



SCHOOL of
GRADUATE STUDIES
EAST TENNESSEE STATE UNIVERSITY

East Tennessee State University
**Digital Commons @ East
Tennessee State University**

Electronic Theses and Dissertations

5-2008

Assessing the Reliability of Scores Produced by the Substance Abuse Subtle Screening Inventory (SASSI).

Joshua A. Woodson
East Tennessee State University

Follow this and additional works at: <http://dc.etsu.edu/etd>

Recommended Citation

Woodson, Joshua A., "Assessing the Reliability of Scores Produced by the Substance Abuse Subtle Screening Inventory (SASSI)." (2008). *Electronic Theses and Dissertations*. Paper 1945. <http://dc.etsu.edu/etd/1945>

This Thesis - Open Access is brought to you for free and open access by Digital Commons @ East Tennessee State University. It has been accepted for inclusion in Electronic Theses and Dissertations by an authorized administrator of Digital Commons @ East Tennessee State University. For more information, please contact dcadmin@etsu.edu.

Assessing the Reliability of Scores Produced by the Substance Abuse Subtle Screening Inventory
(SASSI)

A thesis
presented to
the faculty of the Department of Psychology
East Tennessee State University

In partial fulfillment
of the requirements for the degree
Master of Arts in Clinical Psychology

by
Joshua Woodson
May, 2008

Alan L. Shields, Ph.D., Chair

Jon B. Ellis, Ph.D.

Wallace E. Dixon, Ph. D

Keywords: Reliability Generalization, SASSI, Substance Abuse, Screening Instruments

ABSTRACT

Assessing the Reliability of Scores Produced by the Substance Abuse Subtle Screening Inventory

(SASSI)

by

Joshua Woodson

The fundamental principle that reliability is a property of scores and not of instruments provides the foundation of a meta-analytic technique called reliability generalization (RG). RG studies characterize the reliability of scores generated by a given instrument and identify methodological and sample characteristics that contribute to the variability in the reliability of those scores. The present study is an RG of the Substance Abuse Subtle Screening Inventory (SASSI). Reliability estimates were obtained from 19.8% of studies using the SASSI. Bivariate correlations revealed strong, positive correlations between SASSI score reliability and score variability of the Subtle Attributes ($r = .877, p < .05$) and Family History ($r = .892, p < .05$) subscales and between score reliability and ethnicity for both the Family History ($r = .683, p < .05$) and Tendency to Involvement in Correctional Setting ($r = .76, p < .05$) subscales.

DEDICATION

I would like to dedicate this thesis project to my mother, Lisa Valentine, for instilling the belief in me that education is the foundation of life and that I can accomplish anything, for which I strive. Additionally, I would like to dedicate this project to my family, especially my grandparents, David and Loretta Roberts, who continued to believe in and support me when I made it more difficult than it ever should have been. I am forever grateful.

CONTENTS

	Page
ABSTRACT	2
DEDICATION	3
LIST OF TABLES	7
Chapter	
1. INTRODUCTION.....	9
Screening Instruments	10
Development of the SASSI	11
Development of the SASSI-2.....	13
Development of the SASSI-3	14
Development of SASSI-A.....	14
Development of SASSI-A2.....	15
Validation of the SASSI	16
Controversy Surrounding the Use of the SASSI.....	18
Discrepancies in the Use of the SASSI for Various Populations.....	19
Evaluating Measures.....	20
Reliability.....	20
Reliability Defined and Interpreted.....	20
Reliability Indices.....	22
Inter-Rater Reliability.....	22
Test-Retest Reliability.....	22
Parallel Forms Reliability.....	23

Internal Consistency Reliability.....	23
The Importance of Reliability.....	24
Reliability as a Property of Scores.....	24
Reporting Reliability.....	25
Reliability Reporting: The State of the Art.....	26
Reliability Generalization.....	29
A Review of the Applications.....	30
Purpose of Study.....	32
2. METHOD.....	33
The SASSI.....	33
Study Procedures.....	33
Reliability Induction and the SASSI.....	33
Reliability Generalization and the SASSI.....	34
Data Collection.....	34
Criteria for Including Studies in Data Analysis.....	35
Criteria for Excluding Studies from Data Analysis.....	35
Categorizing Studies.....	35
Criterion or Dependent Variable of Interest.....	36
Predictor Variables.....	36
Data Analysis.....	36
3. RESULTS.....	38
4. DISCUSSION.....	42
Reliability Induction and the SASSI.....	42

Reliability Generalization and the SASSI.....	44
Study Limitations.....	46
Recommendations.....	47
REFERENCES	48
APPENDIXES.....	61
Table 1. Scales Found on Differing Versions of the SASSI.....	61
Table 2. Reliability Reporting Across SASSI Subscales.....	63
Table 3. Descriptive Statistics for Reliability Coefficients of SASSI Total and Subscale Scores.....	64
Table 4. Descriptive Statistics for FVA Reliability Coefficients and Predictor Variables.....	65
Table 5. Descriptive Statistics for FVA Reliability Coefficients and Predictor Variables.....	66
Table 6. Descriptive Statistics for FVA Reliability Coefficients and Predictor Variables.....	67
Table 7. Descriptive Statistics for FVA Reliability Coefficients and Predictor Variables.....	68
Table 8. Descriptive Statistics for FVA Reliability Coefficients and Predictor Variables.....	69
Table 9. Descriptive Statistics for FVA Reliability Coefficients and Predictor Variables.....	70
Table 10. Descriptive Statistics for FVA Reliability Coefficients and Predictor Variables.....	71

Table 11. Descriptive Statistics for FVA Reliability Coefficients and Predictor	
Variables.....	72
Table 12. Descriptive Statistics for FVA Reliability Coefficients and Predictor	
Variables.....	73
VITA.....	74

LIST OF TABLES

Table	Page
1. Scales Found on Differing Versions of the SASSI.....	61
2. Reliability Reporting Across SASSI Subscales.....	63
3. Descriptive Statistics for Reliability Coefficients of SASSI Total and Subscale Scores.....	64
4. Descriptive Statistics for FVA Reliability Coefficients and Predictor Variables.....	65
5. Descriptive Statistics for FOD Reliability Coefficients and Predictor Variables.....	66
6. Descriptive Statistics for SYM Reliability Coefficients and Predictor Variables.....	67
7. Descriptive Statistics for OAT Reliability Coefficients and Predictor Variables.....	68
8. Descriptive Statistics for SAT Reliability Coefficients and Predictor Variables.....	69
9. Descriptive Statistics for DEF Reliability Coefficients and Predictor Variables.....	70
10. Descriptive Statistics for SAM Reliability Coefficients and Predictor Variables.....	71
11. Descriptive Statistics for FAM Reliability Coefficients and Predictor Variables.....	72
12. Descriptive Statistics for COR Reliability Coefficients and Predictor Variables.....	73

CHAPTER 1

INTRODUCTION

The *Diagnostic and Statistical Manual of Mental Disorders (4th ed. Revised*; American Psychiatric Association [APA], 2000) divides substance use disorders (SUDs) into two categories, dependence and abuse. Substance dependence is broadly characterized as “a cluster of cognitive, behavioral, and physiological symptoms indicating that the individual continues use of the substance despite significant substance-related problems” (APA, p.192). Substance abuse is broadly characterized as a “maladaptive pattern of substance use manifested by recurrent and significant adverse consequences related to the repeated use of substances” (p.198). Therefore, the primary distinction between dependence and abuse is dependence reflects a more severe substance use problem than abuse.

Substance use disorders plague the United States population and together form one of the most prevalent mental health disorders, with an estimated lifetime prevalence rate of 9.35% in the general population (Grant et al., 2004). In other words, approximately 1 of every 10 people in the U.S. will meet diagnostic criteria for a SUD at some point in their life. Furthermore, Gray (2001) states that due to the tendency of SUDs to mimic features of other disorders, failure to identify SUDs as a separate problem is a likely outcome. The consequence of not identifying SUDs as an independent disorder and the fact that SUDs frequently co-occur with other mental health problems (Gray et al., 2004) is that SUDs are under-identified and under-diagnosed in clinical settings.

Addiction research has examined multiple ways to identify substance use and abuse. However, due to negative repercussions that may occur from disclosing current substance use and substance use history, some individuals with an SUD may withhold information relating to

their substance use. For example, stricter punishments via the criminal justice system may discourage individuals from openly acknowledging substance use patterns and the consequences of their use. Accordingly, obtaining accurate information pertaining to individuals' substance use patterns can be a real challenge. As a result, effective screening instruments are needed to identify individuals with SUDs, especially among individuals motivated to minimize their substance use history.

Screening Instruments

Substance use screening instruments can be helpful in determining if an individual has a problem with substance use, is at risk for developing one, or is in need of further assessment. Screening tools, like many psychological assessment tools, are often developed with a strict adherence to scripted questions in order to improve the validity of scoring. The way an individual responds to the questions frequently indicate areas for further exploration. There have been various and ongoing attempts to design and develop sensitive and efficient screening tools to identify individuals with substance use problems. The resulting instruments can be divided into two types, direct and indirect (Miller, 1976). The direct instrument is one designed to identify SUDs through rationally derived (i.e., face valid) questions that reference particular signs and symptoms thought to be indicative of a substance abuse problem. By contrast, an indirect instrument uses empirically keyed items and factor analysis when deciding which questions should be included. That is, items are selected for inclusion because the designated population used to construct the measure tended to differentially endorse those items. The most widely recognized example of an indirect instrument is the Minnesota Multiphasic Personality Inventory (Butcher, Dahlstrom, Graham, Tellegen, & Kaemmer, 1989), which is one of the most commonly used screening instruments within the behavioral sciences. For individuals not

motivated to or not capable of honestly reporting substance use patterns, the indirect approach to screening may be necessary (Ingersoll, 2003).

