



Received: 24 October 2015
Accepted: 09 August 2016
Published: 09 September 2016

*Corresponding author: Shehroo B. Pudumjee, Department of Psychology, Central Michigan University, Mount Pleasant, MI, USA
E-mail: pudum1sb@cmich.edu

Reviewing editor:
Peter Walla, University of Newcastle, Australia

Additional information is available at the end of the article

CLINICAL PSYCHOLOGY & NEUROPSYCHOLOGY | RESEARCH ARTICLE

Development of a q-set for MMPI-2 research with substance abusing populations in the USA

Kevin R. Young¹, Nathan C. Weed², Brian J. McCabe², Shehroo B. Pudumjee^{2*} and John E. Williams³

Abstract: Research and clinical applications of q-methodology, in which a set of descriptive statements are sorted into a forced, quasi-normal distribution, have recently been extended to validation of integrated interpretations of the Minnesota Multiphasic Personality Inventory-2nd Edition (MMPI-2). This study aimed to develop a q-set customized for individuals diagnosed with substance abuse or those within a substance abuse treatment setting. 229 q-items, covering 11 content areas, were written to cover both the content area of the MMPI-2 and the variables affecting substance abuse treatment outcome. On the basis of their reliability, measured by inter-rater total correlation and variance components, 98 items were selected for final testing. Scores on this final q-set showed acceptable levels of reliability when forced into both a quasi-normal and flat distribution, with pairwise inter-rater reliability estimates of .519 and .517, respectively. Results of the study suggest the new q-set might be better suited to assessing a substance abusing population than the original Midwestern q-set that inspired it.

Subjects: Behavioral Sciences; Careers in Psychology; General Psychology; Health Psychology; Mental Health; Personality Tests & Assessments; Psychiatry & Clinical Psychology - Adult; Psychological Disorders; Psychometrics/Testing & Measurement Theory; Testing, Measurement and Assessment

Keywords: substance abuse; q-methodology; MMPI-2; assessment

ABOUT THE AUTHORS

The co-authors of this project are members of the Psychological Assessment Laboratory at Central Michigan University. Led by Kyunghye Han, PhD, and Nathan Weed, PhD, the Psychological Assessment Laboratory conducts research on psychometric measures used in applications of clinical psychology. Most commonly, we conduct research on aspects of assessment with the MMPI, the most widely used psychological test in the world. Recent projects have focused on the substance abuse scales of the MMPI-2-RF, the Hindi translation of the MMPI-2, and q-sort applications of MMPI-2-RF research. The present manuscript is adapted from the master's thesis of the first author, Kevin R. Young, PhD, currently affiliated with the Louis Stokes VAMC located in Cleveland.

PUBLIC INTEREST STATEMENT

The Minnesota Multiphasic Personality Inventory-2nd Edition (MMPI-2) is one of the most widely used psychological tests in the world, employed in a variety of clinical and occupational settings. It is used to evaluate a wide range of personality and psychopathological phenomena, and to plan treatment. Because of its widespread use and real world impact, research evaluating its applications and improving its utility is important. This study reports on the development of a measure that can be used to evaluate the usefulness of the MMPI in substance abuse treatment settings. This new measure is in the form of a q-set, a set of statements that can be used to characterize an individual based on the scores obtained on the MMPI-2. Our hope is that this new measure may be useful in research in improving test use for individuals with substance abuse problems.

1. Introduction

Substance abuse and dependence have long been areas of intense research interest and concern because of their high incidence rates. According to the 2014 National Survey on Drug Use and Health (NSDUH) conducted by the Substance Abuse and Mental Health Services Administration (SAMHSA) within the United States of America, 17 million people, amounting to 6.4% of the population over 12 years, were classified as having alcohol abuse or dependence (SAMHSA, 2014). In addition, marijuana use disorder was found to be prevalent in 4.2 million individuals, comprising 1.6% of the population being classified as having marijuana abuse or dependence (SAMHSA, 2014). Marijuana was followed by pain reliever abuse or dependence (1.9 million people) and cocaine abuse or dependence (913,000 people) (SAMHSA, 2014).

Given these statistics, it is not surprising that many assessment tools have been designed for substance abuse or dependence, including the Substance Use Disorders Diagnostic Schedule (SUDDS; Harrison & Hoffmann, 1989) and the Michigan Alcoholism Screening Test (MAST; Selzer, 1971). A measure used as part of a clinical assessment should be able to provide clinicians with enough information to create efficient and effective treatment plans (Hayes, Nelson, & Jarrett, 1987). Due to their more limited scope, instruments like the SUDDS and MAST, which are focused on the diagnosis of abuse or dependence and other related behaviors, might not be as suitable for providing information that pertains to treatment planning, as compared to a broader measure of psychopathology and personality, such as the Minnesota Multiphasic Personality Inventory-2nd Edition (MMPI-2; Butcher et al., 2001).

1.1. Factors influencing outcomes in substance abuse treatment

There are a number of psychosocial factors that are related to substance abuse treatment outcome. Although it is often useful to review the research literature concerning alcohol abuse separate from that concerning the abuse of other drugs, there are numerous similarities in a number of respects (Ciraulo, Piechniczek-Buczek, & Iscan, 2003), including their assessment using the MMPI-2 (Stein, Graham, Ben-Porath, & McNulty, 1999). Accordingly, pertinent literature is reviewed in one combined section below.

Project MATCH, a large-scale, multi-site study, followed 1,726 patients of alcohol dependence in the United States of America who were assigned to one of three manualized treatment formats: Cognitive Behavioral Therapy, Motivational Enhancement Therapy, or Twelve-Step Facilitation Therapy, based on personal attributes and substance-related variables (Fuller & Allen, 2000; Kadden, Longabaugh, & Wirtz, 2003; Mattson, Babor, Cooney, & Connors, 1998). Results showed that only one of the ten primary hypotheses; that outpatients with less severe psychiatric conditions in the Twelve-Step Facilitation Therapy drank less than those in the Cognitive Behavioral Therapy group; showed significant results, and this result was no longer significant at the 39-month follow-up and was considerably diminished at one year post treatment. Of the secondary hypotheses, only three hypotheses pertaining to different attributes and substance use patterns were supported. Considering that of a total of 21 hypotheses, pertaining to differential improvement based on matching variables to treatment conditions, only four were supported, this study seems to argue against the efficacy of matching clients to treatments by client characteristics. However, it is important to observe that each characteristic was tested independent of the others, and that any matching effects that may have existed could have been obscured by the fact that nearly all of the clients who participated in Project MATCH showed remarkable improvement in terms of treatment outcome.

Gottheil, Thornton, and Weinstein (2002) compared the effectiveness of high structure and low structure counseling for substance abusing individuals. High structure counseling included more behaviorally based interventions characterized by the therapist playing a more directive role and reinforcing goal achievement, while low structure counseling was derived from existential, client centered therapies, and those more focused on emotional support and expression, thereby characterizing the therapist as more passive and amenable to being led by the client. The results of the study showed that although neither treatment style produced significantly better results than the

other on a number of outcome measures (patient- and counselor-rated benefit, mean drug free urine samples, attendance), both a main effect and an interaction effect were found for the client's level of depression as assessed by the Beck Depression Inventory (Beck & Steer, 1987). Patients who displayed higher levels of depression did significantly better in high structure counseling, whereas less depressed patients did significantly better in low structure counseling. The authors explain that this finding is supported by the findings of Project MATCH mentioned earlier wherein severity of psychopathology was found to impact progress in different therapeutic modalities.

Ciraulo et al. (2003) provided an insightful review of the characteristics of substance abusers known to impact treatment outcome. One common factor mentioned across multiple substances such as alcohol, cocaine, opioids, and tobacco is that the resultant cognitive impairment that can occur as a result of excessive substance abuse may preempt the formulation of effective coping strategies by the abusing individual and can consequently impact treatment outcomes. This is particularly pertinent as the cognitive effects of the drugs differ in intensity. Other psychosocial factors that are of interest are severity of dependence or withdrawal, negative affect, personality traits and disorders (particularly risk-taking and anti-social tendencies), self-efficacy, and psychiatric comorbidity (Ciraulo et al., 2003).

Differences in personality factors within substance abusing populations has also long been an area of research interest, with scientific exploration beginning in the early 1940s (Babor, 1996) with the search for the "alcoholic" personality, and continuing through to contemporary multidimensional typologies (Morey & Blashfield, 1981). These typologies seek to classify substance abusers—typically alcoholics—into relatively homogeneous groups with similar familial backgrounds, severity of abuse, and personality characteristics (Babor, Dolinsky, et al., 1992; Babor, Hofmann, et al., 1992; Litt, Babor, DelBoca, Kadden, & Cooney, 1992). These typologies of substance abusers may help identify groups of variables that are relevant to treatment planning. For example, Litt et al. (1992) found differential responses in treatment based on the alcoholism typology. Type A was characterized by milder aspects of severity and susceptibility, i.e. a later onset of drinking behavior, less family history of alcoholism, less alcohol dependence, and less pathological personality traits such as anti-social traits as compared to Type B. This typology, and by extension, these characteristics were found to impact treatment outcomes; interactional interventions were more effective with Type A while intervention based on coping skills development were more effective with Type B (Litt et al., 1992).

