AN INVESTIGATION OF THE CONSTRUCT VALDIDITY OF SELECTED ADAPTIVELY ADMINISTERED MMPI-2 SUBSTANTIVE SCALES

A THESIS

SUBMITTED TO THE GRADUATE SCHOOL

IN PARTIAL FULFILLMENT OF THE REQUIREMENTS

FOR THE DEGREE

MASTER OF ARTS IN CLINICAL PSYCHOLOGY

BY

ALEX C. NYQUIST

DR. JOHNATHAN D. FORBEY – ADVISOR

BALL STATE UNIVERSITY

MUNCIE, INDIANA

JULY 2014

Acknowledgements

I would like to acknowledge and thank my thesis chair, Dr. Forbey, for his mentoring and advice throughout this project and my graduate career at *the* Ball State University. I would also like to extend special thanks to my research assistants who were essential to data collection and making sure the lab ran smoothly this year. Finally, I would like to thank my family and dear friends who supported me throughout the completion of my thesis.

ABSTRACT

THESIS: An Investigation of the Construct Validity of Selected Adaptively Administered MMPI-2 Substantive Scales

STUDENT: Alex C. Nyquist

DEGREE: Master of Arts

COLLEGE: Sciences and Humanities

DATE: July, 2014

PAGES: 48

Prior research into adaptive testing with the MMPI-2 has demonstrated significant time- and item-savings with little or no loss of validity (Forbey & Ben-Porath, 2007; Forbey, Ben-Porath, & Gartland, 2009; Forbey, Ben-Porath, & Arbisi, 2012). The current study investigated the utility and validity of both a computerized adaptive and non-adaptive "depression" module of the MMPI-2 utilizing a college student sample. Participants completed one of three MMPI-2 test-retest administrations (i.e., conventional-conventional, conventional-module, or conventional-adaptive module) as well as 15 criterion measures across two testing sessions exactly one week apart. The findings pointed to statistically significant and clinically meaningful time-savings in administering selected MMPI-2 scales (adaptively and non-adaptively). Criterion measures rationally selected to represent similar (depression, anhedonia, anxiety) and dissimilar (behavioral, thought, and somatic dysfunction) psychological constructs were administered to assess the convergent and discriminant validity of the depression module. The criterion correlations suggested minimal differences in discriminant and convergent validity across administration modes, pointing to good construct validity.

An Investigation of the Construct Validity of Selected Adaptively Administered MMPI-2 Substantive Scales

Recent research has suggested that the millennial generation of college students report significantly more psychological distress and dysfunction than earlier generations. For example, Benton and colleagues (2003) investigated data derived from therapist prospective reports spanning 13 years (from academic year 1988-1989 thru 2000-2001) of college student client problems at a university counseling center. Their results suggested significant increases in 14 of 19 client problem areas over time, including increases in depression and suicidal ideation. Along these lines, in their meta-analysis of MMPI (Hathaway & McKinley, 1943) and MMPI-2 (Butcher et al., 2001) data, Twenge and colleagues (2010) reported mean scale scores on the MMPI and MMPI-2 clinical scales, including Clinical scale 2 (which is a measure of depression), have increased among college students between 1938 and 2007. Given the well-established trend for increasing psychological difficulties, such as depression, among college-age adults, there is a growing need for efficient and accurate methods of early detection of such psychological difficulties among the college student population, and the MMPI family of tests offer tremendous potential toward such a goal. This study investigated the utility (time efficiency) and criterion validity of a "depression" module of the MMPI-2 among a college student sample. In addition, this study further explored the potential advantages of adaptive testing with MMPI-2 modules, or sets of selected scales.

The MMPI-2 has a long history of use with college students (Butcher, Atlis, & Fang, 2000; Butcher, Graham, Dahlstrom, & Bowman, 1990; Twenge et al., 2010). While frequently utilized with such individuals, a common criticism of the MMPI-2 is its item-length (567 items), and corresponding duration of test administration (Handel & Hostetler, 1990). To address such

criticisms, the MMPI-2 has been modified experimentally in a number of studies (Ben-Porath, Slutske, & Butcher, 1989; Forbey & Ben-Porath, 2007; Forbey, Ben-Porath, & Arbisi, 2012; Forbey, Ben-Porath, & Gartland, 2009; Handel, Ben-Porath, & Watt 1999) to explore the relative validity of computerized adaptive administrations to conventional administrations of the measure. In general the aim of adaptive testing is to reduce the number of items administered on a scale without impacting the construct validity of the measure. Two approaches to adaptive personality testing, item-response theory (IRT; Reise & Waller, 2009) and the Countdown method (Butcher, Keller, & Bacon, 1985) have been used with the MMPI-2.

IRT is a set of psychometric models used for constructing and/or administering psychological scales (Reise & Waller, 2009). These models aim to estimate the relation between an underlying trait of a measure and the respondent's item responses. IRT also assumes that a unidimensional trait is represented in the item content and has been utilized successfully for ability and aptitude testing (Reise & Waller, 2009). In computerized adaptive ability testing, the test-taker's response to an item or set of items (correct or incorrect) determines the level of difficulty for the next item or set of items (Butcher et al., 1985), and this technique increases ability test administration efficiency while maintaining precision and accuracy of the test-taker's resulting test score.

Butcher et al. (1985) addressed several theoretical issues in applying adaptive IRT to personality testing. An item-level criticism of IRT personality testing was that the responses on personality measures are more complex than on ability measures, that is, there is not a "right" or "wrong" answer to an item regarding a personality trait (as is the case for ability test questions) which immediately complicates construction of an adaptive IRT program for personality testing. Even for a measure with dichotomous, forced choice response options, like the MMPI-2, two

respondents can respond to an item in the same direction (e.g., both answer *True*) for completely different reasons. On the scale-level, Butcher and colleagues (1985) explain IRT assumes the items on a given scale reflect a unidimensional construct; however, the MMPI-2 scales are multidimensional in content, that is, the scales reflect heterogeneous personality constructs. Additionally, IRT assumes scale items vary in level of difficulty. Though this is not the case for personality scale items, it is arguably comparable to assume some items are better discriminators of the underlying trait than others, but Butcher and colleagues (1985) argued that this hypothesis of comparability was not empirically supported.

Childs, Dahlstrom, and Panter (2000) investigated the validity of adaptive IRT for the MMPI-2, and their results gave empirical support for Butcher et al.'s (1985) criticisms. These researchers hypothesized that the MMPI-2 Clinical scale 2 represented the test taker's level of underlying depression and investigated the item properties and full scale performance of Scale 2 using the two-parameter IRT model. They evaluated data derived from a large adult volunteer sample who were administered the "adult experimental" MMPI form, which contains all items from the original MMPI and MMPI-2 plus experimental items that were not included in the final MMPI-2. Their results suggested multidimensionality for both Clinical scale 2 and its corresponding Harris-Lingos subscales. Another problematic finding was evidence to suggest one of the Harris-Lingos subscales (Psychomotor Retardation) reflected different latent traits for men and women. The authors also reported response patterns were more consistent for obvious (face valid) items than subtle items on MMPI-2 Clinical scale 2. Overall their analysis suggested Clinical scale 2 is multidimensional, thus the measure violated a crucial assumption of IRT. This study provided empirical evidence against the utility and validity of IRT for MMPI-2 administrations.

Because MMPI-2 scales do not meet several assumptions of IRT, adaptive testing with the MMPI-2 has relied upon another adaptive administration approach, the Countdown Method (Butcher et al, 1985). The countdown method, as originally described by Butcher et al. (1985), categorizes test takers as either elevated or not elevated on the MMPI-2 scales based on whether individuals' scale scores reach the cutoff for clinical significance. For example, if the threshold for clinical elevation on a 20-item MMPI-2 scale is 10 items endorsed as "true" (i.e., the keyed direction), then if an individual endorses 11 items on the scale as "false" (i.e., the non-keyed direction), it is impossible for that individual to obtain a clinically elevated score for that scale. Under the countdown method, the remaining 9 items are not administered because clinical elevation has been ruled out for that scale. In practice then, only the minimum number of items necessary to elevate a scale (i.e., reach the cutoff) need be administered to categorize an individual in one of the groups. This original strategy was termed the "Classification" method, but Ben-Porath, Slutske, and Butcher (1989) later proposed and tested an alternative "Full Scores on Elevated Scales" (FSES) countdown method. The Classification and FSES countdown methods differ in how they handle cases when test takers achieve the cutoff for clinical elevation on the MMPI-2 scales. The Classification method terminates item administration once the possibility of clinical elevation has been determined. Using the above example, if an individual endorses 10 items in the keyed direction, the remaining items are not administered because the threshold for clinical elevation has been reached. In contrast, the FSES method continues item administration after the threshold for clinical elevation is reached so that the degree of an individual's level of elevation (i.e., severity) is available to the test interpreter.

Several researchers have compared the amount of item-savings, time-savings, and validity of these two adaptive techniques for the MMPI-2. Ben-Porath et al. (1989) conducted a

real data simulation using the MMPI-2 profiles from two personnel and two clinical samples to compare item-savings of the Classification and FSES methods. The results demonstrated no significant differences in terms of mean number of items administered between the methods for the personnel samples; however, fewer items were administered using the Classification method (*Ms*= 302, 310) than the FSES method (*Ms*= 341, 324) among the two clinical samples. Because Ben-Porath et al. (1989) utilized simulation data, a measure of the degree of time-savings and comparable validity was not available.

