A Dissertation
Submitted to the Faculty
Of
Xavier University
In Partial Fulfillment of the
Requirements for the Degree of
Doctor of Psychology
By
Nicole R. Bosse, M.A.
August 22, 2013

Approved:

[Signature]
Karl Stuckenbrug, Ph.D., ABPP
Chair, Department of Psychology

[Signature]
Kathleen J. Hart, Ph.D., ABPP
Dissertation Chair
Detecting Over- and Under-reporting of Symptoms on the MAYSI-2:

Development of a Validity Scale
Dissertation Committee

Chair
Kathleen Hart, Ph.D., ABPP
Professor of Psychology

Member
W. Michael Nelson III, Ph.D., ABPP
Professor of Psychology

Member
Nicholas Salsman, Ph.D., ABPP
Associate Professor of Psychology

Member
Dalia Diab, Ph.D.
Assistant Professor
Acknowledgements

I would first like to thank my dissertation chair for all of her support throughout the dissertation process, suggestions of articles to include, help with statistical analyses, her time spent reading and re-reading sections and providing comments and feedback, and her willingness to go above and beyond what is ordinarily expected of a dissertation chair. Without Dr. Hart’s guidance throughout the process I would have been lost. I also wanted to especially thank my dissertation committee members. Dr. Diab came onto the committee mid-way through the project and she provided an enormous amount of information on a topic that I knew little about, item development and factor analysis. Dr. Salsman and Dr. Nelson were on my dissertation committee from the beginning and they were a great source of support and provided excellent feedback and suggestions at the proposal and also at the defense meeting. Thank you again to all the faculty members involved, I couldn’t have done this without you! I also would like to thank my family, friends, and my boyfriend Jeremy for sticking with me through this process and being patient with all of the long hours I needed to put in in order to complete the dissertation!
Table of Contents

<table>
<thead>
<tr>
<th>Chapter</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>Acknowledgements</td>
<td>i</td>
</tr>
<tr>
<td>Table of Contents</td>
<td>ii</td>
</tr>
<tr>
<td>List of Tables</td>
<td>iii</td>
</tr>
<tr>
<td>List of Figures</td>
<td>iv</td>
</tr>
<tr>
<td>List of Appendices</td>
<td>v</td>
</tr>
<tr>
<td>Chapter</td>
<td></td>
</tr>
<tr>
<td>I. Review of the Literature</td>
<td>1</td>
</tr>
<tr>
<td>II. Rationale and Hypotheses</td>
<td>30</td>
</tr>
<tr>
<td>III. Method</td>
<td>32</td>
</tr>
<tr>
<td>IV. Proposed Analyses</td>
<td>37</td>
</tr>
<tr>
<td>References</td>
<td>39</td>
</tr>
<tr>
<td>Appendices</td>
<td>48</td>
</tr>
<tr>
<td>V. Dissertation</td>
<td>56</td>
</tr>
<tr>
<td>References</td>
<td>74</td>
</tr>
<tr>
<td>Tables</td>
<td>78</td>
</tr>
<tr>
<td>Figures</td>
<td>85</td>
</tr>
<tr>
<td>Appendices</td>
<td>86</td>
</tr>
<tr>
<td>Summary</td>
<td>95</td>
</tr>
</tbody>
</table>
List of Tables

Proposal

<table>
<thead>
<tr>
<th>Table</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. Description of SIRS Scales.</td>
<td>21</td>
</tr>
</tbody>
</table>

Chapter V

<table>
<thead>
<tr>
<th>Table</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. Descriptive Demographics for the Overall Sample and for Each</td>
<td>78</td>
</tr>
<tr>
<td>Randomization Group</td>
<td></td>
</tr>
<tr>
<td>2. Validity Items Developed for the MAYSi-2</td>
<td>79</td>
</tr>
<tr>
<td>3. Extracted Factors and Eigenvalues from the 20 Items Added to the</td>
<td>80</td>
</tr>
<tr>
<td>MAYSi-2</td>
<td></td>
</tr>
<tr>
<td>4. Added MAYSi-2 Item Loadings for Three Factors Identified through</td>
<td>81</td>
</tr>
<tr>
<td>Exploratory Factor Analysis</td>
<td></td>
</tr>
<tr>
<td>5. Means, Standard Deviations, and One Way ANOVA Results for the</td>
<td>83</td>
</tr>
<tr>
<td>Under-Reporting and Over-Reporting Scales Across the Three</td>
<td></td>
</tr>
<tr>
<td>Instruction Conditions</td>
<td></td>
</tr>
<tr>
<td>6. Means, Standard Deviations, and F-Values for Randomization Groups</td>
<td>84</td>
</tr>
<tr>
<td>Each MAYSi-2 Scale in a Multivariate Analysis of Variance</td>
<td></td>
</tr>
</tbody>
</table>
Lists of Figures

Chapter V

1. Scree Plot..........................................................85
# List of Appendices

<table>
<thead>
<tr>
<th>Appendix</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>A - Demographic Questionnaire</td>
<td>48</td>
</tr>
<tr>
<td>B - Validity Items Developed</td>
<td>49</td>
</tr>
<tr>
<td>C - Script for Under-Reporting</td>
<td>51</td>
</tr>
<tr>
<td>D - Script for Over-Reporting</td>
<td>52</td>
</tr>
<tr>
<td>E - Script for Control Group</td>
<td>53</td>
</tr>
<tr>
<td>F - Manipulation Check</td>
<td>54</td>
</tr>
<tr>
<td>G - Debriefing Form</td>
<td>55</td>
</tr>
</tbody>
</table>

**Chapter V**

<table>
<thead>
<tr>
<th>Appendix</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>A- Demographic Questionnaire</td>
<td>86</td>
</tr>
<tr>
<td>B- MMAYSI-2 Questionnaire</td>
<td>87</td>
</tr>
<tr>
<td>C- IRB Approval</td>
<td>89</td>
</tr>
<tr>
<td>D- Script for Under-Reporting</td>
<td>91</td>
</tr>
<tr>
<td>E- Script for Over-Reporting</td>
<td>92</td>
</tr>
<tr>
<td>F- Script for Control Group</td>
<td>93</td>
</tr>
<tr>
<td>G- Manipulation Check</td>
<td>94</td>
</tr>
</tbody>
</table>
Chapter I

Review of the Literature

According to the Office of Juvenile Justice and Delinquency Prevention (OJJDP, 2009), 2.11 million juveniles, individuals younger than age 18, were arrested in 2008. This rate reflects a decline in juvenile arrests over the last two decades. In 2008, there were 3% fewer arrests than in 2007 and juvenile violent crime arrests fell 2%. However, despite these decreases, juveniles accounted for 16% of all violent crime arrests and 26% of all property crime arrests in 2008. The juvenile murder arrest rate in 2008 was 3.8 arrests per 100,000 juveniles, which was 17% higher than the low rate of 3.3 in 2004. Further, juveniles under the age of 15 accounted for more than one-fourth of all juvenile arrests for Violent Crime Index offenses and Property Crime Index offenses (OJJDP, 2009). This shows that while juvenile crime rates appear to be declining, there is still a large portion of crime that is attributed to juveniles.

The goal of rehabilitation and treatment in the juvenile justice system has been overshadowed by the large number of delinquent youth in the system and the severity of the crimes they commit. According to national statistics, one out of every five youths who go before the court for delinquent offenses is placed in a juvenile facility. A recent national survey reported that 26% of juvenile offenders are held in detention facilities, while 32% are held in correctional facilities (OJJDP, 2010). Issues that have been offered to account for the high numbers and severity of delinquent crime include ineffective parenting, poverty, unavailability of social services, and the number of youth who are mentally ill (OJJDP, 2010).
Beginning in the 1990s, research began to reveal a high prevalence of mental disorders among juvenile offenders when compared to youths in the general population (Abram, Teplin, McClelland, & Dulcan, 2003; Grisso & Barnum, 2001). Accurately identifying psychiatric disorders when juveniles enter the juvenile justice system is important. State and county juvenile justice facilities have both a moral and legal responsibility to respond to the mental health needs of youths in their custody, both to maintain their safety and to protect the public. According to Grisso and Barnum (2001), identifying these needs early could improve the efficacy of interventions to reduce future crime, to prevent development of serious mental disorders, and provide for the safety of juveniles in the facility.

**Mental Illness in the Juvenile Justice System**

Epidemiological studies of psychiatric disorders among juveniles within the juvenile justice system indicate that about 40% to 60% of this population meets criteria for a psychiatric diagnosis, even when excluding substance abuse and conduct disorder (Teplin, Abram, McClelland, Dulcan, & Mericle, 2002). Girls in the juvenile justice system are significantly more likely than boys to be diagnosed with anxiety disorders and/or major depression, and are more likely to have three or more comorbid disorders (Abram, Teplin, McClelland, & Dulcan, 2003). Given that the rate of female juvenile offenders is on the rise, this may eventually result in a higher overall prevalence and severity of psychiatric diagnoses in the juvenile justice population.

A meta-analysis conducted by Edens and Otto (1997) included studies that assessed specific disorders for children and adolescents in juvenile justice systems,
DETECTING OVER- AND UNDER-REPORTING

which allowed them to provide tentative estimates of the prevalence of specific disorders in youths in the juvenile justice system. They estimated that between 50-90% of juveniles had conduct disorder; between 6-41% had an anxiety disorder; 20-50% had a substance abuse/dependence disorder (although rates of up to 69% have been reported); 32-78% had a mood disorder; and between 1-6% had a psychotic disorder (Edens & Otto, 1997).

Ulzen and Hamilton (1998) interviewed 49 incarcerated and 49 community adolescents using the Diagnostic Interview for Children and Adolescents-Revised (DICA-R). They found that 22.4% of the incarcerated youth met criteria for at least one psychiatric disorder and 63.3% met criteria for two or more disorders; only 14.3% of the incarcerated youths did not meet enough criteria for any diagnoses. This is compared to the community sample where 18.4% met criteria for at least one psychiatric disorder, 12.2% met criteria for two or more disorders, and 69.4% met no criteria. Common diagnoses for males included Oppositional Defiant Disorder, Conduct Disorder, Attention Deficit Hyperactivity Disorder, mania, and Post-Traumatic Stress Disorder. Among the females, common diagnoses included Post-Traumatic Stress Disorder and Oppositional Defiant Disorder (Ulzen & Hamilton, 1998).

In a recent study conducted by the National Center for Mental Health and Juvenile Justice (NCMHJJ, 2006), 1,400 youth from 29 different juvenile justice programs and facilities in Louisiana, Texas, and Washington were assessed to determine mental illness prevalence. Mental health disorders were identified using the Diagnostic Interview Schedule for Children-Voice Version IV (Voice DISC-IV). The study revealed
that 70.4% of juveniles in the sample were found to meet criteria for at least one mental health disorder. Disruptive disorders (46.5%) such as Conduct Disorder were the most common, followed by Substance Use disorders (46.2%), Anxiety Disorders (34.4%), and Mood Disorders (18.3%).

Examining comorbidity, the authors found that 79% of juveniles in this sample met criteria for two or more diagnoses, and 60% of the sample met criteria for three or more disorders. When examining the differences between genders, the study found that more than 80% of girls in the sample met criteria for at least one mental health disorder compared to 67% of boys. These juveniles also met criteria for severe mental health disorders (either meeting criteria for certain severe disorders or had been hospitalized for a mental disorder) at a fairly high rate: 27% of the overall sample had a mental disorder severe enough to require significant and immediate treatment (NCMHJJ, 2006).

**Risk of suicide.** Keeping juveniles safe while they are incarcerated is an area that has received high levels of attention. Once juveniles enter custody, their safety and well-being fall into the hands of the facility. The National Commission on Correctional Health Care (NCCHC) recommends that all juvenile facilities develop and implement a comprehensive suicide prevention program, regardless of the type or size of facility (NCCHC, 2007). The NCCHC recommends staff training in suicide prevention; ongoing identification of risks in juveniles; communication among all employees when juveniles display certain behavioral signs; safe physical environments where juveniles are housed; monitoring of at-risk juveniles based on their needs and not solely on the facility's resources; and early identification and development of intervention plans for at risk
juveniles. Finally, this report recommended that every completed or serious attempted suicide undergo a morbidity/mortality review process.

Suicidal behavior in the juvenile detention population has been studied by numerous researchers. Rohde, Seeley, and Mace (1997) asked 555 adolescents in a county juvenile detention center to answer questions regarding current suicide ideation, lifetime thoughts of death and suicide, lifetime suicide attempts and exposure to suicide events. They found that over one-third of the adolescents had thought about committing suicide at some point in their life and that 14.2% reported current suicidal ideation (i.e., within the past week). In addition, almost one-fifth reported having made at least one past suicide attempt (9.1% reported one past attempt and 10.3% reported two or more attempts). For boys, past attempts were associated with suicidal ideation and ineffective coping, while for girls past attempts were associated with major life events and impulsivity. In addition, not living with at least one biological parent prior to detention was associated with past attempts for both boys and girls. In terms of current suicidal behavior, decreased social connections was a strong predictor for boys, whereas impulsivity and instability were strong predictors for the girls. Lastly, current depression was most significantly associated with current suicidal ideation (Rohde et al., 1997).