Currently, a wide array of screening instruments is used to assess SUD. Some screening measures focus solely on alcohol use, such as the Alcohol Use Disorders Identification Test (AUDIT; Saunders, Aasland, Babor, De la Fuente, & Grant, 1993), the Michigan Alcohol Screening Test (MAST; Selzer, 1971), the CAGE questionnaire (Ewing, 1984), and the MacAndrew Alcoholism Scale-Revised (MAC-R; MacAndrew, 1965). Other screening instruments such as the Addiction Severity Index (ASI; McLellan, Luborsky, Woody, & O'Brien, 1980; McLellan et al., 1992), the Addiction Acknowledgement and Addiction Potential Scales (AAS; APS; Weed, Butcher, McKenna, & Ben-Porath, 1992), the Drug Abuse Screening Test (DAST; Skinner, 1982), and the Substance Abuse Subtle Screening Inventory (SASSI; Miller, 1985) are used primarily to identify problems with substance use more broadly.

Of these instruments, the AUDIT, MAST, CAGE, ASI, and DAST employ the direct approach. On the other hand, the SASSI, MAC-R, AAS, and APS take an indirect approach to screening for SUDs. As noted, the indirect measures are designed to identify individuals with alcohol and other drug problems who are unwilling or unable to acknowledge substance misuse or symptoms associated with misuse. In those instances where substance users are unable or unwilling to be forthcoming about their substance use, an indirect approach may be useful in identifying SUDs (Feldstein & Campbell, 2006; Ingersoll, 2003).

Development of SASSI

In 1970, Glenn A. Miller began work as a consultant for the first federally funded driving while intoxicated (DWI) program in the U. S. (Miller, 1985). As the number of DWI offenders increased, so did the need to develop an objective, easy to administer screening instrument to

identify individuals likely to have substance use problems. Accordingly, Miller began research in attempts to create a screening instrument to identify individuals in the court system who were likely to have substance use problems (i.e., abuse or dependence) so that they might be referred for appropriate treatment. Moreover, these individuals were believed to not be straightforward about their substance use; therefore Miller wanted to create an instrument to detect substance use problems in individuals whom underreport substance use. Lastly, Miller wanted to design an instrument that offered clear, objective decision rules for the classification of subjects and to be able to detect an SUD in its early stages (1985). After 15 years of clinical research, the Substance Abuse Subtle Screening Inventory (SASSI; Miller, 1985), an efficient, cost effective, and easy-to-administer tool to identify those with a substance dependence problem, was created.

The original SASSI consisted of 52 empirically derived true or false questions taken from a pool of over 1,000 questions. These 52 items were established through a series of systematic trials. These questionnaires were originally administered to 2 groups: (1) “Abusers”- a large group of outpatient substance abusers, and (2) “Non-Abusers”- non-abusing family members of the former group. Furthermore, the abusers were retested and directed to respond as though they were a non-abuser, or to “fake good.” Data obtained from these samples were used to construct the original three scales of the SASSI: (1) Obvious Attributes (OAT)-designed to best differentiate between abusers and non-abusers, (2) Subtle Attributes (SAT)-designed to best differentiate between non-abusers and faking good abusers, and (3) Denial (DEN; renamed Defensiveness, DEF)-designed to best differentiate faking good from both abusers and non-abusers (Miller, 1985). Separate norms for the instrument were established by gender.

An additional three scales were later added, although they provide little discriminative value: (1) Defensiveness (DEF2)-designed to further detect underreporting, (2) Alcohol and

Other Drug (ALD)-designed to distinguish between alcohol and drug abuse, and (3) Family History of Substance Abuse (FAM)-designed to assess family history of SUD (Miller, 1985). Furthermore, a set of Risk Prediction Scales (RPS) was created to accompany the six former scales and increase face validity. These Risk Prediction Scales were divided into 2 categories: (1) Risk Prediction Scale for Alcohol (RPSA) and (2) Risk Prediction Scale for Other Drug (RPSD). The Risk Prediction Scales are composed of 62 true or false questions and both the Risk Prediction Scale for Alcohol and Risk Prediction Scale for other Drug are composed of scaled items with the answers ranging from 0 (“never”) to 3 (“repeatedly”). The scales found on the SASSI, and the scales on each subsequent version, are summarized in Table 1.

Development of SASSI-2

The SASSI-2 was released as a modified version of the original SASSI (Miller, 1994). This instrument was designed to increase the probability of identifying underreported substance misuse and replaced the original SASSI. Four questions on the original SASSI were omitted and 14 new questions were added (Miller). Additionally, the Alcohol and Other Drug scale was removed, the Risk Prediction Scales were renamed Face Valid Alcohol (FVA) and Face Valid Other Drug (FVD), and the Defensiveness 2 scale was replaced by the Supplemental Addiction Scale (SAM; Miller, 1994). In contrast to the original Risk Prediction Scales, designed to identify symptoms and behaviors in the previous 6 months, the modified version focuses on lifetime incidents.

Two new scales were also added, Random Answering Pattern (RAP) and Tendency to Involvement in Correctional System (COR; Miller, 1994). The Random Answering Pattern was designed to detect a pattern of random answering by a client presenting himself or herself as a non-user and was taken from the adolescent version (i.e., SASSI-A). The Tendency to

Involvement in Correctional System scale was intended to recognize those individuals who may have a tendency to become involved with the criminal justice system.

Development of SASSI-3

Miller, Roberts, Brooks, and Lazowski revised the SASSI-2 and released the SASSI-3 (1997). Work on the SASSI-3 was undertaken to develop a version that was even more accurate than SASSI-2 in identifying SUDs. The SASSI-3 maintained most of the items within the SASSI-2, but some differences between the two exist. The biggest change was the addition of a new face valid or direct scale, Symptoms (SYM), designed to increase the accuracy of individuals with SUDs. The SASSI-2 contained many true or false items that were not scored but included as research items with the intent of increasing the accuracy of future versions of the SASSI. These items are now included as scored items because they did, in fact, increase the accuracy of identifying individuals likely to have problems with substance use (Miller et al.). When creating the new scale Symptoms, Miller et al., integrated the existing questions that directly assess substance misuse with these new, face valid true or false items. The goal of adding this new scale was an attempt to keep the structure of the SASSI-3 scales as clear as possible and leave the overall scale structure in tact. Miller et al. maintain that this new scale is useful for determining the extent to which the respondent had been a part of a family or social system that was heavily involved in drugs. Additionally, four questions were deleted and nine questions added to the SASSI-3. The main distinction, when scoring the SASSI-2 and SASSI-3 is the fewer numbers of questions found on each scale of the SASSI-3.

Development of SASSI-A

Twenty-five adolescent residential treatment programs was the normative sample in developing the SASSI-A, the adolescent version of the SASSI (Miller, 1990). One drawback to

the SASSI-A is that it only distinguishes between substance dependent persons and non-substance dependent persons (i.e., the SASSI-A does not distinguish between abuse and dependence). There are some differences between the SASSI-A and all three versions of the adult SASSI. The SASSI-A has seven more questions than all three adult versions of the SASSI even though it has fewer scales than both the SASSI-2 and SASSI-3. The scales found on the SASSI-A include the Obvious Attributes, Subtle Attributes, Defensiveness, Defensiveness2, Face Valid Alcohol, Face Valid Other Drug, Tendency to Involvement in Correctional System, and Random Answering Pattern scales (see Table 1).

Development of SASSI-A2

The Adolescent SASSI-A2 is composed of 40 of the true or false items found on the original SASSI-A, and as part of the ongoing validation process, an additional 32 new research items. Congruent with all editions of the SASSI, the SASSI-A2 includes subtle (i.e., indirect) items and face valid (i.e., direct) questions about substance use. Although they have been somewhat modified, all of the scales that were on the original Adolescent SASSI are included in the SASSI-A2 – Face Valid Alcohol, Face Valid Other Drug, Obvious Attributes, Subtle Attributes, Defensiveness, Supplemental Addiction Measures (replaces DEF2), and Tendency to Involvement in Correctional System. The SASSI-A2 also includes five new scales: (1) Family and Friends Risk Scale (FRISK) – designed to determine to what degree the adolescent is part of a family or social system likely to misuse substances, (2) Attitudes Toward Substance Use (ATT) – which assess the individuals attitudes and beliefs pertaining to substance use, (3) Symptoms of Substance Misuse (SYM) – which measures the consequences of out of control substance use, (4) Validity Check (VAL) – purports to identify individuals who may need further evaluation despite the SASSI-A2 indicating a low probability of a SUD, and (5) Secondary

Classification Scale (SCS) – designed to distinguish between substance abuse and substance dependence.

Validation of the SASSI

Test validation is the ongoing substantiation of a given instrument and can be achieved through a variety of procedures and unique research designs. After the objectives or measurement goals (e.g., detect substance use) of an instrument are specified and items are generated, that instrument is typically subjected to reliability and validity assessment (Lincoln, Liebschutz, Chernoff, Nguyen, & Amaro, 2006). Samples used for these validation procedures should be of adequate size and sufficiently representative to substantiate validity and reliability claims, establish appropriate norms, and support conclusions regarding the use of the instrument for the intended purpose (Lincoln et al.).