Handel, Ben-Porath, and Watt (1999) built upon Ben-Porath et al. (1989) in their real-data study with a clinical sample. Patients completed a computerized conventional (CC) administration of the MMPI-2 and criterion measures upon intake to Veteran's Administration addiction recovery clinic. Two to four days later, individuals who agreed to participate in the research study were assigned to either a CC or computerized adaptive (CA) condition (either Classification or FSES strategies were used in the CA condition). The results showed more items were saved (i.e., not administered) under the Classification method (M = 178.5, SD = 31.8) than under the FSES method (M = 148.2, SD = 50.0) compared to CC administrations. These item-savings corresponded to time-savings as well; more minutes were saved under the classification method (M = 16.1, SD = 5.1) than the FSES method (M = 13.4, SD = 5.9) compared to CC administrations. Although the Classification method resulted in slightly greater item- and time-savings than the FSES method, FSES testing provided more information about the respondents' severity of dysfunction and distress (Handel et al., 1999). However, in this study, only the original validity scales (L, F, and K), the 10 clinical scales, and 15 content scales were administered, while additional validity scales (e.g., VRIN, TRIN, and Fp) and other substantive scales (e.g., Substance abuse and Personality Psychopathology-Five (PSY-5))

developed after the 1989 MMPI-2 revisions were omitted from administration. This lack of inclusion of updated validity scales in particular was a methodological shortcoming as information concerning test-takers' level of content non-responsiveness in their approach to the MMPI-2 was completely omitted (i.e., VRIN and TRIN).

Within the last decade, a computerized adaptive version of the MMPI-2, the MMPI-2-CA, has been developed and empirically investigated. The MMPI-2-CA is based on a Windows Graphical User Interface (GUI) and administers all of the standard MMPI-2 scales (Forbey & Ben-Porath, 2007). This software allows test-takers to respond to the MMPI-2 items using either a mouse or keyboard (Forbey & Ben-Porath, 2007). Investigators have consistently reported that computerized adaptive (CA) administrations of the MMPI-2 provide significant item-savings and time-savings, with equal to or modestly improved external validity with criterion measures compared to computerized conventional (CC) administrations among Midwestern college students (Forbey & Ben-Porath, 2007), male inmates (Forbey, Ben-Porath, & Gartland, 2009), and male veterans recruited from an outpatient VA medical center (Forbey, Ben-Porath, & Arbisi, 2012).

With their sample of college students, Forbey and Ben-Porath (2007) investigated the comparative utility and validity of the MMPI-2-CA in CC and CA-Classification and CA-FSES administrations. This study expanded upon the Handel et al. (1999) methodology, as all of the MMPI-2 validity and substantive scales were included. Using a test-retest design, respondents participated in at least one CC administration for Time 1 (T_1) and/or Time 2 (T_2). Both CA-Classification ($t(719) = 9.31$, $p \leq .001$) and CA-FSES ($t(710) = 6.44$, $p \leq .001$) approaches led to significant time-savings compared to CC administrations. Similarly, CA-Classification ($t(719) = 61.17$, $p \leq .001$) and CA-FSES ($t(710) = 23.34$, $p \leq .001$) yielded significant item-savings

compared to CC administrations. Forbey and Ben-Porath (2007) also reported correlations between the three administration modalities with conjointly administered criterion measures. Overall the authors reported correlations between FSES and criterion measure scores were similar to or in some cases greater than those in the CC condition. However, in a number of cases, the Classification procedure led to significantly lower correlations with criterion measures compared to the CC condition. This pattern of results, as well as the fact that the FSES method yields a full T-Score led the authors to suggest that the FSES method may be preferable to the classification method.

Building upon the results of Forbey and Ben-Porath (2007), Forbey et al. (2009) investigated the comparative utility and validity of the MMPI-2-CA with audio augmentation and standard paper-and-pencil audio version of the MMPI-2 using a sample of male inmates. The audio augmentation was added to the MMPI-2-CA for this study to enhance the test's administration with individuals with potential reading difficulties. As part of their standard intake process, inmates completed the paper-and-pencil audio supported MMPI-2 and were afterward recruited to participate in the study. Participants were then assigned to either CC or CA (FSES) administrations with audio augmentation and also concurrently administered a set of criterion measures. Results demonstrated the mean number of items administered in the CA condition (M = 450.34, SD = 38.36) was significantly lower than the standard 557 items in the CC administration with a large effect size (d = 3.96; $t(322) = 35.61$, $p \leq .001$). Similarly, mean administration times were shorter for the CA (M = 31.75; SD = 7.56) than CC (M = 37.74, SD = 9.59) administrations and this difference reflected a medium effect size (d = .69; $t(322) = 6.24$, $p \leq .001$). Thus the CA administrations yielded meaningful item- and time-savings compared to CC administrations. Moreover correlations between various MMPI-2 administration modes and

criterion measures were greater in the CA condition than CC condition (and no correlations were greater in the paper-and-pencil administrations) which supported the authors' hypothesis for the convergent validity of MMPI-2 CA administrations and even suggested an advantage, in terms of validity, for adaptive administrations. Thus criterion validity appeared not to be lost with a CA administration of the MMPI-2; in fact, the findings suggest validity was possibly enhanced by the adaptive testing approach.

Finally, Forbey et al. (2012) further expanded upon the above findings supporting the utility and validity of the MMPI-2-CA with audio augmentation among a sample of male veterans undergoing outpatient medical treatment. In this test-retest procedure, respondents either participated in two CC administrations or one CC and one CA-FSES administration (and the later condition was counterbalanced), and testing appointments were six to ten days apart. Counterbalanced criterion measures were also administered, half during each testing session. Findings demonstrated fewer items were administered in the FSES (M = 453.24) than CC administrations leading to briefer administration times in the CA administration (M = 37.20, SD = 8.06) compared to CC administrations (M = 49.17, SD = 12.47) with a large effect size (d = 1.04). Correlations between MMPI-2 scores in CC and CA administrations and conceptually related criterion measures scores were indicative of convergent validity. Consistent with the findings of Forbey et al. (2009), this study suggested a moderate advantage in construct validity for CA FSES compared to CC administrations. Once again the researchers reported evidence to support the notion that adaptive testing with the MMPI-2 yields meaningful item- and time-savings with little or no cost to the construct validity of the measure. Overall, across these studies (Forbey & Ben-Porath, 2007; Forbey et al., 2009; Forbey et al., 2012), mean CA administration times were shorter than CC administrations, with the clinical settings requiring

longer administration times to complete the MMPI-2. It appears adaptive administrations of the MMPI-2 offer increased time-efficiency with no significant costs to the validity of the resulting profiles.

While a number of studies have demonstrated the relative validity of the full MMPI-2 adaptive administration to the conventional administration of the MMPI-2, far fewer studies have investigated computerized adaptive module (CAM) administrations of the MMPI-2. CAM administrations of the MMPI-2 could include selected scale set(s) (e.g., Validity, Clinical, and/or Content scales), population specific scale sets (e.g., scales for inpatient, outpatient or forensic populations), or construct specific scale sets (e.g., depression, anxiety, paranoia, etc.). For example, one could argue Handel et al. (1999) in effect examined a selected scale CAM (as the authors did not administer all of the MMPI-2 scales); however they were limited by the scales available via the MMPI-2 adaptive program that they were utilizing. Forbey, Ben-Porath, Graham, and Black (2004) were the first (and only) authors to purposely explore a population specific CAM of the MMPI-2 via their creation of a "correctional module", or a population specific scale set, composed of MMPI-2 validity scales and selected clinical, content, and other scales that they considered most relevant to an adult incarcerated population.

To explore the validity and utility of the "correctional module", Forbey and colleagues (2004) conducted two investigations with archival inmate data and original college student data. Using archival MMPI-2 profiles collected from male and female prison inmates upon intake to a Midwestern correctional facility, the authors investigated the potential item savings of the "correctional" CAM administration. Mean number of items administered to men was 296.76 and 306.49 for women. Because the researchers used archival simulation data, the time-savings of the "correctional module" could not be empirically investigated. Using an undergraduate sample

in a test-retest design, the researchers randomly assigned participants to the CC-CC or "correctional" CAM-CC condition to investigate both item- and time-savings of the "correctional" CAM. Module administrations yielded fewer items administered (M = 316.11, SD = 29.03) compared to CC administrations. The CAM condition also had shorter administration times (M = 22.71, SD = 7.03) compared to CC (M = 38.03, SD = 10.80). Analysis of the correlations between MMPI-2 scale scores and jointly administered criterion measures suggested the MMPI-2 scores derived from the "correctional" CAM conceptually correlated with external criterion measures, and these correlations were roughly equivalent to scores derived from a conventional administration (all scales) of the MMPI-2.

Given the potential for CAM administrations in terms of greatly abbreviated administration times without a corresponding loss of validity as indicated by Forbey et al. (2004), additional CAMs of the MMPI-2 should be explored with other methodologies and populations as well. As indicated, the MMPI-2-CA (Forbey & Ben-Porath, 2007) allows for the creation of such CAMs, that is, a set of selected scales to be administered either adaptively or non-adaptively. Though the MMPI-2-CA software is capable of administering such adaptive and non-adaptive modules, to date there are no additional published explorations of potential CAMs of the MMPI-2-CA other than the Forbey et al.(2004) study. Therefore, one aim of the current study was to address this gap in the adaptive testing literature by empirically investigating both adaptive and non-adaptive module administration options in terms of utility and validity. By including both adaptive and non-adaptive administrations of the module, the current study improved upon the Forbey et al. (2004) methodology, which only compared CC full-scale to CAM administrations (i.e., the researchers did not include a conventional module (CM) condition).