The issue of safety is especially critical: each year, a small number of incarcerated youths successfully commit suicide. According to OJJDP (2009), 110 incarcerated juveniles committed suicide between 1995 and 1999. In addition, a national survey administered to 1,800 confined youth found that in the previous 12 months almost 22% had seriously considered suicide, 20% made a plan, 16% made at least one
DETECTING OVER- AND UNDER-REPORTING

attempt, and 8% injured themselves during a suicide attempt (OJJDP, 2009). A number of factors have been shown to be related to suicide risk such as past behavior, substance abuse, exposure to family conflict and aggression, presence of psychiatric disorders, accessibility of weapons, and the presence of situational triggers that provoke violent acts (Archer, 2005).

Statistics for non-incarcerated juveniles’ suicidal behavior is also surprisingly high. According to the American Academy of Child and Adolescent Psychiatry (AACAP, 2008), suicide is the third leading cause of death for 15-24 year olds, and the sixth leading cause of death for 5-14 year olds. One out of 10 teens contemplates suicide, and almost half a million youths make a suicide attempt each year. According to the Center for Disease Control (CDC; 2004), 6% of high school males and 11% of high school females reported at least one suicide attempt. They also found that nationwide, 16.9% of youth reported having thought seriously about attempting suicide and 16.5% reported making a detailed plan.

Penn, Espositio, Schaeffer, Fritz, and Spirito (2003) examined the lifetime history and current suicidal behavior of 289 adolescents who were incarcerated in a state correctional facility using a systematic random sample chart review. This is based on chart review, so they did not identify “undetected” attempts. The rate of juveniles’ prior suicide attempts was 12%, which is lower than other studies reported, but higher than rates typically found in community samples (e.g., 9%; New York University Child Study Center Letter, 2001). Of the prior suicide attempts, two thirds of the juveniles attempted suicide by violent means (e.g., cutting, hanging), whereas adolescents from psychiatric
samples typically attempt suicide by overdose. The researchers identified 78 adolescents (27% of the 289 random sample) who were referred for psychiatric evaluation upon admission and/or during incarceration for a variety of problems such as suicidal behavior, self-mutilating behavior, sleep problems, psychotropic medication maintenance, disruptive behaviors, and/or by youth’s request. They found that this subsample of adolescents were more likely than a national school-based sample to have seriously considered attempting suicide (28.2% versus 19.3%); made a plan for attempting suicide (21.8% versus 14.5%); made at least one suicide attempt (32% versus 8.3%); and made a suicide attempt requiring medical attention (12.8% versus 2.6%) in the prior year. The researchers reported that of the 78 clinically referred adolescents, 30% reported suicidal ideation/behavior and 30% reported self-mutilating behavior while incarcerated. Adolescents who were suicidal while incarcerated reported higher rates of prior suicide attempts and self-mutilating behavior, prior to and during incarceration. The researchers also reported that the referred suicidal adolescents displayed more affective disturbance, were more depressed, angry, and anxious when compared to the referred nonsuicidal adolescents.

Together these studies indicate that incarcerated adolescents display high rates of suicidal behavior/ideation, and that they typically choose more lethal methods, pointing to the need to screen for suicidal risk in juvenile facilities.

**Screening for Mental Illness in the Juvenile Justice System**

Given the high prevalence of mental illness that has been documented in the juvenile justice system and the risk for suicidal behavior while incarcerated, there is a
DETECTING OVER- AND UNDER-REPORTING

need to screen youths who enter this system for symptoms of a psychiatric disorder in order to assure safety and timely intervention. Screening also addresses the requirement that most states have to assess every youth who enters the justice system for suicide risk, mental and emotional health, and substance abuse (Lexcen, Vincent, & Grisso, 2004; Stewart & Trupin, 2003).

Unfortunately, intake workers at detention facilities rarely have access to mental health records, and they usually do not have sufficient training to adequately assess for the presence of clinically significant symptoms (Grisso, Barnum, Fletcher, Cauffman, & Peuschold, 2001). Two comprehensive instruments frequently are used by trained professionals to assess adolescent psychopathology, the Minnesota Multiphasic Personality Inventory-Adolescent (MMPI-A; Butcher, Dahlstrom, Graham, Tellegen, & Kaemmer, 1989) and the Millon Adolescent Clinical Inventory (MACI; Millon, 1993), but these measures are lengthy and often too complex to use during standard initial screenings. There are also brief symptom-focused questionnaires such as the Beck Depression and Anxiety Inventories (Beck & Steer, 1987; Beck & Steer, 1990), but these focus on discrete areas of psychopathology and do not provide a broad-spectrum of information about current psychiatric symptoms. Further, none of these measures can be used independently by paraprofessionals.

Anderson, Townsend, Everly, and Lating (1995) outlined the primary strategies that have been used in assessing suicide risk: identifying sociodemographic factors, determining the presence of relevant clinical signs and symptoms, and using standardized psychological assessment measures. In addition, Rittner, Smyth and
DETECTING OVER- AND UNDER-REPORTING

Wodarski (1995) suggested using assessment instruments in addition to interview techniques in order to create the most effective strategy when working with suicidal adolescents. These researchers highlighted several variables that should be examined when using these two methods: depression, gender, suicidal ideation, psychiatric comorbidity, social dysfunction, prior attempts, self-esteem and cognitive factors, hopelessness, alcohol and other substance use and sexual orientation. As previously mentioned, many of these pieces of information are not available, as a youth enters a juvenile detention facility.

MAYSI-2

To meet the perceived need for an appropriate screening tool for mental illness and suicide risk for juveniles entering residential juvenile justice settings, Grisso and Barnum developed the Massachusetts Youth Screening Instrument (MAYSI) in 1994 (Grisso & Barnum, 2001). The authors hoped to create a brief, easy-to-read, self-report measure that can be administered to youths entering detention or other correctional facilities and that could be used and interpreted by personnel who had no specific clinical training (Grisso & Barnum, 2001). The purpose of the MAYSI was to identify significant levels of psychiatric symptoms, and most importantly, to identify youths in crisis who may require immediate and more detailed evaluation and/or intervention. A pool of 52 items was developed based on an extensive review of the literature on child psychopathology and juvenile correction facilities. These items were administered in pilot studies followed by administration to over 1,300 youths ages 12-17 in intake probation departments, pretrial detention centers, and youth commission assessment
DETECTING OVER- AND UNDER-REPORTING

centers in Massachusetts. The majority of the children were 15-17 years of age. The sample included boys (67%) and girls (33%) who had been charged with a broad range of offenses. Factor analysis of the items resulted in nine scales that corresponded to nine types of mental, emotional, and behavioral problems (Grisso & Barnum, 2001).

In 2003, Grisso and Barnum refined the MAYSI through additional factor analysis and other psychometric methods, and published the second edition of the instrument, the MAYSI-2. This version was normed on 1,279 youths, primarily from eastern and central Massachusetts. The new version consists of seven scales for boys and six scales for girls: Alcohol/Drug use, Angry-Irritable, Depressed-Anxious, Somatic Complaints, Suicide Ideation, Thought Disturbance (boys only), and Traumatic Experiences (Grisso & Barnum, 2003). The MAYSI-2 is a paper and pencil self-report instrument that met the other goals Grisso and Barnum established: it is brief (taking about 8-10 minutes to complete), easy to score, and can be interpreted without clinical specialization. It contains 52 items in a “yes-no” answer format; the respondent answers each item based on his/her experiences over the last few months. The score on an individual scale is the sum of all items answered “yes.” Six of the seven scales contain “Caution” or “Warning” categories to evaluate the youth’s scores. The ‘caution’ cut-off scores capture juveniles who would most likely score in the “clinically significant” range on other, more extensive measures of mental and emotional disturbances, such as the Child Behavioral Checklist-Youth Self-Report (CBCL-YSR) or MACI. According to Stewart et al. (2003), scores between the 85th and 94th percentile receive the ‘caution’ label, and scores at or above the 95th percentile receive the ‘warning’ label.
DETECTING OVER- AND UNDER-REPORTING

Grisso and Barnum (2003) reported that all internal consistency and reliability scales are at least comparable to or better than the leading comprehensive psychological tests that measure adolescents' mental and emotional problems. The MAYSI-2 was found to accurately identify 75% of youths who reported clinically significant symptoms compared with the general population on the MACI and the CBCL-YSR (Grisso et al., 2001). Archer, Stredny, Mason, and Arnau (2004) examined the psychometric properties of the MAYSI-2 using scores generated by 704 adolescents in detention settings in Virginia and found that the factor structure of the MAYSI-2 (construct validity) was consistent with what Grisso and Barnum (2003) reported. The internal consistency coefficients for this sample ranged from .61 to .90, and the test-retest reliability coefficients ranged from a low of .60 on the Suicide Ideation scale to a high of .82 on the Thought Disturbance scale; the average test-retest reliability coefficient was .73. In terms of concurrent validity, the MAYSI-2 correlated substantially (.45 or above) with the corresponding scales on the MACI and CBCL-YSR (Archer et al., 2004). Many studies have examined the reliability and validity of the MAYSI-2. Meyers (2006) examined MAYSI-2 scores gathered over a 10 year period at a detention facility located in an urban county in Ohio in order to develop local norms and establish cut-off scores for this facility. Given that the original MAYSI-2 cut-off scores were established from norms collected primarily from juveniles in the eastern and central parts of Massachusetts, Grisso & Barnum (2003) recommended establishing local norms.

Dunphy (2002) used test scores generated by juvenile offenders who completed the MAYSI-2 on intake to a juvenile detention facility in an urban county in the Midwest
to examine the relationship between the MAYS1-2 Suicidal Ideation (SI) scale and occurrence of suicidal behavior while in the facility. The results indicated a significant but modest relationship ($r = .20$) between the MAYS1-2 SI scale and a subsequent suicide attempt. Dunphy also examined how well the other MAYS1-2 scores predicted suicide attempts, and found that the Angry/Irritable scale was the only scale to correctly identify juveniles who engaged in suicidal behavior and those who did or did not make a suicide attempt after being admitted to a juvenile facility; 70% of attempters and nonattempters were correctly classified by the Angry/Irritable scale. Another study conducted by Butler, Loney, and Kistner (2007) investigated the relationship between institutional maladjustment and scores on the MAYS1-2 obtained on admission. Daily Incident Reports were searched for evidence of maladjustment in a sample of 104 male juvenile offenders during the first 90 days of confinement. They found that the Angry-Irritable subscale uniquely predicted severe rule violations and intensive supervision placement (ISP). However, the researchers reported a high level of false negatives (47%), which suggested that almost half of the youth who did not show an elevation on the Angry-Irritable subscale later engaged in severe rule violation and aggression, limiting the predictive and clinical utility of the scale.

Chavez-Frigon (2003) examined how well the MAYS1-2, administered at intake, predicted behavioral infractions in 1394 youths (age range 9 to 19 yrs.) admitted to a county juvenile detention facility. Behavioral infractions consist of conflict with authority such as defiance and running away; covert actions such as lying and stealing; and overt actions such as aggressive or violent behavior. Juveniles completed the
MAYS1-2 upon admission to the facility and behavioral infractions were tracked while the youths were in the facility. Chavez-Frigon found that the Suicide Ideation scale scores were the best predictors of behavioral infractions for girls, and that the Thought Disturbance and Suicide Ideation scales are the best predictors for boys.

Meyers (2006) examined the psychometric properties of juvenile offenders’ MAYS1-2 scores at an urban juvenile detention facility that included 7847 incarcerated youth (age range 8 to 19 yrs): 5696 male (73%) and 2151 female (27%). Meyers found that inter-item correlations ranged from .52 to .86 and split-half reliability ranged from .60 to .89. In this sample, girls had significantly higher scale scores when compared to boys. Age differences were also observed. Juveniles aged 14 to 19 (n = 6719) scored significantly higher on the Alcohol/Drug Use (AD) scale, while juveniles aged 8 to 13 (n = 1128) had significantly higher scores on the remaining scales. Two scales did not differ significantly by age, the Traumatic Experiences-Girls (TEG) and Traumatic Experiences-Boys (TEB) scales. Caucasians scored significantly higher on all seven scales than African Americans. The “Warning” and “Caution” cut-off scores for this sample were found to be similar for all scales when compared to the cut-offs established for the Grisso and Barnum (2003) standardization group. Meyers noted limitations similar to those mentioned by Grisso and Barnum: because there are no indicators of validity, it cannot be determined whether the scores are actually a reflection of the juveniles’ symptoms or a product of response bias. Meyers suggests that for future research, endeavors should be focused on developing a validity index to build into the measure.
Archer, Simonds-Bisbee, Spiegel, Handel, and Elkins (2010) conducted a more recent examination of the concurrent and discriminant validity of the MAYSI-2. The sample consisted of 1,192 adolescents (1,082 boys and 110 girls between the ages of 13 to 18 years) who were admitted to Virginia juvenile correction facilities. This study used a variety of criterion variables, including offense history; social history; psychological history; drug and/or alcohol use; educational information; medical history; behavioral ratings during admission; staffing team opinions; mandatory and recommended treatment; and additional service needs (i.e., behavior management, problem solving skills) against MAYSI-2 scores were compared. Results were compared to external data such as the Substance Abuse Subtle Screening Inventory-Adolescent-2, Face Valid Alcohol Scale and Face Valid Other Drugs scale, and DSM-IV diagnoses based on a clinical interview. They found that concurrent validity varied by scale, with the strongest evidence for concurrent validity for both boys and girls from the Alcohol/Drug Use and Suicide Ideation scales. For the sample of girls, good concurrent validity was established for the Depression-Anxious scale and Angry-Irritable scale; the Traumatic Experiences and Somatic Complaints scales showed relatively weaker concurrent validity. For the sample of boys, only half of the conceptually-related extra-test variables (i.e., number of prior psychiatric placements, suicidal ideation, parasuicidal behaviors, suicide attempts, pattern of self-mutilation, suicidal urges, self-harm behavior, suicide alert level, meeting DSM-IV criteria for depressive disorder, past/current antidepressant medication, and need for mental health services) correlated with the Depressed-Anxious scale, which shows adequate but limited concurrent validity. The Angry-Irritable, Thought
Disturbance, Traumatic Experiences, and Somatic Complaints scales showed few meaningful correlations with extra-test variables for the sample of boys. Evidence of discriminant validity for the MAYS1-2 included information about constructs that were unrelated to the MAYS1-2 scales. For example, the Alcohol/Drug Use scale did not correlate with more serious drug-related problems such as drug trafficking and sex for drugs, but it did correlate with related extratest variables such as drug use and history of drug use. The Angry-Irritable scale scores correlated strongly with an individual's tendency to display trait-based anger, and it was not strongly related to state-dependent aggressive behavior such as daily anger control ratings by staff, history of assault against others, diagnosis of conduct or bipolar disorder in boy offenders, and oppositional defiant disorder for boy or girl offenders. The Depressed-Anxious scale scores were related to a history of suicidal ideation, self-harm behaviors, and diagnosis of a depressive disorder, and were unrelated to appetite disturbances, attention or concentration difficulties, and family history of mental health problems for boys and girls. The Suicide Ideation scale was related to self-harm behaviors and suicide alert level, and unrelated to outpatient treatment in boys and girls. Archer et al., (2010) concluded that, with the exception of Angry-Irritable scale for the girl sample, all scales showed adequate discriminant validity by demonstrating low correlations with variables judged to be unrelated to the scale construct.

