JOURNAL OF PERSONALITY ASSESSMENT, *87*(2), 217–222

Syndromal Complexity, Paradigm Shifts, and the Future of Validation Research: Comments on Nichols and Rogers, Sewell, Harrison, and Jordan

Nathan C. Weed
Department of Psychology
Central Michigan University

In this comment, I address a number of the points raised in the reviews of the MMPI–2 Restructured Clinical (RC; Tellegen et al., 2003) Scales by Nichols (2006/this issue) and Rogers, Sewell, Harrison, and Jordan (2006/this issue), and I advocate for changes in assessment validation research. There is little evidence that the "syndromal complexity" Nichols ascribes to the original MMPI–2 (Butcher, Dahlstrom, Graham, Tellegen, & Kaemmer, 1989) Clinical Scales is worth preserving. Although their construction does not constitute the paradigm shift claimed by Rogers et al., the RC Scales are promising, psychometrically defensible measures of core features of the original MMPI–2 Clinical Scales. However, validation of inferences from multiscale inventories such as the MMPI–2 is limited at present by a disconnection between the integrative manner in which MMPI–2 profiles are interpreted and the scale-by-scale nature of most MMPI–2 validation studies. Q-sort procedures show promise for operationalizing integrated MMPI–2 interpretations, with both research and teaching applications.

It is all too easy to forget the tremendous achievement represented by the original construction of the Minnesota Multiphasic Personality Inventory (MMPI; Hathaway & McKinley, 1943). In an age in which waiting 10 seconds for one's computer to invert a large correlation matrix constitutes legitimate cause for frustration, it might seem churlish to hold the psychometric products of the 1930s unforgivingly to modern standards. Far worse than churlishness, however, is dogged insistence that the old ways are the best ways, especially in the face of clear progress. The development of the MMPI–2 Restructured Clinical (RC; Tellegen et al., 2003) Scales represents significant theoretical and psychometric advance for the MMPI–2 (Butcher, Dahlstrom, Graham, Tellegen, & Kaemmer, 1989); Nichols's (2006/this issue) critique of these scales, although raising a number of important issues, fails to appreciate just how badly the restructuring was needed.

SYNDROMAL COMPLEXITY

Nichols acknowledges that the high intercorrelation among the original Clinical Scales is problematic, and he helpfully reviews some alternative approaches to identifying and removing the shared variance. Nichols's preference is for empirical solutions (such as use of Welsh's, 1956, A scale) that disclaim understanding of the nature and source of the intercorrelation. Nichols distrusts Tellegen et al.'s understanding of the shared variance as comprising the Pleasantness–Unpleasantness axis of the affective circumplex and worries that their theoretically grounded solution distorted the nature of the resulting RC Scales. Surprisingly, Nichols's worries on this score are not eased by the fact that 14 of the 24 items comprising Tellegen et al.'s operationalization of this dimension (Demoralization, abbreviated RCd) are found on the A scale, nor that the correlation between RCd and A was .95 in Nichols's own large sample. Nichols's attempts to show that RCd is "depressively biased" notwithstanding, it is difficult to imagine that scale restructuring using A would have produced meaningfully different results.

Nichols does not, however, acknowledge a second major failing of the original Clinical Scales: their possession of numerous poorly performing items. Such items are sometimes excused as "subtle," a label that suggests that despite their poor face validity they are nevertheless effective. Empirical research does not support this suggestion. Aggregates of subtle items such as the Wiener–Harmon Subtle Scales (Wiener, 1948), scoring keys for which are no longer distributed by the University of Minnesota Press, rarely performed well in validation studies, although some advocates have attempted to

explain their empirical failures. Some (e.g., Hollrah, Schlottmann, Scott, & Brunetti, 1995) speculated that subtle items failed empirically because the criteria with which they were related were too "obvious," an argument that has never to my knowledge been accompanied by explanation of how the concept of subtle "criterion" might reasonably be defined. Others (e.g., Nichols, 2001) have argued that the role of subtle items, rather than helping to predict criteria, has been to augment the discriminant validity of the Clinical Scales (i.e., reduce their intercorrelation). Unfortunately, this argument provides no hint of how subtle items serving this role might be distinguished from useless items, which would of course also serve to reduce intercorrelation among scales. Whatever the explanation for the empirical failure of these item aggregates, few are now willing to advocate the use of these subtle scales in isolation.