Three diverse populations were chosen to conduct initial reliability and validity studies of the SASSI. These groups consisted of an outpatient treatment group, patients in a detoxification program, and individuals who were on probation (Miller, 1985). The probation sample was drawn from three separate geographical regions. The reliability estimates from these original validation studies were not entirely encouraging. Cronbach's alpha coefficients were consistently below .60 on most scales for all three of these populations. However, the Risk Prediction Scales and Obvious Attributes scales were more promising. Cronbach's alpha coefficients for the Risk Prediction Scales were .90 and above, while the values for the Obvious Attributes ranged from .71 - .83. Subsequently, validity estimates for sensitivity (i.e., the ability of test scores to correctly identify an individual with a disorder as having the disorder) and specificity (i.e., the ability of test scores to correctly identify an individual without a disorder as not having the

disorder) were .88 and .92, respectively. However, this cannot necessarily be attributed to the performance of the entire instrument. Rather, it is likely these values were a product of the more reliable scores taken from the Risk Prediction Scales.

The SASSI has been subjected to a variety of other validation procedures and evidence on the usefulness of the instrument has emerged. For example, Teslak (2000) suggested the SASSI-3 is effective in identifying SUDs in individuals reluctant to admit they have a problem. Furthermore, Klein (2000) supported using the SASSI to identify malingerers, those individuals feigning an SUD or feigning not having an SUD. Moreover, Cooper and Robinson (1987) identified the SASSI as a promising device for discriminating between nonabusers, moderate abusers, and severe abusers. Likewise, support for using adult versions of the SASSI has been generated. Schmidt (2001) demonstrated the SASSI-2 is not only helpful in identifying dependence, but also in identifying recovery when the instrument is administered on two separate occasions. Furthermore, Horrigan and Piazza (1999) suggested that combining the SASSI with other self-report measures was the most effective way of identifying SUDs. Nevertheless, other studies suggested that only the face valid items were useful in determining whether or not an individual has a substance use problem (Bauman, Merta, & Steiner, 1999; Myerholtz & Rosenberg, 1998; Sweet & Saules, 2003; Teslak, 2000).

Many studies have validated the use of the SASSI-A with an adolescent population. (Fisher & Harrison, 1990; Fox, 1992; Hansink, 1992; Yeh, 1993). Subsequently, Coll, Juhnke, Thobro, and Haas (2003) concluded that the SASSI-A can be used to plan treatment and as a measure to evaluate treatment outcomes. Due to their findings that alcohol or other drug-abusing residents in a treatment facility exhibited a reduction in symptomatology, as measured by the SASSI-A after 6 months, Coll et al., suggest the SASSI-A is useful in evaluating treatment

outcomes. In addition, the SASSI-A has been found to detect substance abuse underreporting in adolescents (Hudson, 1997; Ingersoll, 2003; Pearson, 2000; Piazza, 1996).

Controversy Surrounding the Use of the SASSI

Recently, the adult versions of the SASSI have been criticized. For example, the SASSI's predictive power and predictive validity at rates as high as the developers obtained could not be replicated (Emanuelson, 2005). Furthermore, Clements (2002) raised the question of the effectiveness of the indirect scales of the SASSI-3, claiming that only the Face Valid Alcohol and Face Valid Other Drug scales were useful in recognizing SUDs. Likewise, Gray (2001) suggested only the direct scales of the SASSI are beneficial in classifying individuals with SUDs.

Based on false positives (i.e., identifies an individual as having a substance use problem when one does not exist) rates approaching 70%, Rogers, Cashel, Johansen, and Sewell, (1997) concluded the SASSI-A should not be used to classify adolescents as substance dependent or non-dependent. Although Bauman et al. (1999) demonstrated that the SASSI-A was useful in identifying groups at risk for SUD from non-risk groups, they also reported that SASSI-A classification did not coincide with DSM-IV diagnoses and furthermore, found the Defensiveness subscale to be a poor predictor of a diagnosis of SUD. Additionally, Sweet and Saules (2003) demonstrated the subtle scales of the SASSI-A are not good at identifying SUDs. Moreover, Nishimura et al., (2005) reported that the SASSI-A was not useful for identifying SUDs in a Hawaiian and Japanese population, and subsequently suggested modifying the SASSI-A for use in non-white populations.

Discrepancies in Use of the SASSI for Various Populations

The SASSI has been applied in a variety of settings. Hudson (1997) proposes using the SASSI in a family counseling agency, claiming it to be the most effective substance use problem screening device in that study. Swartz (1998) demonstrated the usefulness of using the SASSI-2 with the criminal justice population, where a quick, easily administered screening device was needed. Also, Horrigan and Piazza (1999) found the SASSI efficacious in identifying SUDs among pregnant women. In addition, Buzzanga (2000) recommends using the SASSI-2 with individuals who have suffered a head trauma. The SASSI-3 was also found useful for identifying SUDs in a population typically challenging to discern, the chronically mentally ill (Pearson, 2000). Lastly, the SASSI-3 was found to be effective in identifying SUDs in college students (Laux, Salyers, & Kotova, 2005).

Still, researchers have questioned the use of the SASSI for other samples of similar populations. Arenth, Bogner, Corrigan, and Schmidt (2001) compared the efficacy of the SASSI-3 to Blood Alcohol Level (BAL) and the diagnosis of staff psychologists for identifying SUDs in individuals suffering from a head injury and found lower accuracy, sensitivity, and specificity for individuals with traumatic brain injury. Teslak (2000) questioned the use of the SASSI among individuals with a diagnosable mental illness and attributes lower identification rates to the length of administration and scoring procedures. This claim is in direct opposition of the one noted above by Pearson (2000). In addition, the usefulness of the SASSI in the college student population has been disputed (Clements, 2000; Myerholtz & Rosenberg, 1998). Myerholtz and Rosenberg demonstrated that a sample of college students can manipulate the SASSI-2 when instructed to respond as faking good or faking bad. Although the SASSI holds promise as a clinically useful tool, more validation and evaluation studies are warranted.

Evaluating Measures

Scientific advancement relies heavily on the psychometric assessment of scientific observations (i.e., measurements and tests) and numerous techniques exist to aid researchers in making these observations (e.g., Cook & Campbell, 1979; Grimm & Yarnold, 1995, 2000; Kazdin, 1992; Keppel, 1991). Two primary psychometric properties that contribute to the theoretical and statistical adequacy of an instrument are reliability and validity. Broadly, reliability is defined as the extent to which scores generated by a test remain constant over various administrations, whereas validity is the extent to which those scores measure what an instrument is designed to evaluate. If a given instrument does not yield scores that are reliable and valid, the effectiveness of the instrument should be called into question. The present query focuses on the property of reliability; nonetheless, these two properties are intertwined.

Reliability

Although the property of reliability plays an essential role in the professional lives of psychologists, aspects of it may not be well understood (Li, Rosenthal, & Rubin, 1996). Indeed, a variety of researchers note that reliability is underreported in the behavioral sciences (e.g., Kieffer, Reese, & Thompson, 2001; Meier & Davis, 1990; Miller, Shields, Campfield, Wallace, & Weiss, *In Press*; Shields & Caruso, 2003, 2004; Vacha-Haase, Henson, & Caruso, 2002; Whittington, 1998). In order to clarify the often misunderstood concept of reliability, the following three sections will define and interpret reliability, describe its indices, and explain its importance.

Reliability Defined and Interpreted

A brief discussion of the components of a score produced by any given measure will aid in understanding reliability coefficients. First, however, it is important to distinguish a “true”

score from an “observed” score. According to classical measurement theory (Lord & Novick, 1968), a true score cannot be observed directly but instead is the theoretical average score a single respondent would get if he or she were observed an infinite number of times. An observed or obtained score, however, is the *actual* score a respondent gets on a given test administration. It is the observed score that one has access to and this score is divided into two parts: true score and error. Error is the component of the observed score that fluctuates due to random error sources. Test administration characteristics and respondent characteristics are examples of possible error sources. Thus,

$$\text{Observed score} = \text{true score} + \text{error}.$$

The reliability coefficient is the proportion of true variability to the total obtained variability. Therefore, a reliability coefficient of .85 roughly suggests that 85% of the variability in the observed scores could be said to represent true individual differences and 15% of the variability is due to random error.

Broadly, high reliability is observed when similar scores are attained through repeated administrations of the same test under varying assessment circumstances. Scores are considered reliable if differences caused by the conditions in which the measurements were gathered (i.e., the variability) are not present or not significant. Reliability coefficients range from 0 to 1. The value of 0 is interpreted as the absence of reliability and the value of 1 is interpreted as perfect reliability (i.e., there is no error variance). In practice, scores from an instrument will vary in degrees of reliability rather than be completely reliable or unreliable. Current standards suggest a minimum reliability score cut-off value of .70 for the initial stage of measure development, .80 for standard research purposes, and .90 when important clinical decisions are being made,

although a stringent rule for what is an adequate reliability coefficient does not exist (Nunnally & Bernstein, 1994).

Reliability Indices

In scientific observation, a reliability index is designed to measure the degree of random or unsystematic error within scores produced by a given measure. Techniques used to calculate reliability coefficients evaluate unique sources of error variance. While there are various ways in which random error can be introduced into a given observation and various methods for demonstrating reliability across observations, the following four are general classes of reliability estimates typically used in the empirical literature: 1) inter-rater, 2) test-retest, 3) parallel-forms or alternative forms, and 4) internal consistency.