**Malingering and Response Bias**

In order to diagnose and evaluate psychopathology, assessors rely heavily on the honesty, accuracy, and completeness of the examinee’s self-report (Rogers, 1997;
DETECTING OVER- AND UNDER-REPORTING

Rogers, 2008; Rogers & Correa, 2008; Rogers & Shuman, 2005). When an examinee fails to participate in the assessment process openly and honestly, whether intentionally or unintentionally, false information may be obtained and distinguishing it from the truth is often difficult. Diagnostic decisions largely affect services the individual will receive, such as where the individual will be placed (e.g., in jail or being transferred to a hospital), treatment options (e.g., medication), and, ultimately, outcomes of treatment (Parker, 1979; Rogers, 1990; Resnick, 1994; Pollock, 1996b; Rogers & Shuman, 2005).

Rogers (1997, 2008) has identified several response styles that describe the range of approaches an examinee might take in responding to a mental health interview or questionnaire. Dissimulation is a term that is used to describe an individual who is deliberately distorting or misrepresenting psychological symptoms (Rogers, 1997, 2008). Dissimulation involves “faking” symptoms and is identified with terms such as: malingering, defensiveness, irrelevant responding, and random responding. According to the American Psychiatric Association (2000), malingering is defined as “the conscious fabrication or gross exaggeration of physical and/or psychological symptoms for an external goal” (p. 739). Defensiveness is the conscious denial or gross minimization of physical and/or psychological symptoms (Rogers, 1984a, 2008). Irrelevant responding involves a response style where the individual is not psychologically engaged in the assessment process. Random responding is a subset of irrelevant responding where a random pattern can be identified (Rogers, 1997, 2008).

Other terms that have been used in the literature to describe an individual’s lack of forthrightness are unreliability, deception (Rogers, 1997, 2008), self-disclosure
DETECTING OVER- AND UNDER-REPORTING

(Jourard, 1971), social desirability (Carskey, Dolan, & Free, 1991), and impression management (Tesser & Paulhus, 1983). Unreliability is a term that describes the clinical characteristics of an individual whose response style is not honest or self-disclosing, but the intention is unknown. Deception refers to any and all attempts by an individual to distort or misrepresent their self-reporting. Self-disclosure refers to how much an individual chooses to reveal about him-or-herself (Jourard, 1971). Social desirability is composed of the denial of negative characteristics and the portrayal of positive qualities (Carskey et al., 1991). Impression management is a construct that explains the attempt to create a positive image and avoid embarrassment and other negative emotions (Tesser & Paulhus, 1983).

Most studies on response style have focused on malingering. Although the true prevalence of malingering is unknown, several studies have sought to provide estimates, which vary across populations. Based on a review of the literature of feigned neuropsychological impairment, Rogers, Harrell, and Liff (1993) reported that as many as half of individuals involved in evaluations involving personal injury litigation were possibly feigning all or part of their cognitive deficits. Rogers Sewell, and Goldstein (1994) surveyed 320 forensic psychologists and estimated that malingering occurred in 15.7% of cases in the forensic setting and in 7.4% of nonforensic cases. In another survey of 221 forensic experts, Rogers, Salekin, Sewell, Goldstein, and Leonard (1998) estimated that 17.4% of cases in a forensic setting involved malingering, in contrast to 7.8% of cases in a nonforensic setting.
DETECTING OVER- AND UNDER-REPORTING

Rogers (1997, 2010) described three explanatory models of malingering: pathogenic, criminological, and adaptational. In the pathogenic model, malingering is a result of a mental disorder. This model holds that a patient attempts to gain control over emerging symptoms by creating the symptoms and portraying them as genuine. When the onset of the mental disorder occurs, the patient loses control over the simulated symptoms and the outcome is the worsening of the mental disorder and appearance of true symptoms. The criminological model is what is described in the DSM-IV (American Psychiatric Association [APA], 2000). This model states that malingering occurs within the broader context of antisocial conduct and involves general deception. The criminological model predicts that malingering is a goal-oriented effort to achieve unworthy gains. The adaptational model was proposed by Rogers (1990a, 1990b) to explain how some individuals respond to adverse situations where they have a substantial personal investment. According to this model, would-be malingerers conduct a cost-benefit analysis when they encounter an assessment they perceive as indifferent to their needs. According to this model, malingering is more likely to occur when the context of the evaluation is perceived as adversarial, when personal stakes are high, and when no other alternatives are viable. Data from research support the adaptational model; prevalence rates are highest in adversarial situations (e.g., forensic vs. nonforensic) and when personal stakes are high (e.g., trying to get out of military combat and trying to win personal injury suits; Rogers, 1997).

According to Resnick (1993), people typically malinger mental illness for one of five purposes. One purpose for malingering includes criminals who are seeking to avoid
punishment by pretending to be incompetent to stand trial, insane at the time of the
crime, worthy of mitigation at sentencing, or incompetent to be executed. Other reasons
for malingering include avoiding military service, seeking financial gain, prisoners
wanting to obtain drugs or be transferred to a psychiatric hospital to escape or serve
lighter time, and being admitted to a psychiatric hospital to avoid arrest or receive free
room and board.

**Measures of Malingering & Response Bias**

The need to identify malingering and consider response bias is well
acknowledged, especially for evaluations conducted in a psycholegal context. In
response to this need, various measures and evaluation strategies have been developed to
assist examiners. These measures fall into two general categories: specific measures of
response style versus measures that imbed the assessment of response style in a test that
also measures clinical symptoms.

Two measures have been used widely in clinical and research settings to detect
malingering. Rogers, Bagby, & Dickens (1992) developed the Structured Interview of
Reported Symptoms (SIRS) to assess malingering of mental illness in adults; this has
recently been updated as the SIRS-2 (Rogers, Sewell, & Gillard, 2010). Another widely
used measure is the Minnesota Multiphasic Personality Inventory-2 (MMPI-2; Butcher
et al., 1989), which has validity indices that specifically measure response bias.
According to Heilburn, Bennet, White, and Kelly (1990), the MMPI-2 provides the most
satisfactory method of reducing errors and false positives when determining whether or
not individuals are feigning mental illness.
The SIRS-2 is a 172-item structured interview designed to assess feigning within the domain of mental disorders and psychopathology (Rogers et al., 2010). Factor analysis of the items resulted in eight primary scales that are used for both clinical descriptions and the classification of response styles, and five supplementary scales that are used for descriptive analysis of scale elevations (Rogers et al., 2010). The scales are described in Table 1.

Validity of the SIRS-2 was evaluated by using three research designs, a simulation design and a known-groups comparison, which were then augmented with a bootstrapping design (Rogers et al., 2010). The simulation design involves comparing honest responders to individuals who were instructed to feign mental illness and with individuals who were suspected of malingering. The known-groups design requires an accurate and independent classification of probable feigning and probable genuine responding groups comprised of individuals involved in real-world evaluations. The bootstrapping design involves using objective cut-off scores from feigning measures that were supplemented by other subjective indicators (e.g., history and clinical presentation) to develop a suspected malingering group. As was true of the original SIRS, the SIRS-2 remains highly accurate in distinguishing between feigning and honest responding using these methods. Rogers et al. (2010) report that false-positive rates were minimized at 2.5% and that classification rules achieve very good sensitivity and specificity ratings of .80 and .975 respectively. Convergent validity evidence was found by comparing the SIRS and the MMPI-2 and validity indexes; the MMPI-2 validity indexes show modest correlations to the L and K scales on the MMPI-2 (.26 and -.35) and moderately high
Table 1

*Description of SIRS Scales*

<table>
<thead>
<tr>
<th>Primary Scale Name</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>Rare Symptoms (RS)</td>
<td>Items that are infrequently reported in clinical populations.</td>
</tr>
<tr>
<td>Symptom Combination (SC) not</td>
<td>Items where symptoms are paired that typically do occur together.</td>
</tr>
<tr>
<td>Improbable or Absurd Symptoms (IA)</td>
<td>Items about symptoms that have a fantastic or preposterous quality to them.</td>
</tr>
<tr>
<td>Blatant Symptoms (BL)</td>
<td>Symptoms that untrained individuals would identify as indicative of major mental illness.</td>
</tr>
<tr>
<td>Subtle Symptoms (SU)</td>
<td>Symptoms that untrained individuals would see as everyday problems and not indicative of mental illness.</td>
</tr>
<tr>
<td>Selectivity of Symptoms (SEL)</td>
<td>Overall measurement of symptom endorsement of 32 symptoms.</td>
</tr>
<tr>
<td>Severity of Symptoms (SEV)</td>
<td>Questions about whether 32 symptoms are unbearable or too painful to stand.</td>
</tr>
<tr>
<td>Reported Versus Observed Symptoms (RO)</td>
<td>Items that ask about observable behavior that is then compared with clinical observations.</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Supplementary Scale Name</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>Direct Appraisal of Honesty (DA)</td>
<td>Items about the honesty and completeness of self-reports, level of self-disclosure to mental health professionals, concerns about being truthful to professionals.</td>
</tr>
<tr>
<td>Defensive Symptoms (DS)</td>
<td>Symptoms that are representative of everyday problems, worries, and situations that most individuals experience to some degree.</td>
</tr>
<tr>
<td>Overly Specified Symptoms (OS)</td>
<td>Symptoms that are endorsed with an unrealistic degree of precision.</td>
</tr>
<tr>
<td>Inconsistency of Symptoms (INC)</td>
<td>Repeats a set of 32 clinical inquiries to measure the stability of the individual's self-report.</td>
</tr>
<tr>
<td>Improbable Failure (IF)</td>
<td>Set of simple cognitive tasks (opposites and rhyming) that should easily be completed by most persons.</td>
</tr>
</tbody>
</table>
correlations between SIRS-2 scales and MMPI-2 fake-bad indicators (.6 to .7; Rogers et al., 2010).

The Minnesota Multiphasic Personality Inventory-Second Edition (MMPI-2) is the most commonly used measure in forensic evaluations (Melton, Petrila, Poythress, & Slobogin, 2007). In contrast to the SIRS-2, the MMPI-2 is designed to assess personality and psychological characteristics, but it also measures test-taking attitudes and provides the clinician with a means of knowing whether the client has cooperated sufficiently with the evaluation to provide an accurate portrayal of his or her personality characteristics and problems. The MMPI-2 is comprised of 567 true-false items that contribute to 10 clinical scales, as well as three standard and four supplementary validity scales. The standard validity scales are F, L, and K, and the four supplementary validity measures are: the Variable Response Inconsistency (VRIN) scale, the True Response Inconsistency (TRIN) scale, and the F₁ and F₂ subscales.

The F scale measures deviant or atypical ways of responding to test items, such as those that resulted when test takers did not read the items and responded randomly (Graham, 2006). Extreme elevations on the F scale could indicate severe psychiatric illness, may reflect an attempt to “fake-bad” or overreport symptoms, or may indicate that the examinee is responding randomly to the questions. The L scale is used to identify adolescents who are deliberate and unsophisticated in their attempt to present themselves in an unrealistically favorable light (Graham, 2006). Extreme elevations on the L scale could indicate that the test taker was not honest and frank in answering items on the inventory and may have claimed virtues and denied negative characteristics to a
DETECTING OVER- AND UNDER-REPORTING

greater extent than most people. The K scale identifies examinees that subtly deny psychopathology and present themselves in a favorable light or, conversely, to exaggerate psychopathology and to try to appear in a very unfavorable light (Graham, 2006). High scores on the K scale are associated with a defensive approach to the test, whereas low K scores are associated with unusual frankness and self-critical attitudes. The additional validity scales provide data about the adolescent’s tendency to respond to the items in a consistent manner. The Variable Response Inconsistency (VRIN) scale detects when similar items are answered inconsistently, which captures random responding (Graham, 2006). The True Response Inconsistency (TRIN) scale assesses an adolescent’s tendency to indiscriminately respond to items as either all true or all false, regardless of the content of items (Graham, 2006).