As discussed earlier, while psychological difficulties in general among college students has increased (Twenge, et al., 2010), depression is among the most common of these difficulties (Baumeister & Härter, 2007; Kessler, et al., 2003). In their analysis of longitudinal data from individuals born between 1900 and 1974, Lewinsohn, Rohde, Seeley, and Fisher (2003) reported significant increases in the prevalence of major depressive episodes among younger cohorts over time. Relatedly, Kessler and colleagues (2010) reported while the duration of major depressive episodes does not vary across age groups, symptom severity is significantly greater among younger cohorts, according to the National Comorbidity Survey Replication. Additionally, Benton et al. (2003) reported the number of college students seen for depression at a university counseling center doubled in 13 years, and the number of students reporting suicidal ideation tripled in that time period. These studies highlight the importance of an efficient, valid method of assessing depression-related distress among college students.

Thus the use of a depression focused CAM (i.e., construct specific) with the MMPI-2 may provide a possible avenue to address this need for an efficient assessment of depression for college students. Building on existing research into previous CAM administrations of the MMPI-2 (i.e., Forbey et al., (2004)), there remains the potential for such MMPI-2 modules to yield valid results in a time-efficient manner. Thus, the current study investigated the utility and validity of a "depression" CAM administration of the MMPI-2 to efficiently and accurately detect significant psychological distress related to symptoms of anhedonia among college students. By including only selected scales of the MMPI-2, administration time is immediately shortened since the number of test items has been reduced. Also, adaptively administering selected scales offers the potential for additional time-savings. The "depression" module, constructed for the current study, contains only the MMPI-2 substantive scales (i.e., Clinical,

Content, RC, and PSY-5) that target the construct of depression and the validity scales (which were designed to assess the respondent's general approach to the test, such as inattention to item-content, fatigue, over- or under-reporting of distress and dysfunction). In the current study, the utility (defined by time and item savings) and comparative validity of both a non-adaptive and adaptive version of the "depression" module were examined in comparison with a full conventional administration of the MMPI-2. Overall it was hypothesized that the module administrations will yield significant item-savings and time-savings over conventional administrations. Further it was hypothesized that the adaptive "depression" module will lead to greater item savings, with no loss of validity, compared to the non-adaptive "depression" module. Given the overlap between the constructs of depression and anxiety, criterion measures that assess depression/anhedonia as well as anxiety will be included to explore the convergent validity of the "depression" module. It was predicted that the depression module would have good convergent validity with conceptually similar criterion measures (e.g., depression/anhedonia) and good discriminant validity with conceptually dissimilar criterion measures (e.g., externalizing, thought disorder, or somaticizing symptoms).

Method

Participants

A total of 282 participants were recruited for the current study, which is part of a larger, ongoing data collection protocol. Potential participants were undergraduate men (n =73) and women (n = 209) from a state university in the Midwestern United States who volunteered in exchange for credit in their undergraduate introductory psychology course. Their ages ranged from 18 to 35 years (M = 18.8, SD = 1.76). They were primarily Caucasian (87.2%, n = 246),

8.2% were African American (n = 23) and 4.6% either had a different ethnicity or did not report their ethnicity (n = 13).

Potential participants were removed from the study if they produced an invalid MMPI-2 profile in either the first or second data collection sessions. Invalid MMPI-2 profiles were defined as having either a Cannot Say raw score (CNS) ≥ 30 or a Variable Response Inconsistency (VRIN) or VRIN-CA[1], True Response Inconsistency (TRIN), Lie (L), or Defensiveness (K) T score > 80, or an F(p) T score > 100. A total of 46 participants (16.3%) produced an invalid profile at Time 1, Time 2 or both sessions**.** There were no significant differences between valid and invalid groups in terms of age, t = -1.665(280), p = .097. However, there were significant differences between groups in terms of gender, χ^2 = 5.025(1), p = .025, and ethnicity, χ^2 = 28.281(2), p < .001, with men producing more invalid profiles than women, and African Americans and those who reported other ethnicities producing more invalid profiles than Caucasians. The remaining sample included a total of 55 men and 181 women, with an overall mean age of 18.75 years (SD = 1.54). The final sample was mainly Caucasian (91.5%, n = 216), with African Americans (6.4%, n = 15) and those reporting other ethnicities (2.1%, n = 5) constituting a smaller portion of the sample.

Measures

Minnesota Multiphasic Personality Inventory-2-Computerized Adaptive Version. Forbey and Ben-Porath (2007) developed and validated the MMPI-2-CA which consists of 557 items administered via computer (ten items of the original 567 are not included because these are not scored on any of the standard MMPI-2 scales). Test-takers responded using either mouse or keyboard. The software provided several administration options including a full conventional

[1] VRIN-CA is a shorter, experimental version of the VRIN scale developed specifically for computer adaptive modules and was used in the "depression" CAM

administration, full scores on elevated scales (FSES) adaptive administration, classification adaptive administration, or the administrator's choice of a set of scales to administer either conventionally or adaptively. The test administrator may also select the cutoffs for determining scale elevation (for adaptive administrations; Forbey & Ben-Porath, 2007). While test administrators can select Classification or FSES adaptive administrations of the MMPI-2-CA, the FSES method was utilized in this study to obtain maximal information on participants' elevation on the selected depression scales. For the depression CAM (either adaptive or non-adaptive), in addition to the standard validity scales, the following substantive scales were administered: Clinical Scale 2, Content Scale Depression (DEP), Restructured Clinical Scales Demoralization (RCd) and Low Positive Emotions (RC2), andPSY-5 Scale Introversion/Low Positive Emotionality (INTR) (see Table 1).

Criterion Measures. Fifteen external measures were rationally selected to reflect the psychological constructs of depression, anhedonia, anxiety, somatization, antisocial tendency, substance abuse, impulsivity, and thought disorder. These measures were used to investigate the convergent and discriminant validity of the MMPI-2 depression module. Two measures of anhedonia are included in addition to two measures of depression because Joiner, Walker, Pettit, Perez, and Cukrowicz (2005) have argued that anhedonia is the best discriminator between the constructs of depression and anxiety.

Antisocial Process Screening Devise – Self-report. The APSD-self-report is a 20-item behavior rating scale (Frick & Hare, 2001). Respondents used a three-point rating scale to rate each item (0= *Not at all*, 2= *Definitely*), with higher scores indicating higher levels of antisocial behaviors. Although originally developed for adolescents, this measure has been utilized by one of the test authors (Kruh, Frick, & Clements, 2005) as well as others (e.g., Ross, Molto, Poy,

Segarra, Pastor & Montanes, 2007) with young adults. A sample item is "You act charming and nice to get what you want." The self-report version of the APSD, full scale score has good internal consistency (α = .78 - .81), though Muñoz and Frick (2007) found less evidence for the internal consistency of the subscale scores. The APSD-self-report has good reliability and validity; the one-year stability estimates were .70 and .72, and good concurrent validity was demonstrated by significant correlations with external measures of antisocial behavior (Muñoz & Frick, 2007). A reliability coefficient of (α) .76 was observed with the sample in this study.

Cardiff Anomalous Perceptions Scale. The CAPS is a 32-item measure used to measure the frequency of perceptual anomalies (Bell, Halligan, Ellis, 2006). Each item is a question concerning if and how often (frequency) one has experienced a variety of perceptual anomalies, for example, "Do you ever hear noises or sounds when there is nothing about to explain them?" Respondents rate the items on a five-point scale (1= *Never*, 5= *Always*), with higher scores indicating a higher frequency of perceived anomalies. The CAPS has good internal consistency (α = .87) and good test-retest reliability (.78). Bell and colleagues (2006) found significant correlations between the CAPS and measures of related constructs which suggests good convergent validity; the researchers also demonstrated good criterion validity as mean CAPS scores for a clinical sample were significantly higher than mean scores for a non-clinical sample. A reliability coefficient (α) of .96 was observed with the sample in this study.

Center for Epidemiologic Studies – Depression Scale. The CES-D is a 20-item self-report measure for current level of depression to be used for the general population (Radloff, 1977). Each item describes a particular feeling and subjects rate how often they have felt that feeling in the past week on a 0 (*rarely or none of the time*) to 3 (*most or all of the time*) scale. Sample items include: "I felt lonely"; "I enjoyed life" (reverse scored). Higher scores on the

CES-D scale indicate more depressive symptoms experienced in the last week. The CES-D scale has excellent internal consistency ($\alpha \geq .84$) and acceptable inter-item and test-retest correlations. This scale has also demonstrated good discriminate validity, for clinical and non-clinical samples, as well as convergent validity with other self-report measures of depression (Radloff, 1977). A reliability coefficient of .91 was observed with the sample in this study.

Clinical Anxiety Scale. The CAS is a 25-item scale used to measure the severity of clinical anxiety (Hudson, 1992). Respondents endorsed items on a five-point scale (1= *Rarely or none of the time*, 5= *Most or all of the time*). One sample item is "I get upset easily or feel panicky unexpectedly." The items are based on the diagnostic criteria for anxiety disorders in DSM-III. The cutoff for clinical significance is 30 (±5). A reliability coefficient (α) of .90 was observed with the sample in this study.

Cognitive-Somatic Anxiety Questionnaire. The CSAQ is a 14-item measure of the cognitive and somatic components of anxiety, but can also be used as a trait measure of anxiety because it assesses enduring patterns (Schwartz, Davidson & Goleman, 1978). Items were scored on a five-point scale (1= *Not at all*, 5= *Very much so*.) "I imagine terrifying scenes," is a sample cognitive item and "I feel jittery in my body" is a sample somatic item. The CSAQ has good concurrent validity as demonstrated with significant correlations with the State-Trait Anxiety Inventory. The reliability coefficient for the composite was (α) .91 in this study, and .87 and .80 for the cognitive and somatic subscales, respectively.