Although the effect of the subtle scales within the full Clinical Scales is a matter of some debate, at least some research (e.g., Weed, Butcher, & Ben-Porath, 1990) has indicated that their presence serves to attenuate the external validity of the Clinical Scales. It is not disputed that subtle items contribute to unreliability in Clinical Scale scores (as evidenced by the difference between reliability estimates of the Clinical Scales with and without the Wiener–Harmon subtle items). Nichols, however, appears to maintain that these items, although noisy, although lacking in face validity, and although not useful in isolation are somehow critical components of the Clinical Scales. Nichols describes the original Clinical Scales as "models" of the clinical syndromes they were developed to measure. Likening the items of the Clinical Scales to the *Diagnostic and Statistical Manual of Mental Disorders* (4th ed. [*DSM–IV*]; American Psychiatric Association, 1994) diagnostic criteria, Nichols praises the heterogeneity of item content and laments the damage done by restructuring to these "multivariate models."

This choice of analogy is surprising, as it is more appropriate to a description of rational scale development. Although the *DSM–IV* diagnostic criteria may be multidimensional, they were carefully and consciously selected and weighted, guided in large part by expert human judgment; MMPI–2 Clinical Scale items were of course selected empirically. The empirical method was chosen for the Clinical Scales, as Nichols (2001) wrote elsewhere, as "a practical means of *avoiding* [italics added] theory and sidestepping rational or intuitive guidance" (p. 2). The MMPI–2 Clinical Scales are not models of psychopathology by any conventional sense of the term. No effort was made, for example, to guarantee that the most important features of depression were reflected within Clinical Scale D, let alone in careful balance. Furthermore, no effort was made to prevent the inclusion of items with content that lacked theoretical relevance to Major Depression. (After all, what model of Major Depression includes as symptoms a marked failure to tease animals, the absence of vomiting blood more days than not, and a clinically significant sweat deficiency?) Scale D is not a neatly ordered multivariate model of depression; it is a dimensional cacophony, probably overrepresenting facets here and underrepresenting there, and certainly comprising both good items and poorly performing items.

Nichols's characterization thus inaccurately recasts a grave flaw of the MMPI–2 Clinical Scales in a benign or even favorable light. Rather than highlighting their internal chaos, Nichols describes the scales as reflecting "syndromal complexity," a phrase that might sound euphemistic if it were not clear that he is serious. One might as well speak of the Clinical Scales as being "charmingly free of typical psychometric restraints," characterized by "sassy heterogeneity," or filled to the brim with "intrascale insouciance." An "interesting" dimensionality may make for an entertaining evening with a component loadings matrix, but these days, one expects more from clinical measurements: more power per item, more dimensional clarity, and more freedom from noise.

Relatedly, Nichols decries Tellegen et al.'s initial external validation efforts as unduly focused on "molecular" ratings of concrete symptoms. These are "soft targets" for the RC Scales, Nichols argues, because they, like the RC Scales, are unidimensional. Nichols predicts that the original Clinical Scales will be discovered to be superior to the RC Scales in the prediction of complex criteria such as psychiatric diagnosis. This seems to me an unlikely proposition, but it is at the least quite testable and does not rely on the concept of "subtle criteria." Whether Nichols's hypothesis is eventually borne out, it is worth considering which set of criteria—(a) the "soft targets" of symptom expression or (b) psychiatric diagnoses—serve more appropriately as criteria in MMPI–2 scale validation studies. A review of MMPI–2 interpretive texts will reveal that clinicians are typically taught to infer from Clinical Scale scores not diagnostic status but symptoms and other observables that are quite similar to the kinds of criteria Tellegen et al. used in their initial RC Scale validation. Whether they are soft targets, they are surely the most appropriate targets for Clinical Scale validation.

It is difficult to share a number of Nichols's other concerns as well. Nichols's complaint of "construct drift" (the extent to which restructuring resulted in shifting a scale away from its clinical core) presupposes knowledge of the nature of the clinical core; one person's drift could be another's "zeroing in" or "fleshing out." Nichols's unelaborated worries about the choice of principal components analysis (PCA) and/or varimax in RC Scale exploration seem also to constitute arguing from a vacuum. There are doubtless hundreds of decision points in a project of this magnitude. Even in the construction of the original MMPI Clinical Scales, a project many of whose decisions were supposedly simplified via the empirical method, Hathaway (1956; as cited in Butcher, 2000) acknowledged the enormity of the task complexity and wrote

> It is impossible to describe fully the steps in selection of the MMPI scale items. … [The] multiple checking of items and scales is probably the most characteristic general procedure

> relative to the derivation. Beyond this, specific steps in scale development were so varied that they cannot be completely described. (p. 49)

Although it is important to ask critical questions about a scale's developmental methodology, it is not particularly constructive to advance a given methodological decision as a source of grave concern without offering an explanation of how the alternatives would have produced meaningfully different results. Why should we expect the use of tetrachorics or quartimax or different loadings criteria to have altered the nature of RC Scales and in what way?