As discussed earlier, the MAYSI-2 was developed in 1998 and refined in 2001 to identify incarcerated youths who may experience signs of mental health problems and/or may pose a higher than typical risk for suicidal or other challenging behavior in the facility. Grisso and Barnum (2001) and Meyers (2006) noted that the MAYSI-2 lacks a validity index that might be sensitive to over-or under-reporting of symptoms. The purpose of this study is to develop validity items to incorporate with the MAYSI-2, with the end result being a first step towards measuring validity of MAYSI-2 profiles.

**Designs Used to Develop Response Bias Measures**

A number of different designs are used when developing measures to detect response bias. Ideally, one would be able to identify a group of individuals whose response bias is known. However, individuals rarely admit to their response bias,
particularly if they are malingering (Tombaugh, 2002). A second approach is to use another test to classify participants into categories, most commonly into malingering and non-malingering groups, then determine the accuracy of the developed measure in correctly classifying members into the two groups.

Widely used measures, such as the Test of Memory Malingering (TOMM; Tombaugh, 2002) and the SIRS-2 (Rogers et al., 2010), have followed a series of research designs to validate the measure. The first design that is typically used is a simulation design. In a simulation design, mentally healthy individuals (typically college students) are randomly divided into simulators, who are instructed to fake a condition, and controls, who are instructed to respond honestly. A study that was used to develop the SIRS randomized participants into a simulator group and an honest responder group (Rogers, Gillis, Dickens, & Bagby, 1991). Results showed that simulators had elevated scores on the SIRS, which indicated a feigning response style. Another study that was used to develop the SIRS randomly assigned 90 college students into either a coached simulator, uncoached simulator, or control/honest condition (Rogers, Gillis, Bagby, & Monteiro, 1991). The outcomes were compared to an inpatient psychiatric sample. Results indicated that coached simulators were able to lower their SIRS scores on scales designed to detect feigning. However, coached simulators continued to show significant differences from honestly responding psychiatric inpatients on the majority of the scales.

The second design that is used examines participants’ actual responses with their malingering responses to evaluate the effects of test sophistication on its ability to discriminate between simulating and control participants (Tombaugh, 2002). This
DETECTING OVER- AND UNDER-REPORTING

involves a within-subject design in which participants are administered the measure under two counterbalanced instructional sets – to simulate and respond honestly. Once it is determined that the two above designs accurately discriminate between simulators and controls, the third design level is used. The third level design embeds the malingering test into a battery of other similar tests. Because the participant has the opportunity to compare the malingering test with other tests measuring the same construct, this situation presents a noticeably different environment to detect malingering. The rationale for this is that when malingering tests are presented in isolation, the true purpose (i.e., to detect inaccurate responding) is less obvious than when administered along with other measures (Tombaugh, 2002).

During the fourth design level, participants are chosen who are “truly knowledgeable” about what the test measures (Tombaugh, 2002). The reasoning for this step is that the sensitivity of the measure to detect malingering may have reflected the participants’ lack of knowledge about mental illness. During this step, individuals who have suffered mental illness are randomly assigned to a mental illness simulation group or a mental illness control group. The scores from the two groups are then compared to a mentally healthy control group.

One final strategy that is useful in validating a malingering measure is to evaluate individuals in situations where the likelihood of malingering is high (Tombaugh, 2002). One example of this is when an opportunity for financial compensation exists, which is typically present in personal liability suits or disability petitions. A useful example of this type of design comes from studies examining the Test of Memory Malingering.
DETECTING OVER- AND UNDER-REPORTING

(TOMM) in which they compared the TOMM performance of individuals with Traumatic Brain Injury (TBI) who were known to be in litigation involving their injury, TBI survivors who were not involved in litigation, and a group of healthy controls (Tombaugh, 2002). The significant differences in TOMM scores across these groups supported the utility of the TOMM’s ability to measure response style.

Test Development

When a test is developed or modified, there are well-established steps that are taken to maximize the likelihood that the measure will accurately assess the desired construct and be psychometrically sound (Clark & Watson, 1995). The first step in test development is to define the behaviors the test represents, the target audience, and the purpose of the test (Miller, McIntire, & Lovler, 2011). In order to define the behaviors the test represents, a working definition of the construct is created that operationally describes the construct in terms of observable and measureable behaviors. The definition also provides limits for the test domain by outlining which content to include and which content to exclude. Information that is gathered by defining behaviors the test represents, the target audience, and the purpose of the test creates the foundation for making other decisions about the test (Miller et al., 2011).

To create a clear conceptualization of the target construct, Cronbach and Meehl (1955) described three steps that are necessary to investigate construct validity: describing a set of theoretical concepts and their interrelations; developing ways to measure the constructs that the theory proposes; and testing the hypothesized relations among constructs empirically. Construct validity evaluates whether the scale being
developed measures what it is supposed to measure. According to Chronbach and Meehl, in order for a measure to have construct validity it must demonstrate content validity, criterion-related validity, and internal consistency. When developing test items, the goal is to develop items that sample the theoretical domain of interest (Hinkin, 1998). However, it is not possible to capture the entire domain of interest, so it is important that the sample of items draws from potential items that adequately represent the construct (Hinkin, 1998).

Having experimental evidence of construct validity is crucial (Borsboom, Mellenbergh & Heerden, 2004). The author’s state that a test is shown to be a valid measure of an attribute if the attribute exists, and if the variations in the attribute causally produce variation in the measurement outcomes. Borsboom et al. (2004) state that validation research should be directed at the relation between the measured attributes and other attributes, but instead at the processes that produce the effect of the measured attribute on the test scores.

When writing items for a test, there are three main questions that should be considered: What range of content should the items cover? Which format should be used? and How many items should be written? (Cohen & Swerdlik, 2005). It is generally recommended to have the initial pool of potential items contain approximately twice the number of the items that the final version will contain (Hinkin, 1998). To compile an item pool, the test developer often writes a large number of items from personal or academic experience, often seeking assistance from others, such as subject matter experts (Cohen & Swerdlik, 2005). Once a pool of items is created, the test developer
will try out the test, with the aim of identifying those test items that provide the greatest discrimination between individuals who reflect the construct and those who do not, so that the most discriminating items will be included on the final test. There are four main strategies to analyze and select test items. The first strategy is the item-endorsement index, which measures the percent of people who endorsed the item. The second strategy is the item-reliability index, which provides an indication of the internal consistency of a set of test items. The third strategy is the item-validity index, which provides an indication of the degree to which a test is measuring what it purports to measure. The fourth strategy is the item-discrimination index, which indicates how adequately an item separates or discriminates between high scorers and low scorers on the test (Cohen & Swederlik, 2005).

As mentioned previously, there is a clear need to screen for mental illness in the juvenile justice system. Although the MAYSI-2 has been shown to be a useful screening measure, it provides no indication of possible response bias, which limits its utility, especially in the psycho-legal context.

The aim of the present study is to address this shortcoming of the widely used MAYSI-2 (Grisso & Barnum, 2003; Meyers, 2006) by developing and evaluating items that will be embedded into the MAYSI-2 to provide information about possible over-or under-reporting of symptoms. Specifically, this study will employ a second level simulator design to assess the proposed additional items to the MAYSI-2. Consistent with the strategies used by previous studies, this simulator study will employ a sample of college students who will be asked to respond to a modified version of the MAYSI-2
under three conditions: an over-reporting condition, under-reporting condition, and a control condition.
Chapter II

Rationale and Hypotheses

Given the high prevalence of mental illness that has been documented in the juvenile justice system, there is a need to screen youths who enter this system for symptoms of mental illness in order to assure safety and appropriate precautions. Juvenile justice facilities have the responsibility to keep the juveniles in their facilities safe; which includes making sure they do not engage in self-harm, endanger other juveniles, or commit suicide. Since 2001, the Massachusetts Youth Screening Instrument-2 (MAYSI-2; Grisso & Barnum, 2001) has become widely used as a brief self-report screening tool that is administered to youths upon entrance to detention or other correctional facilities. This measure has been shown to identify significant levels of psychiatric symptoms and identify youths in crisis who may require immediate and more detailed evaluation. However, a major limitation of the MAYSI-2 is that it lacks a validity index or other means to assess response style, so there is no indication whether or not the youth’s self-report is a reflection of true symptoms, or whether the youth is under-reporting or over-reporting symptoms. As a result, juveniles who are under-reporting their symptoms are going without appropriate precautions that could be put in place to ensure their safety and the safety of others, and juveniles who are over-reporting their symptoms are using up unnecessary resources and allocation of staff time, which could be spent helping youth who are actually in crisis.

The purpose of this study is to add additional items to the MAYSI-2 and determine if these items might assess under and over-reporting of psychological distress
in juveniles, with the end result being a first step towards measuring response style on the MAYSI-2. A total of 20 items will be incorporated into the MAYSI-2, with 10 items assessing under-reporting and 10 items assessing over-reporting. The MAYSI-2 with the 20 additional items will be labeled the Modified MAYSI-2 (MMAYSI-2). This study will use an experimental design using a simulation methodology, which has been widely used in the development of measures to assess response style (Rogers, 1988). This will involve randomly assigning participants to one of three groups: simulation of mental illness, which involves over-reporting symptoms; simulating perfect mental health, which involves under-reporting symptoms; and a control group, which involves responding to questions on the measure as one normally would.

The following hypotheses are proposed:

H1. The mean under-reporting score of participants in the under-reporting group will be statistically significantly higher than the mean under-reporting scores of the over-reporting and control groups.

H2. The mean over-reporting score of participants in the over-reporting group will be statistically significantly higher than the mean over-reporting scores of the under-reporting and control groups.

H3. The mean clinical scale scores of the MMAYSI-2 will be statistically significantly higher for the over-reporting group than for the under-reporting and control groups.
Chapter III

Method

Participants

The sample will consist of 200 undergraduate students, between the ages of 18 and 19, recruited from a private Midwestern university. This age restriction is imposed to create a sample whose age is close to that of the intended, adolescent sample of the MAYSI-2. Participants will be recruited through the Psychology Department participant pool, which provides a mechanism through which students enrolled in Psychology courses can receive course-related credit for their participation.

Power

In order to determine the appropriate sample size needed to achieve an adequate level of power, statistical analyses were conducted using the G POWER software (Erdfelder, Faul, & Buchner, 1996). Using a one-way ANOVA analysis, results indicate that in order to achieve a medium effect size of 0.25 a sample of 159 participants is needed, with 53 participants in each of the three groups. However, in order to conduct a factor analysis to examine the structure of the scales, it is recommended to use at least 200 participants (Clark & Watson, 1995).

Measures

Demographic Questionnaire. Participants will complete a Demographic Questionnaire (Appendix A) prior to completing the MAYSI-2. The questionnaire requests background information that includes the individual’s age, ethnicity, year in college, and major. The questionnaire also requests information regarding whether the
individual has taken or is currently taking a course in Abnormal Psychology. This information is being sought to see if there is a difference between the way students who are familiar with abnormal psychology respond to under/over reporting.

**Massachusetts Adolescent Youth Screening Instrument-2 (MAYS1-2; Grisso & Barnum, 2000).** The MAYS1-2 is a 52 item, yes/no paper-and-pencil self-report questionnaire that produces scores on seven scales (six for girls) that address behavioral, mental and/or emotional problems (Alcohol/Drug Use, Angry-Irritable, Depressed-Anxious, Somatic Complaints, Suicide Ideation, Thought Disturbance, Traumatic Experiences and Thought Disorder). The MAYS1-2 was developed for use on juveniles from age 9 to 17; however local norms have been developed for 8 to 19 year olds (Meyers, 2006). The MAYS1-2 is used during entrance to a juvenile facility or at a point of transitional placement to identify youth who are experiencing a significant degree of psychological distress. It is not intended to provide a psychiatric diagnosis. Six of its seven scales utilize “Caution,” or “Warning” categories to identify those youth who may be in need of some psychological or behavioral services. If a youth’s score is in the Caution category, he/she has scored at a level that is indicative of “possible clinical significance” (Grisso & Barnum, p. 30). Similarly, a score in the Warning category is “exceptionally high in comparison to other youth in the juvenile system” (p. 31) and merits further evaluation or some type of intervention. Grisso and Barnum note that a score in the Warning category suggest that the respondent “should be considered most likely to be in need of attention for mental problems because they are reporting problems at a level that far exceed the average youth in juvenile justice settings” (p. 31). The
DETECTING OVER- AND UNDER-REPORTING

Traumatic Experiences scale does not utilize cutoff scores, as it primarily focuses on information about past incidents of trauma. In contrast to instructions for the other items, respondents are asked to consider his/her entire life experience, rather than “the past few months” (p. 25) when responding to the Traumatic Experience items. In addition, Grisso and Barnum indicate that a high score on this scale “should not be interpreted as indicating the presence of PTSD” (p. 26). Of note, the Thought Disturbance scale is limited to boys, as it was not found to be relevant to the girls in the normative sample.