Inter-Rater Reliability. Unsystematic error can occur from different individuals administering a measure. This type of reliability is estimated by having two or more observers watching the same event and independently recording the variables according to a pre-determined coding system. A correlation coefficient is computed with the equation,

$$\frac{\text{Total Agreements}}{\text{Total Agreements} + \text{Total Disagreements}}$$

to demonstrate the strength of the relationship between each observer's ratings. The extent to which different observers provide uniform ratings of a given feature is assessed via inter-rater reliability estimates. Due to the kappa coefficients (κ) ability to account for the possibility of random chance agreement (i.e., the likelihood individuals agree by chance), κ is a commonly reported estimate of inter-rater reliability.

Test-Retest Reliability. Test-retest reliability is defined as the degree to which an assessment yields similar results from one testing instance to another. It is assumed that the individual under examination will have no coaching between evaluations. Intelligence,

personality, and other constructs believed to remain relatively constant over time should yield similar scores over repeated measurements. Unsystematic differences generated by discrepancies in the testing conditions represent the error of interest in test-retest reliability methods.

Parallel-Forms Reliability. Parallel-forms methods of calculating reliability are designed to understand the error in measurement due to the characteristics of the test itself. Specifically, a parallel-forms coefficient is the correlation of scores generated from two separate but equivalent measures of the same construct. If items are truly parallel, they will yield identical true scores and identical error variances. Responses to parallel items will differ only with respect to random fluctuations.

Internal Consistency Reliability. Internal consistency reliability refers to the extent to which individual item responses correlate with one another and with the score produced by the instrument as a whole. High levels of correlation among items indicate the instrument is measuring the same construct and is expected when all the items making up an instrument are drawn from the domain of all possible items that actually assess that construct. Low levels of correlation among items suggest sources of unexplained error in the measurement. Subsequently, if the purpose of an instrument is to assess multiple dimensions of more complex constructs, one would expect items used to identify one dimension of the construct to strongly correlate amongst each other. Similarly, one would not necessarily expect items used to assess one dimension of the construct to strongly correlate with items used to identify a separate dimension of the construct. Common methods of internal consistency are the split-half method (Spearman, 1910), Kuder-Richardson method (Kuder & Richardson, 1937), and coefficient alpha (α ; Cronbach, 1951).

The Importance of Reliability

Attenuation theory (Lord & Novick, 1968) states that the observed correlation between two scores on two different measures is unavoidably lower than their true score or actual correlation. Moreover, the actual correlations among variables (i.e., what truly exists in nature) are typically higher than what their observed correlations are (i.e., what the behavioral scientist is able to measure). Because scores produced by two measures will always exhibit less than perfect reliability, there remains a difference between observed and true score correlations. For example, if an instrument designed to assess symptoms of depression generated score X and an instrument designed to assess symptoms of substance abuse generated score Y , their true score correlation would be higher than the observed correlation. This unreliability of measures of depression, X , and substance abuse, Y , may result in researchers concluding the variables are unrelated. Errors of inference and loss of the ability to defend against a Type II error (i.e., not being able to identify a significant relationship where one exists) can result from unreliable scores. Another drawback that stems from unreliable scores being generated is that clinically significant relationships may be unnoticed. Unreliable test scores result from undetected and unsystematic errors in measurement. Furthermore, the ability to draw accurate conclusions about relationships between events, variables, and constructs are impeded when test scores are unreliable. This fact has often led to the belief that reliability precedes validity.

Reliability as a Property of Scores

Feldt and Brennan (1989) stressed the importance of recognizing reliability as a property of *scores* on a test for a given population, and that a test is neither reliable, nor unreliable. Because a test can neither be reliable nor unreliable, it is important to identify those factors that influence test score reliability (i.e., contribute to error). Similarly, Strube (2000) stated reliability

is influenced by the conditions surrounding when the measurement is given. Furthermore, Thompson (1994) noted too few researchers recognize reliability as a characteristic of scores obtained from their current data. Additionally, Thompson specifically noted that researchers should consider the conditions surrounding the administration of an instrument and be aware of the influence of sample- and method-level testing characteristics on the reliability of scores produced by that test. Indeed, numerous methodological and participant characteristics can affect the reliability of scores produced by a given measure. Understanding this concept is of great importance to researchers because the reliability of scores from samples with different characteristics (e.g., differing diagnostic groups, ages, or gender representation) will vary. Nevertheless, it is still common to hear clinicians and researchers indicate a *test* as being sufficiently reliable or that one *test* is more or less reliable than another. Additionally, one can review the empirical literature and find researchers claiming an instrument is reliable without providing any reliability coefficients generated from their obtained data (cf. Vacha-Haase et al., 2002). Although this may appear hypercritical, Thompson (1994) concisely noted that, “Such language is both incorrect and deleterious in its effects on scholarly inquiry, particularly given the pernicious consequences that unconscious paradigmatic beliefs can exact” (p. 839).

Reporting Reliability

Guidelines for reporting reliability in the empirical literature have emerged due to the increasing attentiveness surrounding reliability as a property of scores and not tests. The editor of *Educational and Psychological Measurement* offered the following provisions for authors interested in submitting articles to the journal based on the principle of reliability as a property of scores: 1) referring to a test as reliable will be considered unacceptable, 2) referring to a test as valid will be considered unacceptable, and 3) rather, authors should report reliability and validity

coefficients of the given instrument (Thompson, 1994). Other journals now have similar requirements.

Moreover, the Task Force on Statistical Inference (TFSI) was brought together by the Board of Scientific Affairs of the American Psychological Association (APA) and asked to address problematic statistical reporting methods in psychological journals. This group was asked to offer guidelines for researchers to consult when deciding how statistical results should be reported in the literature. Guided by Leland Wilkinson, the TSFI (Wilkinson & the APA/TFSI, 1999) stated:

The TFSI offers the following guidelines to aid authors in their statistical reporting: 1) tests are neither reliable nor unreliable but rather the scores produced by tests, 2) effect sizes are influenced by the accurate assessment of the reliability coefficient produced by a set of scores, and 3) at a minimum, reliability coefficients and validity estimates from the data being analyzed should be reported each and every time scores of a test are reported (Wilkinson & The APA/TFSI, 1999).

Reliability Reporting: The State of the Art

In spite of the importance of reporting reliability estimates in research, even a quick review of the literature to date suggests that scientists are inadequately reporting such data. Vacha-Haase et al. (2002) documented reliability underreporting in their review of over 20 studies that independently evaluated score reliability reporting practices for various measures. Among all the measures under evaluation ($n = 28$), fewer than 25% of them reported reliability information based on current data and a large discrepancy between reporting practices across substantive research domains (range = 0%-71%) was observed. Consistent with these findings,

reliability reporting in the *Journal of Counseling Psychology (JCP)* has also been found to be lacking (Meier & Davis, 1990). In 1967, only 5% of studies reported reliability measures. Although, this percentage increased to 23% in 1987, reporting practices are still low. The more recent reliability reporting practices of studies published in *JCP* have also been examined. Kieffer et al. (2001) found that among articles published in *JCP* from 1988 to 1997, 43.9% of studies adequately reported reliability, 15.9% cited previously reported reliability estimates, and 40.2% did not mention reliability at all. Reviews of other journals have yielded similar results (Thompson & Snyder, 1998; Whittington, 1998).

Reliability reporting inquiries among unique instruments have yielded similar results to those reviews applied across journal volumes. For example, Youngstrom and Green (2003) conducted a reliability generalization (Vacha-Haase, 1998) and found that none of the articles (0%) using the Differential Emotions Scale reported adequate reliability data. Similarly, Graham, Liu, and Jeziorski (2006) reviewed reliability reporting practices of researchers using the Dyadic Adjustment Scale and found only 91 of 403 (22.6%) included reliability estimates from the data at hand and 34% made no mention of reliability at all. A review of articles published on the Geriatric Depression Scale indicated only 98 of 338 (29%) reported reliability data (Kieffer & Reese, 2002).

Importantly in the present context, researchers have documented similar reliability reporting practices among substance use scales. For example, in a review of eight adolescent alcohol screening measures, Shields et al. (*in press*) reported that only 34% of evaluated studies provided usable reliability data. Additionally, Miller, Shields, Campfield, Wallace, and Weiss (*in press*) reported that fewer than 10% of studies using the substance abuse subscales of the Minnesota Multiphasic Personality Inventory-2 (MMPI-2; Butcher, Dahlstrom, Graham,

Tellegen, & Kaemmer, 1989) provided data on reliability. Additionally, Kieffer, Cronin, and Fister (2006) reported that 26 of 71 (36.62%) studies using the Alcohol Expectancy Questionnaire reported reliability coefficients from the present data. Nevertheless, some improvement in reliability reporting practices has been noted. In their review of studies using the Survey of Perceived Organizational Support, Hellman, Fuqua, and Worley (2006) reported that 56 of 62 (90.3%) articles provided reliability estimates from the current data. O'Rourke (2004) reported that 106 of 155 (68.4%) studies on the Center for Epidemiologic Studies-Depression Scale contained reliability measures from the collected data. Furthermore, Dunn, Smith, and Montoya (2006) examined the reliability reporting practices within the area of multicultural competency. They found 65% (89 of 137) of studies included reliability estimates. Ross, Blackburn, and Forbes (2005) estimated that nearly 50% of articles published on the Patterns of Adaptive Learning Survey Goal Orientation Scales contained reliability data from the current samples.