1.2. Using self-report measures with substance abusing populations

Researchers or clinicians who wish to screen for substance abuse problems or measure treatment-relevant personality constructs often choose to do so with self-report measures. Feit, Fisher, Cummings, and Peery (2015) note, in their chapter on screening and assessment for substance use and abuse, that the use of assessment measures in these contexts has increased and that they have advantages of easy quantification, time efficiency, and psychometric reliability. Still, it is important that these measures be evaluated carefully, as self-reports may be susceptible to various types of distortion. Babor, Steinberg, Anton, and Del Boca (2000) used the data collected from Project MATCH to compare self-report measures of substance abusing behavior with biological indicators of alcoholism assessed using liver enzyme levels obtained from blood samples and collateral reports from individuals aware of the participant's substance use habits and nominated by the participants. Results from this study indicated that self-report measures showed moderate correlations with both collateral reports and biological tests but comparatively there were greater correlations between self-report and collateral measures. Self-report measures were also found to be a more complete source of information as compared to biochemical and collateral data with the authors stating that the two latter sources may only add peripheral data. In addition, self-report measurements showed higher correlations with each of the other two data collection methods as the intensity of alcohol abuse increased. The authors also cite socially desirable responding as being one of the core limitations of self-report measures. However, these can be well assessed by the MMPI-2 validity scales,

which evaluate the response patterns and styles in order to identify social desirability, over or under reporting and inconsistent responses (Butcher et al., 2001).

Personality variables relevant for substance abusing populations have also been investigated using the MMPI-2. The MMPI-2 has been shown to be significant predictor of Axis I and Axis II diagnoses in substance abusing populations (Butcher et al., 2001) per the Diagnostic and Statistical Manual of Mental Disorders (4th ed.; *DSM-IV*; American Psychiatric Association, 1994), even above the NEO-PI-R (Quirk, Christiansen, Wagner, & McNulty, 2003).

Although the literature clearly indicates that there is a number of relevant personality or demographic characteristics that impact treatment outcome (Ciraulo et al., 2003; Gottheil et al., 2002), and that reliably differ among substance abusers (Turiano, Whiteman, Hampson, Roberts, & Mroczek, 2012), the ability of the MMPI-2 to assess these factors has yet to be discussed. There is extensive research literature on the ability of the MMPI-2 to assess presence of substance abuse (Gantner, Graham, & Archer, 1992; MacAndrew, 1981; Morey & Blashfield, 1981; Stein et al., 1999), principally focused on the ability of scores on the MacAndrew Alcoholism Scale (MAC; MacAndrew, 1965), the Revised MacAndrew Alcoholism Scale (MAC-R; Butcher et al., 2001), the Addiction Potential Scale (APS; Weed, Butcher, McKenna, & Ben-Porath, 1992), and the Addiction Acknowledgement Scale (AAS; Weed et al., 1992) to accurately predict substance abuse and dependence. In a study comparing the efficacy of these measures, Clements and Heintz (2002) examined the diagnostic accuracy of MAC-R, AAS, and APS and investigated the factor structure of AAS and APS to identify item content sub-dimensions. Results indicated that scores on AAS significantly outperformed those on APS and MAC-R in terms of sensitivity, specificity, and discriminative ability. Factor analysis of AAS and APS item scores suggested two- and four-factor structures, respectively. AAS items consisted of two primary factors, labeled “Acknowledgement of Alcohol/Drug Problems” (internal consistency $\alpha = .60$) and “Positive Alcohol Expectancies” ($\alpha = .58$) by the authors. APS items produced a four-factor structure, labeled “Satisfaction with Self” ($\alpha = .60$), “Cynicism/Pessimism” ($\alpha = .31$), “Impulsivity” ($\alpha = .50$), and “Risk Taking” ($\alpha = .25$). These results are consistent with the multi-dimensional factor structures found in previous factor analyses of AAS and APS (Weed, Butcher, & Ben-Porath, 1995) and suggest that although AAS scores were shown to have greater diagnostic utility than APS scores, all three scales might be able to provide useful information on differing facets of personality of substance abusers.

A study by Levenson et al. (1990) highlights the sensitivity of the MMPI-2 substance abuse scales to behavioral dimensions other than substance abuse. The authors drew a large sample of men from a normative aging study and compared those participants who reported having a pattern of drinking seen as being problematic to those who did not report having a history of drinking problems. Participants who had been arrested and had drinking problems received the greatest average scores on the MAC scale, whereas participants who had been arrested but did not report a history of problem drinking and those who had not been arrested but did report a history of problem drinking were virtually indistinguishable from each other. All three of these groups produced significantly higher scores on the MAC than those who reported neither a history of arrests nor drinking problems. These results seem to indicate that MAC is measuring personality traits that correlate highly, but not exclusively, with substance abuse, such as a tendency toward risk-taking behaviors.

Although the potential ability of the MMPI-2 to provide clinicians with meaningful information about clients with substance abuse problems has been documented, previous research has typically focused on individual MMPI-2 scale functionality and has not evaluated the ability of an integrated interpretation of MMPI-2 scale scores to describe examinees accurately (Weed, 2006). Q-methodology (Stevenson, 1953) provides a means for such an integrated interpretation to be evaluated. Q-methodology in this application involves a set of descriptive statements about a client that are sorted by a clinician on the basis of information inferred from an MMPI-2 profile. For example, a q-set of 100 statements could be sorted into a quasi-normal distribution of seven categories ranging from *most descriptive* to *least descriptive*, with statements that are thought by the clinician not to be applicable or of neutral

applicability being placed in the middle of the distribution. Such a q-sort might be considered an “interpretive q-sort” because it characterizes the client on the basis of MMPI-2 interpretation.

Meanwhile, a source of information (e.g. clinician, spouse, chart reviewer) blind to the MMPI-2 results might complete a q-sort using the same items to describe the client. Such a q-sort might be termed a “descriptive q-sort.” The correspondence, indexed by a correlation coefficient, between the interpretive q-sort based on information from the MMPI-2 only and the descriptive q-sort based on non-MMPI-2 information only can serve to provide an estimate of the validity of the MMPI-2 profile in describing the client’s symptoms (Harrington, 2000). This methodological design offers researchers a great deal of flexibility, as a set of q-statements can be written to pertain either to broad and general personality traits, or tailored to the salient characteristics of a particular clinical population, such as substance abusing clients.

In research applying q-methodology, a fixed distribution is typically used to reduce response bias on the part of the interpreter, ensure overall item variability within a given subject, and reduces the probability of obtaining cumbersome and disorganized data (Block, 1961). Using a fixed quasi-normal distribution is theoretically defensible provided the distribution resembles that of a T-distribution, with degrees of freedom equal to one less than the maximal number of independent (non-correlated) factors that can be derived from the q-set itself. These factors, being uncorrelated, would be distributed randomly in any given individual, and therefore, as the number of factors increased, the distribution would approach normality in shape. Hence, most q-sorts are forced into quasi-normal distributions based on the “best guess” of the instrument developer. One consequence of the forced distribution in a q-sort is that item scores produced by a single individual are interdependent upon one another (Serfass & Sherman, 2013).

1.3. Using Q-methodology in interpretation of the MMPI-2

The Midwestern q-set, a 100-item set of common MMPI-2 clinical inferences, was originally developed to help students achieve proficiency in integrated interpretation of the MMPI-2 (Weed, 1997). Revisions of the Midwestern q-set augmented the original 73-items; McNeal (1999) altered items based on their correlations with an external criterion across raters and independent observers, and made the vocabulary more accessible to multiple settings. These items were written to capture typical interpretations made using the MMPI-2, and were examined for their contributions to estimates of reliability and validity in later research (Robinson, 2004). Furthermore, since the items were designed to cover the breadth of the MMPI-2, some may be redundant or not specific to a given clinical sample. Although it has been used successfully in a variety of MMPI validation studies (e.g. Deskovitz, Weed, McLaughlan, & Williams, 2016), there have not been any attempts to develop MMPI-2 interpretive q-sets that are specific to clients with substance abuse or dependence. A q-set of items constructed explicitly for clients with substance abuse or dependence, and which showed similar levels of inter-rater reliability to the Midwestern q-set, may provide superior and more specific information regarding the relevant personality characteristics, treatment, and prognosis.

1.4. Objective of the study

Given both the breadth of the MMPI-2 item pool and the large number of potentially relevant personality variables that meaningfully differ across clients with substance abuse problems, this study sought to develop and evaluate a set of q-sort statements pertaining specifically to substance abuse. The statements within the q-set were aimed at comprehensively describing and addressing the behavioral and personality correlates of substance abuse that can be inferred from an MMPI-2 profile. This q-set could ideally be used in validation studies of the MMPI-2 with populations engaging in or prone to substance abuse. A secondary goal of the study was to determine the optimal distribution for the derived q-set, i.e. whether it is served better by statements rated into forced quasi-normal distribution or a flat distribution.