A strength of the MAYSI-2 is its psychometric properties. The MAYSI-2 yielded alpha coefficients of .61 to .86, with an average of .75. All items correlate .33 or above with the scale to which they contribute. The average correlation between items and total scores ranged from .37 to .63, while correlations between the scales were in the .24 to .61 range, with an average of .39 for boys and .41 for girls. The inter-scale correlations are not sufficiently high, consistent with the intent that the MAYSI-2 is designed to measure multiple concepts, not a singular one. Test-retest reliability varied across scales. Coefficients ranged from .53 to .89 for boys (average, .74) and .66 to .85 for girls (average, .74). The validity of the MAYSI-2 was analyzed by administering the Millon Adolescent Clinical Inventory (MACI) and the Child Behavior Checklist-Youth Self Report (CBCL-YSR) to 561 boys and 198 girls. An additional 73 youth completed the CBCL-YSR, resulting in 882 valid CBCL-YSR protocols. The authors of the MAYSI-2 found that most scales were correlated with scales on other instruments at .45 or greater (Grisso & Barnum, 2000).
The MAYSI-2 will be adapted by adding 20 items to comprise a validity scale, creating the MMAYSI-2. Ten of the items will consist of questions designed assess over-reporting of mental illness symptoms, while the other 10 items are designed to assess under-reporting.

Procedure

Item development. Construction of the new items for the MMAYSI-2 was done by reviewing items on the SIRS and MMPI-A that have been shown to be sensitive to response style. The content of select items was adapted to conform to the format of the MAYSI-2.

Originally, 20 items (each) were developed for the under-reporting and over-reporting scales (see Appendix B). These 40 items will be reviewed by five mental health professionals who have experience in the field of detecting malingering and response bias. Professionals will be instructed to rank order the items, 1-20 for each scale. Based on agreement between the reviewers, the top 10 under-reporting items and top 10 over-reporting items will be added to the MAYSI-2.

Simulation study. Prior to the recruitment of participants, and data collection, approval will be obtained from the Institutional Review Board at Xavier University to conduct the second phase of this study. Undergraduate students who are 19-years old or younger will be recruited from the Psychology Department participant pool; all participation will be on a voluntary basis. This age restriction is imposed to create a sample whose age is close to that of the intended, adolescent sample of the MAYSI-2. Students will complete the MMAYSI-2 and Demographic Form on-line via Survey
Gizmo. Students will be randomly assigned to one of three conditions: an under-reporting simulator group, an over-reporting simulator group, or a control group.

Once participants are randomly assigned to either the experimental or control group, they will receive a simulation script and directions to complete the MMAYSI-2. The simulation scripts were modeled after previous studies using simulation designs (Tombaugh, 2002). In the under-reporting simulator group, participants will be instructed to under-report the presence of mental illness (Appendix C), whereas the over-reporting simulator group participants will be instructed to feign mental illness (Appendix D). Participants in the control condition will be instructed to respond honestly (Appendix E). After completion of the MMAYSI-2 and Demographic Form, participants will complete a manipulation check, which will involve indicating which instructions they were given and stating whether or not they responded honestly, over-reported their symptoms, or under-reported their symptoms (see Appendix F). Data from participants who fail the manipulation check will be eliminated. Following participation, each participant will receive a written debriefing form (Appendix G).
Chapter IV

Proposed Analyses

Prior to hypothesis analysis, an exploratory factor analysis using principal axis factoring with an oblique direct oblimin rotation will be conducted (Conway & Huffcutt, 2003). Based upon the results of the factor analysis, the four to five items that have the highest factor loading will serve as the items for each scale. As recommended by Hinkin (1998), internal consistency reliability will also be assessed using Cronbach’s alpha after determining the factor structures of the scales. The prediction is that the over-reporting and under-reporting test items will result in two underlying factors that represent an over-reporting test approach and an under-reporting test approach.

The first two hypotheses assess the ability of the Under- and Over-reporting scales to discriminate between honest and simulating responders. The first hypothesis states that the under-reporting score of participants in the under-reporting group will be significantly higher than the scores of the over-reporting and control groups. The second hypothesis states that the over-reporting score of participants in the over-reporting group will be significantly higher than the scores of the under-reporting and control groups. To test these hypotheses, one-way ANOVAs with post-hoc analyses will be used to test differences across the three groups on the over- and under-reporting scales. A significance level of .05 will be used.

The third hypothesis states that the clinical scales of the MMAYSI-2 generated by the over-reporting group will be significantly different from those generated by the
under-reporting and control groups. One-way ANOVAs with post-hoc analyses will be used to test this hypothesis, with a significance level of .05.
DETECTING OVER- AND UNDER-REPORTING

References


DETECTING OVER- AND UNDER-REPORTING


DETECTING OVER- AND UNDER-REPORTING


DETECTING OVER- AND UNDER-REPORTING


DETECTING OVER- AND UNDER-REPORTING


DETECTING OVER- AND UNDER-REPORTING


DETECTING OVER- AND UNDER-REPORTING


Appendix A

Demographic Questionnaire

Sex: ____________
Age: ____________
Year in college: ________________
Major: ________________

Have you ever taken or are you currently taking Abnormal Psychology? ___ Yes ___ No

Ethnicity: ___ Caucasian ___ African American ___ Asian ___ Hispanic

Other (please specify) ____________________________________________

Prefer not to answer _______
Appendix B

Validity Items Developed

Under-reporting Items:
1. Have you noticed that you are immediately comfortable with all the new people you meet?
2. Have you found that it is very easy to concentrate even when you are upset or depressed?
3. Have you noticed that it is easy for you to make decisions both big and small?
4. Have you found that it is very easy to forgive others even if they have really hurt you?
5. Have you ever had thoughts you do not feel comfortable talking about?*
6. Have you ever felt like punching something?*
7. Have you ever lied?*
8. Have you ever put off doing important jobs or tasks?*
9. Have you ever felt sad?*
10. Have you ever kept money that you found on the ground?*
11. Have you ever talked about others behind their backs?*
12. Have you ever argued with anyone in your family?*
13. Have you ever felt like swearing?*
14. Have you ever felt sad when someone criticized you?*
15. Have you ever been concerned about how other people think about you?*
16. Have you ever noticed that your mind is filled with worries?*
17. Have you ever worried about how you are doing in school?*
18. Have you ever been mad, but you weren’t sure why?*
19. Have you ever felt that you are smarter than everyone else?*
20. Have you ever had problems speaking up for yourself?*

Over-Reporting Items:
1. Have you noticed that you rhyme your words when talking to others?
2. Have you noticed that you have to hold your breath when you pass through doorways?
3. Have you noticed a strange color when you look at the moon?
4. Have you lost exactly 4 lbs in the last 4 weeks?
5. Have you ever had periods of times when you fainted often?
6. Have you ever felt that aliens have over your body?
7. Have you ever been afraid to go to church?
8. Have you ever felt that the world would be a better place without laws?
9. Have you ever seen animals that others cannot see?
10. Have you ever felt that the government is plotting against you?
11. Have you ever felt that you are an evil person?
12. Have you ever felt that everything smells the same?
13. Have you ever felt that you are no longer in your physical body?
14. Have you ever been afraid of using the stove or anything very hot?
15. Have you ever felt that other people insert thoughts in your head?
16. Have you ever felt there are times when you were under a spell?
17. Have you ever noticed special messages just for you while watching TV?
18. Have you ever felt you needed others to plan your life?
19. Have you ever noticed that you get angry when you are complimented?
20. Have you ever felt that someone was trying to take credit for your innermost thoughts?
21. Have you ever noticed that when you are around others, you are troubled by seeing weird things?

In the MMAYSI-2, all items that are answered "Y" get one point. Items that have a * are reverse-scored, answering "N" indicates either over- or under-reporting.
I want you to pretend that you are in the process of being interviewed for a job. As part of the interview process, you are instructed to take a self-report measure that screens for mental illness. You really want and need this job, and in order to obtain it you need to present yourself in the best light possible. Complete the attached measure in a way that you think will give you a clean bill of mental health.
I want you to pretend that you are in danger of being expelled from the University because of failing grades. You have learned that the University will allow you to remain at school if you are seriously mentally ill and possibly in need of hospitalization. The measure you are taking is a screening measure to detect mental illness and doctors use it to determine what sort of symptoms individuals have. You are taking this measure as part of the review process at the University regarding your status. Keep in mind that it is very important that you fool the doctors into thinking you are too mentally ill to be expelled.
DETECTING OVER- AND UNDER-REPORTING

Appendix E

Script for Control Group

The measure you are taking is a screening measure to detect mental illness. Please fill this out in a way that reflects your current emotional status and experiences.
Appendix F

Manipulation Check

What were the instructions that you were given prior to completing the questionnaire?

Did you respond to the questions honestly?

Did you respond to the questions trying to portray a mentally ill person?

Did you respond to the questions trying to portray yourself in the best light possible?
Appendix G

Debriefing Form

This study is assessing over-and-under-reporting on the MAYSI-2, a screening measure used in juvenile detention facilities to detect mental illness. The current study is looking to improve the original MAYSI-2 by adding items designed to assess over-and-under-reporting. Some items that you completed assess over-reporting, some under-reporting, and some were from the original MAYSI-2. For this study, you were given instructions about how you were to fill out the measure: some of you were told to answer based on your real experiences, some of you were asked to try to appear mentally ill, and some of you were asked to try to appear as if you have no psychological issues.

Sometimes when people complete mental health measures, they become aware of difficulties they are having or become concerned that they might have difficulties that warrant intervention. If you feel you would like to speak with someone about your feelings or experiences, you can access counseling and psychological services through the Psychological Services Center at 513-745-3531 or the McGrath Health and Wellness Center at 513-745-3022. Most services at these centers are provided at no additional cost to full-time Xavier students. Part-time students can access services at a reduced cost.

Because the MAYSI-2 is designed to be used with adolescents, we restricted the sample to undergraduates who are closest in age to the intended group. If you participated, but are above the age of 20, please click “do not submit survey” below and your answers will not be saved. We will check the ages of all participants through the publicly available information in the registrar’s office; participants who misrepresented their age will be reported as having missed a study session, and may be docked participation credit.

_____ submit responses

_____ do not submit responses
The Massachusetts Youth Screening Instrument – Second Version (MAYSI-2) is a self-report questionnaire designed to identify the mental health needs of juveniles in correctional facilities. This study evaluated the scores of a Modified MAYSI-2 (MMAYSI-2) that added 10 items capturing over-reporting and 10 items capturing under-reporting response bias in a sample of college students who were instructed to over-report, under-report, or answer honestly. A factor analysis established the new items to capture over- and under-reporting and ANOVAs identified statistically significant differences across the instruction groups in both the original MAYSI-2 clinical scale scores and the proposed validity scales in the expected directions. The results are discussed in reference to further development of measuring response style on the MAYSI-2.
Detecting Over- and Under-reporting of Symptoms on the MAYSI-2: Development of a Validity Scale

Several studies have documented high rates of mental illness among children who come into contact with the juvenile justice system (Fazel et al., 2008; Rohde et al., 1997; Teplin et al., 2012; Wasserman et al., 2002), which has led to a need to screen youths who enter this system for symptoms of mental illness in order to assure safety and appropriate precautions. Screening also addresses the requirement that most states have to assess every youth who enters the justice system for suicide risk, mental and emotional health, and substance abuse (Lexcen, Vincent, & Grisso, 2004; Stewart & Trupin, 2003). To meet this need, Grisso and Barnum (2001) created the Massachusetts Youth Screening Instrument (MAYSI-2). The MAYSI-2 is a brief self-report screening tool that is administered to youth as they enter a detention center or other correctional facilities. Grisso and Barnum designed the MAYSI-2 to meet several goals: it is brief (taking about 8-10 minutes to complete), easy to score, and can be interpreted without clinical specialization.

The MAYSI-2 has well established, strong psychometric properties (Archer et al., 2010; Butler et al., 2007; Ford et al., 2008; Grisso & Barnum, 2003; Kerig et al., 2011; Lexcen et al., 2004; Meyers, 2006; Wasserman et al., 2005) and has been shown to identify youths with psychiatric symptoms and/or youths in crisis who may require immediate and more detailed evaluation and intervention. However, a major limitation of the MAYSI-2 is that it lacks a measure of validity or response style that would detect
whether or not the youth’s self-report is a reflection of his or her true functioning (Grisso & Barnum, 2003; Meyers, 2006). Recommendations for future research included developing a validity index to build into the measure to assess response bias and over/under-reporting (Grisso & Barnum 2003; Meyers, 2006).

The need to accurately identify youths in need of mental health services is obvious, especially when they are entering the custody of a juvenile facility, which can be stressful and upsetting. Juveniles who under-report symptoms potentially place themselves and others at risk for harm, whereas juveniles who over-report symptoms may use unnecessary resources and staff time. The issue of misrepresentation of one’s true psychological status is particularly relevant in a forensic setting, where one’s report of psychological difficulties could be influenced by secondary gain, either in the form of being able to leave a corrections facility for a hospital, or to avoid prosecution, such as by being found incompetent to stand trial. According to Resnick (1993), people typically over-report mental illness for one of five purposes. One purpose for over-reporting includes individuals who are seeking to avoid punishment by pretending to be incompetent to stand trial, insane at the time of the crime, worthy of mitigation at sentencing, or incompetent to be executed. Other reasons for over-reporting include avoiding military service, seeking financial gain, prisoners wanting to obtain drugs or be transferred to a psychiatric hospital to escape or serve lighter time, and being admitted to a psychiatric hospital to avoid arrest or receive free room and board.