The concept of reliability induction (RI) is another noteworthy addition to the literature, and pertains to the importance and implications of reliability reporting. An RI occurs when a study cites prior reliability estimates to infer or induce that the current data is equally reliable (Vacha-Haase, Kogan, & Thompson, 2000). Additionally, Vacha-Haase et al. (2000) examined whether researchers were aware of the underlying concepts that allow RI to be empirically defensible by comparing studies that induced reliability to the sample characteristics and score variability of scores produced by those previous samples that originally generated reliability estimates. They discovered that of the 270 articles featuring the Bem Sex Role Inventory and Rosenberg Self Esteem Instrument that induced reliability, only 20% provided sufficient information necessary to determine if induction was justified. These authors concluded that even

when adequate information was available to compare samples, there were no instances when the induction was seen as plausible.

Other researchers have suggested that the omission of reliability coefficients altogether constitutes a distinct type of RI. Deditius-Island and Caruso (2002) argue that even though it may not be as apparent as inducing reliability from previously obtained data, neglecting to report any reliability information indicates the researcher “either (1) does not care about reliability or (2) assumes that the scores will be reliable because they have been so in previous samples, with the former indicating poor research and the latter indicating that a reliability induction has taken place” (p.732). Approximately 88% of articles examining Zuckerman’s Sensation Seeking Scale, Form V, failed to mention the property of reliability at all (Deditius-Island & Caruso). Shields and Caruso (2004) proposed the titles of reliability induction by report and reliability induction by omission to distinguish between the two reliability induction scenarios.

Based on reviews of reliability reporting practices, suggestions for improving psychometric reporting practices, and RI studies, one can conclude: 1) most published manuscripts do not include measurement reliability information; 2) even when studies include reliability information, it is often presented as previously obtained coefficients rather than coefficients obtained from a current sample; and 3) while there are certain assumptions that allow for this practice of reliability reporting to be defensible, most studies that induce reliability do not consider these assumptions.

Reliability Generalization

Vacha-Haase (1998) introduced the term Reliability Generalization, explained the purpose of it, and described the usefulness of this meta-analytic technique. Reliability Generalization studies typically have two main objectives: 1) To provide the typical reliability of

scores obtained from a given instrument's measurements and 2) to explore the variability in score reliability across samples and identify to what extent this variability can be attributed to sample and method characteristics.

RG: A Review of Applications.

By examining score reliability of the Bem Sex Role Inventory, Vacha-Haase (1998) also provided the first example of a RG study. In her evaluation, she noted that only 13.8% of studies provided reliability estimates from the data at hand, while 14.7 % of studies induced reliability by report of previous estimates and 67.3% made no mention of reliability. Furthermore, she demonstrated that the reliability of scores can be predicted by various sample and method characteristics (Vacha-Haase).

Educational and Psychological Measurement (Thompson, 2000) dedicated a section to RG once Vacha-Haase (1998) documented its usefulness. In this special section were RG studies assessing the reliability of scores generated from the Beck Depression Inventory (Yin and Fan, 2000), the “Big Five Factors” of personality (Viswesvaran and Ones, 2000), and the NEO Personality Scales (Caruso, 2000). These studies support the usefulness of viewing reliability in terms of scores by recording significant variability in score reliability estimates and the increased need for reporting reliability. Similar to historical and contemporary views on psychometric reporting practices, these findings also support a growing effort aimed at reminding researchers, journal editors, and professional groups of the necessity of score reliability reporting (e.g., Kieffer et al., 2001; Vacha-Haase, Nilsson, Reetz, Lance, & Thompson, 2000; Willson, 1980).

A practical advantage to RG studies is that they help to inform clinicians’ decisions on which instrument to use in a given situation. In accordance, Caruso (2000) stresses that RG studies should provide an empirically based set of guidelines on when to use the specific

instrument under review. That is, Caruso (2000) presents his results in a way that provides the clinician with a straightforward list of those situations in which score reliability was too low and use of the NEO personality scales may be questionable.

Following the special section in *Educational and Psychological Measurement* (Thompson, 2000) other RG studies have been conducted, and include investigations of the children's Junior Eysenck Personality Questionnaire (Caruso & Edwards, 2001), the adult Eysenck Personality Questionnaire (Caruso, Witkiewitz, Brlcourt-Dittloff, & Gottlieb, 2001), the Mathematics Anxiety Rating Scale (Capraro, Capraro, & Henson, 2001), the Teacher Efficacy Rating Scale and its related instruments (Henson, Kogan, & Vacha-Haase, 2001), and all the validity (Vacha-Haase, Tani, Kogan, Woodall, & Thompson, 2001) and clinical (Vacha-Haase, Kogan, Tani, & Woodall, 2001) scales of the Minnesota Multiphasic Personality Inventory.

As the importance of score reliability in measurement has gained increased attention, *Educational and Psychological Measurement* has chosen to follow up their special section on RG (Thompson, 2000) by devoting an entire issue to the method. This issue features score reliability examinations of the Spielberger State-Trait Anxiety Inventory (Barnes, Har, & Jung, 2002), Marlow-Crown Social Desirability Scale, (Beretvas, Meyers, & Leite, 2002), Myers-Briggs Type Indicator (Capraro & Capraro, 2002), Zuckerman's Sensation Seeking Scales (Deditius-Island & Caruso, 2002), Working Alliance Inventory (Hanson, Curry, & Bandalos, 2002), Learning Style Inventory (Henson & Hwang, 2002), Geriatric Depression Scale (Kieffer & Reese, 2002), Coopersmith Self-Esteem Inventory (Lane, White, & Henson, 2002), Career Decision-Making Self-Efficacy Scale (Nilsson, Schmidt, & Meek, 2002), Adult Attachment Scale and related instruments (Reese, Kieffer, & Briggs, 2002), and the Differential Emotions Scale (Youngstrom & Green, 2002). Still other RG studies have been performed on the Geriatric

Depression Scale (Kieffer & Reese, 2002), the Center for Epidemiologic Studies – Depression Scale (O’Rourke, 2004), the Revised Children’s Manifest Anxiety Scale (Ryngala, Shields, & Caruso, 2005), the Patterns of Adaptive Learning Survey Goal Orientation Scales (Ross, Blackburn, & Forbes, 2005), the Psychopathy Checklist (Campbell, Pulos, Hogan, & Murry, 2005), the Dyadic Adjustment Scale (Graham, Liu, & Jeziorski, 2006), the Survey for Perceived Organizational Support (Hellman, Fuqua, & Worley, 2006), and the Self-Description Questionnaire (Leach, Henson, Odom, & Cagle, 2006).

There have also been RG studies on research conducted with alcohol and substance use screening instruments. Kieffer, Cronin, and Fister (2006) examined the Alcohol Expectancy Questionnaire. Similarly, RG studies have been conducted on the Alcohol Use Disorders Identification Test (Shields & Caruso, 2003), the CAGE questionnaire (Shields & Caruso, 2004), the substance abuse subscales of the MMPI-2 (Miller, Shields, Campfield, Wallace, & Weiss, *in press*) and eight adolescent alcohol screening instruments (Shields et al., *in press*).

Purpose of Study

Based on the findings of the aforementioned RG studies, one may conclude research on the psychometric properties, in particular reliability estimates, of scores produced by substance use screening instruments is generally lacking. In part, this may be attributed to the lack of reliability estimates reported in the empirical literature. The SASSI is one of the most commonly used screening devices to detect SUDs; however, it has been argued this instrument has not been adequately studied to merit its extensive use (Gray, 2001). Therefore, the purpose of the current study is to analyze the psychometric property of reliability in scores produced by the SASSI, identify possible sources of variance within those scores, and evaluate reliability reporting practices within the empirical literature.

CHAPTER 2

METHOD

The aims of this study are two-fold. First, to evaluate the reliability reporting practices observed in peer reviewed articles that employ the SASSI. Secondly, to evaluate the reliability of scores produced by the SASSI and attempt to identify sources of variance for those reliability estimates. Both aims can be met via meta-analytic techniques, specifically through RG.

The SASSI

The SASSI has been revised two times, as the current version in use is the SASSI-3 (Miller, Roberts, Brooks, & Lazowski, 1997). The SASSI-3 is compiled of 10 scales, with a total of 93 items. These items include both face valid items and subtle, or indirect, items that appear to have no relationship to substance use. The subtle items are designed to identify those individuals with substance use problems that are unwilling or unable to acknowledge substance misuse or problems associated with misuse. The SASSI-3 can be administered individually or collectively, via pencil and paper, interview, or computer and takes approximately 15 minutes to administer. Brief training is required of administrators. Scoring takes 5 to 10 minutes and is conducted by an administrator; however, there is now computerized scoring also available.

Study Procedures

Reliability Induction and the SASSI.

As noted, reliability is a property of scores produced by a test and not a property of the actual test itself (Gronlund & Linn, 1990). In instances in which novel test scores are obtained, it follows that an evaluation of the reliability of that particular set of scores is needed. Ideally, researchers would report reliability estimates based on their data (i.e., best practice; Kallen, 2005). However, this is not always the case and, instead of reporting reliability estimates

based on data obtained from their current study, researchers often report reliability obtained from previous studies or make no mention of reliability at all. Reporting reliability obtained in previous studies is known as RI because researchers are assuming reliability measures obtained in one sample will generalize to other samples or situations (Vacha-Haase, Kogan, & Thompson, 2000). Assumptions that previously reported reliability estimates will generalize across samples are potentially erroneous. To evaluate reliability reporting practices among studies using the SASSI, frequency counts will be kept for how often studies 1) report reliability estimates based on data from their own study, 2) induce reliability by report, and 3) induce reliability by omission.