Dissociative Experiences Scale. The DES is a 28-item scale used to measure dissociation (Bernstein & Putman, 1986). Items were created based on data from interviews with individuals who met DSM-III criteria for dissociative disorders. The items were scored on a ten-point scale (1= *Very little*, 10= *Very much*), with higher scores indicating greater levels of

dissociation. A sample item is "Some people have the experience of finding new things among their belongings that they do not remember buying." This instrument has good split-half reliability, with coefficients ranging from .71 to .96, and good test retest reliability demonstrated by a coefficient of .84 (Bernstein & Putman, 1986). The DES has good construct validity because it was not significantly correlated with unrelated variables, but significantly correlated with theoretically related variables (Bernstein & Putnam, 1986). A reliability coefficient (α) of .96 was observed with the sample in this study.

Green Paranoid Thoughts Scales. The GPTS is a 32-item scale used to measure paranoid thoughts (Green, Freeman, Kuipers, Bebbington, Fowler, Dunn, & Garety, 2008). The instrument assesses two constructs related to paranoia, ideas of persecution and social reference. Respondents rate statements about possible experiences on a five-point scale (1= *Not at all*, 5= *Totally*) with higher scores indicating more paranoid thoughts. "I was stressed out by people watching me" is a sample item from the social reference subscale, and "Certain individuals have had it in for me" is a sample item from the ideas of persecution subscale. Green et al. (2008) demonstrated evidence of excellent internal consistency (αs = .90 - .95) and calculated a test-retest reliability coefficient (.87) for the GPTS. These researchers found good evidence of the convergent validity of the GPTS among clinical and non-clinical samples; the GPTS-total was significantly correlated with measures of depression (ρs = .56, .50), anxiety (ρs = .41, .49), and paranoia (ρs = .81, .71; Green et al., 2008). A reliability coefficient of (α) .95 was observed with the sample in this study.

Index of Alcohol Involvement. The IAI is a 25-item scale used to measure a respondent's degree of alcohol abuse problems (MacNeil, 1991). Each item is a statement that reflects presence or absence of difficulties with alcohol use. Items were scored on a seven-point

scale (1= *Never*, 7= *Always*). A sample item is "When I have a drink with friends, I usually drink more than they do." The IAI has excellent internal consistency (α = .90) and has good construct validity (MacNeil, 1991). A reliability coefficient (α) of .82 was observed with the sample in this study.

Index of Drug Involvement. The IDI is a 25-item scale used to measure a respondent's degree of drug abuse problems (Faul & Hudson, 1997). Each item is a statement that reflects presence or absence of drug abuse. Items were scored on a seven-point scale (1= *Never*, 7= *Always*). A sample item is "My drug use causes problems with my work." The IDI has excellent internal consistency (α = .97). This instrument has good construct validity and a cutoff of 30 for clinical significance has been established (Faul & Hudson, 1997). A reliability coefficient (α) of .72 was observed with the sample in this study.

Mood Disorder Questionnaire. The MDQ is a 13-item measure for bipolar spectrum disorder (Hirshfield et al., 2000). The *yes/no* questions are based on the DSM-IV diagnostic criteria. The MDQ has excellent internal consistency (α = .90). The measure has also demonstrated good concurrent validity with the Sheehan Disability Scale and the Social Adjustment Scale. Hirshfield and colleagues (2000) demonstrated good sensitivity (.73) and specificity (.90) when a score of seven is used as the cutoff for clinical significance. A reliability coefficient (α) of .84 was observed with the sample in this study.

UPPS Impulsive Behavior Scale. The UPPS was created through exploratory factor analysis which identified four factors of impulsivity: (lack of) premeditation (11 items), urgency (12 items), sensation seeking (12 items) and (lack of) perseverance (10 items; Whiteside & Lynam, 2001). There is a subscale for each of the four factors; there is no composite score. Respondents rate each item on a five-point (1= *Strongly disagree*, 5= *Strongly agree*), with

higher scores indicating greater levels of impulsive behaviors. A sample item from the sensation seeking subscale is "I sometimes like doing things that are a bit frightening." The UPPS is a reliable measure of impulsivity as shown by good internal consistency (α = .87) and self-reports were significantly correlated with peer reports (Kämpfe & Mitte, 2009). Whiteside, Lynam, Miller and Reynolds (2005) found evidence to support the construct validity of the UPPS, and their discriminant function analysis showed the UPPS scales could discriminant between respondents with borderline personality disorder, pathological gambling, or alcoholism and those in the control group. Reliability coefficients (α) in this study were .79, .87, .62, and .69 for the urgency, premeditation, perseverance, and sensation seeking scales, respectively.

Patient Health Questionnaire-15. The PHQ-15 is the somatic symptom subscale taken from the full Patient Health Questionnaire (Kroenke, Spitzer, Williams, 2002). The scale items account for 14 of the most prevalent somatization disorder somatic symptoms in the DSM-IV. Respondents rate a list of possible somatic complaints, for example "Dizziness," on a three-point scale (0= *Not bothered at all*, 2= *Bothered a lot*). The PHQ-15 has good internal consistency (α = .80) and good convergent validity as demonstrated by significant correlations with functional status, disability days, and symptom-related difficulties (Kroenke et al., 2002). A reliability coefficient (α) of .84 was observed with the sample in this study.

Rimon's Brief Depression Scale. The RBDS is a seven-item scale used to measure depressive symptoms (Keltkangas-Järvinen & Rimon, 1987). Items are based on diagnostic symptoms of depression, and respondents rate their answer to questions about various symptoms on a four-point scale (1= *No*, 4= *Severe*). A sample item is "Have you noticed a recent decrease in your interest in your work and/or your hobbies?" The RBDS has good reliability and convergent validity as shown by significant correlations with the Beck Depression Inventory

(Keltkangas-Järvinen & Rimon, 1987). A reliability coefficient (α) of .85 was observed with the sample in this study.

Snaith-Hamilton Pleasure Scale. The SHAPS is a 14-item measure of anhedonia (Snaith,1993). Items consist of statements about situations an individual might enjoy and respondents rate each statement on a four-point scale (1= *Definitely Agree*, 4= *Strongly Disagree*), with higher scores indicating greater levels of anhedonia. A sample item is "I would enjoy being with family or close friends." The SHAPS has good internal consistency (α = .91) and good convergent validity with measures of depression and good discriminant validity with measures unrelated to depression (Nokonezny, Carmody, Morris, Kurian, & Trivedi, 2010). Snaith et al. (1995) reported evidence to suggest the SHAPS is sensitive to changes in client status, and for this reason, test-retest reliability coefficients were not reported. The non-parametric Kuder-Richardson formula (comparable to the parametric Cronbach's α) was used to estimate internal consistency; .86 suggests good internal consistency of the SHAPS (Snaith et al., 1995). A reliability coefficient (α) of .93 was observed with the sample in this study.

Tripartite Pleasure Inventory. The TPI is a 12-item measure of trait anhedonia (Leventhal, 2012). Though this instrument contains three subscales for hedonic responsivity, hedonic engagement and hedonic desire, only scores for hedonic responsivity will be utilized in the present research. The TPI items describe 12 types of experiences and for each experiences, respondents rate how much pleasure, happiness, or enjoyment they usually feel with respect to that experience on a five-point scale (1= *No Pleasure,* 5= *Extreme Pleasure*;) with lower scores indicating greater anhedonia. The TPI has demonstrated good internal consistency (α = .77) and good convergent validity with the SHAPS (r = .44, p <.0001; Leventhal, 2012). The TPI has also demonstrated good convergent validity with depressive symptoms (r = -.34, p < .01;

Meinzer, Pettit, Leventhal, & Hill, 2012). A reliability coefficient (α) of .81 was observed with the sample in this study.

Procedure

A test-retest design was used to investigate the validity of adaptively administering the depression module. Participants, in groups of up to five individuals, completed the full MMPI-2-CA conventional administration and the depression module, either conventional or adaptive administration. In addition, criterion measures rationally selected to reflect various psychological constructs were administered via online survey to investigate the convergent and discriminant validity of the proposed MMPI-2 depression module. Participants' scores on criterion measures were excluded from analysis if they responded to fewer than 90% of the items on a given measure or if their scores were outside a four-standard deviation range of the mean for the sample. The criterion measures were split into two "sets," and one set was administered at each testing session. There were eight scales in "Set A" – 166 items, and seven scales in "Set B" – 175 items (see Table 2). Measures within each set were presented in randomized orders by the online survey. Also included in Set B were the ten MMPI-2 items that do not load on any MMPI-2 scales and therefore were omitted from the MMPI-2-CA program. However, because some of these items load on MMPI-2-RF scales, these items were included for future research investigations.

During the first testing session (T_1), participants completed a randomly assigned version of the MMPI-2 and one half of the collateral measures. Participants returned for the second testing session (T_2) exactly seven days after T_1 and completed another MMPI-2 and the other half of the collateral measures. Each participant completed at least one full computerized conventional version (CC) of the MMPI-2; roughly half at T_1 and half at T_2. Approximately one

third of participants also completed the full conventional version a second time, thus these respondents were considered the control condition. Another third of the respondents completed the conventional module (CM) for depression, and the final third completed the computerized adaptive module (CAM) for depression. The testing session order was counterbalanced across all three MMPI-2 conditions. For each testing session the presentation order of the MMPI-2 and set of collateral measures was counterbalanced, so that roughly half the participants completed an MMPI-2 first and collateral measures second, and vice-versa. Presentation order of the fifteen collateral measures (eight scales administered under "Set A" and seven scales administered under "Set B") was randomized by the online survey software.

Results

To test the utility hypothesis, t-tests were computed to determine if the differences in mean administration times and mean items-administered were significant between the conventional and the depression module (non-adaptive and/or adaptive) administration modalities. Table 3 provides a summary of the time and item savings per administration modality. The computerized conventional modality administered 557 items (as 10 items are not score on any MMPI-2 scales). In the depression CM, 320 items were administered which is 42.6% fewer items administered than in the conventional modality. In the depression CAM a mean of 226.6[2] items were administered. As predicted, the depression CM (237 fewer items) and CAM ($t(79) = 233.50$, $p < .001$)[3] yielded shorter administration times than the CC modality. In addition, the adaptive administration resulted in increased item savings compared to the

[2] The CAM modality utilized a different version of the VRIN scale, VRIN-CA, which contains fewer item pairs than the original VRIN scale. Further implications are addressed in the discussion.