The need to identify and consider response bias is well acknowledged, especially for evaluations conducted in a psycholegal context (Resnick, 1993; Rogers et al., 2010).
In response to this need, various measures and evaluation strategies have been developed to assist examiners. These measures fall into two general categories: specific measures of response style and measures that imbed the assessment of response style in a test that also measures clinical symptoms. A specific measure of response style is a measure that solely assesses response style, such as over/under-reporting. Examples of such measures include the Test of Memory Malingering (TOMM, Tombaugh, 2002; designed to assess feigned memory deficits) and the Structured Interview of Reported Symptoms-2 (SIRS-2, Rogers et al., 2010; designed to assessed feigned or exaggerated psychiatric symptoms). In contrast, a measure that imbeds the assessment of response style into a test measures personality and psychological characteristics/symptoms, but it also measures test-taking attitudes and provides the clinician with a means of knowing whether the client has cooperated sufficiently with the evaluation to provide an accurate portrayal of his or her personality characteristics and problems (Graham, 2006).

Widely used measures, such as the TOMM (Tombaugh, 2002) and the SIRS-2 (Rogers et al., 2010), have followed accepted strategies in their development and validation. Specifically, the use of simulation of various response styles has become a common means to develop measures of response bias, and is usually the first step taken by researchers. In a simulation design, mentally healthy individuals (often college students) are randomly divided into simulators, who are instructed to fake a condition, and controls, who are instructed to respond honestly or perform the task with strong effort. Rogers, Gillis, Dickens and Bagby (1991) used this design as a phase of development of the SIRS, and they found that simulators had elevated scores on the
SIRS, which indicated a feigning response style that could be differentiated from the responses of people who were honestly and openly reporting on their experiences.

Another common design to develop and validate a measure is the second-level design that examines participants’ actual responses with their malingering responses to evaluate the effects of test sophistication on its ability to discriminate between simulating and control participants (Tombaugh, 2002). This involves a within-subject design in which participants are administered the measure under two counterbalanced instructional sets – to simulate and respond honestly. Once it is determined that the two above designs accurately discriminate between simulators and controls, the third design level is used. The third-level design embeds the malingering test into a battery of other similar tests. Because the participant has the opportunity to compare the malingering test with other tests measuring the same construct, this situation presents a noticeably different environment to detect malingering. The rationale for this is that when malingering tests are presented in isolation, the true purpose (i.e., to detect inaccurate responding) is less obvious than when administered along with other measures (Tombaugh, 2002).

A fourth-level design can also be used to validate a measure, which is when participants are chosen who are “truly knowledgeable” about what the test measures (Tombaugh, 2002). The reasoning for this step is that the sensitivity of the measure to detect malingering may be influenced by the participants’ lack of knowledge about mental illness. During this step, individuals who have suffered mental illness are
randomly assigned to a mental illness simulation group or a mental illness control group. The scores from the two groups are then compared to a mentally healthy control group.

One final strategy that is useful in validating a malingering measure is to evaluate individuals in situations where the likelihood of malingering is high (Tombaugh, 2002), such as when there is an opportunity for financial compensation, such as in personal liability suits or disability petitions. A useful example of this type of design comes from studies examining the TOMM in which they compared the TOMM performance of individuals with Traumatic Brain Injury (TBI) who were known to be in litigation involving their injury, TBI survivors who were not involved in litigation, and a group of healthy controls (Tombaugh, 2002). The significant differences in TOMM scores across these groups supported the utility of the TOMM’s ability to measure response style.

As mentioned previously, there is a clear need to screen for mental illness in the juvenile justice system. Although the MAYSI-2 has been shown to be a useful screening measure, it provides no indication of possible response bias, which limits its utility given its use in the psycho-legal context. Therefore, the aim of the study was to address this shortcoming of the widely used MAYSI-2 (Grisso & Barnum, 2003) by developing and evaluating items to be embedded into the MAYSI-2 to provide information about possible over-or under-reporting of symptoms. Specifically, the study employed a first level simulator design to assess the proposed additional items to the MAYSI-2. Consistent with the strategies used by previous studies, the simulator study employed a sample of college students who were asked to respond to the Modified MAYSI-2
DETECTING OVER- AND UNDER-REPORTING

(MMAYSI-2) under one of three conditions: an over-reporting condition, under-reporting condition, and a control condition.

Method

Participants

A total sample of 234 college students completed study measures after being recruited through a psychology department participant pool at a private university in the Midwest. The only stipulation for participation was that students be 19-years old or younger; 36 student did not meet this inclusion criterion and their data were excluded from the data analysis, leaving a total sample of 197 participants. Further, 19 participants failed the manipulation check, leaving a total of 178 participants.

Demographic information for the sample is presented in Table 1. The students ranged in age from 17 to 19 years old with a mean age of 18.71 (SD = 0.46). As can be seen, the majority of the total sample were Caucasian (77.5%) and women (68.5%).

The online survey collection site randomly assigned participants into one of three instruction conditions; the demographic information for each of these groups is also presented in Table 1. There were no significant age, F(2, 177) = .160, p = .85, sex, X^2 (1, N = 178) = .324, p = .85, or race, X^2 (3, N = 178) = 5.24, p = .88, differences across the three instruction conditions.

Measures

Demographic Questionnaire. Prior to completing the MMAYSI-2, participants provided basic personal information: age, ethnicity, year in college, and college major (see Appendix A). Because two of the instruction conditions asked participants to under-
DETECTING OVER- AND UNDER-REPORTING

or over-report mental illness, the questionnaire also requested information regarding whether the individual had taken or was currently taking a course in Abnormal Psychology.

Massachusetts Adolescent Youth Screening Instrument-2 (MAYSI-2; Grisso & Barnum, 2000). The MAYSI-2 is a 52 item, yes/no self-report questionnaire that produces scores on seven scales (six for girls) that address behavioral, mental and/or emotional problems: Alcohol/Drug Use; Angry-Irritable; Depressed-Anxious; Somatic Complaints; Suicide Ideation; Thought Disturbance; Traumatic Experiences; and Thought Disorder. The MAYSI-2 was developed for use with juveniles from age 9 to 17; however local norms have been developed for 8 to 19 year olds (Meyers, 2006). The MAYSI-2 is used during entrance to a juvenile facility or at a point of transitional placement to identify youth who are experiencing a significant degree of psychological distress that would benefit from further evaluation. Six of its seven scales utilize “Caution” or “Warning” categories to identify those youth who may be in need of psychological or behavioral services. The Traumatic Experiences scale does not utilize cutoff scores, as it primarily focuses on information about past incidents of trauma. Of note, the Thought Disturbance scale is scored only for boys, as it was not supported by factor analysis for girls in the normative sample.

The MAYSI-2 manual (Grisso & Barnum, 2000) reports alpha coefficients of .61 to .86, with an average of .75. All items correlate .33 or above with the scale to which it contributes. The average correlation between items and total scores ranged from .37 to .63, whereas correlations between the scales ranged from .24 to .61, with an average of
.39 for boys and .41 for girls. The inter-scale correlations are not high, consistent with
the intent to measure distinct areas of difficulty. Test-retest reliability coefficients ranged
from .53 to .89 for boys (average = .74) and .66 to .85 for girls (average = .74). The
validity of the MAYSI-2 was established by comparing its scores to the Millon
Adolescent Clinical Inventory (MACI) and the Child Behavior Checklist-Youth Self
Report (CBCL-YSR) in a sample of 561 boys and 198 girls. Most of the MAYSI-2 scale
scores correlated with comparable scales on other instruments at .45 or greater (Grisso &
Barnum, 2000).

MAYSI-2 modification. In the current study, the MAYSI-2 was adapted by
adding 20 items to comprise two validity scales; 10 of the items consisted of questions
designed to assess over-reporting of mental illness symptoms, whereas 10 other items
were designed to assess under-reporting. The new items are listed in Table 2, creating the
Modified MAYSI-2 (MMAYSI-2). The strategy used to develop these items is described
below.

Procedure

Item development. The new items for the MMAYSI-2 were written by the
principal investigator after reviewing items from the SIRS and the MMPI-A that had
already demonstrated the ability to identify response bias. An original pool of 20 items
for each proposed scale was developed. Five mental health professionals who have
experience in the field of detecting malingering and response bias reviewed these 40
items and rank ordered the items from 1-20 for each scale, assigning the item that best
assessed over- or under-reporting as 1, the next best item as 2, and the so forth. The top
10 agreed upon under-reporting items and top 10 over-reporting items were added to the original 52 MAYS1-2 items, resulting in a 72-item scale (see Appendix B). Although there were some differences in the rankings made by the five professionals, the majority of items that were retained for the MMAYSI-2 were ranked in the top 10 best items by all five professionals; items that were ranked by four professionals to be in the top 10 were chosen next, until the list of 10 under-reporting items and 10 over-reporting items was complete.

The added items were embedded into the original MAYS1-2, rather than added to the end so that detection of these items was less likely. It was thought that if the items were in one group at the end of the measure the scale would be less sensitive when detecting over-under-reporting, although we did not empirically test this assumption.

Simulation study. Prior to the recruitment of participants and data collection, approval was obtained from the Institutional Review Board at Xavier University to conduct the second phase of this study (see Appendix C). Undergraduate students who met the eligibility requirement of being 19-years old or younger were invited from the Psychology Department participant pool; all participation was on a voluntary basis. This age restriction was imposed to create a sample whose age was close to that of the intended, adolescent sample of the MAYS1-2. Students completed the MMAYSI-2 and Demographic Form on-line via Survey Gizmo. Students were randomly assigned to one of three conditions by Survey Gizmo: an under-reporting simulator group, an over-reporting simulator group, or a control group.
Once participants were randomly assigned to a group, they received a simulation script and directions to complete the MMAYSI-2. The simulation scripts were modeled after previous studies using simulation designs (e.g., Tombaugh, 2002). In the under-reporting simulator group, participants were instructed to under-report the presence of mental illness (Appendix D), whereas the over-reporting simulator group participants were instructed to feign mental illness (Appendix E). Participants in the control condition were instructed to respond honestly (see Appendix F). After they completed the MMAYSI-2 and Demographic Form, participants completed a manipulation check, which involved indicating which instructions they were given and stating whether or not they responded honestly, over-reported their symptoms, or under-reported their symptoms (see Appendix G). Out of the 197 participants, 178 passed the manipulation check. The 19 participants who did not pass the manipulation check were excluded from analyses.

**Results**

Prior to evaluating differences in responses across the three instruction conditions, an exploratory factor analysis (EFA) using principal axis factoring (PAF) with an oblique direct oblimin rotation was conducted on the 10 over-reporting items and the 10 under-reporting items that were written for this study (Conway & Huffcutt, 2003). A total of three factors emerged using Kaiser's criterion to retain factors with eigenvalues greater than 1 (Kaiser, 1961; see Table 3), but some researchers have stated that this method is problematic and have recommended using other criteria, such as examining the scree plot (e.g., Schultz & Whitney, 2005). Schultz and Whitney (2005)
stated that the number of factors could be determined by using the factor number
associated with the "elbow" (where the line in the plot flattens) minus one. As seen in
Figure 1, it is clear that there are two factors underlying the over-reporting and under-
reporting items. The total amount of variance accounted for in the EFA was 49.84% (see
Table 3). The factors are presented in order and are assigned a number label ordered by
decreased contribution to total variance. The factor loadings for the two factors are
presented in Table 4. Upon inspection of the factor loadings, there were two items that
should be monitored in future studies. One item, "Have you ever been mad but you
weren't sure why?" did not meet the criteria of having a factor loading equal to or
greater than .40, the cut point at which items are considered to load strongly onto factors
(Pallant, 2001). Another item, "Have you ever lied?" loaded more strongly on the over-
reporting scale rather than the under-reporting scale, for which it had been chosen. Given
that this was just the first level of item development, the two items were not discarded
and the proposed scales were analyzed with all 20 items. It was decided to continue to
include these two items to see how they hold up during the next levels of item
development; however, they should continue to be monitored closely.

The mean scores of participants who completed the MMAYSI-2 in the under-
reporting, over-reporting and honestly responding conditions were examined using one-
way between-groups analysis of variance (ANOVA) on the Under-Reporting and Over-
Reporting scales. The means, standard deviations, and F value comparing the three
groups are presented in Table 5. As can be seen, there was a statistically significant
difference in the Under-Reporting scores for the three groups, \( F(2, 175) = 10.39, p = \)
.001. Post-hoc comparisons using the Tukey HSD test indicated that the mean score for the under-reporting group ($M = 13.10$, $SD = 2.81$) was significantly different from the over-reporting group ($M = 11.82$, $SD = 1.85$) and the control group ($M = 11.43$, $SD = 1.53$); the means of the over-reporting and the control group did not differ significantly.

Similarly, there was a statistically significant difference in the Over-Reporting scores across the three groups, $F(2, 175) = 122.34$, $p = .001$. Post-hoc comparisons using the Tukey HSD test indicated that the mean score for the over-reporting group ($M = 15.69$, $SD = 3.41$) was significantly higher than the under-reporting group ($M = 10.52$, $SD = 1.00$) and the control group ($M = 10.54$, $SD = .90$); the under-reporting and the control group scores did not differ significantly from each other.