Fortunately, one misgiving of Nichols that is quite easy to share is his concern that scores on RC Scales might be mistakenly interpreted as having the same meaning as scores on the original Clinical Scales. Differences in correlate patterns between the original scales and the RC Scales are certain to be discovered, whether due to "defects in design and composition" of the RC Scales, as Nichols alleges, or more charitably, due to the excision of poorly performing items and liberation of the core clinical constructs from the heavy burden of demoralization. Clearly, additional research will be necessary to help guide interpretation of the new scales (e.g., Simms, Casillas, Clark, Watson, & Doebbeling, 2005). This is a theme also sounded in the article by Rogers, Sewell, Harrison, and Jordan (2006/this issue), who emphasize a particular need for the identification of clinical characteristics uniquely associated with the RC Scales.

PARADIGM SHIFTS

Although there is much more with which to agree in Rogers et al., these authors go too far to credit the RC Scales with representing a "paradigmatic shift in scale development" necessitating "fundamental changes in interpretation." To be sure, the Tellegen et al. strategy for decontamination of the Clinical Scales was quite clever, using a set of demoralization items within PCAs as a sort of bait to lure away from each scale items that were not sufficiently specific to the scale core. This strategy could probably also be considered unique in that aside from the MMPI–2, there are few other personality inventories whose scales would be appropriate candidates for such scale purification. However, although it might be appropriate to speak of the development of the RC Scales, the methods used do not properly constitute a general scale development strategy in that the main goal of the RC project was to modify existing scales rather than to create completely new scales. Use of the Tellegen et al. procedures for creating a novel inventory would be unnecessarily cumbersome and roundabout; scale repair should not be necessary if scales are constructed from the start according to modern psychometric standards.

Similarly, there should be no expectations that the RC Scales will bring about fundamental changes in interpretation strategy. To be sure, RC Scales should be interpreted differently from the original scales in accordance both with their theoretical underpinnings and the empirical scale–correlate relationships that accumulate. However, attention to the empirical validation literature is not a novel practice, nor should a "dimensional approach" to interpretation be considered new to the MMPI–2. Although diagnostic categories drove the construction of the original Clinical Scales, many of the empirical research studies that have guided MMPI–2 scale interpretation have consisted of linear correlations between scale scores and external criteria. Few modern validation studies have relied exclusively on dichotomized distributions of MMPI–2 scores; fewer still provide any evidence whatsoever of curvilinearity in the scale–correlate relationship. A dimensional approach to interpretation has always been appropriate whether this has been recognized in clinical lore.

THE FUTURE OF MMPI–2 VALIDATION RESEARCH

Perhaps one context in which something approximating a genuine paradigm shift is needed is research procedures employed for validating interpretations from multiscale inventories such as the MMPI–2. Evident in Nichols's and Rogers et al.'s reviews and in the RC Scales manual, is the traditional emphasis on validating each scale of the MMPI–2 one at a time. It is important, of course, to understand how single scales function, both during development and for interpretation. On the other hand, MMPI–2 interpretation is rarely executed one scale at a time. Instead, MMPI profiles are interpreted via the integration of inferences deriving from a large number of scales. Sometimes these inferences are additive, sometimes they involve reconciling conflicting pieces of information, and sometimes (as Rogers et al. point out) the sources of inference are redundant. If the typical product of MMPI–2 interpretation, then, is not a scale-by-scale recitation of correlates but a descriptive account of the examinee's personality and clinical symptom expression, it is this descriptive narrative that ought to be at the center of validation efforts. How, though, can one operationalize, let alone validate, a descriptive account?