[3] Levene's test for homogeneity of variances was significant for all tests of the utility hypothesis, and the "equal variances not assumed" corrected *t* statistics and degrees of freedom are reported.

depression module ($t(79) = 66$, $p < .001$). The mean differences between the CAM and other administration modalities had large effect sizes (CC: $d = 36.2$, CM: $d = 10.50$) indicating the depression CAM administration modality yields substantive items savings over CC administrations as well as the non-adaptive depression module, respectively.

Means for CC administration times ranged from 38.1 to 40.2 minutes. The mean administration time in the depression CM was 21.1 minutes ($SD = 5.11$) and 14.8 minutes ($SD = 3.41$) in the CAM. The CM modality ($t(154) = 11.72$, $p < .001$) and CAM modality ($t(152) = 9.25$, $p < .001$) resulted in significantly shorter administration times compared to the CC administration. In addition the adaptive module yielded greater time savings over the non-adaptive depression module ($t(160) = 9.29$, $p < .001$). Mean differences in the administration times between the CC and CM ($d = 1.88$), CC and CAM ($d = 3.10$), and CM and CAM modalities ($d = 1.46$) yielded large effect sizes.

Test-retest zero-order correlations for the selected MMPI-2 scales used in the depression module are reported in Table 4. Fisher's *R* to Z transformations were used to allow for significance tests between test-retest correlation coefficients. Cohen's *q* statistic for effect size was calculated to indicate the magnitude of *differences* between transformed test-retest correlations with .10, .30, and .50 reflecting small, medium and large effect sizes, respectively. Ten significant differences between test-retest correlations were observed in this sample, and the effect sizes for these differences ranged from medium to large. Nine of these differences reflected that the test-retest correlation was higher in the control condition (CC-CC) than either the module or adaptive module condition.

Correlations between the criterion measure and selected MMPI-2 depression scales are reported in Tables 5, 6, and 7 for each administration condition. Because the analysis was

predicted to support the null hypothesis (i.e., no significant differences in scale scores across conditions), an alpha correction was not employed for comparisons in order to reduce the possibility of Type II error. In each modality condition there were 95 correlate comparisons, which implied five comparisons would be statistically significant by chance. Cohen's *d* values were used to measure the effect sizes of significant differences between zero-order correlations (.20, .50, and .80 reflect small, medium, and large effects, respectively). As indicated in Table 5, for the CC-CC condition, three criterion test-retest correlates were significantly different. One difference was observed for a convergent correlation (i.e., CESD with Clinical scale 2, d = .65) which was significantly higher at CC Time 2. Two differences were observed among correlations for discriminant validity, with one higher at Time 1 (i.e., Impulsivity-Urgency with PSY-5- Introversion, d = .90) and one higher at Time 2 (i.e., CSAQ-Somatic and RC Scale Demoralization, d = .50).

Table 6 reports the results of the CC-CM comparisons. In the CM condition, eight criterion correlates were significantly different between the CC and CM administrations; half of these differences were significantly higher in the CC and half in the CM administration. Five differences were observed among correlations for convergent validity; two of these differences were higher in the CC administration (i.e., CAS and RBDS with PSY-5-Introversion, ds = .65, .63) and three differences were higher in the CM administration (i.e, CAS and SHPS with Clinical scale 2, ds = .46, .53, and TPI with RC Scale Demoralization, d = .49). Three differences among discriminant correlations were observed; two of these differences were higher in the CC administration (i.e., CAPS with Clinical Scale 2, d = .64, and IAI with PSY-5-Introversion, d = .45) and one difference was higher in the CM administration (i.e., DES with Content Scale Depression, d = .61).

Table 7 reports the results of the CC-CAM comparisons. In the CAM condition, seven criterion correlates were significantly different between the CC and CAM administrations. Among the convergent correlations, four significant differences were observed; two differences were higher in the CC administration (i.e., RBDS with RC Scale 2, d = .50, and TPI with PSY-5-Introversion d = .46) and two were higher in the CAM administration (i.e., CESD and CSAQ-Cognitive with PSY-5-Introversion ds = .56, .64). Three differences occurred among discriminant correlations and all were significantly higher in the CAM administration (i.e., DES with Clinical Scale 2, d = .61, DES with PSY-5-Introversion, d = .59, and APSD with RC Scale Demoralization, d = .48).

Discussion

Researchers have noted increases in depression among the millennial generation of college students (Benton et al., 2003; Twenge et al., 2010) and that depression is the most common psychological difficulty for this population (Baumeister & Härter, 2007; Kessler et al., 2003). The present study sought to address the potential for time efficient and valid means of assessing emotional dysfunction related to depression among the college student population by examining the utility and validity of a depression module of the MMPI-2 administered in both an adaptive (CAM) and non-adaptive (CM) manner. With respect to utility, the results of the current study supported the advantages of an MMPI-2 depression module in terms of time and item savings over computerized conventional administrations. As predicted, the adaptive depression module yielded significant time and item savings over the non-adaptive module which supports the notion that adaptive testing with the MMPI-2 increases the efficiency of computer and selected-scale administrations of the MMPI-2. There was limited evidence to suggest a loss of criterion validity in the CM and CAM administrations (i.e., significant differences in criterion

correlations between administration modalities). Because evidence of validity threats in the depression CM and CAM was minimal, it is argued that any loss of validity is negligible in light of the evidence from criterion measures to support the validity of CM and CAM administrations. The findings in this study also support the construct validity of the depression module in terms of both convergent and discriminant validity. Overall, this study supports the use of a depression module among college students as both a time efficient and valid means of assessing emotional distress and dysfunction.

While both modules provided item and time savings over the CC, the CAM administration modality yielded the greatest time and item savings, as the CAM administered approximately half of the items and in half the time of the CC modality, on average, and also led to significant item savings over the CM administration. Though the absolute time savings in minutes may seem insignificant, in a clinical setting ten or fifteen minutes saved can be quite meaningful for both clinician or test administrator and test-taker alike. Often in clinical settings clients are given a battery of assessments which can be tedious and tiring for individuals; thus, to a client, saving ten minutes on a test may not seem insignificant or trivial. For test administrators, saving ten or more minutes per test-taker adds up quickly. For example, assume a practitioner can administer the computerized conventional MMPI-2 to six clients in roughly four hours. If the practitioner switches to module administrations of the MMPI-2, s/he could assess eight clients in the same length of time. Thus the utility of adaptive and non-adaptive MMPI-2 modules has meaningful implications for psychological assessments conducted in clinical as well as other settings (e.g., health, forensic, etc.).

With respect to general scale validity, over one quarter (10 out of 39) of the test-retest correlate comparisons among administration modes on the selected MMPI-2 scales were

significantly different, and in all but one case the CC-CC administration resulted in higher test-retest correlations than either of the module administrations. The majority of the differences (7 out of 10) in the test-retest correlates were observed on the validity scales. The significant differences in test-retest correlations favored the CC-CC condition; however, the observed test-retest correlations in the CC-CM and CC-CAM conditions were good which suggested good test-retest reliability between CC and module administrations of the MMPI-2. Further three of five depression related scales exhibited significant differences between conditions on the substantive scales (all of which favored the CC-CC condition); however, all of the test-retest correlations on these scales were very good (ranging from .75 to .95). Overall the lack of significant differences for one week test-retest correlations on the selected substantive scales and the magnitude of these correlations suggested stability and limited loss of validity in either module administration.

With respect to construct validity, criterion correlations within administration modalities were quite similar between T_1 and T_2, CC-CM and CC-CAM administrations. Among nearly 300 correlation comparisons across the three modality conditions, 18 of these comparisons were significantly different across the three procedures. The low frequency of significant differences suggests some or all of the observed differences in this sample occurred by chance (as 15 correlations would be expected to be significant by chance alone out of 300 correlations comparisons at the .05 level) and thus most likely does not point to a pattern of differential criterion correlations among the administration modalities. The consistency of the criterion correlations over a one-week interval points to construct validity of the depression module. There was no pattern of differences in the criterion correlations among the administration conditions which suggests the depression CM and CAM are equivalent in terms of criterion validity with full scale administrations of the MMPI-2.

The evidence from the criterion correlations suggested good construct validity of the depression CM and CAM in terms of convergent validity and discriminant validity. Overall the MMPI-2 depression scales were moderately correlated with the criterion measures of emotional dysfunction (i.e., depression, anhedonia, and anxiety) across all administration conditions. This pattern pointed to the convergent validity of the MMPI-2 substantive (depression) scales. Where there were observed differences in the convergent correlations between administration modes, the differences did not consistently favor (i.e., greater in magnitude) any modality in particular. Rather it could be argued some of the differences in convergent correlations occurred by chance. Though a few correlation differences were observed, these did not seem to negatively impact the convergent validity of the MMPI-2 scales.

Overall there was a consistent pattern of weaker correlations between the MMPI-2 depression scales and criterion measures of thought, behavioral and somatic dysfunction across all administration conditions, which suggested good discriminant validity for the MMPI-2 substantive scales. The correlations with thought and behavioral dysfunction and the MMPI-2 depression scales were small in magnitude. In terms of somatic dysfunction, some of the criterion correlations tended to be moderate, but it could be argued this reflected the comorbidity of emotional/mood and somatic complaints of individuals who report depressive symptoms. Also, these moderate correlations with measures of somatization were observed in the "original" MMPI-2 scales (i.e., Clinical scale 2 and DEP) and not in the "newer" scales (i.e., RC2 and INTR) whose correlations with somatization measures were much smaller in magnitude. Though significant differences among correlations appeared among discriminant measures, the magnitudes of these correlations were consistently small; thus, these instances of significant differences did not seem to subtract from the discriminant validity of the MMPI-2 substantive

scales. The consistency of discriminant correlations observed in both CC and module administrations suggested there were no threats to discriminant validity in the CM and CAM conditions.