Lastly, we examined the mean scores of the MMAYSI-2 clinical scales across the three instruction groups using a MANOVA with a Bonferroni adjusted alpha level of .006. The decision to use a MANOVA, rather than several one-way ANOVAs, was due to the fact that the correlations among the MAYSI-2 clinical scale scores are generally moderate. There was a significant effect, $F(16, 336) = 19.43$, $p = .001$; Wilks’ Lambda = .27, partial eta squared = .48. The means, standard deviations and $F$ values for the ANOVAs for each scale are presented in Table 6. Each clinical scale score differed significantly across the three instruction groups, with the over-reporting group obtaining significantly higher scores (endorsing significantly greater difficulties) than both the under-reporting group and the control group. In addition, the under-reporting group obtained significantly lower scores (endorsing much less pathology) than the control group for every clinical scale except the thought disturbance, female traumatic
experiences, and male traumatic experiences scales. These differences indicate that the groups generally responded in a manner that is consistent with the instructions provided.

Discussion

The purpose of this study was to create scales for the MAYSI-2 that would identify response style, with the hope of increasing the utility of the MAYSI-2 by including measurement of the respondent’s approach to the measure. A total of 20 items were incorporated into the MAYSI-2, with 10 items assessing under-reporting and 10 items assessing over-reporting, creating the MMAYSI-2. This study used an experimental design with a simulation methodology, which has been widely used in the development of measures to assess response style (e.g., Rogers, 1988; Tombaugh, 2002), and recommended.

Results from the analyses supported the over- and under-reporting items as a means to identify those response styles. Using factor analysis to examine the items, there were two interpretable factors, one that represented all of the over-reporting items and one that represented all of the under-reporting items. There were two items that should continue to be monitored in future studies; one item loaded just below the recommended cut off point and one item loaded more strongly on the over-reporting scale rather than its intended under-reporting scale. Similarly, the results of the one-way ANOVA showed that the participants assigned to the over-reporting group endorsed significantly more items on the over-reporting scale compared to the control and under-reporting groups, whereas the participants assigned to the under-reporting group endorsed significantly more items on the under-reporting scale compared to the control and over-reporting
DETECTING OVER- AND UNDER-REPORTING

groups. Together, this provides support that the new items measured two unified but
distinct features, and that they correctly identified participants who were instructed to
over-report and under-report symptoms.

Lastly, the MANOVA that was performed on the clinical scales of the MMAYSI-2 showed that the clinical scales accurately reflected each group's response style,
meaning that the over-reporting group had significantly higher scores on the clinical
scales and the under-reporting group had significantly lower scores. This provides
further support that the newly developed scales validly measure over- and under-
reporting response styles.

These results are consistent with other research that has been conducted on
simulation designs and scale development. For example, a study on the Personality
Inventory for Youth (PIY) Validity Scales that used a simulation design found that the
control group (psychiatric inpatients) had statistically significantly elevated clinical
scales and a non-elevated defensiveness scale compared to the fake-good group, who
showed non-elevated clinical scales and a statistically significantly elevated
defensiveness scale (Wrobel et al., 1999). Slaughter et al. (2004) used EFA as part of the
development of a scale to measure organization personality that resulted in five distinct
factors, eliminating factors that had cross-loadings greater than .40 or that did not have
loadings of at least .40 on any factors with minimum eigenvalues of 1.0. Their study also
showed that the measure was sensitive to experimental manipulation; participants
randomly assigned to read one of five articles about an organization reflecting high
levels of a single personality dimension rated that organization as reflecting that single
DETECTING OVER- AND UNDER-REPORTING

personality dimension. In other words, the ratings corresponded to the experimental manipulation of personality in the articles.

Limitations

Although these results show great promise for the newly developed scales, there were several limitations with the current study. One limitation is that, due to the sample size of the current study, we were not able to conduct a full factor analysis using all 72 items. Since we were only able to conduct the factor analysis on the 20 newly developed items, an important next step in the development of the adapted MAYS1-2 is to conduct the full 72-item factor analysis to make sure the factor loadings for this sample are similar to those found in the original standardization (Archer et al., 2010) and subsequent norming of the MAYS1-2 (Meyers, 2006). Another limitation of the current study is that a convenience sample of college students was utilized for this initial step of scale development. It is important to note, however, that since a simulation design is a recommended step in scale development (TOMM; Tombaugh, 2002; Rogers et al., 2010), it would have been difficult to carry this out with juveniles in a detention center. Since college students were used, the study was limited to 17- to 19-year-olds, which is slightly older than the normative age for the MAYS1-2. In addition, most of the participants in this study were Caucasian females, which does not reflect the general population of the local juvenile detention facility, where offenders are largely African American and male. Even though the developed items significantly discriminated the simulation groups, an important next step could be to use another simulation design with
DETECTING OVER- AND UNDER-REPORTING

adolescents within the normative age-range before testing the items on juvenile offenders.

Implications

The findings from the present study support the utility and predictive validity of the MMAYSI-2 with 17-19 year old college students, who were primarily Caucasian females. Given that the MMAYSI-2 was developed for eventual use by juveniles in detention facilities, additional research is recommended. Other possibilities of future studies include implementing the next level of study design that is typically used when validating a scale. According to Tombaugh (2002), the next design level that is used has participants complete a measure twice, once responding how they typically would and once either over-under reporting. Participants’ honest responses are then compared with their over-under reporting responses to evaluate the effects of test sophistication on its ability to discriminate between simulating and control participants. Given this next design level recommendation, a potential next step would be to have college students complete the MMAYSI-2 responding honestly and completing it according to instructions to over-under report. Participants’ honest responses would then be compared with their over or under reporting responses to determine if the added items could detect the simulation responses as another means to examine discriminant validity. Once it is determined that the above design accurately discriminates between simulators and controls, the third design level could be used. The third level design embeds the malingering test into a battery of other similar tests. Because the participant has the opportunity to compare the malingering test with other tests measuring the same
References


DETECTING OVER- AND UNDER-REPORTING


DETECTING OVER- AND UNDER-REPORTING


DETECTING OVER- AND UNDER-REPORTING


### Table 1

*Descriptive Demographics for the Overall Sample and for Each Randomization Group*

<table>
<thead>
<tr>
<th></th>
<th>Overall Sample (n = 178)</th>
<th>Control (n = 69)</th>
<th>Overreport (n = 51)</th>
<th>Underreport (n = 58)</th>
<th>F/X²</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Age</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Mean</td>
<td>18.71</td>
<td>18.70</td>
<td>18.71</td>
<td>18.74</td>
<td>.16 (ns)</td>
</tr>
<tr>
<td>SD</td>
<td>0.47</td>
<td>0.49</td>
<td>0.46</td>
<td>0.44</td>
<td></td>
</tr>
<tr>
<td><strong>Sex</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Male</td>
<td>31.5%</td>
<td>29.0%</td>
<td>33.3%</td>
<td>32.8%</td>
<td>.32 (ns)</td>
</tr>
<tr>
<td>Female</td>
<td>68.5%</td>
<td>71.0%</td>
<td>66.7%</td>
<td>67.2%</td>
<td></td>
</tr>
<tr>
<td><strong>Race</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Caucasian</td>
<td>77.5%</td>
<td>78.3%</td>
<td>80.4%</td>
<td>74.1%</td>
<td>5.24 (ns)</td>
</tr>
<tr>
<td>African American</td>
<td>8.4%</td>
<td>8.7%</td>
<td>9.8%</td>
<td>6.9%</td>
<td></td>
</tr>
<tr>
<td>Asian/Pacific</td>
<td>4.5%</td>
<td>5.8%</td>
<td>2.0%</td>
<td>5.2%</td>
<td></td>
</tr>
<tr>
<td>Hispanic</td>
<td>2.8%</td>
<td>2.9%</td>
<td>0.0%</td>
<td>5.2%</td>
<td></td>
</tr>
<tr>
<td>Other/Multi</td>
<td>3.9%</td>
<td>2.9%</td>
<td>3.9%</td>
<td>5.2%</td>
<td></td>
</tr>
<tr>
<td>Declined</td>
<td>2.8%</td>
<td>1.4%</td>
<td>3.9%</td>
<td>3.4%</td>
<td></td>
</tr>
</tbody>
</table>
Table 2

Validity Items Developed for the MMAYSI-2

Under-Reporting Items:

1. Have you ever lied?*
2. Have you ever put off doing important jobs or tasks?*
3. Have you ever felt sad?*
4. Have you ever kept money that you found?*
5. Have you ever talked about others behind their backs?*
6. Have you ever argued with anyone in your family?*
7. Have you ever felt like swearing?*
8. Have you ever felt sad when someone criticized you?*
9. Have you ever been concerned about how other people think about you?*
10. Have you ever been mad, but you weren’t sure why?*

Over-Reporting Items:

1. Have you noticed that you have to hold your breath when you pass through doorways?
2. Have you noticed a strange color when you look at the moon?
3. Have you ever seen animals that others cannot see?
4. Have you ever felt that everything smells the same?
5. Have you ever been afraid of using the stove or anything very hot?
6. Have you ever felt that other people insert thoughts in your head?
7. Have you ever felt there are times when you were under a spell?
8. Have you ever noticed that you get angry when you are complimented?
9. Have you ever felt that someone was trying to take credit for your innermost thoughts?
10. Have you ever noticed that when you are around others, you are troubled by seeing weird things?

Note. In the MMAYSI-2, all items that are answered “Yes” get one point; for these items, a “Yes” indicates over- or under-reporting. The starred items are reverse-scored; answering “N” indicates under-reporting.
Table 3

*Extracted Factors and Eigenvalues from the 20 Items Added to the MMAYSI-2*

<table>
<thead>
<tr>
<th>Factors</th>
<th>Eigen Values</th>
<th>% of Variance</th>
<th>Cumulative %</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>6.50</td>
<td>32.52</td>
<td>32.52</td>
</tr>
<tr>
<td>2</td>
<td>3.46</td>
<td>17.32</td>
<td>49.84</td>
</tr>
<tr>
<td>3</td>
<td>1.05</td>
<td>5.25</td>
<td>55.08</td>
</tr>
</tbody>
</table>
Table 4

*Added MMAYSI-2 Item Loadings for Three Factors Identified through Exploratory Factor Analysis*

<table>
<thead>
<tr>
<th>Under-Reporting Items</th>
<th>Factor 1</th>
<th>Factor 2</th>
<th>Factor 3</th>
</tr>
</thead>
<tbody>
<tr>
<td>1.</td>
<td>.468</td>
<td>.429</td>
<td>.114</td>
</tr>
<tr>
<td>2.</td>
<td>-2.52</td>
<td>.499</td>
<td>-.112</td>
</tr>
<tr>
<td>3.</td>
<td>.182</td>
<td>.553</td>
<td>.007</td>
</tr>
<tr>
<td>4.</td>
<td>-.057</td>
<td>.417</td>
<td>.150</td>
</tr>
<tr>
<td>5.</td>
<td>.063</td>
<td>.610</td>
<td>-.036</td>
</tr>
<tr>
<td>6.</td>
<td>.055</td>
<td>.565</td>
<td>.063</td>
</tr>
<tr>
<td>7.</td>
<td>-.100</td>
<td>.590</td>
<td>.040</td>
</tr>
<tr>
<td>8.</td>
<td>-.053</td>
<td>.602</td>
<td>-.093</td>
</tr>
<tr>
<td>9.</td>
<td>.011</td>
<td>.564</td>
<td>-.100</td>
</tr>
<tr>
<td>10.</td>
<td>-.385</td>
<td>.399</td>
<td>-.033</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Over-Reporting Items</th>
<th>Factor 1</th>
<th>Factor 2</th>
<th>Factor 3</th>
</tr>
</thead>
<tbody>
<tr>
<td>1.</td>
<td>.730</td>
<td>.084</td>
<td>-.256</td>
</tr>
<tr>
<td>2.</td>
<td>.630</td>
<td>.024</td>
<td>-.592</td>
</tr>
<tr>
<td>3.</td>
<td>.850</td>
<td>.081</td>
<td>.035</td>
</tr>
<tr>
<td>4.</td>
<td>.716</td>
<td>-.007</td>
<td>-.129</td>
</tr>
<tr>
<td>5.</td>
<td>.521</td>
<td>-.083</td>
<td>.015</td>
</tr>
</tbody>
</table>
### Table 4 Continued

<p>| | | | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>6.</td>
<td>.660</td>
<td>-.086</td>
<td>.270</td>
</tr>
<tr>
<td>7.</td>
<td>.893</td>
<td>.002</td>
<td>.067</td>
</tr>
<tr>
<td>8.</td>
<td>.699</td>
<td>-.069</td>
<td>-.142</td>
</tr>
<tr>
<td>9.</td>
<td>.773</td>
<td>-.057</td>
<td>-.057</td>
</tr>
<tr>
<td>10.</td>
<td>.869</td>
<td>.062</td>
<td>.068</td>
</tr>
</tbody>
</table>

*Note.* Highest Loading for Each Item are Presented in Bold Typeface. Item numbers (left column) reflect the items as they are presented in Table 2.
Table 5

*Means, Standard Deviations, and One Way ANOVA Results for the Under-Reporting and Over-Reporting Scales Across the Three Instruction Conditions.*