Reliability Generalization and the SASSI.

Reliability Generalization studies have two main objectives: 1) To provide the typical reliability of scores obtained from a given instrument's measurements and 2) to recognize the variability and identify study characteristics that predict or explain variation in score reliability across samples and administration protocols. Reliability Generalization techniques will be applied to the SASSI in order to obtain the average reliability of scores generated from the SASSI and to identify sources of variation within those reliability estimates. In order to characterize average reliability across samples, descriptive statistics will be employed. Furthermore, to explore variability in those reliability estimates (i.e., predict or explain variation in score reliability) correlation techniques will be used.

Data Collection

Analysis of the psychometric properties of scores produced by an instrument can be more encompassing when using an archival data collection approach as opposed to performing a single study and thus, we will employ archival methods. Our initial literature search began with a

review of literature housed in the PsychArticles, PsycINFO, and PUBMED databases. These searches were not limited to a specific year, but rather spanned from earliest to latest available. The key words searched for included *substance abuse screening instruments, substance abuse screening inventories, substance abuse subtle screening inventory, SASSI, SASSI-2, SASSI-3, SASSI-A, and SASSI-A2*. Additionally, a search was conducted for G.A. Miller, the creator of the SASSI. Lastly, manual bibliography searches were conducted among all the obtained articles.

Criteria for Including Studies in the Data Analysis

Only studies that reported using the SASSI, or one of its revised versions, were included in the analysis. The frequency of reliability induction occurrences will be taken from those studies that make no mention of reliability (i.e., RI by omission) or provide reliability estimates from previous studies (i.e., RI by report). Subsequently, each study that reports reliability coefficients from their present sample will be considered for inclusion in the RG.

Criteria for Excluding Studies in the Data Analysis

Any of the studies that do not report reliability data based on the data at hand will be excluded from the RG part of this project. Additionally, studies found in more than one database (e.g., PUBMED and PsycINFO) will only be included in the analysis once.

Categorizing Studies

The obtained studies were categorized on a scale from one to seven. A study is coded as a one when it is considered to be a false positive (i.e., the article is not relevant to the measure of interest). When a study is coded as a two, it indicates the study could not be located. A study coded as a three indicates that the study is not a primary empirical study (i.e., reviews, meta-analyses, letter to the editor, etc). Studies that are coded as four do not contain reliability information of any kind (i.e., reliability induction by omission). Similarly, a study is coded as a

five if the study mentions reliability but provides no empirical support (i.e., reliability induction by report). In addition, a study is coded as a six if the study mentions reliability and provides reliability estimate from previous research (i.e., reliability induction by report). Lastly, a study is scored as a seven when the study reports a reliability coefficient that is based on the data obtained in that study.

Criterion or Dependent Variable

The criterion variable of interest in this study is Cronbach's coefficient alpha (α). Cronbach's α was selected as the index of reliability for two primary reasons. First, Cronbach's α is a commonly reported reliability estimate in the empirical literature (Hogan, Benjamin, & Brezinski, 2000). Second, Cronbach's α is one of the most frequently used estimates of reliability in previous RG studies (Miller et al., *in press*).

Predictor Variables

This study will also examine methodological and sample characteristics possibly influencing the reliability of scores produced by the SASSI. Although the number of potential factors that could contribute to the variance found within score reliability estimates is likely countless, the number and type of predictor variables that can be coded are limited by their availability within the literature. We selected the following five predictor variables for investigation in the present study: score variability, age, gender, ethnicity, and sample type.

Data Analysis

Characterizing the typical reliability of SASSI scores can be accomplished via traditional descriptive statistics. The second goal of RG, to recognize the variability and identify sample characteristics that predict or explain variation in score reliability across samples and administration protocols, can be accomplished via conventional correlation procedures and

multiple regression and correlation procedures. The lack of reliability reporting in the larger SASSI literature, however, precluded the use of multiple regression techniques and, therefore, we rely on simple correlations.

CHAPTER 3

RESULTS

Reliability reporting practices within the SASSI literature are reported in Table 2. At least one reliability estimate from the sample under study was reported for each scale in 27.1% ($n = 13$) of studies using the SASSI. Reliability Induction occurred in 72.9% ($n = 35$) of SASSI studies. Specifically, RI by report was observed at least once for 16.7% ($n = 8$) of those scales, and RI by omission was observed at least once in 56.2% ($n = 27$) of studies. Table 2 further breaks down the reliability reporting practices by SASSI subscale.

The descriptive statistics characterizing reliability coefficients of the SASSI total scale and subscale scores are presented in Table 3. Relative to recommended guidelines for use in a clinical setting, (cf. Cicchetti & Sparrow, 1990) only the SASSI total scale ($M = .87$) and Face Valid (Alcohol $M = .88$; Other Drug $M = .92$) scales approached or met suggested cutoff values. The wide range of reliability coefficients reported indicates considerable variability in the reliability of SASSI total scale and subscale scores across samples. For example, reliability estimates ranged from a low of .01 on the SAT subscale to a high of .98 on the FVOD subscale.

Table 4 includes descriptive statistics for the FVA subscale score reliability and the other predictor variables examined in the present study. Reported reliability coefficients ranged from .78 to .93 (range = .15), and generated an unweighted mean of .88 and median of .89. Together, these findings suggest that the FVA subscale tends to produce adequately reliable scores. The relationship between score reliability of FVA subscale scores and age, gender, ethnicity, sample type, and FVA score variability were examined via bivariate correlations and one statistically significant relationship was detected between FVA score reliability and age ($r = .761$, $p < .006$). These findings are also summarized in Table 4.

Descriptive statistics for the FVOD subscale and all other variables examined in the current study are summarized in Table 5. Across samples, reliability coefficients ranged from .78 to .98 (range = .20) and generated an unweighted mean of .92 and median of .93. Collectively, these values indicate that the FVOD subscale tends to produce acceptably reliable scores across a variety of measurement circumstances. Bivariate correlations were conducted to examine the relationships between score reliability of FVOD subscale scores and age, gender, ethnicity, sample type, and FVOD score variability but no statistically significant relationships were detected.

The descriptive statistics for the SYM subscale score reliability are presented in Table 6. Additionally, Table 6 includes descriptives for the other sample and method level characteristics. For this subscale, reliability estimates ranged from .14 to .81 (range = .67) and produced an unweighted mean of .67 and median of .76, indicating wide variability in the obtained reliability estimates. Jointly, these findings indicate the SYM subscale produces modestly reliable scores. The associations between score reliability of SYM subscale score and age, gender, ethnicity, sample type, and SYM score variability were examined with bivariate correlations. No statistically significant relationships were detected.

Table 7 summarizes the descriptive statistics for the OAT subscale score reliability and the other variables under investigation in this study. The OAT's reliability estimates ranged from .24 to .76 (range = .52) indicating large variability in these scores. Also, the unweighted mean (.60) and median (.63) are of interest. Taken together, these values indicate that the OAT subscale yields score with reliability below conventionally accepted cut-off values (i.e., below .70; Cicchetti & Sparrow, 1990). Bivariate correlations were performed in order to determine the

relationships between score reliability and age, gender, ethnicity, sample type, and OAT score variability and no statistically significant relationships were detected.

Descriptive statistics for the SAT subscale score reliability and the remaining predictor variables examined in the study are presented in Table 8. Reliability coefficients ranged from .01 to .49 (range = .48) and generated an unweighted mean of .21 and median of .25, again indicating high variability for the scores on this scale. Similarly, bivariate correlations were conducted to examine the associations between score reliability of SAT subscale scores and age, gender, ethnicity, sample type, and SAT score variability. A statistically significant relationship was identified between mean SAT subscale score reliability and score variability ($r = .877, p < .022$) which are presented in Table 8. Collectively and relevant to current guidelines, these findings indicate that SAT subscale scores exhibit poor reliability (cf. Cicchetti & Sparrow, 1990).

Table 9 presents the descriptive statistics for the DEF subscale reliability and the identified predictor variables. Reliability estimates ranged from .29 to .66 (range = .37) and produced an unweighted mean of .53 and median of .55. Together, these findings indicate that the DEF subscale tends to produce unreliable scores. The relationships between DEF subscale score reliability and age, gender, ethnicity, sample type, and DEF score variability were examined via bivariate correlations and no statistically significant relationships were detected.

The descriptive statistics for the SAM subscale reliability, along with all other examined variables are summarized in Table 10. The reliability coefficients for this scale ranged from .22 to .57 (range = .35), generating an unweighted mean of .35 and median of .32. Collectively, these findings indicate that SAM subscale scores are generally unreliable. Bivariate correlations were conducted to explore the relationships between score reliability and sample age, gender,

ethnicity, type, and SAM score variability and also presented in Table 10. No statistically significant relationships were detected.

Descriptive statistics for the FAM subscale score reliability and the identified predictor variables examined in this study are displayed in Table 11. The reliability estimates obtained from this scale ranged from .15 to .36 (range = .21) yielding an unweighted mean of .26 and median of .27. These findings indicate that the FAM subscale typically produces unreliable scores. The degree of association for score reliability and sample age, gender, ethnicity, type, and FAM score variability were assessed via bivariate correlations. A statistically significant relationship was detected between FAM subscale score reliability and both FAM score variability ($r = .892, p < .017$) and proportion non-White ($r = .683, p < .043$).