The Interpretive Q-Sort

An approach with considerable potential for the validation of integrated inferences from multiscale personality inventories, including the MMPI–2, involves the use of interpretive Q-sorts (Weed & Noland, 1998). The interpretive Q-sort is a standardized collection of personality descriptors that are sorted into a preset distribution to interpret of the results of a personality test. For example, a set of items ("Q-set") currently in use in research and teaching applications involving the MMPI–2, the Midwestern Q-sort (Williams & Weed, 2003), comprises

100 statements designed to reflect the most common inferences made on the basis of MMPI–2 results. In the development of this Q-set, items were judged as likely to reflect common MMPI–2 inferences if they appeared prominently in Graham's (1999) interpretive text as descriptive correlates of scores on frequently used MMPI–2 scales. Attempts were made to express items in terms of observable behaviors (e.g., "Cries frequently," and "Complains about aches and pains"), to use nontechnical language (e.g., "Believes things that obviously are not true" rather than "Has delusions"), to cover presumed correlates of a breadth of MMPI–2 scales (all Validity, Clinical, Content, and substance abuse scales), and to balance positive and negative keying direction for desirability (e.g., the items "Obeys authority" and "Manipulates others" were both written to reflect common inferences about scores on MMPI–2 Scales Pd and ASP).

To complete the Midwestern Q-sort (a Web-accessible Authorware application), the 100 statements are sorted by the test interpreter into seven categories representing 7 points on a continuum from 1 (the item is *most descriptive* of the individual producing the MMPI–2 profile) to 7 (*least descriptive*). The shape of the imposed distribution is symmetric, with fewer items permitted as extreme descriptors; in category order, the sorted distribution is 5, 10, 20, 30, 20, 10, and 5 items. The resulting Q-sort description is quantified by assigning each of the 100 items a number corresponding to the category into which the item was sorted. For example, judgments that are *most descriptive* of the individual producing the MMPI–2 profile are assigned a score of 1. This vector of 100 category scores may then be compared to other Q-sort vectors via a "Q-correlation," which is simply the product–moment correlation between the pair of 100 element vectors. For example, a Q-correlation of +.60 between interpretive Q-sorts obtained by interpreting two different MMPI–2 profiles would indicate that the two examinees' MMPI–2 results suggest quite similar personality descriptions. A Q-correlation of +.80 between interpretive Q-sorts obtained from two independent sorters of a single MMPI–2 profile would indicate strong agreement between the interpreters about the descriptive meaning of the MMPI–2 profile. Note that such information is not available at all in the context of traditional single-scale validation. Although it is common enough to wonder about the interclinician reliability of personality test interpretation, this important property of clinical assessment is rarely assessed formally (Deskovitz, Weed, & Williams, 2005), probably for lack of familiarity with an appropriate method. (I discuss related teaching applications of the interpretive Q-set below.)

Research Applications of the Interpretive Q-Sort

Validation research using an interpretive Q-sort requires at least two sets of Q-sorts: one providing personality descriptions based on interpretation of personality test results (e.g., Q-sorts based on MMPI results) and one providing personality descriptions based on actual knowledge about the personality features of the examinees (e.g., a Q-sort description provided by a spouse, a therapist, or a close friend). The latter set of Q-sorts, which are completed without knowledge of the test results, serves as a criterion measure against which to gauge the accuracy of the test interpretation. To quantify the validity of test inferences, correlations are computed between pairs of corresponding Q-sorts (description based on test results with description based on knowledge of the examinee), thereby producing a Q-correlation for each examinee. The mean, median, range, and standard deviation of these Q-correlations can be taken to characterize the validity of test interpretation in the obtained sample.

This Q-sort validation technique played an important role in the early history of the MMPI (e.g., see Little & Shneidman, 1959; Meehl, 1956, 1960; Sines, 1959) and sees periodic surges of interest (e.g., Meyer, Mihura, & Smith, 2005; Ozer, 1993; Shedler & Westen, 1998). It is unclear, however, why this technique has not become the dominant methodology for validating clinical inferences from personality test results given its unique ability to evaluate test inferences in a form so compatible with that of actual clinical interpretation. Like traditional single-scale validation, this method permits test validation against descriptions from multiple sources: nonprofessionals such as peers and spouses; service providers such as therapists, teachers, and clergy; as well as self-report.

Different types of test inferences may also be validated by using descriptive Q-set items that are specific to a test application. For example, rather than validating general MMPI–2 personality descriptions, an investigator may choose to develop a set of Q-items that restricts personality description to salient characteristics or concerns of a particular examinee group such as substance abuse populations (e.g., "Owns up to social consequences of substance abuse"), psychiatric inpatients (e.g., "Threatens or intimidates other patients"), or college counselees (e.g., "Prepares for exams conscientiously") or that restricts personality description to statements about response to treatment (e.g., "Appears comfortable in groups" or "Benefits from to muscle relaxation training"). Young, Weed, and Williams (2005) recently described the multistage development of a Q-set designed to evaluate MMPI–2 interpretations of substance abusing clients. First, items were generated based on a review of assessment considerations most commonly cited in the substance abuse research literature. Next, items were reviewed and rated by experts according to (a) their relevance to substance abuse and (b) accessibility via MMPI–2 interpretation. Finally, the Q-set was refined by statistical methods at the both item and instrument levels.