The current study had several limitations, such as the generalizability of the findings based on a non-clinical college student sample. While the selected sample was appropriate for the purpose of the current study, depression module MMPI-2 administrations may not be as useful or relevant among other groups. However this caveat could inspire further research into MMPI-2 modules. For example, a drug and alcohol module might be relevant in community mental health settings, or a thought dysfunction module in an in-patient psychiatric treatment setting. Additionally, recruiting a clinical, college student sample (e.g., university counseling center consumers) for assessment research with clinical measures, like the MMPI-2, would be a logical follow-up to the current study.

A second limitation centers on the selection of scales for the depression module. Specifically, the criterion correlations raised some concerns about the PSY-5-Introversion/Low Positive Emotionality (INTR) scale selected for the depression module. Of the 18 observed differences in the criterion correlations, eight related to the INTR scale, while the remaining differences were more evenly dispersed through the four other MMPI-2 scales. The PSY-5 scales, including INTR, were constructed as measures of personality/psychopathology *traits* (Harkness, Finn, McNulty, & Shields, 2012), that is, stable enduring patterns. It could be argued the INTR scale does not reflect quite the same underlying construct as other MMPI-2 scales that measure a *state* of emotional dysfunction. Alternatively, the criterion measures selected to assess various dimensions of emotional dysfunction (e.g., depression, anxiety, anhedonia) could be responsible for the observed inconsistencies among correlations with INTR in this study.

Future research into MMPI-2 depression modules should explore removing the INTR scale from the module and/or including additional criterion measures (particularly measures of low positive emotions or trait introversion) of emotional dysfunction to clarify the current findings.

Another limitation of the current study was that the depression CM and CAM utilized different versions of the VRIN scale; specifically VRIN-CA was used in the CAM. This experimental abbreviated version, VRIN-CA includes fewer item pairs than its original counterpart. With this in mind, if VRIN-CA was used in the CM, this would reduce the total number of items administered to 300 (rather than the reported 320 using VRIN). It can be estimated that the CAM modality saves 24.5% (226.6 of 300 items) of the items administered in the CM modality. Unfortunately a similar adjustment cannot be made to estimate a comparison of the time savings in the CAM over the CM. For the sake of simplicity, the original VRIN scale should be used to allow for more direct comparisons of time and item savings between adaptive and non-adaptive module administrations. MMPI-2 administrators seeking to maximize item- and time-savings could utilize VRIN-CA.

While the current study had shortcomings, it was not without strengths. This study was the first to explore construct-specific MMPI-2 modules. Handel et al. (1999) reported a scale-specific module and Forbey et al. (2004) introduced a setting-specific "correctional" module. The findings of this study supported the notion of construct-specific MMPI-2 modules, thus future researchers could explore the construct validity of additional modules that target other constructs, such as anxiety or substance abuse. The design of this study allowed for comparison of adaptive and non-adaptive module – a gap in the MMPI-2 module literature. Additional research into MMPI-2 modules should continue to compare adaptive and non-adaptive modalities as adaptive testing with the MMPI-2 (particularly with modules) remains relatively

young, but promising.

The findings reported in this study provide additional empirical support for the countdown method (Butcher et al., 1985) as a means of adaptive testing with the MMPI-2 for significant time- and item-savings without a meaningful loss of validity. This study also provided additional evidence of the utility of FSES rules for the countdown method. As noted in the literature, FSES provided test administrators with richer MMPI-2 profiles when test takers report clinically significant distress or dysfunction (Handel et al., 1999). In addition this research demonstrated the continued usefulness of the MMPI-2-CA (Forbey & Ben-Porath, 2007) for research into adaptive testing with the MMPI-2. The findings reported here support the potential for adaptive personality testing used in clinical practice settings given the efficiency and accuracy offered by computerized adaptive measures. Previous research has supported the utility and validity of adaptive testing with the MMPI-2 (Forbey et al., 2009; Forbey et al., 2012) and the current findings likewise support the item- and time-savings without threats the validity of the MMPI-2.

The current study investigated the utility, in terms of time- and item-savings, and construct validity of an experimental MMPI-2 depression module for both adaptive and non-adaptive computerized administrations. Emotional dysfunction, particularly depression, is a growing concern among the college student population (Benton et al., 2003; Twenge et al., 2010), and this study offered a means of assessing these psychological difficulties. The results supported the hypotheses and suggest the CM and CAM administration techniques are time efficient and valid. The findings also pointed to the potential for further research into this area of personality assessment as well as the utility of modules and adaptive testing with the MMPI-2 in clinical practice.

References

Baumeister, H., & Härter, M. (2007). Prevalence of mental disorders based on general population surveys. *Social Psychiatry & Psychiatric Epidemiology, 42,* 537-546.

Bell, V., Halligan, P.W., & Ellis, H.D. (2006). The Cardiff Anomalous Perceptions Scale (CAPS): A new validated measure of anomalous perceptions. *Schizophrenia Bulletin, 32,* 366-377.

Benton, S.A., Robertson, J.M., Tseng, W-C., Newton. F.B., & Benton, S.L. (2003). Changes in counseling center client problems across 13 years. *Professional Psychology: Research and Practice, 34,* 66-72.

Ben-Porath, Y.S., Slutske, W.S., & Butcher, J.N. (1989). A real-data simulation of computerized adaptive administration of the MMPI. *Psychological Assessment, 1,* 18-22.

Bernstein, E.M., & Putnam, F.W. (1986). Development, reliability, and validity of a dissociation scale. *Journal of Nervous and Mental Disease, 174,* 727-735.

Black, M.S., Forbey, J.D., Ben-Porath, Y.S., Graham, J.R., McNulty, J.L., Anderson, S.V., & Burlew, A.K. (2004). Using the Minnesota multiphasic personality inventory- (MMPI-2) to detect psychological distress and dysfunction in a state correctional setting. *Criminal Justice and Behavior, 31,* 734-751.

Butcher, J.N., Atlis, M.M., & Fang, L. (2000). Effect of altered instructions on the MMPI-2 profiles of college students who are not motivated to distort their responses. *Journal of Personality Assessment, 75,* 492-501.

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

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

Butcher, J.N., & Hostetler, K. (1990). Abbreviating MMPI item administration: What can be learned from the MMPI for the MMPI-2? *Psychological Assessment, 2,* 12-21.

Butcher, J.N., Keller, L.,S., & Bacon, S.F. (1985). Current developments and future directions in computerized personality assessment. *Journal of Consulting and Clinical Psychology, 6,* 803-815.

Butcher, J.N., Williams, C.L., Graham, J.R., Archer, R.P., Tellegen, A., Ben-Porath, Y.S., & Kaemmer, B. (1992). *Minnesota Multiphasic Personality Inventory—Adolescent (MMPI-A): Manual for administration, scoring, and interpretation.* Minneapolis, MN: University of Minnesota Press.

Childs, R.A., Dahlstrom, W.G., & Panter, A. T. (2000). Item response theory in personality assessment: A demonstration using the MMPI-2 depression scale. *Assessment, 7,* 37-54.

Clark, L.A., & Watson, D. (1991). Tripartite model of anxiety and depression: Psychometric evidence and taxonomic implications. *Journal of Abnormal Psychology, 100,* 316-336.

Faul, A.C. & Hudson, W.W. (1997). The Index of Drug Involvement: A partial validation. *Social Work, 42,* 565-572.

Feigelson, M.E., & Dwight, S.A. (2000). Can asking questions by computer improve the candidness of responding? A meta-analytic perspective. *Consulting Psychology Journal: Practice and Research, 52,* 248-255.

Forbey, J.D., & Ben-Porath, Y.S. (2007). Computerized adaptive personality testing: A review and illustration with the MMPI-2 computerized adaptive version. *Psychological Assessment, 19,* 14-24.

Forbey, J.D., Ben-Porath, Y.S., & Arbisi, P.A. (2012). The MMPI-2 computerized adaptive version (MMPI-2-CA) in a veterans administration medical outpatient facility. *Psychological Assessment, 24*, 628-639.

Forbey, J.D., Ben-Porath, Y.S., Graham, J.R., & Black, M.S. (2004). The potential of computerized adaptive testing with the MMPI-2 in a correctional facility. *Ohio Corrections Research Compendium, 2*, 201-207.

Forbey, J.D., Ben-Porath, Y.S., & Gartland, D. (2009). Validation of the MMPI-2 computerized adaptive version (MMPI-2-CA) in a correctional intake facility. *Psychological Services, 6*, 279-292.

Frick, P.J., & Hare, R.D. (2001). *The antisocial process screening device.* Toronto: Multi-Health Systems.

Graham, J.R. (2011). *MMPI-2: Assessing personality and psychopathology.* New York: Oxford University Press.

Green, C.E.L., Freeman, D., Kulpers, E., Bebbington, P., Fowler, D., Dunn, G., & Garety, P.A. (2008). Measuring ideas of persecution and social reference: The Green et al. Paranoid Thought Scales (GPTS). *Psychological Medicine, 38*, 101-111.

Handel, R. W., Ben-Porath, Y.S., & Watt, M. (1999). Computerized adaptive assessment with the MMPI-2 in a clinical setting. *Psychological Assessment, 11*, 369-380.

Harknesss, A.R., Finn, J.A., McNulty, J.L., & Shields, S.M. (2012). The personality psychopathology-five (PSY-5): Recent constructive replication and assessment literature review. *Psychological Assessment, 24*, 432-443.