This q-set has the potential to facilitate the investigation of the interpretive validity of the MMPI-2 in the context of substance abuse and has the advantage of being specifically tailored to address substance abuse related concerns.

2. Method

2.1. Participants

2.1.1. Q-sorters

Four raters each interpreted the 74 MMPI-2 protocols. Each rater had, at minimum, graduate-level training with regard to both MMPI-2 interpretation and the use of the q-sort program. Raters were informed that the MMPI-2 protocols represented results from an examinee from a substance abuse treatment setting.

2.2. Instruments

2.2.1. MMPI-2

The MMPI-2 protocols that were selected for use in the study were scored and interpreted using Pearson's Minnesota Report, a computer-based test interpretation program. Information from the Minnesota Report served as the only basis on which the raters generated their descriptions, thereby eliminating any idiosyncrasies of MMPI-2 profile interpretation across the sorters, and standardizing the interpretive information provided to them. Sorters utilized a web-based computer program that facilitates item rating according to the desired q-distribution.

2.2.2. Archival data-set

Seventy-four anonymous MMPI-2 protocols collected from the Hazelden Clinic (Weed et al., 1992) were used in this study: 50 in Stage Three (described below) and 24 in Stage Four. These protocols were selected randomly from among the valid profiles in the data-set. Protocols were considered invalid if any of the following conditions were met: 36 or more missing or multiple responses; 95% or more of the items marked True; 95% or more of the items marked False; F-K (raw) score greater than or equal to 14; TRIN T-score greater than or equal to 90; VRIN T-score greater than or equal to 90; or FT-score greater than or equal to 110. Only protocols that contained an Addiction Acknowledgement Scale (AAS) T-score of at least 55 were included to ensure that all protocols included were typical of clients with substance abuse.

2.3. Procedures and analyses

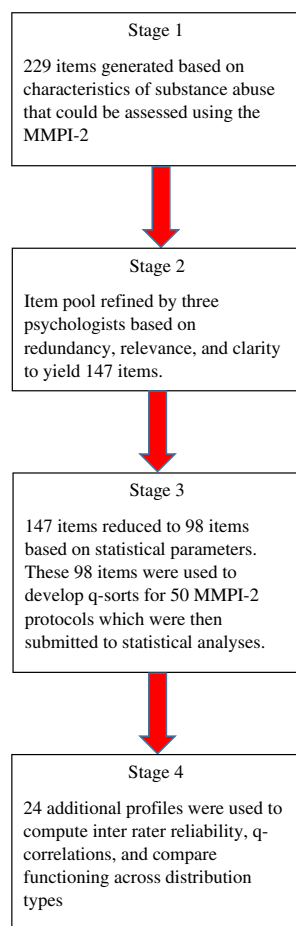
The development and validation of the q-sort proceeded in a four-stage format summarized in Figure 1.

2.3.1. Stage 1: Item generation/composition

Based on a comprehensive review of the available literature pertaining to characteristics and functioning of clients with substance abuse, psychosocial factors pertaining to treatment efficacy, and the ability of the MMPI-2 to assess these factors an initial item pool was developed. Item generation began with the items used in the original Midwestern q-sort (Weed, 1997) and was subsequently altered to cover the constructs specific to substance abuse, operationalizing substance abuse through prevalent literature and formal taxonomies such as the DSM-IV. A number of items were developed to be relevant to the Cloninger (1987) and Babor, Dolinsky, et al. (1992) typologies of substance abusers, specifically relating to impulse control and sensation seeking. Items were also added to account for other research findings (e.g. Gottheil et al., 2002). This initial list of items was designed to be relevant to both the MMPI-2 and the substance abuse population and comprised 216 items.

Items were selected to represent the information clinicians assess with regard to substance abuse and to be representative of what can be reliably provided by the MMPI-2. Previous studies (see

Figure 1. Flow chart showing the stages followed in developing the q-sort.



Deskovitz, 2003 for a review) have mainly focused on developing statements that pertained to the MMPI-2 only, or statements that pertained to general psychopathology. Additional items were also added to maintain the balance of overall item content and category representation. This brought the total item pool to 229 items. Each of these items fell into one of twelve content categories, which were derived based on a broad review of the literature: Depression/Demoralization (18 items), Anger/Anti-social Attitudes (26 items), Emotional Control (18 items), Explicit Substance Use (10 items), Mania/Impulse Control (14 items), Treatment Prognosis (23 items), Interpersonal Style (21 items), Family History/Problems (8 items), Miscellaneous Pathology (11 items), Anxiety/Neuroticism (24 items), Response Style (25 items), and General Personality (18 items). A complete list of the original pool of items can be obtained upon request from the authors.

2.3.2. Stage 2: Rational item selection

The use of raters has been used extensively in research involving q-methodology (Deskovitz, Weed, Chakranarayan, & Williams, 2016; Deskovitz, Weed, McLaughlan, et al., 2016). Three doctoral-level psychologists who had expertise in both substance abuse treatment or assessment, and interpretation of the MMPI-2 evaluated the item pool in terms of clarity, relevance to substance abuse, and whether or not the information could reasonably be obtained from the MMPI-2. Items that were not relevant to substance abusing clients were unlikely to provide much useful information for treatment purposes and would have little functional value. Items that could not be reliably assessed using the MMPI-2 would not likely reliably vary across raters and would only add error to the instrument. Items rated poorly by the panel on either of these dimensions were eliminated at this stage. Each item was rated on a 5-point Likert scale for two domains; how much the item assessed factors important to substance abuse, and how well the MMPI-2 could assess the item content. Based on the

expert ratings, it was decided that items should be removed from the potential item pool if they received an average rating of below 3 on either the substance abuse specificity scale or the MMPI-2 specificity scale. This threshold of a rating of 3 was used in order to select the items that were rated highest and were above the threshold of 60% which suggested that on average, the raters considered this item to be suitable, i.e. descriptive of substance abuse and assessable by the MMPI-2. This reduced the overall item pool down to 148. Finally, to bring the total number of items down to a multiple of 7 for ease of sorting purposes, the item “manipulates others for own gain” was eliminated due to redundancy with “manipulates others.” Thus, a potential item pool of 147 items was developed, and these items moved on to the third stage of item selection.

2.3.3. Stage 3: Statistical item selection/evaluation

Statistical item selection incorporated the item’s performance on five different indices. The four raters completed q-sorts for each of the 50 protocols using the items resulting from Stages One and Two. Raters q-sorted the items into a 7-point (8-15-29-45-29-15-8) quasi-normal distribution and a 7-point flat distribution (14-14-14-14-14-14-14) using an online q-sort application created using Authorware for web-based research and teaching applications involving MMPI interpretation (Williams & Weed, 2003). In this phase, to further narrow the item pool, first, each item’s mean and kurtosis were calculated.

In addition, the Within Protocol/Between Rater (WP/BR) variance of each item was also calculated by averaging the ratings of each of the four raters on a given protocol. The sum of squares was also calculated, giving an estimate of the error inherent in any one item. The Between Protocol/Within Rater (BP/WR) variance for each was calculated in complementary fashion: the ratings of each rater across all 50 protocols were averaged, and the sum of squares was calculated. Averaging these scores across the four raters produced the BP/WR statistic, which was used as a measure of item discrimination of one protocol from another.

The Variance Index was computed simply by dividing the average BP/WR variance by its WP/BR variance. The Variance Index serves as a measure of the tradeoff between inter-rater reliability and variability of the item itself. Thus, if an item varied more, on average, between raters on the same protocol than it varied within raters on different protocols, it would have a Variance Index of less than 1 and would be adding more error to the instrument than discriminability.

Additionally, by individually removing each item from the item pool and calculating the inter-rater reliability without the item, the overall contribution each item was making to reliability (inter-rater total correlation) was calculated. Each q-sort item was then evaluated in terms of its mean, WP/BR variance, BP/WR variance, Variance Index (a ratio of the preceding variances), distribution shape, and overall contribution to inter-rater reliability.

Based on these statistical measures, the item’s convergent and discriminative properties were determined. It was concluded that items discriminating well between examinees should display distributions across protocols that are approximately normal or platykurtic. Items that fail to discriminate between protocols should display a distribution that is heavily skewed and/or leptokurtic, and/or have a restricted range. Items were then selected from among this pool of items based jointly on considerations of inter-rater agreement, descriptive statistics, and theoretical relevance, until a final item pool of 98 items was achieved.

2.3.4. Stage 4: Instrument evaluation

An additional 24 protocols were selected from the archival pool of MMPI-2 protocols and interpreted using the Minnesota Report. Four interpreters sorted the final set of 98 q-statements. In addition to the item statistics calculated in Stage Three, q-correlations were computed between raters across items to determine the inter-rater reliability of the q-ratings as a whole. Finally, to provide an indication of whether the q-sort items are best distributed quasi-normally or platykurtically, the raters again completed the q-sorts according to a flat distribution on the same 24 protocols. Item statistics

and q -reliabilities for these protocols were compared by distribution type to determine the most appropriate q -distribution shape for these items.