Perhaps the greatest potential benefit of Q-sort methodology for test validation research is its capacity to facilitate examination of relative interpretive accuracy under a variety of interpretive conditions. By varying the procedures under which test interpretation is performed, researchers may learn

which interpretive practices lead to the best integrative products. For example, Kwon, Noland, and Weed (2006) recently examined the relative validity of MMPI–2 interpretations when based on either (a) the original Clinical Scales, (b) the RC Scales, or (c) both sets of scales. MMPI–2 profiles from 24 inpatients were each interpreted under these three conditions by three raters whose Q-sorts were aggregated and correlated with Q-sorts performed by therapists of the MMPI–2 examinees. Results revealed no substantial differences between either the reliability or the validity of interpretations obtained under the three conditions, tentatively suggesting that despite the different organizations, similar information can be obtained from either set of scales.

Teaching Applications of the Interpretive Q-Sort

Another important application of this validation methodology is to the teaching of personality assessment interpretation. Presently, it appears that the most common way to teach appropriate test interpretation to prospective test users is by providing trainees interpretive practice and feedback in the form of written reports. Written reports, however, have a number of instructional limitations, including the following: (a) their preparation and evaluation are very time consuming, thereby limiting the number of opportunities for practice; (b) they reflect more than raw interpretive skills, confounded as they are with such ingredients as writing mechanics, style, and organization; (c) they allow limited feedback about interpretive accuracy, which may be delayed, inconsistent, or too general to be of corrective use; and (d) they encourage bad interpretive habits such as assuming all interpretive statements in a narrative are equally descriptive and disconnecting interpretation from clinical hypothesis testing.

The interpretive Q-sort can be used efficiently to address some of these limitations of written reports in teaching personality test interpretation. For a given profile of personality test results, the instructor can complete a Q-sort to serve as the expert criterion against which students' interpretations are evaluated via Q-correlation. This provides an exercise with the following benefits: (a) it can be completed relatively quickly; (b) it is independent of writing skills, permitting clearer instructional focus on interpretive fidelity; (c) a single sort by the instructor serves as feedback for multiple trainees; (d) it yields detailed and prompt feedback both in the form of a reliably computed Q-correlation between teacher and student and in the form of relative ratings on 100 descriptive statements, thus focusing discussion and directing further instruction; (e) it encourages good interpretive habits such as synthesis of information from a variety of scales to weigh in on specific clinically relevant questions, formulation and testing of specific clinical hypotheses, and recognition of gradients of confidence in descriptive inferences; and (f) it is flexible in that the Q-set can be modified to reflect varying concerns across clinical populations, the magnitude of the Q-correlation can be used to index improvement in a student's interpretive accuracy over time, and the source of the criterion Q-sort can be altered to demonstrate to students the magnitude of descriptive convergence that can be expected with, say, therapist judgments or spouse descriptions.

Williams and Weed (2003) provided an example of how interpretive Q-sorting can be applied to graduate training in MMPI–2 interpretation. A total of 10 MMPI–2 profiles, selected on the basis of their similarity to frequently occurring profiles in the MMPI–2 research literature and ordered by difficulty level (primarily in terms of the number of MMPI–2 scales reported), were interpreted using the Midwestern Q-sort by instructors of several graduate personality assessment courses. Aggregate instructor sorts for the 10 profiles served as criteria against which student interpretive performance was evaluated, with student/instructor Q-correlations serving as quantitative evidence of mastery. For each profile interpretation, students were permitted to submit as many Q-interpretations as necessary to meet mastery criteria (provided immediately at the Web site), but excellence in interpretation was motivated both by recognition on the Web site on which the interpretations were conducted and by competition between universities participating in the "MMPI–2 Interpretation Competition" (http://www.psych.uni.edu/mic/). Discrepancies between student and instructor interpretations (provided at the item level) served as the basis for discussion and instruction during class hours.[1]

SUMMARY

Although neither their origins nor their interpretation require invocation of a paradigm shift, the introduction of the MMPI–2 RC Scales represent significant progress in clarifying the constructs targeted by the original Clinical Scales and for measuring them without the burden of poorly performing items and excessive scale intercorrelations. Additional external validation research will be necessary for providing guidelines for clinical interpretation of the RC Scales and for distinguishing them from their parent scales. Guidelines for clinical interpretation (and instruction in clinical interpretation) can be particularly well informed by validation research designed to operationalize integrated MMPI–2 inferences.