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

Hirschfield, R.M, et al. (2000). The development and validation of a screening instrument for bipolar spectrum disorder: The Mood Disorder Questionnaire. *American Journal of Psychiatry, 157*, 1873-1875.

Hudson, W.W. (1992). *The WALMYR assessment scales scoring manual.* Tempe, AZ: WALMYR Publishing Co.

Joiner, T.E., Brown, J.S., & Metalsky, G.I. (2003). A test of the tripartite model's prediction of anhedonia's specificity to depression: Patients with major depression versus patients with schizophrenia. *Psychiatry Research, 119*, 243-250.

Joiner, T.E., Walker, R.L, Pettit, J.W., Perez, M., & Cukrowicz, K.C. (2005). Evidence-based assessment of depression in adults. *Psychological Assessment, 17,* 267-277.

Kämpfe, N., & Mitte, K. (2009). A German validation of the UPPS Impulsive Behavior Scale. *European Journal of Psychological Assessment, 25*, 252-259.

Keltkangas-Järvinen, L. & Rimon, R. (1987). Rimon's Brief Depression Scale, A rapid method for screening depression. *Psychological Reports, 60*, 111-119.

Kessler, R.C., Bergland, P., Delmer, Olga, D., Jin, R., Koretz, D., Merikangas, K.R., Rush, J.A. Walters, E.E., & Wang, P.S. (2003). The epidemiology of major depressive disorder: Results from the National comorbidity Survey Replication (NCS-R). *Journal of the American Medical Association, 289*, 3095-3105.

Kessler, R.C., Birnbaum, H., Bromet, E., Hwang, L., Sampson, N., & Shahly, V. (2010). Age differences in major depression: Results from the National Comorbidity Survey Replication (NCS-R). *Psychological Medicine, 2010*, 225-237.

Kroenke, K., Spitzer, R.L., &Williams, J.B.W. (2002). The PHQ= 15: Validity of a new instrument for evaluating the severity of somatic symptoms. *Psychosomatic Medicine,*

64, 258-266.

Kruh, I.P., Frick, P.J., & Clements, C.B. (2005). Historical and personality correlates to the violence patterns of juveniles tried as adults. *Criminal Justice and Behavior, 32,* 69-96.

Leventhal, A.M. (2012). Relations between anhedonia and physical activity. *American Journal of Health Behavior, 36,* 860-872.

Lewinson, P.M., Rohde, P., Seeley, J.R., & Fisher, S.A. (1993). Age-cohort changes in the lifetime occurrence of depression and other mental disorders. *Journal of Abnormal Psychology, 102,* 110-120.

MacNeil, G. (1991). A short-form scale to measure alcohol abuse. *Research on Social Work Practice, 1,* 68-75.

Meinzer, M.C., Pettit, J.W., Leventhal, A.M., & Hill, R.M. (2012). Explaining the covariance between attention-deficit hyperactivity disorder symptoms and depressive symptoms: The role of hedonic responsivity. *Journal of Clinical Psychology, 68,* 1111-1121.

Muñoz, L.C., & Frick, P.J. (2007). The reliability, stability and predictive utility of the self-report version of the Antisocial Process Screening Device. *Scandinavian Journal of Psychology, 48,* 299-312.

Nakonezny, P.A., Carmody, T.J., Morris, D.W., Kurian, B.T., & Trivedi, M.H. (2010). Psychometric evaluation of the Snaith-Hamilton Pleasure Scale (SHAPS) in adult outpatients with major depressive disorder. *International Clinical Psyopharmacology, 25,* 328-333.

Radloff, L.S. (1977). The CES-D Scale: A self-report depression scale for research in the general population. *Applied Psychology Measurement, 1,* 385-401.

Reise, S.P., & Waller, N.G. (2009). Item response theory and clinical measurement. *Annual*

Review of Clinical Psychology, 5, 27-48.

Ross, S.R., Molto, J., Poy, R., Segarra, P., Pastor, M.C., & Montanes, S. (2007). Gray's model and psychopathy: BIS but not BAS differentiates primary from secondary psychopathology in noninstitutionalized young adults. *Personality and Individual Differences, 43*, 1644-1655.

Schwartz, G.E., Davidson, R.J., & Goleman, D.J. (1978). Patterning of cognitive and somatic processes in the self-regulation of anxiety: Effects of meditation versus exercise. *Psychosomatic Medicine, 40*, 321-328.

Snaith, R.P. (1993). Anhedonia: A neglected symptom of psychopathology. *Psychological Medicine, 23*, 957-966.

Snaith, R.P., Hamilton, M., Morley, S., Humayan, A., Hargreaves, D., & Trigwell, P. (1995). A scale for the assessment of hedonic tone the Snaith-Hamilton pleasure scale. *British Journal of Psychiatry, 167*, 99-103.

Twenge, J.M., Gentile, B., Dewall, C.N., Ma, D., Lacefield, K., & Schurtz, D.R. (2010). Birth cohort increases in psychopathology among young Americans, 1938-2007: A cross-temporal meta-analysis of the MMPI. *Clinical Psychology Review, 30*, 145-154.

Whiteside, S.P. & Lynam, D.R. (2001). The five factor model and impulsivity: using a structural model of personality to understand impulsivity. *Personality and Individual Differences, 30*, 669-689.

Whiteside, S.P., Lynam, D.R., Miller, J.D., & Reynolds, S.K. (2005). Validation of the UPPS Impulsive Behavior Scale: A four-factor model of impulsivity. *European Journal of Personality, 19*, 559-574.

Table 1

Selected MMPI-2 Scales for the Depression Module

Validity Scales		Substantive Scales	
Content Non-Responsiveness		Clinical Scale	
	VRIN		2
	TRIN	Content Scale	
Over-reporting			DEP
	F	Restructured Clinical (RC) Scales	
	F_b		RCd
	F(p)		RC2
	FBS	PSY-5 Scale	
Under-reporting			INTR
	L		
	K		
	S		

Note. VRIN = Variable Response Inconsistency, TRIN = True Response Inconsistency, F = Infrequency, F_b = Back Infrequency, F(p) = Infrequency Psychopathology, FBS = Fake Bad Scale, L = Lie, K = Correction, S = Superlative, 2 = Depression, DEP = Depression, RCd = Demoralization, RC2 = Low Positive Emotions, INTR = Introversion/Low Positive Emotionality.

Table 2
Internal consistencies for Criterion Measures in the Current Study

Criterion Measures in Set A	*n* (items)	*α*	Criterion Measures in Set B	*n* (items)	*α*
Emotional Dysfunction					
Clinical Anxiety Scale	25	.90	Centers for Epidemiologic Studies Depression Scale	20	.91
Rimon's Brief Depression Scale	7	.85	Cognitive-Somatic Anxiety Questionnaire	14	.91
Snaith-Hamilton Pleasure Scale	14	.93	Cognitive Subscale		.87
			Somatic Subscale		.80
			Tripartite Pleasure Inventory	12	.81
Thought Dysfunction					
Dissociative Experiences Scale	28	.96	Cardiff Anamolous Perceptions Scale	32	.96
Green Paranoid Thoughts Scales	32	.95			
Behavioral Dysfunction					
Antisocial Process Screening Devise – Self Report	20	.76	Index of Alcohol Involvement	25	.82
Index of Drug Involvement	25	.72	Mood Disorder Questionnaire	17	.84
			UPPS Impulsive Behavior Scales	45	
			Urgency		.79
			Premeditation		.87
			Perseverance		.62
			Sensation Seeking		.69
Somatic Dysfunction					
Patient Health Questionnaire	15	.84	MMPI-2 RF items	10	
Total in Set A	166		Total in Set B	175	

Table 3

Number of Items Administered and Administration Times per Administration Modality

	n (items)						Time (min)				
Administration Condition	*n*	Min	Max	M	SD	Average not administered (%)	Min	Max	M	SD	Average Savings (%)
Conventional$_1$	74	557	557	557	0		24	69	40.2	9.21	
Conventional$_2$	74	557	557	557	0		19	58	33.6	8.00	
Conventional	82	557	557	557	0		23	62	39.6	8.82	
Module	82	320	320	320	0	42.55	12	39	21.1	5.11	36.90
Conventional	80	557	557	557	0		23	60	38.1	8.50	
Adaptive Module	80	201	256	226.6	12.66	59.14	8	28	14.8	3.41	55.95

Table 4.

Test-Retest Correlations for Validity and Depression Scales of the MMPI-2 and q *Effect Sizes for Significant Differences*

	Test-retest Correlations *r*			Effect Size Cohen's *q*		
	CC – CC *n* = 74	CC – CM *n* = 82	CC – CAM *n* = 80	CC – CC / CC – CM	CC – CM / CC – CAM	CC – CC / CC – CAM
Validity						
VRIN	.51[e]	.29	.21			.35
TRIN	.41	.24	.57[d]		.40	
F	.87[a,e]	.76	.76	.34		.34
Fb	.88[a]	.70	.81	.51		
F(p)	.71	.55	.70			
L	.79[e]	.68	.62			.35
K	.80	.68	.69			
S	.90[e]	.83	.74			.52
Depression						
2	.81	.83	.76	.44		
DEP	.93[a]	.84	.90	.50		.50
RCd	.95[a,e]	.87	.87			
RC2	.85	.75	.84			
INTR	.88	.85	.84			

Note. CC = computerized conventional administration, CM = computerized module administration, CAM = computerized adaptive module administration, VRIN = Variable Response Inconsistency, TRIN = True Response Inconsistency, F = Infrequency, F_b = Back Infrequency, F(p) = Infrequency Psychopathology, FBS = Fake Bad Scale, L = Lie, K = Correction, S = Superlative, 2 = Depression, DEP = Depression, RCd = Demoralization, RC2 = Low Positive Emotions, INTR = Introversion/Low Positive Emotionality.
[a] denotes CC-CC greater than CC-CM, [b] denotes CC-CM greater than CC-CC, [c] denotes CC-CM greater than CC-CAM, [d] denotes CC-CAM greater than CC-CM, [e] denotes CC-CC greater than CC-CAM, [f] denotes CC-CAM greater than CC-CC.