2.3.5. Principal components analysis

In an attempt to identify the factors within the new q -set, a Principal components analysis (PCA) was conducted. The four ratings for each of the 74 profiles were collapsed across each profile using the mean of the four ratings. These means were then submitted to a PCA using SPSS software, where factors were extracted using the criteria of eigenvalue greater than one.

3. Results

3.1. Stages 1 and 2: Rational item selection

In keeping with the goal of generating and selecting items that addressed the content areas relevant to substance abuse and were also interpretable from MMPI-2 profiles, three doctoral-level psychologists rated each of the 229 items on these parameters. Their ratings on each of these parameters were submitted to further analyses. The average inter-rater correlation on the Substance Abuse dimension was .35, while the average inter-rater correlation on the MMPI-2 dimension was .46. These ratings were relatively high compared to the average random rater correlation, which was .10. Thus, correlations between the scales were likely due to item content and not random variance between the raters. The average correlation within raters between the two scales was .21.

After removing 82 items that had an average rating less than 3 on either of the two domains of relevance to substance abuse or inclusion in the MMPI-2 profiles, the overall item pool was reduced to 147. Average expert ratings on the substance abuse relevance scale increased from 3.46 to 3.81 after the 82 items were removed, and the MMPI-2 relevance ratings similarly increased from 3.55 to 3.90.

3.2. Stage 3: Statistical item selection

Statistical item selection proceeded in five phases based on the functioning of the item in terms of five different parameters; Inter-rater total correlations, WP/BR variance, BP/WR variance, Variance Index, and Item Means and Item Content. The degree to which a given item was agreed upon by multiple raters was jointly determined by the item's between rater/within protocol variability, Variance Index, and contribution to total inter-rater reliability. The degree to which a given item was discriminating within the substance abusing population was determined by that item's BP/WR variability, Variance Index, mean, and overall contribution to inter-rater reliability. Items were first identified on the basis of their contribution to reliability and their subsequent performance on other statistical parameters was used as criteria for removal. Thus, the same set of 62 items was submitted to further examination after Phase 1. The number of items removed due to each one of the following parameters is displayed in Table 1.

3.2.1. Phase 1: Inter-rater total correlations

The first set of items removed were those that displayed the highest item-removed inter-rater total correlations. In addition to their negative contribution to overall reliability, these items also frequently displayed other undesirable characteristics, such as high scores on the average BR/WP

Table 1. Items removed at each step of statistical selection

Step of removal	No. of items
Relevance to substance abuse and/or assessment through the MMPI-2	82
Inter-rater total correlations	13
Inter-rater total correlations and Variance Index	6
Inter-rater total correlations and BP/WR variance	10
Inter-rater total correlations and WP/BR variance	13
Inter-rater reliability, item content, and item means	7

variance and low Variance Index statistics. After removal of 12 items, average inter-rater reliability rose from .482 to .510.

3.2.2. Phase 2: Variance Index

After removing these items, 62 items remained that were negatively affecting overall reliability. Of these 62 candidates which also performed poorly in terms of inter-rater total correlations, items were removed in this phase if they displayed Variance Indices lower than 1. These were items likely adding more error than discriminative power to the overall instrument. Removing these 6 items increased the inter-rater reliability from .510 to .517.

3.2.3. Phase 3: BP/WR variance

Again, from within the subset of 62 candidates for removal, 10 items removed in this phase of item selection displayed unacceptably low BP/WR variance (<1.00). Removal of these items caused the overall inter-rater reliability to increase from .517 to .525.

3.2.4. Phase 4: WP/BR variance

Thirteen items removed in the fourth phase of item refinement (again, from among the 62 item candidates) were removed due to excessively high WP/BR variance (>1.30). Items that display high levels of variability across raters within the same protocol will likely contribute noticeably higher proportions of error when used in settings where all raters are not given the same interpretation on which to base q-sorts. If the item is unreliable when interpretation is held constant, it will likely become even more unreliable when interpretation is allowed to vary. Removal of these items caused the overall inter-rater correlation to increase from .525 to .540.

3.2.5. Phase 5: Item means and item content

In the final phase of item selection, seven items were removed to balance of the overall instrument mean and balance the item content within the set. “Is motivated to avoid failure,” “Is likely to punch or kick others,” and “Acts disoriented” were removed because they had the lowest item means among the 62 item candidates for removal. “Has lost interest ...” was removed due to an abundance ($n = 11$) of items from the depression content area that were already included in the final instrument. “Is frequently belligerent without reason” was removed due to a sufficient number of items from the antisocial clusters already in the final item pool. Finally, “Crumbles under pressure” was removed because it was judged to be the least specific remaining item on the anxiety item cluster, which also had sufficient representation in the final item pool.

Through these five phases of item refinement, removing 49 items from the overall item pool increased the average inter-rater correlation from .482 to .547. The aggregate four-rater reliability (corrected with the Spearman–Brown formula) increased from .788 to .828. The average standard deviation of the remaining items was 1.3, a value of interest in Stage Four below. Table 2 comprises a complete list of the 98 items of the final q-set. Item statistics for these items are available on request from the authors.

3.3. Stage 4: Instrument evaluation

For the final stage of the project, two forced distributions were pitted against each other across an identical set of 24 protocols. The order of distribution completion was counter-balanced, and protocols were randomly distributed relative to the order in which they were presented across the two distributions.

The quasi-normal 98-item distribution showed an essentially negligible difference in inter-rater reliability (.524) when compared to the flat distribution (.521) across the 24 protocols. Furthermore, the average variability of the inter-rater correlations across all 24 protocols (4 raters by 24 protocols = $3! * 24$ correlations = 144 correlations per distribution shape) was more than twice as high on the flat distribution (.05) as it was on the quasi-normal distribution (.02). This suggests that the flat distribution might have provided an increase in discriminability across certain types of protocols, but

Table 2. List of the final 98 items included in the q-set

Item name	Item name
1. Abuses alcohol	50. Is currently abusing a drug other than alcohol
2. Actively copes with most of life's problems	51. Is currently abusing alcohol
3. Acts angry	52. Is deliberate and careful with work
4. Acts anxious	53. Is depressed often
5. Acts confused	54. Is exaggerating symptoms
6. Acts depressed	55. Is hostile toward others
7. Acts nervous	56. Is hyperactive
8. Acts relaxed	57. Is jaded toward life
9. Acts tense	58. Is likely to abuse prescribed medications
10. Argues with others frequently	59. Is likely to improve with treatment
11. Attempts to avoid dangerous situations	60. Is motivated primarily by short-term rewards and pleasure
12. Attends (or would attend) treatment regularly	61. Is motivated to succeed
13. Avoids conflict	62. Is often alone
14. Behaves in a psychotic manner	63. Is optimistic
15. Behaves in an unpredictable manner	64. Is paranoid
16. Believes other people are "out to get them"	65. Is perfectionistic
17. Believes treatment is unnecessary	66. Is ready and willing to change
18. Blames own problems on others	67. Is realistic
19. Complains of physical illnesses	68. Is satisfied with life
20. Does not demonstrate much personal efficacy	69. Is self-confident
21. Does not see any strengths within self	70. Is social
22. Engages in risky behaviors	71. Lies frequently
23. Experiences many and frequent problems	72. Likes self
24. Fits in well with mainstream society	73. Makes decisions that seem ill-advised
25. Frequently seems euphoric	74. Makes friends easily
26. Gets along well with family	75. Manipulates others
27. Gets into trouble with the law	76. Often acts without thinking
28. Habitually denies own problems	77. Often breaks rules or laws
29. Handles emotions well	78. Often cannot function due to stress
30. Handles stress well	79. Often loses control of emotions
31. Has a juvenile history of delinquency	80. Overestimates own abilities
32. Has a rigid belief system	81. Perceives others as threats to own security
33. Has adequate emotional control	82. Pretends not to have problems
34. Has current family conflict	83. Pretends to have problems
35. Has difficulty solving problems	84. Problems will remain despite treatment
36. Has few problems	85. Seems comfortable in most social situations
37. Has frequent headaches	86. Seems content with life
38. Has many fears or phobias	87. Seems to be driven by a motor
39. Has many friends	88. Seems to be out of step with society
40. Has many hobbies, interests, and goals	89. Seems to be under pressure frequently
41. Has periods of depression that last for a month or more	90. Sees or hears things others can not

(Continued)

Table 2. (Continued)

Item name	Item name
42. Has potential to be a contributing member of society	91. Takes good care of daily responsibilities (e.g. child care, work, school)
43. Has problems at work	92. Takes unnecessary risks
44. Has regular sleep patterns	93. Thinks highly of self
45. Has stable relationships with others	94. Views self as having few flaws
46. Has trouble restraining self	95. Was raised in benign environment
47. Is a "thrill-seeker"	96. Worries frequently
48. Is agreeable	97. Would be a difficult therapy client
49. Is careful	98. Would show improvements with therapy

caused a decrease in reliability on others. This is also suggested by the WR/BP variance of the flat distribution, which was twice as high as that of the quasi-normal distribution (3.27 to 1.40), and by the BR/WP variance, which was also higher on the flat (1.87) when compared to the quasi-normal (.80). Interestingly, the ratio of BR/WP variance to WR/BP variance was nearly identical in the flat (1.81) and quasi-normal (1.77) distribution shapes.

