-2: Assessing personality and psychopathology* (2nd ed.). New York, NY: Oxford University Press.
- Graham, J. R., Ben-Porath, Y. S., & McNulty, J. L. (1999). *MMPI-2 correlates for outpatient community mental health settings*. Minneapolis: University of Minnesota Press.
- Greene, R. L. (1987). Ethnicity and MMPI performance: A review. *Journal of Consulting and Clinical Psychology, 55*, 497–512.
- Greene, R. L. (2000). *The MMPI-2: An interpretive manual* (2nd ed.). Boston, MA: Allyn & Bacon.
- Greene, R. L., Robin, R. W., Albaugh, B., Caldwell, A., & Goldman, D. (2003). Use of the MMPI-2 in American Indians: II. Empirical correlates. *Psychological Assessment, 15*, 360–369.
- Handel, R. W., & Ben-Porath, Y. S. (2000). Multicultural assessment with the MMPI-2: Issues for research and practice. In R. H. Dana (Ed.), *Handbook of cross-cultural and multicultural personality assessment* (pp. 229–245). Mahwah, NJ: Erlbaum.
- Hathaway, S. R., & McKinley, J. C. (1940). A multiphasic personality schedule (Minnesota): I. Construction of the schedule. *Journal of Psychology, 10*, 249–254.
- Hathaway, S. R., & Monachesi, E. D. (1953). *Analyzing and predicting juvenile delinquency with the MMPI*. Minneapolis: University of Minnesota Press.
- Hill, J. S., Pace, T. M., & Robbins, R. R. (2010). Decolonizing personality assessment and honoring indigenous voices: A critical examination of the MMPI-2. *Cultural Diversity and Ethnic Minority Psychology, 16*, 16–25.

- Hoffman, T., Dana, R. H., & Bolton, B. (1985). Measured acculturation and MMPI-168 performance of Native American adults. *Journal of Cross-Cultural Psychology*, 16, 243–256.
- Katz, M. M., & Lyerly, S. B. (1963). Methods for measuring adjustment and social behavior in the community: I. Rationale, description, discriminative validity and scale development. *Psychological Reports*, 13, 503–535.
- Long, K. A. (1993). *The relations between socioeconomic status and the MMPI-2 scores of men and women* (Unpublished doctoral dissertation). Kent State University, Kent, OH.
- Long, K. A., & Graham, J. R., & Timbrook, R. E. (1994). Socioeconomic status and MMPI-2 interpretation. *Measurement and Evaluation in Counseling and Development*, 27, 158–177.
- McNulty, J. L., Graham, J. R., Ben-Porath, Y. S., & Stein, L. A. R. (1997). Comparative validity of MMPI-2 scores of African American and Caucasian mental health center clients. *Psychological Assessment*, 9, 464–470.
- Meehl, P. E. (1945). The dynamics of “structured” personality tests. *Journal of Clinical Psychology*, 1, 296–303.
- Nichols, D. S. (1992). Review of the Minnesota Multiphasic Personality Inventory-2. In J. J. Kramer & J. C. Conoley (Eds.), *The eleventh mental measurements yearbook* (pp. 562–565). Lincoln: University of Nebraska Press.
- Nichols, D. S., Padilla, J., & Gomez-Maqueo, E. L. (2000). Issues in the cross-cultural adaptation and use of the MMPI-2. In R. H. Dana (Ed.), *Handbook of cross-cultural and multicultural personality assessment* (pp. 247–266). Mahwah, NJ: Erlbaum.
- Pace, T. M., Robbins, R. R., Choney, S. K., Hill, J. S., Lacey, K., & Blair, G. M. (2006). A cultural contextual perspective on the validity of the MMPI-2 with American Indians. *Cultural Diversity and Ethnic Minority Psychology*, 12, 320–333.
- Reed, M. K., Walker, B., Williams, G., McLeod, S., & Jones, S. (1996). MMPI-2 patterns in African American females. *Journal of Clinical Psychology*, 52, 437–441.
- Robin, R. W., Greene, R. L., Albaugh, B., Caldwell, A., & Goldman, D. (2003). Use of the MMPI-2 in American Indians: I. Comparability of the MMPI-2 between two tribes and with the MMPI-2 normative group. *Psychological Assessment*, 15, 351–359.
- Smith, L. T. (1999). *Decolonizing methodologies: Research and indigenous peoples*. New York, NY: Zed Books.
- Sue, S., Keefe, K., Enomoto, K., Durvasula, R., & Chao, R. (1996). Asian American and White college students’ performance on the MMPI-2. In J. N. Butcher (Ed.), *International adaptations of the MMPI: Research and clinical applications* (pp. 206–220). Minneapolis: University of Minnesota Press.
- Timbrook, R. E., & Graham, J. R. (1994). Ethnic differences on the MMPI-2? *Psychologica Assessment*, 6, 212–217.