Table 5

Criterion Correlations within Conventional MMPI-2 Administrations

		Control Condition									
		2		DEP		RCd		RC2		INTR	
Criterion Scale	*n*	CC_1	CC_2	CC_1	CC_2	CC_1	CC_2	CC_1	CC_2	CC_1	CC_2
Emotional Dysfunction											
Clinical Anxiety Scale	72	.58	.66	.69	.63	.71	.70	.45	.53	.46	.53
Centers for Disease Control – Depression Scale	70	.69	.81[b]	.81	.77	.79	.80	.49	.55	.52	.53
Cognitive-Somatic Anxiety Questionnaire - Cognitive	71	.60	.63	.64	.66	.74	.70	.51	.48	.40	.40
Rimon's Brief Depression Scale	72	.59	.60	.71	.70	.65	.70	.51	.52	.52	.51
Snaith Hamilton Pleasure Scale	71	.23	.24	.39	.37	.33	.34	.34	.41	.31	.39
Tripartite Pleasure Inventory	71	- .26	- .23	- .33	- .29	- .24	- .24	- .42	- .42	- .51	- .49
Thought Dysfunction											
Cardiff Anamolous Perceptions Scale	72	.20	.26	.26	.25	.32	.27	.01	.07	- .03	- .03
Dissociative Experiences Scale	71	.34	.32	.35	.33	.44	.38	.22	.22	.10	.11
Green Paranoid Thoughts Scale	72	.25	.23	.32	.32	.42	.43	.05	.08	.03	.04
Behavioral Dysfunction											
Antisocial Processes Screening Device – Self Report	72	.09	.16	.21	.20	.34	.31	.06	.00	.00	- .10
Index of Alcohol Involvement	72	.15	.17	.26	.23	.29	.28	.06	.06	.00	.00
Index of Drug Involvement	71	.21	.18	.18	.10	.22	.15	.05	.07	.03	-.02
Impulsivity: Urgency	72	.05	.10	- .06	-.05	- .10	- .08	.07	.11	.02	.22[b]
Impulsivity: Perseverance	71	- .31	- .32	- .27	- .29	- .23	- .28	- .33	- .39	- .41	- .50
Impulsivity: Premeditation	72	.34	.44	.48	.52	.52	.56	.37	.40	.20	.30
Impulsivity: Sensation Seeking	72	- .25	- .34	- .25	- .19	- .26	- .29	- .28	- .34	- .45	- .37
Mood Disorder Questionnaire	69	.02	.01	.16	.16	.26	.20	.04	.03	- .16	- .15
Somatic Dysfunction											
Cognitive-Somatic Anxiety Questionnaire - Somatic	71	.37	.41	.36	.32	.45[a]	.38	.25	.27	.20	.25
Patient Health Questionnaire	71	.44	.48	.47	.48	.55	.59	.31	.27	.24	.23

Note. CC = computerized conventional administration

[a] denotes the correlation at CC_1 was greater than at CC_2, [b] denotes the correlation at CC_2 was greater than at CC_1. Participant scores were excluded from these calculations if they responded to fewer than 90% of the items on a given criterion measure.

Table 6

Criterion Correlations within Depression Computerized Module (CM) MMPI-2 Administrations

		Depression Module Condition									
		2		DEP		RCd		RC2		INTR	
Criterion Scale	*n*	CC	CM	CC	CM	CC	CM	CC	CM	CC	CM
Emotional Dysfunction											
Clinical Anxiety Scale	80	.58	.65[b]	.56	.53	.59	.56	.51	.39	.53[a]	.38
Centers for Disease Control – Depression Scale	80	.59	.59	.66	.65	.64	.65	.49	.47	.42	.38
Cognitive-Somatic Anxiety Questionnaire - Cognitive	76	.50	.52	.51	.44	.54	.49	.36	.36	.38	.32
Rimon's Brief Depression Scale	80	.46	.47	.51	.50	.50	.49	.41	.36	.33[a]	.17
Snaith Hamilton Pleasure Scale	79	.03	.18[b]	.08	.19	.04	.15	.20	.27	.26	.35
Tripartite Pleasure Inventory	75	- .08	- .09	- .16	- .25	- .09	- .21[b]	- .23	- .23	- .38	- .36
Thought Dysfunction											
Cardiff Anamolous Perceptions Scale	79	- .10[a]	.08	.04	.02	.10	.05	- .04	.06	- .06	- .04
Dissociative Experiences Scale	80	.10	.12	.07	.24[b]	.14	.21	.09	.06	.06	.10
Green Paranoid Thoughts Scale	80	.34	.34	.45	.42	.44	.42	.39	.30	.30	.24
Behavioral Dysfunction											
Antisocial Processes Screening Device – Self Report	79	- .06	- .17	.15	.15	.16	.18	.08	.07	.06	.05
Index of Alcohol Involvement	78	.12	.10	.24	.21	.29	.23	.16	.04	.15[a]	.03
Index of Drug Involvement	77	.07	.10	.11	.06	.18	.09	.08	.04	.01	- .07
Impulsivity: Urgency	80	.22	.22	.30	.22	.28	.25	.17	.29	.19	.23
Impulsivity: Perseverance	78	- .34	- .30	- .19	- .17	- .19	- .17	- .23	- .17	- .41	- .30
Impulsivity: Premeditation	79	.04	.00	.16	.12	.20	.16	.02	- .02	.03	.04
Impulsivity: Sensation Seeking	80	.00	.01	.12	.04	.13	.08	.01	-.01	- .07	- .07
Mood Disorder Questionnaire	80	- .05	- .16	.07	.03	.19	.10	- .17	- .16	- .23	- .22
Somatic Dysfunction											
Cognitive-Somatic Anxiety Questionnaire - Somatic	76	.30	.38	.29	.32	.36	.37	.21	.21	.24	.22
Patient Health Questionnaire	80	.38	.47	.20	.24	.29	.30	.23	.32	.21	.11

Note. CC = computerized conventional administration, CM = computerized module administration

[a] denotes the correlation at CC was greater than at CM, [b] denotes the correlation at CM was greater than at CC. Participant scores were excluded from these calculations if they responded to fewer than 90% of the items on a given criterion measure.

Table 7

Criterion Correlations within Depression Computerized Adaptive Module (CAM) MMPI-2 Administrations

		Depression Module – Adaptive Condition									
		2		DEP		RCd		RC2		INTR	
Criterion Scale	*n*	CC	CAM	CC	CAM	CC	CAM	CC	CAM	CC	CAM
Emotional Dysfunction											
Clinical Anxiety Scale	78	.39	.41	.49	.52	.48	.48	.31	.30	.14	.23
Centers for Disease Control – Depression Scale	78	.44	.55	.69	.64	.66	.66	.43	.49	.32	.46[b]
Cognitive-Somatic Anxiety Questionnaire – Cognitive	77	.29	.36	.49	.48	.54	.50	.32	.42	.09	.26[b]
Rimon's Brief Depression Scale	77	.55	.52	.73	.71	.66	.65	.53[a]	.41	.31	.39
Snaith Hamilton Pleasure Scale	75	- .02	.01	.27	.30	.16	.22	.22	.13	.34	.30
Tripartite Pleasure Inventory	76	- .23	- .18	- .24	- .24	- .26	- .23	- .31	- .25	- .39[a]	- .27
Thought Dysfunction											
Cardiff Anamolous Perceptions Scale	77	- .05	.05	.28	.22	.32	.28	.04	.03	- .11	- .04
Dissociative Experiences Scale	76	.11	.31[b]	.31	.35	.29	.38	.16	.18	.07	.23[b]
Green Paranoid Thoughts Scale	78	.22	.28	.36	.39	.36	.38	.16	.14	.07	.07
Behavioral Dysfunction											
Antisocial Processes Screening Device – Self Report	78	.12	.26	.36	.44	.34	.45[b]	.30	.27	.15	.19
Index of Alcohol Involvement	78	-. 02	- .02	.17	.15	.20	.26	- .02	.00	- .14	- .06
Index of Drug Involvement	78	- .07	- .12	- .03	.01	- .08	.02	- .07	- .10	- .11	- .19
Impulsivity: Urgency	78	.19	.20	.15	.14	.20	.09	.19	.15	.06	.08
Impulsivity: Perseverance	76	- .15	- .12	- .08	- .04	- .01	- .01	- .12	- .17	- .31	- .27
Impulsivity: Premeditation	78	.27	.27	.41	.42	.47	.52	.17	.12	.09	.10
Impulsivity: Sensation Seeking	78	- .14	.00	.02	- .02	- .01	- .06	- .19	- .26	- .30	- .28
Mood Disorder Questionnaire	76	- .15	- .11	.15	.14	.23	.17	- .10	- .06	- .27	- .24
Somatic Dysfunction											
Cognitive-Somatic Anxiety Questionnaire - Somatic	77	.09	.09	.21	.20	.25	.21	.07	.17	- .09	.03
Patient Health Questionnaire	77	.54	.53	.50	.49	.49	.48	.31	.31	.16	.25

Note. CC = computerized conventional administration, CAM = computerized adaptive module administration

[a] denotes the correlation at CC was greater than at CAM, [b] denotes the correlation at CAM was greater than at CC. Participant scores were excluded from these calculations if they responded to fewer than 90% of the items on a given criterion measure.