Table 3 shows inter- and intra-rater correlations between and across the two distribution types. For each rater, along with the average correlation with the other three raters for each protocol within distribution type, the Cross Distribution Correlation (CDC), an average correlation of the rater with oneself on identical protocols across distribution shapes, was calculated. This provided an estimate of the differences purely due to the distribution shape, with the rater and protocol not offering any variance. This value is shown on the main diagonal. It is interesting to note here the difference in this ratio for Rater 1, whose ratings seemed to be most affected by the difference in distribution. In addition, these values were then compared to the average correlation each rater had with themselves across distributions to the average correlation with the other raters across protocols. This ratio, labeled the Cross Distribution Ratio (CDR) in Table 3, allows us to estimate the amount of unique variance being contributed by the raters, independent of the protocol being rated.

3.4. Principal components analysis

In addition to the inter-rater statistics discussed previously, exploratory principal components analysis (PCA) was also conducted using SPSS on the averaged sorts of the four raters on the 74 protocols completed in the initial and final, quasi-normal distribution shapes. The flat distribution shape was not included because these 24 protocols were already represented in the quasi-normal distribution sorts and there was not enough data to conduct a reduced space analysis within the flat distribution shape alone. Ten components with eigenvalues (EV) greater than 1 were extracted from the matrix

Table 3. Inter-rater correlations between and across distribution shapes

Rater	1	2	3	4	C.D.C.	C.D.R.
1	0.54	0.37	0.48	0.42	0.45	0.82
2	0.43	0.74	0.58	0.66	0.51	0.69
3	0.43	0.60	0.71	0.59	0.55	0.78
4	0.45	0.63	0.59	0.69	0.55	0.80

Notes: C.D.C. = Cross distribution correlation; the mean correlation of each rater with sorts done by the other three raters, on the same protocol, on the opposite distribution type. C.D.R. = Cross distribution ratio; the ratio of a rater's average correlation with themselves across distributions to their average correlation with the other three raters across distributions. Values below the diagonal represent correlations within the quasi-normal distribution shape, Values above the diagonal represent correlations within the flat distribution shape, and Items on the diagonal represent average rater correlation with themselves on identical protocols across distribution shapes.

Table 4. Total variance explained by extracted components

Component	Initial eigenvalues		
	Total	% of variance	Cumulative %
1	38.82	39.61	39.61
2	24.87	25.38	64.99
3	5.76	5.88	70.87
4	3.69	3.76	74.63
5	2.71	2.76	77.39
6	2.31	2.36	79.75
7	2.22	2.26	82.01
8	1.85	1.90	83.90
9	1.20	1.22	85.12
10	1.14	1.17	86.29

of correlations (Table 4). Components were then rotated using Varimax with Kaiser Normalization. Examination of the item content represented by the components (Table 5) suggests that the first component is tapping a dimension of positive emotionality, with items like “Is optimistic” loading extremely highly in the positive direction, and items like “Is depressed often” loading extremely highly in the negative direction. Of the initial item content clusters, the first component appears to be composed primarily of items from the depression, anxiety, and general pathology clusters. The second component, however, appears to be tapping a combination of anger/anti-social attitudes, emotional control, and treatment prognosis item content. In comparison with the two main components, the remaining eight were represented by relatively few (17) items between them.

4. Discussion

The main goal of this project was to develop a q-set that could potentially be used in validation research with the MMPI-2 in substance abusing populations. This goal was accomplished, resulting in a q-set of 98 items. Both the flat and quasi-normal distribution shapes for the final instrument were associated with acceptable pairwise inter-rater reliability, within and between rater variance figures, and high average Variance Index scores. The pairwise inter-rater reliability estimate is somewhat lower ($\Delta r = -.08$ to $-.18$) than that obtained in similar studies using the traditional Midwestern q-sort (Robinson, 2004), but not substantially so. Furthermore, this q-set may be expected to show somewhat lower correlations with MMPI-2 interpretations, as it was not written to pertain only to the MMPI-2, but also to the variables relevant to substance abuse treatment. Such a q-set might be expected to show lower inter-rater reliability between MMPI-2 raters, but higher inter-observer reliability between significant others with explicit knowledge of the person being tested, considering the high observability of the item content included. Furthermore, the reliability of the mean interpretation, over four raters, was .81 (calculated via Spearman-Brown) for both the quasi-normal and flat distributions, an acceptable reliability figure by any standards.

However, the secondary goal of the project, to determine the most appropriate shape for the q-distribution, was not so easily determined. In terms of their overall inter-rater correlation, the flat and quasi-normal distribution shapes did not meaningfully differ. However, the flat distribution shape has the advantage of potentially being able to provide a greater amount of protocol-specific information than the quasi-normal distribution shape does. This is due to the fact that the total number of discriminations between q-items that a rater must make is maximized at a flat distribution shape (Ozer, 1993). Furthermore, since all of the raters had been previously primed with a normal shape, it is possible that upon retest with the flat distribution, their inter-rater reliability would improve. Based only on the results of the current study, it cannot be concluded that either distribution is superior. Therefore, it is suggested that the quasi-normal sort, which did display a slightly higher inter-rater reliability and is the distribution shape to which most raters are accustomed due

Table 5. Varimax rotated, sorted item loadings

Item content	Components									
	1	2	3	4	5	6	7	8	9	10
Is optimistic	.93	-.21	.17	.00	-.06	.08	.01	-.03	-.04	-.01
Is self-confident	.92	.07	.17	.04	-.10	.17	.03	-.02	.01	-.01
Is satisfied with life	.92	-.22	.11	.05	-.06	-.02	.07	.02	.04	.07
Acts nervous	-.91	-.20	-.19	-.03	.02	-.05	-.05	-.06	-.12	.04
Is depressed often	-.91	-.07	-.05	-.04	.05	-.10	.04	-.05	-.03	.09
Often cannot function due to stress	-.91	-.12	-.08	-.19	.02	.02	.05	.12	-.02	-.10
Likes self	.91	-.12	.17	.03	-.16	.13	.06	.05	.05	.08
Has periods of depression that last for a month or more	-.90	-.01	.07	-.05	.15	-.05	.07	-.09	.09	.04
Seems content with life	.90	-.23	.18	.12	-.03	-.06	.05	-.06	.02	.10
Does not demonstrate much personal efficacy	-.90	.05	-.08	-.14	.10	-.02	-.02	.10	.12	.02
Acts anxious	-.89	-.20	-.01	-.04	.00	-.07	-.04	-.08	-.20	-.01
Acts depressed	-.89	.00	.00	.01	.04	-.22	-.11	-.04	-.07	.08
Handles stress well	.89	-.18	.11	.14	-.07	-.16	.00	-.18	-.01	.08
Thinks highly of self	.88	.01	.09	.02	-.16	.25	.10	.06	.10	.02
Acts relaxed	.87	-.06	.24	.09	-.10	-.23	.03	.06	.08	.00
Worries frequently	-.87	-.29	-.11	-.01	.07	-.03	-.06	-.01	-.16	-.10
Acts tense	-.86	-.22	-.21	.06	.14	.05	.00	-.03	-.13	.01
Does not see any strengths within self	-.86	-.24	-.02	-.08	.16	-.27	.01	-.07	-.02	.03
Views self as having few flaws	.86	.03	.08	.13	-.25	.12	.05	.15	.07	.06
Experiences many and frequent problems	-.83	.36	-.19	-.15	.17	.07	.03	.13	.04	-.11
Handles emotions well	.80	-.45	.17	.14	-.02	-.19	.03	-.11	-.08	.04
Has many hobbies, interests, and goals	.79	-.31	.28	.06	-.11	.10	.14	.02	-.08	.04
Has few problems	.79	-.45	.17	.22	-.04	-.17	.01	-.07	-.03	.11
Actively copes with most of life's problems	.78	-.44	.17	.25	.07	-.09	.11	-.14	-.09	.04
Seems to be under pressure frequently	-.77	-.40	-.09	.03	.15	.15	.03	.10	-.07	-.14
Seems comfortable in most social situations	.76	.12	.49	.17	-.04	.14	.04	.06	.11	-.19
Makes friends easily	.76	.00	.48	.14	-.06	.10	.10	.07	.11	-.22
Has difficulty solving problems	-.75	.32	-.12	-.31	-.19	.02	.09	.00	.17	-.04
Has many fears or phobias	-.74	-.28	-.09	-.05	.03	-.03	.24	.06	-.10	-.18
Has many friends	.74	-.05	.53	.13	-.09	.13	.04	.04	.08	-.18
Is social	.72	.19	.46	.15	-.10	.22	.02	.12	.13	-.24
Believes treatment is unnecessary	.71	.18	.04	.23	-.27	-.04	-.03	.40	-.08	.18
Has adequate emotional control	.70	-.56	.19	.23	-.02	-.20	.02	.03	-.05	.08
Has frequent headaches	-.70	-.36	.13	.00	.13	-.18	-.04	.24	.17	.08
Has problems at work	-.70	.51	-.16	-.08	.04	-.07	.07	.09	.14	-.08