<table>
<thead>
<tr>
<th></th>
<th>Control</th>
<th>Over-reporting</th>
<th>Under-reporting</th>
<th>F</th>
<th>(p)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>M (SD)</td>
<td>M (SD)</td>
<td>M (SD)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Under-Reporting</td>
<td>11.43 (.52)</td>
<td>11.82 (1.85)</td>
<td>13.10 (2.81)</td>
<td>10.39*</td>
<td></td>
</tr>
<tr>
<td>Scale</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Over-Reporting</td>
<td>10.54 (.90)</td>
<td>15.68 (3.41)</td>
<td>10.51 (1.00)</td>
<td>122.34*</td>
<td></td>
</tr>
<tr>
<td>Scale</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Note: * p = .001
Table 6

Means, Standard Deviations, and F-Values for Instruction Groups on Each MMAYSI-2 Scale in a Multivariate Analysis of Variance

<table>
<thead>
<tr>
<th>Scale</th>
<th>Control M (SD)</th>
<th>Overreport M (SD)</th>
<th>Underreport M (SD)</th>
<th>F</th>
</tr>
</thead>
<tbody>
<tr>
<td>AD</td>
<td>9.54 (1.98)</td>
<td>11.47 (3.17)</td>
<td>9.10 (2.01)</td>
<td>14.97</td>
</tr>
<tr>
<td>AI</td>
<td>12.49 (2.11)</td>
<td>15.86 (2.43)</td>
<td>10.81 (2.00)</td>
<td>75.44</td>
</tr>
<tr>
<td>DA</td>
<td>11.70 (2.26)</td>
<td>15.75 (2.37)</td>
<td>10.47 (1.55)</td>
<td>94.23</td>
</tr>
<tr>
<td>SC</td>
<td>9.55 (1.75)</td>
<td>11.04 (1.70)</td>
<td>9.07 (1.93)</td>
<td>17.59</td>
</tr>
<tr>
<td>SI</td>
<td>5.59 (1.14)</td>
<td>8.10 (2.11)</td>
<td>5.07 (0.26)</td>
<td>78.54</td>
</tr>
<tr>
<td>TD&lt;sup&gt;a&lt;/sup&gt;</td>
<td>5.60 (0.94)</td>
<td>8.53 (1.62)</td>
<td>5.26 (0.45)</td>
<td></td>
</tr>
<tr>
<td>TEG&lt;sup&gt;b&lt;/sup&gt;</td>
<td>6.45 (1.28)</td>
<td>8.21 (1.34)</td>
<td>6.46 (1.17)</td>
<td></td>
</tr>
<tr>
<td>TEB&lt;sup&gt;a&lt;/sup&gt;</td>
<td>7.45 (1.19)</td>
<td>9.76 (2.02)</td>
<td>7.68 (1.45)</td>
<td></td>
</tr>
</tbody>
</table>

Note. AD=Alcohol/Drug Use, AI=Angry/Irritable, DA=Depressed/Anxious, SC=Somatic Complaints, SI=Suicide Ideation, TD=Thought Disturbance, TEG=Traumatic Experiences Girls, TEB=Traumatic Experiences Boys

All comparisons are significant at a .01 or greater; bolded means are significantly different.

<sup>a</sup>This scale only for boys; total n= 56; Control n=20; Overreport n=17; Underreport n=19

<sup>b</sup>This scale only for girls; total n=122; Control n=49; Overreport n=34; Underreport n=39
Figure 1

Scree Plot

Eigenvalue

Factor Number
Appendix A

Demographic Questionnaire

Sex: ___________

Age: ___________

Year in college: _____________________

Major: _____________________________

Have you ever taken or are you currently taking Abnormal Psychology? ___Yes ___No

Ethnicity: ___Caucasian ___African American ___Asian ___Hispanic

Other (please specify) _____________________________

Prefer not to answer ______
Appendix B

MMAYSI-2 QUESTIONNAIRE

Name_________________________ Male____ Female____
Date of Birth___________________ Today’s Date____________

These are some questions about things that sometime happen to people. For each question, please circle YES or NO to answer whether that question has been true for you IN THE PAST FEW MONTHS. Please answer these questions as well as you can.

2. Have you ever lied?*
   Y  N

5. Have you noticed that you have to hold your breath when you pass through doorways?
   Y  N

8. Have you ever put off doing important jobs or tasks?*
   Y  N

11. Have you noticed a strange color when you look at the moon?
    Y  N

14. Have you ever felt sad?*
    Y  N

17. Have you ever seen animals that other people cannot see?
    Y  N

20. Have you ever kept money that you found?*
    Y  N

23. Have you ever noticed that everything smells the same?
    Y  N

26. Have you ever talked about others behind their back?*
    Y  N

29. Have you ever been afraid of using the stove or anything very hot?
    Y  N
32. Have you ever argued with anyone in your family?*  
   Y  N

36. Have you ever felt that other people insert thoughts in your head?  
   Y  N

39. Have you ever felt like swearing?*  
   Y  N

45. Have you ever felt there are times when you were under a spell?  
   Y  N

48. Have you ever felt sad when someone criticized you?*  
   Y  N

51. Have you ever noticed that you get angry when you are complimented?  
   Y  N

54. Have you ever been concerned about how other people think about you?*  
   Y  N

57. Have you ever felt that someone was trying to take credit for your innermost thoughts?  
   Y  N

60. Have you ever been mad, but you weren't sure why?*  
   Y  N

63. Have you ever noticed that when you are around others, you are troubled by seeing weird things?  
   Y  N

* Note: The starred items are reverse-scored; answering “N” indicates under-reporting.

The MAYS I-2 is a copyrighted measure and is not reproduced here. The items listed above were developed by the author and her dissertation advisor; the numbering of the items reflects their position in the Modified MAYS I-2 constructed for this study. Information about the MAYS I-2, including purchasing, can be found at http://nysap.us/MAYS2.html
Appendix C

IRB Approval
February 6, 2012

Nicole Bosse
11787 Thistlehill Dr.
Loveland, OH 45140

Dear Ms. Bosse:

Re: Protocol #1139, Detecting Over-and-Under Reporting of Symptoms on the MAYSII-III:
Development of a Validity Scale

The IRB has reviewed the revised materials regarding your study, referenced above, and has
determined that it meets the criteria for the Exempt from Review category under Federal
Regulation 45CFR46. Your protocol is approved as exempt research, and therefore requires no
further oversight by the IRB.

If you wish to modify your study, including the addition of data collection sites, it will be
necessary to obtain IRB approval prior to implementing the modification. If any adverse events
occur, please notify the IRB immediately.

Please contact our office if you have any questions. We wish you success with your project!

Sincerely,

[Signature]

Morell E. Mullins, Jr., Ph.D.
Chair, Institutional Review Board
Xavier University

MEM/sb

c: Kathy Hurt, advisor
DETECTING OVER- AND UNDER-REPORTING

Appendix D

Script for Under-Reporting

I want you to pretend that you are in the process of being interviewed for a job. As part of the interview process, you are instructed to take a self-report measure that screens for mental illness. You really want and need this job, and in order to obtain it you need to present yourself in the best light possible. Complete the attached measure in a way that you think will give you a clean bill of mental health.
DETECTING OVER- AND UNDER-REPORTING

Appendix E

Script for Over-Reporting

I want you to pretend that you are in danger of being expelled from the University because of failing grades. You have learned that the University will allow you to remain at school if you are seriously mentally ill and possibly in need of hospitalization. The measure you are taking is a screening measure to detect mental illness and doctors use it to determine what sort of symptoms individuals have. You are taking this measure as part of the review process at the University regarding your status. Keep in mind that it is very important that you fool the doctors into thinking you are too mentally ill to be expelled.
DETECTING OVER- AND UNDER-REPORTING

Appendix F

Script for Control Group

The measure you are taking is a screening measure to detect mental illness. Please fill this out in a way that reflects your current emotional status and experiences.
What were the instructions that you were given prior to completing the questionnaire?

Did you respond to the questions honestly?

Did you respond to the questions trying to portray a mentally ill person?

Did you respond to the questions trying to portray yourself in the best light possible?
Title: Detecting Over- and Under-reporting of Symptoms on the MAYS1-2: Development of a Validity Scale

Problem: The MAYS1-2 (Grisso & Barnum, 2003) has well established, strong psychometric properties (Archer, Simonds-Bisbee, Spiegel, Handel, & Elkins, 2010; Butler, Loney, & Kistner, 2007; Ford, Chapman, Pearson, Borum, & Wolpaw, 2008; Grisso & Barnum, 2003; Kerg, Arnzen-Moedell, & Becker, 2011; Lexcen, Vincent, & Grisso, 2004; Meyers, 2006; Wasserman, Vilhauer, McReynolds, Shoai, & John, 2005) and has been shown to identify youths with psychiatric symptoms and/or youths in crisis who may require immediate and more detailed evaluation and intervention. However, a major limitation of the MAYS1-2 is that it lacks a measure of validity or response style that would detect whether or not the youth’s self-report is a reflection of his or her true functioning (Grisso & Barnum, 2003; Meyers, 2006). Therefore, the aim of the present study was to address this shortcoming by developing and evaluating items to be embedded into the MAYS1-2 to provide information about possible over-or under-reporting of symptoms.

Method: The MAYS1-2 was adapted by adding 20 items to comprise two validity scales, creating the Modified MAYS1-2 (MMAYS1-2); 10 of the items consisted of questions designed to assess over-reporting of mental illness symptoms, whereas 10 other items were designed to assess under-reporting. The added items were modeled after well-known tests specifically measuring over-under-reporting. A handful of mental health professionals ranked 40 items designed to assess over and under-reporting, the 20 added
items were among the highest ranked. A sample of 178 college students met study inclusion criteria and completed study measures after being recruited through a psychology department participant pool at a private university in the Midwest. The students ranged in age from 17 to 19 years old with a mean age of 18.71 (SD = .47); the majority of the sample were Caucasian (77.5%) and women (68.5%). Using a simulation design modeled after previous, similar studies (e.g., Tombaugh, 2002), students were randomly assigned to one of three instruction groups for responding to the MMAYSI-2: under-report the presence of mental illness, over-report (or feign) mental illness, or respond honestly. Participants filled out a demographic questionnaire and the MMAYSI-2 online via Survey Gizmo.

Findings: Prior to evaluating differences in responses across the three instruction conditions, an exploratory factor analysis (EFA) using principal axis factoring (PAF) with an oblique direct oblimin rotation was conducted on the 10 over-reporting items and the 10 under-reporting items that were written for this study (Conway & Huffcutt, 2003). Using EFA, a total of 2 factors emerged after examining the Scree Plot (Shultz & Whitney, 2005). The amount of variance accounted for in the EFA was 49.84%.

One-way analysis of variance (ANOVA) comparing the three instruction conditions on the Under-Reporting and Over-Reporting scales revealed a statistically significant difference in the Under-Reporting scores for the three groups, \( F(2, 175) = 10.39, p = .001 \). Post-hoc comparisons using the Tukey HSD test indicated that the mean score for the under-reporting group (\( M = 13.10, SD = 2.81 \)) was significantly different from the over-reporting group (\( M = 11.82, SD = 1.85 \)) and the control group (\( M = 11.43, \).
DETECTING OVER- AND UNDER-REPORTING

$SD = 1.53$; the means of the over-reporting and the control group did not differ significantly. Similarly, there was a statistically significant difference in the Over-Reporting scores across the three groups, $F(2, 175) = 122.34$, $p = .001$. Post-hoc comparisons using the Tukey HSD test indicated that the mean score for the Over-Reporting group ($M = 15.69$, $SD = 3.41$) was significantly higher than the under-reporting ($M = 10.52$, $SD = 1.00$) and the control groups ($M = 10.54$, $SD = .90$), which did not differ significantly from each other.

Lastly, we examined the mean scores of the MMAYSI-2 clinical scales across the three instruction groups using MANOVA with a Bonferroni adjusted alpha level of .006. There was a significant effect, $F(16, 336) = 19.43$, $p = .001$; Wilks’ Lambda = .27, partial eta squared = .48. Each clinical scale score differed significantly across the three instruction groups, with the over-reporting group obtaining significantly higher scores (endorsing significantly greater difficulties) than both the under-reporting group and the control group. In addition, the under-reporting group obtained significantly lower scores (endorsing much less pathology) than the control group for every clinical scale except the thought disturbance, female traumatic experiences, and male traumatic experiences scales. These differences indicate that the groups generally responded in a manner that is consistent with the instructions provided.

*Implications:* Results from the analyses supported the over- and under-reporting items as a means to identify those response styles. Similarly, the results of the one-way (ANOVA) showed that the participants assigned to the over-reporting group endorsed significantly more items on the over-reporting scale compared to the control and under-
reporting groups, whereas the participants assigned to the under-reporting group endorsed significantly more items on the under-reporting scale compared to the control and over-reporting groups. Together, this provides support that the new items measured two unified but distinct features, and that they correctly identified participants who were instructed to over-report and under-report symptoms. Lastly, the MANOVA that was performed on the clinical scales of the MMAYSI-2 showed that the clinical scales accurately reflected each group's response style, meaning that the over-reporting group had significantly higher scores on the clinical scales and the under-reporting group had significantly lower scores. This provides further support that the newly developed scales validly measure over- and under-reporting response styles.

The findings from the present study support the utility and predictive validity of the adapted MMAYSI-2 with 17-19 year old college students, who were primarily Caucasian females. Given that the MMAYSI-2 was developed for eventual use by juveniles in detention facilities, additional study is recommended. Other possibilities of future studies include implementing the next level of study design that is typically used when validating a scale. A potential next step would be to have participants complete the MMAYSI-2 responding honestly and completing it according to instructions to over-under report.