Lastly, Table 12 includes descriptive statistics for the COR subscale score reliability and the remaining predictor variables. Reliability estimates ranged from .47 to .78 (range = .34) generating an unweighted mean of .62 and median of .65. These values indicate that the COR subscale scores tend to demonstrate low reliability but may be capable of generating adequate score reliability for research purposes. Bivariate correlations were performed in order to assess the relationships between score reliability and sample age, gender, ethnicity, type and COR score variability and a statistically significant relationship emerged between COR subscale score reliability and proportion non-White ($r = .76, p < .017$). These findings are summarized in Table 12.

CHAPTER 4

DISCUSSION

Reliability Induction and the SASSI

Stemming from a lack of empirical data pertaining to score reliability of the SASSI, a systematic reliability investigation was conducted. The first area investigated was to what extent researchers that have used the SASSI have also reported reliability information. Reliability reporting is lacking with alcohol and substance abuse measures (Shields & Caruso, 2003, 2004; Shields et al., *in press*) but also in a variety of other areas of measurement (Kieffer, Reese, & Thompson, 2001; Vacha-Haase, Henson, & Caruso, 2002; Vacha-Haase, Ness, Nilsson, & Reetz, 1999; & Whittington, 1998). Results from the present study indicate the SASSI is no exception. The frequency of researchers inducing reliability from past research especially warrants consideration, as does the lack of reliability reporting in general. The criteria for RI by report or RI by omission were met in 80.2% of SASSI studies. In other words, less than 20% of studies using the SASSI reported reliability information consistent with current guidelines. All eight studies that induced reliability by report for the total scale provided those reliability estimates obtained in previous samples; however, of those eight studies, only three reported reliability estimates for the subscales. Even more troublesome is the extent to which studies employing the SASSI have induced reliability by omission. Of 48 studies evaluated in the present study, the vast majority (i.e., 34 to 40, depending on the scale) made no mention of reliability.

In RGs on a variety of alcohol and substance use measures, reliability underreporting has also been noted (Shields & Caruso, 2003, 2004). More specifically, in a review of adolescent alcohol screening measures (Shields et al. *in press*) researchers report that only 34% provided usable reliability data. Similarly, in a recent study (Miller, Shields, Campfield, Wallace, &

Weiss, *in press*) on the substance abuse subscales of the MMPI-2, the authors report that 90.6% of these studies did not provide data on reliability. Hence, the extent to which studies that have used the SASSI did not provide usable reliability is consistent within the alcohol use screening literature.

The findings from this study indicate that researchers using the SASSI may be operating under a false assumption that arises when citing psychometric information (e.g., reliability estimates) obtained from previous studies (Meier & Davis, 1990). More specifically, the assumption is that reliability estimates obtained from previous studies will accurately reflect the reliability of scores produced by that same instrument upon some future administration. Because score reliability can and will vary across administrations, past reliability estimates may, in fact, not reflect score reliability estimates obtained in future studies.

This scarcity of reliability reporting in SASSI research is not surprising and may even be expected based on findings from other areas that report a lack of reliability reporting. Dating back to 1980, Willson reported that only 37% of studies published in the *American Education Research Journal* reported reliability estimates from the data at hand. Similarly, Meier and Davis (1990) report that of studies published in the *JCP* from 1967-1987, reliability estimates were provided only 5%-40 % of the time. More recently, in a special section of *Educational and Psychological Measurement* devoted to RG, guest editors (Vacha-Haase, Henson, & Caruso, 2002) report that, depending on the instrument under evaluation, reliability estimates were provided from 0% to 71.4% of the time with the majority (i.e., 24 of 25, 96%) reporting reliability data less than 50% of the time.

The present findings contribute to a growing body of empirical literature documenting the underreporting of reliability information in general and furthermore, support scientific journals

providing authors with reliability reporting guidelines when submitting papers for publication consideration. Reliability information was provided for 13 of 48 (27%) studies employing the SASSI. However, of these 13, only 3 reported reliability for the total scale and the instances where reliability estimates for each subscale were reported ranged from 6 (SYM) to 11 (FVA, FVOD, OAT, SAT, DEF, and SAM). Although only 3 of 48 SASSI studies provided reliability information for the total scale, none of these concurrently provided all of the desired dependent variable information (i.e., mean sample age, gender, ethnicity, sample type, and score variability). Therefore, our RG findings are limited in two important ways. One, only 3 studies reported reliability estimates for the SASSI total scale. Two, there was limited information provided for our other variables of interest. This holds true for the subscales as well. For example, of the 11 studies reporting reliability estimates for the FVOD scale, only 6 provided each of the desired dependent variable information.

Reliability Generalization and the SASSI

In order to characterize SASSI score reliability estimates across studies and identify characteristics that may contribute to variability in those estimates, an RG (Vacha-Haase, 1998) was conducted. Cicchetti and Sparrow (1990) suggested the following guidelines when evaluating score reliability for a particular test: (a) if internal consistency is below .70, the level of clinical significance is unacceptable; (b) if internal consistency is between .70 and .79, the level of clinical significance is fair; (c) if internal consistency is between .80 and .89, the level of clinical significance is good; and (d) if internal consistency is above .90, the level of clinical significance is excellent. In these instances, clinical significance can be operationalized as the confidence one has when using these estimates to make a diagnosis or inform clinical judgment. The mean value of the total scale reliability ($M = .87$) suggests adequate proportions of true score

variance for most research purposes and may be considered for use in clinical settings. However, this value is taken from only three studies and reliability coefficients ranged from .74 to .95 within these studies. Of interest are the reliability coefficients provided for the SASSI's subscales. The only subscales with a mean reliability value of .80 or higher were the FVA ($M = .88$) and FVOD ($M = .92$) scales, which is ironic for an instrument marketed as "subtle" or indirect. In other words, although the SASSI is marketed as an indirect screening instrument, only the direct scales of the SASSI (i.e., Face Valid Alcohol and Face Valid Other Drug) consistently yield reliable scores. The next highest mean reliability value obtained was .75 for the Symptoms subscale. Of the SASSI's total scale and subscale ($n = 10$) scores, only 2 had a level of good clinical significance and 1 had an excellent level of clinical significance.

The SASSI is an alcohol and drug screening measure and, therefore, was designed to distinguish individuals who may have an alcohol or drug use problem and expose the need for further assessment. Screening measures are not typically used to make diagnoses pertaining to an individual's substance use disorder status (Connors, 1995). This holds especially true for the SASSI, because SASSI subscale score reliability estimates fall short of the frequently recommended cutoff value of .90 needed for use in clinical settings. Therefore, individuals using the SASSI in clinical situations are encouraged to do so with caution and in a prediagnostic fashion.

Bivariate correlations were conducted to evaluate the relationship between score reliability estimates and select sample and method characteristics. While small sample size limits confidence in these results, bivariate correlational analyses identified several statistically significant relationships. Specifically, there was a strong, positive correlation between SASSI score reliability and score variability for both the Subtle Attributes ($r = .877$, $p < .022$) and

Family History ($r = .892, p < .017$) subscales. This is a noteworthy finding, indicating that, at least for these two subscales, score variability is positively related to score reliability. This is consistent with classical test theory which states increases in score variability typically indicates increases in score reliability (Lord & Novick, 1968) and with prior research evaluating alcohol screening tools (Shields & Caruso, 2003, 2004).

Additionally, there was a strong positive correlation between score reliability and ethnicity (i.e., proportion non-white) for both the Family History subscale score ($r = .683, p < .043$) and Tendency to Involvement in Correctional Setting subscale score ($r = .76, p < .017$). These findings are of a particular interest for several reasons. These findings imply that non-White or minority samples tend to generate more reliable scores for each of these scales than do Caucasians. This finding may be attributed to the overrepresentation of non-White ethnic groups in the correctional system.

The remaining statistically significant correlation was between sample age and the Face Valid Alcohol subscale score ($r = .761, p < .006$). The sample age ranged from 19 to 32 years, with older respondents typically producing more reliable scores than younger respondents. Past research has identified a similar relationship between the CAGE, an alcohol screening measure and sample age (cf. Shields & Caruso, 2004). However, with insufficient score reliability reported, conclusions drawn from these analyses should be drawn with caution.

Study Limitations

As noted, the obtained results should be considered cautiously. One reason to remain tentative is that reliability reporting practices observed within the SASSI literature fell short of Wilkinson and the TFSI's (1999) recommendations for reporting reliability. This lack of reliability reporting resulted in low sample size. Additionally relevant, is the notion of the "file-

drawer” problem (Rosenthal, 1979) which indicates that only higher quality studies are likely to be published while studies of less quality remain in the file drawers of researchers. This notion indicates that the obtained coefficients may be higher than what would appear if all research employing the SASSI could be obtained. This is disheartening for the SASSI’s indirect (i.e., subtle) subscales. Because one can assume that the reliability of scores in unpublished studies is lower than what appears in the empirical literature, this subpar index of score reliability would likely decrease.

Internal consistency was observed as the primary method for reporting reliability estimates within the empirical SASSI literature, which resulted in a separate limitation. While internal consistency was the index of reliability we used, other reliability indices were reported in some instances. Test-retest reliability coefficients obtained by Lazowski et al. (1998), indicate that SASSI total scale and Face Valid subscale scores can have excellent test-retest reliability for a clinical population over a period of 2 weeks. Nevertheless, these other indices were also not often reported.