(Continued)

Table 5. (Continued)

Item content	Components									
	1	2	3	4	5	6	7	8	9	10
Complains of physical illnesses	-.69	-.28	.18	.04	.13	-.20	.02	.33	.13	.03
Acts confused	-.68	.06	-.08	-.62	.17	.01	.03	.04	.01	.04
Often loses control of emotions	-.65	.52	-.10	-.20	.08	.27	.02	-.05	-.05	-.11
Has regular sleep patterns	.64	-.37	-.04	.04	.06	-.46	.03	-.05	-.21	-.13
Is often alone	-.62	-.20	-.52	-.09	.11	-.25	-.02	-.13	-.09	.30
Seems to be out of step with society	-.61	.26	-.27	-.41	.20	-.11	.15	.07	.19	.04
Frequently seems euphoric	.61	.27	.27	-.09	.03	.56	-.04	.00	-.05	.08
Overestimates own abilities	.59	.40	.03	-.09	-.26	.46	.05	.10	.03	.08
Was raised in benign environment	.59	-.59	.23	.09	-.05	-.04	.00	.05	.06	.24
Is jaded toward life	-.56	.13	-.46	.07	-.03	-.03	.13	.04	.21	.23
Has current family conflict	-.55	.46	-.10	-.05	-.01	-.09	.00	.13	-.50	.00
Is realistic	.53	-.50	.32	.38	.13	-.20	.13	-.04	-.05	-.01
Often breaks rules or laws	.07	.93	.06	.06	.04	-.08	-.07	-.01	.03	-.04
Gets into trouble with the law	.12	.91	.02	.04	.05	-.03	-.04	-.05	.10	-.06
Is deliberate and careful with work	.00	-.89	-.10	.18	-.03	-.18	.11	-.01	-.10	-.11
Attempts to avoid dangerous situations	.06	-.87	.03	.13	.01	-.23	.09	.15	.12	.04
Has a history of juvenile delinquency	.15	.86	.15	.11	.09	-.09	-.13	.05	.04	.10
Makes decisions that seem ill-advised	-.16	.86	-.06	-.17	.00	.15	.07	-.11	.09	.16
Often acts without thinking	-.02	.85	.25	.06	.11	.15	-.12	-.04	-.08	.12
Engages in risky behaviors	.32	.84	.21	.03	.07	.11	-.11	-.10	-.02	.16
Is careful	-.17	-.84	-.20	.15	-.07	-.24	.16	.04	.03	-.17
Has trouble restraining self	-.22	.82	-.02	-.01	.08	.18	-.03	-.07	-.01	-.13
Would show improvements with therapy	-.06	-.82	.14	.09	-.01	-.07	.03	-.42	.08	.03
Avoids conflict	-.06	-.82	.21	-.02	.04	-.08	-.02	.05	.13	.33
Is motivated primarily by short-term rewards and pleasure	.19	.80	.21	.03	-.08	.08	-.02	-.13	.20	.03
Takes unnecessary risks	.11	.80	.27	.00	.10	.20	-.01	-.28	-.05	.07
Is likely to improve with treatment	-.13	-.79	.18	.05	.18	.01	.09	-.40	.15	.08
Takes good care of daily responsibilities (e.g. child care, work, school)	.38	-.79	.05	.29	-.01	-.07	.06	-.06	-.09	.04
Manipulates others	.11	.77	.28	.19	-.08	.25	.13	.14	.07	.03
Has stable relationships with others	.51	-.73	.21	.24	-.01	-.12	-.01	-.07	-.02	.09
Is perfectionistic	-.40	-.73	-.14	.24	-.04	.05	.02	-.07	-.13	-.12
Problems will remain despite treatment	-.12	.72	-.09	-.20	-.07	-.15	-.12	.36	-.17	-.09

(Continued)

Table 5. (Continued)

Item content	Components									
	1	2	3	4	5	6	7	8	9	10
Has potential to be a contributing member of society	.46	-.71	.20	.38	.01	-.03	-.08	.00	-.09	.11
Attends (or would attend) treatment regularly	-.21	-.71	.17	-.05	.17	-.09	-.02	-.36	.05	-.04
Would be a difficult therapy client	-.21	.70	-.26	-.07	.01	-.08	-.11	.43	-.18	-.06
Is motivated to succeed	.46	-.70	.09	.31	.00	.02	.09	-.18	-.12	-.01
Is ready and willing to change	-.17	-.70	.22	-.19	.19	-.05	-.05	-.46	.13	.07
Blames own problems on others	.16	.69	-.45	.20	-.18	-.15	.09	.16	.02	-.15
Is a “thrill-seeker”	.47	.69	.29	.03	.07	.23	-.09	-.17	-.12	.18
Argues with others frequently	-.25	.69	-.47	.01	.06	.09	.02	-.05	-.09	-.39
Fits in well with mainstream society	.50	-.68	.19	.31	-.10	-.10	.05	-.03	-.16	-.01
Behaves in an unpredictable manner	-.27	.68	-.12	-.47	.19	-.04	-.05	.04	-.07	.00
Acts angry	-.41	.65	-.39	.07	.17	.06	.08	-.06	.01	-.27
Is agreeable	.45	-.64	.47	.11	-.11	-.02	.11	-.05	.08	.22
Is hostile toward others	-.30	.64	-.49	-.06	.09	.00	.06	.00	-.01	-.34
Gets along well with family	.45	-.57	.16	.05	.01	-.06	.06	-.11	.49	.03
Believes other people are “out to get them”	-.37	.05	-.83	-.16	-.09	-.06	.11	.00	-.04	-.11
Is paranoid	-.33	.07	-.82	-.24	.02	-.01	.15	.02	.10	-.10
Perceives others as threats to own security	-.39	-.11	-.81	-.05	-.04	-.04	.12	.11	-.06	.09
Has a rigid belief system	-.32	-.45	-.56	.17	-.29	.01	.01	.11	-.16	-.15
Sees or hears things others cannot	-.18	.12	-.08	-.84	.23	.06	-.01	-.05	-.10	.00
Behaves in a psychotic manner	-.39	.21	-.20	-.77	.21	.07	.05	.06	.04	-.01
Pretends to have problems	-.36	.13	.01	-.10	.81	.04	.12	.09	.03	.05
Pretends not to have problems	.41	-.23	-.01	.20	-.78	-.13	.01	.01	-.01	.04
Is exaggerating symptoms	-.57	.14	-.06	-.18	.65	.10	.16	.03	.10	-.01
Habitually denies own problems	.29	.11	-.24	.20	-.63	-.05	.17	.31	.02	.02
Lies frequently	.19	.46	.08	.08	-.55	-.14	.23	-.02	.32	.00
Seems to be driven by a motor	.22	.35	.10	.03	.18	.80	.01	-.05	-.01	-.08
Is hyperactive	.34	.32	.06	-.14	.24	.75	.05	-.02	-.02	-.02
Is currently abusing alcohol	-.15	.25	.07	.05	.02	.11	-.81	-.02	.00	.06
Abuses alcohol	.12	.13	.12	.17	-.08	-.11	-.75	-.27	-.09	-.11
Is likely to abuse prescribed medications	-.10	.03	.16	-.13	-.01	-.16	-.71	.43	.03	.06
Is currently abusing a drug other than alcohol	-.16	.40	.01	-.27	.17	.14	-.54	.38	.06	.00

Note: Bold text indicates the primary loading for each item on the respective component.

to it being used in other established q-sort measures such as the Midwestern q-sort (Williams & Weed, 2003) and the California q-sort (Block, 2008), should be used until sufficient evidence exists to suggest the flat distribution shape can outperform it. It is of interest, however, to note that all four

raters complained of feeling as if they were hard-pressed to find enough items to fill the extreme bins for several of the protocols. This suggests that appropriate distribution shape may vary from protocol to protocol.

A *post hoc* analysis was also conducted to determine if rater reliability varied consistently as a function of specific scale elevations on the MMPI-2. Since the flat distribution allows for a greater amount of variance, and thus a greater likelihood of higher inter-rater reliability, one might expect the reliability of the flat distribution to be higher on protocols that displayed elevations on constructs that were well-represented in the item set. The quasi-normal distribution was expected to outperform the flat distribution when relatively uncommon MMPI-2 scales were elevated in the protocol. This was not the case, as the quasi-normal distribution reliability was observed to correlate higher than the flat distribution with the three validity scales (L, F, K) and with the clinical scales. What is of most interest here, however, is that nearly all correlations of the reliability estimates with the MMPI-2 scales, whatever the magnitude, tended to be positive, with the average correlation for the quasi-normal distribution being .17 and the average correlation for the flat distribution being .09. Further, the strongest correlations were always positive, with MMPI-2 scales F and Hy being strongest for the quasi-normal, with correlations of .38, and Scales L and F being strongest for the flat, with correlations of .29 and .25, respectively. This suggests that higher overall levels of scale elevations may produce increases in rater reliability.