REFERENCES

American Psychiatric Association. (2000). *Diagnostic and statistical manual of mental disorders* (4th ed.). Washington, DC: Author.

Butcher, J. N. (Ed.). (2000). *Basic sources on the MMPI–2.* Minneapolis: University of Minnesota Press.

[1]The MMPI–2 Interpretation Competition is now in its fourth year; prospective participating graduate instructors are encouraged to contact the author of this comment.

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.

Deskovitz, M. A., Weed, N. C., & Williams, J. E. (2005, April). *Interpretive reliability of six MMPI–2 computer-based test interpretation programs.* Paper presented at the 40th Annual Symposium on Recent Developments in the Use of the MMPI, MMPI–2, and MMPI–A, Ft. Lauderdale, FL.

Graham, J. R. (1999). *MMPI–2: Assessing personality and psychopathology* (3rd ed.). New York: Oxford University Press.

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

Hollrah, J. L., Schlottmann, R. S., Scott, A. B., & Brunetti, D. G. (1995). Validity of the MMPI subtle items. *Journal of Personality Assessment, 65,* 278–299.

Kwon, S., Noland, K. A., & Weed, N. C. (2006). *Interpretive reliability and validity of the MMPI–2 Restructured Clinical Scales.* Manuscript in preparation.

Little, K. B., & Shneidman, E. S. (1959). Congruences among interpretation of psychological test and anamnestic data. *Psychological Monographs, 73*(6, Whole No. 476).

Meehl, P. E. (1956). Wanted: A good cookbook. *American Psychologist, 11,* 263–272.

Meehl, P. E. (1960). The cognitive activity of the clinician. *American Psychologist, 15,* 19–27.

Meyer, G. J., Mihura, J. L., & Smith, B. L. (2005). The interclinician reliability of Rorschach interpretation in four data sets. *Journal of Personality Assessment, 84,* 296–314.

Nichols, D. S. (2001). *Essentials of MMPI–2 assessment.* New York: Wiley.

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

Ozer, D. J. (1993). The Q-sort method and the study of personality development. In C. Tomlinson-Keasey & K. Widaman (Eds.), *Studying lives through time* (pp. 147–168). Washington, DC: American Psychological Association.

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

Shedler, J., & Westen, D. (1998). Refining the measurement of Axis II: A Q-sort procedure for assessing personality pathology. *Assessment, 5,* 335–355.

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

Sines, L. K. (1959). The relative contribution of four kinds of data to accuracy in personality assessment. *Journal of Consulting Psychology, 23,* 483–492.

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.

Weed, N. C., Butcher, J. N., & Ben-Porath, Y. S. (1990). Failure of Wiener and Harmon Minnesota Multiphasic Personality Inventory (MMPI) Subtle scales as personality descriptors and as validity indicators. *Psychological Assessment, 2,* 281–285.

Weed, N. C., & Noland, K. A. (1998, March). *The MMPI Research Applications of the Mississippi Q-sort.* Paper presented at the 33nd Annual Symposium on Recent Developments in the Use of the MMPI, MMPI–2, and MMPI–A, Minneapolis, MN.

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

Wiener, D. N. (1948). Subtle and obvious keys for the MMPI. *Journal of Consulting Psychology, 12,* 164–170.

Williams, J. E., & Weed, N. C. (2003, June). *Using a web-based Q-sort program to teach MMPI interpretation.* Paper presented at the 38th Annual Symposium on Recent Developments in the Use of the MMPI, MMPI–2, and MMPI-A, Minneapolis, MN.

Young, K. R., Weed, N. C., & Williams, J. E. (2005, April). *Development of a Q-set to interpret the MMPI–2 with substance abusing clients.* Paper presented at the 40th Annual Symposium on Recent Developments in the Use of the MMPI, MMPI–2, and MMPI–A, Ft. Lauderdale, FL.

Nathan C. Weed
Department of Psychology
227 Sloan Hall
Central Michigan University
Mount Pleasant, MI 48859
Email: weed1nc@cmich.edu

Received February 10, 2006
Revised April 28, 2006