4.1. Instrument factor structure and divergence

With regard to the PCA conducted for this study, it is important to note that these observed components should be viewed with a certain degree of caution. Because the placement of each item within the q-set is interdependent on the placement of all the items that have been sorted before it, the correlations between items within a q-set are necessarily linked to the order in which the items are sorted. Consequently, it is suggested that only the first two components be interpreted, as they displayed eigenvalues so large that it is unlikely they could be attributed to item placement order alone. Furthermore, these two components seemed to comprise several aspects of personality such as conscientiousness and neuroticism (Löckenhoff, Terracciano, Costa, Bienvenu, & Crum, 2008), impulsivity and risk taking traits (Eskandari & Helmi, 2014), and emotionality and anger (Stringer, 2011), that have been thought to correlate with differences among substance abusing clients. Because only 13 of the 98 items did not have either a primary or secondary loading on one of these two components, and many had high loadings on one or the other, these components would likely remain relatively stable upon retest. The same could not necessarily be said of the other eight components.

That said, the main two components of the current q-set resemble two main components of the Mississippi q-set, as reported by Robinson (2004) when she analyzed q-ratings of a set of MMPI-2 protocols: “Emotional Stability” and “Anti-social Tendencies.” However, the other two components observed by Robinson (“Rationality” and “Somatization”) were not observed among our components. This suggests that the current set might be more specific, as designed, to a substance abusing population. The components observed in the current study are somewhat different from those observed by Robinson in that they appear to be broader, incorporating aspects of response style, treatment prognosis, and sensation seeking.

4.2. Limitations

Although the main goals of the study were achieved, there were some limitations to note. The items selected for the current q-set were chosen based on their relevance both to a substance abusing population and to phenomena accessible via the MMPI-2. This of course does not ensure that the items represent the most important clinical phenomena in such a setting. It is probable that other instruments and sources of information are needed to represent fully the assessment of problems associated with substance abuse.

The use of expert raters in item selection has been used in the development of other q-sets such as the Defense Mechanism Rating Scales q-sort (Di Giuseppe, Perry, Petraglia, Janzen, & Lingiardi, 2014). However, their use may also be a limitation to the extent that our experts do not reflect the judgment of users of the instrument. Also, at each stage, there was a single rater who produced somewhat different results than the other raters. Specifically, at multiple stages, one rater failed to rate a number of the items differently across protocols, later stating that they were poorly worded; these items received midpoint ratings for each protocol. The uniqueness of this rater's responses may have given them undue influence in selecting items. Further, this pattern of responses contributed to lower inter-rater correlations across all distribution shapes and all item sets, due to the overall reduction in variance. The extent to which item selection was influenced is not clear.

4.3. Conclusions and future directions

The q-set developed in this study comprises 98 items after refining and culling items from a pool of 229 items. The content of the resultant q-set covers (1) important features of substance abuse that (2) can be assessed using the MMPI-2. The instrument has the potential to be a useful tool in personality assessment research in substance abuse settings, but future research will be necessary to determine its incremental use over competing measures with similar aims, such as the Midwestern q-sort (Weed, 1997, 2006). Our finding that a flat distribution performed as well as a quasi-normal distribution is interesting, and somewhat counterintuitive. Future work might investigate more finely graded differences in distribution shape to discover any general rules for optimizing q-sort distribution. Finally, several statistical techniques were introduced and developed for use with evaluating prospective q-set items; future research should examine their utility in other contexts.

Funding

The authors received no direct funding for this research.

Competing Interests

The authors declare no competing interest.

Author details

Kevin R. Young¹

E-mail: kevin.young3@va.gov

Nathan C. Weed²

E-mail: weed1nc@cmich.edu

ORCID ID: <http://orcid.org/0000-0003-4195-8074>

Brian J. McCabe²

E-mail: yirzmbrian@gmail.com

Shehroo B. Pudumjee²

E-mail: pudum1sb@cmich.edu

ORCID ID: <http://orcid.org/0000-0002-4969-9321>

John E. Williams³

E-mail: john.eustis@gmail.com

¹ Louis Stokes Cleveland VA Medical Center, Cleveland, OH, USA.

² Department of Psychology, Central Michigan University, Mount Pleasant, MI, USA.

³ Canmore, Alberta, Canada.

Citation information

Cite this article as: Development of a q-set for MMPI-2 research with substance abusing populations in the USA, Kevin R. Young, Nathan C. Weed, Brian J. McCabe, Shehroo B. Pudumjee & John E. Williams, *Cogent Psychology* (2016), 3: 1223900.

References

American Psychiatric Association. (1994). *Diagnostic and statistical manual of mental disorders* (4th ed.). Washington, DC: Author.

Babor, T. F. (1996). The classification of alcoholics: Typology theories from the 19th century to the present. *Alcohol*

Health & Research World, 20, 6–17.

Babor, T. F., Dolinsky, Z. S., Meyer, R. E., Hesselbrock, M., Hofmann, M., & Tennen, H. (1992). Types of alcoholics II: Concurrent and predictive validity of some common classification schemes. *Addiction*, 87, 1415–1431. doi:10.1111/j.1360-0443.1992.tb01921

Babor, T. F., Hofmann, M., DelBoca, F. K., Hesselbrock, V., Meyer, R. E., Dolinsky, Z. S., & Rounsaville, B. (1992). Types of alcoholics, I: Evidence for an empirically derived typology based on indicators of vulnerability and severity. *Archives of General Psychiatry*, 49, 599–608. doi:10.1001/archpsyc.1992.01820080007002

Babor, T. F., Steinberg, K., Anton, R., & Del Boca, F. (2000). Talk is cheap: Measuring drinking outcomes in clinical trials. *Journal of Studies on Alcohol*, 61, 55–63. doi:10.15288/jsa.2000.61.55

Beck, A. T., & Steer, R. A. (1987). *BDI, Beck depression inventory: Manual*. San Antonio, TX: The Psychological Corporation.

Block, J. (1961). *The Q-sort method in personality assessment and psychiatric research*. Springfield, IL: Thomas. <http://dx.doi.org/10.1037/13141-000>

Block, J. (2008). *The Q-sort in character appraisal: Encoding subjective impressions of persons quantitatively*. American Psychological Association. doi:10.1037/11748-000

Butcher, J. N., Graham, J. R., Ben-Porath, Y. S., Tellegen, A., Dahlstrom, W. G., & Kaemmer, B. (2001). *MMPI-2 (Minnesota Multiphasic Personality Inventory, 2nd edition): Manual for administration, scoring, and interpretation* (Revised ed.). Minneapolis, MN: University of Minnesota.

Ciraulo, D. A., Piechniczek-Buczek, J., & Iscan, E. N. (2003). Outcome predictors in substance use disorders. *Psychiatric Clinics of North America*, 26, 381–409. doi:10.1016/S0193-953X(02)00106-5

Clements, R., & Heintz, J. M. (2002). Diagnostic accuracy and factor structure of the AAS and APS scales of the MMPI-2. *Journal of Personality Assessment*, 79, 564–582. doi:10.1207/S15327752JPA7903_10

Cloninger, C. R. (1987). Neurogenetic adaptive mechanisms in alcoholism. *Science*, 236, 410–416. <http://dx.doi.org/10.1126/science.2882604>

- Deskovitz, M. (2003). *Interpretive reliability of two common MMPI-2 code types* (Unpublished masters' thesis). Central Michigan University, Mount Pleasant, MI.
- Deskovitz, M. A., Weed, N. C., Chakranarayan, C., & Williams, J. E. (2016). Interpretive reliability of two common MMPI-2 profiles. *Cogent Psychology*, 3(1), 1161287. doi:10.1080/23311908.2016.1161287
- Deskovitz, M. A., Weed, N. C., McLaughlan, J. K., & Williams, J. E. (2016). Interpretive reliability of six computer-based test interpretation programs for the Minnesota Multiphasic Personality Inventory-2. *Assessment*, 23, 250–261. <http://dx.doi.org/10.1177/1073191115584970>
- Di Giuseppe, M., Perry, J. C., Petraglia, J., Janzen, J., & Lingiardi, V. (2014). Development of a Q-sort version of the Defense Mechanism Rating Scales (DMRS-Q) for clinical use. *Journal of Clinical Psychology*, 70, 452–465. doi:10.1002/jclp.22089
- Eskandari, H., & Helmi, S. (2014). The role of impulsivity, risk-taking, and sensation-seeking in people prone to substance abuse. *Applied Psychology*, 2, 48–60.
- Feit, M. D., Fisher, C., Cummings, J., & Peery, A. (2015). Substance use and abuse: Screening Tools and assessment instruments. In J. S. Wodarski, M. J. Holosko, & M. D. Feit (Eds.), *Evidence-Informed Assessment and Practice in Child Welfare* (pp. 123–133). Cham: Springer International Publishing.
- Fuller, R. K., & Allen, J. P. (2000). Patient to treatment matching. In G. Zerning, A. Saria, M. Kurz, & S. S. O'Malley (Eds.), *Handbook of alcoholism* (pp. 363–368). Boca Raton, FL: CRC. <http://dx.doi.org/10.1201/IHCHBKPHATOX>
- Gantner, A. B., Graham, J. B., & Archer, R. A. (1992). Usefulness of the MAC scale in differentiating adolescents in normal, psychiatric, and substance abuse settings. *Psychological Assessment*, 4, 133–137. doi:10.1037/1040-3590.4.2.133
- Gottheil, E., Thornton, C., & Weinstein, S. (2002). Effectiveness of high versus low structure individual counseling for substance abuse. *American Journal on Addictions*, 11, 279–290. doi:10.1080/10550490290088081
- Harrington, A. R. (2000). *The relative validity of computer-based interpretations of the MMPI-2* (Order No. AA19956068). Available from PsycINFO. (619555903; 2000-95012-074). Retrieved from <http://cmich.idm.oclc.org/login?url=http://search.proquest.com/cmich.idm.oclc.org/docview/619555903?accountid=10181>
- Harrison, P. A., & Hoffmann, N. G. (1989). *SUDDS: Substance Use Disorder Diagnostic Schedule manual*. St. Paul, MN: New Standards.
- Hayes, S., Nelson, R., & Jarrett, R. (1987). The treatment utility of assessment: A functional approach to evaluating assessment quality. *American Psychologist*, 963–974. <http://dx.doi.org/10.1037/0003-066X.42.11.963>
- Kadden, R. M., Longabaugh, R., & Wirtz, P. W. (2003). The matching hypothesis: Rationale and predictions. In T. Babor & F. K. DelBoca (Eds.), *Treatment matching in alcoholism* (pp. 81–102). New York, NY: Cambridge University.
- Levenson, M. R., Aldwin, C. M., Butcher, J. N., De Labry, L., Workman-Daniels, K., & Bossé, R. (1990). The MAC scale in a normal population: The meaning of “false positives.” *Journal of Studies on Alcohol*, 51, 457–462. doi:10.15288/jsa.1990.51.457
- Litt, M. D., Babor, T. F., DelBoca, F. K., Kadden, R. M., & Cooney, N. L. (1992). Types of alcoholics, II: Application of an empirically derived typology to treatment matching. *Archives of General Psychiatry*, 49, 609–614. doi:10.1001/archpsyc.1992.01820080017003
- Löckenhoff, C., Terracciano, A., Costa, P., Bienvenu, O. J., & Crum, R. (2008). Five-factor model personality profiles of drug users. *BMC Psychiatry*, 8, doi:10.1186/1471-244X-8-22
- MacAndrew, C. (1965). The differentiation of male alcoholic outpatients from nonalcoholic psychiatric outpatients by means of the MMPI. *Quarterly Journal of Studies on Alcohol*, 26, 238–246.
- MacAndrew, C. (1981). What the MAC scale tells us about men alcoholics: An interpretive review. *Journal of Studies on Alcohol*, 42, 604–625. <http://dx.doi.org/10.15288/jsa.1981.42.604>
- Mattson, M. E., Babor, T., Cooney, T., & Conners, G. (1998). Matching patients with alcohol disorders to treatments: Clinical implications from project MATCH. *Journal of Mental Health*, 7, 589–602. doi:10.1080/09638239817743
- McNeal, T. P. (1999). *The longitudinal impact of personality assessment on understanding clients during the course of psychotherapy* (Unpublished Manuscript). University of Mississippi, Oxford, MI.
- Morey, L. C., & Blashfield, R. K. (1981). Empirical classifications of alcoholism: A review. *Journal of Studies on Alcohol*, 42, 925–937. <http://dx.doi.org/10.15288/jsa.1981.42.925>
- Ozer, D. J. (1993). The q-sort method and the study of personality development. In D. C. Funder, R. D. Parke, C. Tomlinson-Keasey, & K. Widaman (Eds.), *Studying lives through time: Personality and development* (pp. 147–168). Washington, DC: APA. <http://dx.doi.org/10.1037/10127-000>
- Quirk, S. W., Christiansen, N. D., Wagner, S. H., & McNulty, J. L. (2003). On the usefulness of measures of normal personality for clinical assessment: Evidence of the incremental validity of the Revised NEO Personality Inventory. *Psychological Assessment*, 15(3), 1–15. doi:10.1037/1040-3590.15.3.311
- Robinson, B. C. (2004). *The reliability and internal structure of the Mississippi Interpretive Q-sort* (Unpublished doctoral dissertation). University of Mississippi, Oxford, MI.
- Selzer, M. L. (1971). The Michigan Alcoholism Screening Test: The quest for a new diagnostic instrument. *American Journal of Psychiatry*, 127, 1653–1658. <http://dx.doi.org/10.1176/ajp.127.12.1653>
- Serfass, D. G., & Sherman, R. A. (2013). A methodological note on ordered Q-sort ratings. *Journal of Research in Personality*, 47, 853–858. <http://dx.doi.org/10.1016/j.jrp.2013.08.013>
- Stein, L. A. R., Graham, J. R., Ben-Porath, Y. S., & McNulty, J. L. (1999). Using the MMPI-2 to detect substance abuse in an outpatient mental health setting. *Psychological Assessment*, 11, 94–100. doi:10.1037/1040-3590.11.1.94
- Stevenson, W. (1953). *The study of behavior: Q-technique and its methodology*. Chicago, IL: University of Chicago.
- Stringer, A. D. (2011). *Substance abuse proclivity and personality traits* (Order No. 3465534). Available from ProQuest Dissertations & Theses Global: Health & Medicine; ProQuest Dissertations & Theses Global: Social Sciences. (883983121). Retrieved from <http://cmich.idm.oclc.org/login?url=http://search.proquest.com/cmich.idm.oclc.org/docview/883983121?accountid=10181>
- Substance Abuse and Mental Health Services Administration. (2014). *Results from the 2014 National Survey on Drug Use and Health: Summary of National Findings*. Retrieved June 2, 2016 <http://www.samhsa.gov/data/sites/default/files/NSDUH-FRR1-2014/NSDUH-FRR1-2014.pdf>
- Turiano, N. A., Whiteman, S. D., Hampson, S. E., Roberts, B. W., & Mroczek, D. K. (2012). Personality and substance use in midlife: Conscientiousness as a moderator and the effects of trait change. *Journal of Research in Personality*, 46, 295–305. doi:10.1016/j.jrp.2012.02.009
- Weed, N. C. (1997). *Using the interpretive Q-sort technique to teach MMPI interpretation* (Unpublished Manuscript). Oxford, MI: University of Mississippi.
- Weed, N. C. (2006). Syndromal complexity, paradigm shifts, and the future of validation research: Comments on Nichols and Rogers, Sewell, Harrison, and Jordan. *Journal*

of *Personality Assessment*, 87, 230–235. doi:[10.1207/s15327752jpa8702_12](https://doi.org/10.1207/s15327752jpa8702_12)
Weed, N. C., Butcher, J. N., & Ben-Porath, Y. S. (1995). MMPI-2 measures of substance abuse. In J. N. Butcher & C. D. Spielberger (Eds.), *Advances in personality assessment* (Vol. 10, pp. 121–145). Hillsdale, NJ: Lawrence Erlbaum.
Weed, N. C., Butcher, J. N., McKenna, T., & Ben-Porath, Y. S. (1992). New measures for assessing alcohol and drug

abuse with the MMPI-2: The APS and AAS. *Journal of Personality Assessment*, 58, 389–404. doi:[10.1207/s15327752jpa5802_15](https://doi.org/10.1207/s15327752jpa5802_15)
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.



© 2016 The Author(s). This open access article is distributed under a Creative Commons Attribution (CC-BY) 4.0 license.

You are free to:

Share — copy and redistribute the material in any medium or format

Adapt — remix, transform, and build upon the material for any purpose, even commercially.

The licensor cannot revoke these freedoms as long as you follow the license terms.

Under the following terms:

Attribution — You must give appropriate credit, provide a link to the license, and indicate if changes were made.

You may do so in any reasonable manner, but not in any way that suggests the licensor endorses you or your use.

No additional restrictions

You may not apply legal terms or technological measures that legally restrict others from doing anything the license permits.



Cogent Psychology (ISSN: 2331-1908) is published by Cogent OA, part of Taylor & Francis Group.

Publishing with Cogent OA ensures:

- Immediate, universal access to your article on publication
- High visibility and discoverability via the Cogent OA website as well as Taylor & Francis Online
- Download and citation statistics for your article
- Rapid online publication
- Input from, and dialog with, expert editors and editorial boards
- Retention of full copyright of your article
- Guaranteed legacy preservation of your article
- Discounts and waivers for authors in developing regions

Submit your manuscript to a Cogent OA journal at www.CogentOA.com