Recommendations

In light of the present findings, we recommend the following: 1) Whenever the SASSI is used in a research setting, researchers should obtain and report an estimate of reliability for the sample under study; 2) When the SASSI is used in a clinical setting, only the Face Valid scales should be considered reliable indicators of problematic substance use/abuse; 3) Conclusions pertaining to one’s substance use or abuse should not be drawn from the SASSI alone, but rather from multiple screening instruments; and 4) Because we used internal consistency as the reliability index for our study, future research could expand on our findings by evaluating other indices of reliability in future research.

REFERENCES

- American Psychiatric Association. (1994). *Diagnostic and statistical manual of mental disorders (4th ed. revised)*. Washington, DC: American Psychiatric Association.
- Arenth, P. M., Bogner, J. A., Corrigan, J. D., & Schmidt, L. (2001). The utility of the Substance Abuse Subtle Screening Inventory-3 for use with individuals with brain injury. *Brain Injury, 15*, 499-510.
- Barnes, L. B., Harp, D., & Jung, W. S. (2002). Reliability generalization of scores on the Spielberger State-Trait Anxiety Inventory. *Educational and Psychological Measurement, 62*, 603-618.
- Bauman, S., Merta, R., & Steiner, R. (1999). Further validation of the adolescent form of the SASSI. *Journal of Child & Adolescent Substance Abuse, 9*, 51-71.
- Beretvas, S. N., Meyers, J. L., & Leite, W. L. (2002). A reliability generalization study of the Marlowe-Crowne Social Desirability Scale. *Educational and Psychological Measurement, 62*, 570-589.
- Bradley, K. A., Boyd-Wickizer, J., Powell, S. H., & Burman, M. L. (1998). Alcohol screening questionnaires in women. *Journal of the American Medical Association, 280*, 166-171.
- Butcher, J. N., Dahlstrom, W. G., Graham, J. R., Tellegen, A., & Kaemmer, B. (1989). *MMPI-2: Manual for administration and scoring*. Minneapolis: University of Minnesota Press.
- Buzzanga, V. L. (2000). The usefulness of the Substance Abuse Subtle Screening Inventory-2 with a head injured population. *Dissertation Abstracts International: Section B: The Sciences and Engineering, 60*, pp. 5828.

- Campbell, J. S., Pulos, S., Hogan, M., & Murry, F. (2005). Reliability generalization of the Psychopathy Checklist applied in youthful samples. *Educational and Psychological Measurement, 65*, 639-656.
- Capraro, M. M., Capraro, R. M., & Henson, R. K. (2002). Measurement error of scores on the Mathematics Anxiety Rating Scale across studies. *Educational and Psychological Measurement, 61*, 373-386.
- Capraro, R. M., & Capraro, M. M. (2002). Myers-Briggs Type Indicator score reliability across studies: A meta-analytic reliability generalization study. *Educational and Psychological Measurement, 62*, 590-602.
- Caruso, J. C. (2000). Reliability generalization of the NEO personality scales. *Educational and Psychological Measurement, 60*, 236-254.
- Caruso, J. C., & Edwards, S. (2001). Reliability generalization of the Junior Eysenck Personality Questionnaire. *Personality and Individual Differences, 31*, 173-184.
- Caruso, J. C., Witkiewitz, K., Belcourt-Dittloff, A., & Gottlieb, J.D. (2001). Reliability of scores from the Eysenck Personality Questionnaire: A reliability generalization study. *Educational and Psychological Measurement, 61*, 675-689.
- Cherpitel, C. J. (1998). Differences in performance of screening instruments for problem Drinking among Blacks, Whites, and Hispanics in an emergency room population. *Journal of Studies on Alcohol, 59*, 420-426.
- Cicchetti, D. V., & Sparrow, S. S. (1990). Assessment of adaptive behaviors in young children. In J.J. Johnson and J. Goldman (Eds.), *Developmental assessment in clinical child psychology: A handbook* (pp. 173-196). New York: Pergamon Press.

- *Clements, R. (2002). Psychometric properties of the Substance Abuse Subtle Screening Inventory-3. *Journal of Substance Abuse Treatment, 23*, 419-423.
- Cohen, J., Cohen, P., West, S. G., & Aiken, L. S. (2003). *Applied multiple regression/correlation analysis for the behavioral sciences (3rd. edition)*. Mahwah, NJ: Lawrence Erlbaum Associates.
- Coll, K. M., Juhnke, G. A., Thobro, P., & Haas, R. (2003). A preliminary study using the Substance Abuse Subtle Screening Inventory-Adolescent version as an outcome measure with adolescent offenders. *Journal of Addictions & Offender Counseling, 24*, 11-22.
- Cook, T. D., & Campbell, D. T. (1979). *Quasi-experimentation: Design and analysis issues for field settings*. Boston, MA: Houghton Mifflin.
- Cooper, S. E., & Robinson, D. A. (1987). Use of the Substance Abuse Subtle Screening Inventory with a college population. *Journal of American College Health, 36*, 180-184.
- Cronbach, L. J. (1951). Coefficient alpha and the internal structure of tests. *Psychometrika, 16*, 297-334.
- Deditius-Island, H. K., & Caruso, J.C. (2002). An examination of the reliability of scores from Zuckerman's sensation seeking scales, form V. *Educational and Psychological Measurement, 62*, 728-734.
- Dunn, T. W., Smith, T. B., & Montoya, J. A. (2006). Multicultural competency instrumentation: A review and analysis of reliability generalization. *Journal of Counseling and Development, 84*, 471-482.
- *Emanuelson, G. (2005). SASSI-3 and the court-ordered population: A preliminary reliability and factor analysis. *Dissertation Abstracts International Section A: Humanities and Social Sciences, 65* pp. 2575.

- Ewing, J. A. (1984). Detecting alcoholism, the CAGE questionnaire. *Journal of American Medical Association, 252*, 1905-1907.
- Feldt, L. S., & Brennan, R. L. (1989). Reliability. In R. H. Linn (Ed.), *Educational measurement (3rd ed.)*. New York: Macmillan.
- Feldstein, S. W., & Campbell, W. (2006). Using sneaky instruments to circumvent denial: A review of the history of substance abuse and dependence self-report measures. Poster presented at the annual conference of the Research Society on Alcoholism, Tucson, Arizona.
- Fisher, G. L., & Harrison, T. C. (1990, April). Assessment of alcohol and other drug abuse with referred adolescents. Paper presented at the annual meeting of the National Association of School Psychologists, San Francisco, CA.
- Fox, M. R. (1993). MMPI and SASSI scores as predictors of sexual abuse and substance abuse in adolescent psychiatric inpatients. Doctoral Dissertation, University of Toledo, 1992. *Dissertation Abstracts International, 53*, pp. 3801.
- Grant, B. F., Stinson, F. S., Dawson, D. A., Chou, P., Dufour, M. C., Compton, W., et al. (2004). Prevalence and co-occurrence of substance use disorders and independent mood and anxiety disorders: Results from the national epidemiologic survey on alcohol and related conditions. *Archives of General Psychiatry, 61*, 807-816.
- *Gray, B. T. (1999) Factor structure of the Substance Abuse Subtle Screening Inventory-2 (SASSI-2). Ph.D. dissertation, Texas A&M University, United States -- Texas.
- **Gray, B. T. (2001). A factor analytic study of the Substance Abuse Subtle Screening Inventory (SASSI). *Educational and Psychological Measurement, 61*, 102-118.

- Graham, J. M., Liu, Y. J., & Jeziorski, J. L. (2006). The Dyadic Adjustment Scale: A reliability generalization meta-analysis. *Journal of Marriage and Family*, 68, 701-717.
- *Greene, R. E. (2001). Validity of the Substance Abuse Subtle Screening Inventory-3 (SASSI-3) for people arrested for driving-under-the-influence of alcohol. Ph.D. dissertation, University of Kansas, United States -- Kansas.
- Grimm, L. G. & Yarnold, P. R. (Eds.). (2000). Introduction to multivariate statistics. *Reading and understanding MORE multivariate statistics* (pp. 437). Washington, DC: American Psychological Association.
- Grimm, L. G., & Yarnold, P. R. (Eds.). (1995). Introduction to multivariate statistics. *Reading and understanding multivariate statistics* (pp. 373). Washington, DC: American Psychological Association.
- Gronlund, N. E., & Linn, R. L. (1990). *Measurement and evaluation in teaching*. (6th ed.). London: Macmillan.
- Hansink, R. L. (1994). Adolescents' perceptions of their relationships with their fathers: A comparison of substance abusers and substance non-abusers. Doctoral Dissertation, United States International University, 1992. *Dissertation Abstracts International*, 54, pp. 3220.
- Hanson, W. E., Curry, K. T., & Bandalos, D. L. (2002). Reliability generalization of Working Alliance Inventory scale scores: *Educational and Psychological Measurement*, 62, 659-673.
- Hellman, C.M., Fuqua, D. R., & Worley, J. (2006). A reliability generalization study on the Survey of Perceived Organizational Support: The effects of *M* age and number of items on score reliability. *Educational and Psychological Measurement*, 66, 631-642.

- Henson, R. K., & Hwang, D-Y. (2002). Variability and prediction of measurement error in a Kolb's Learning Style Inventory Scores: A reliability generalization study. *Educational and Psychological Measurement, 62*, 712-727.
- Henson, R. K., Kogan, L. R., & Vacha-Haase, T. (2001). A reliability generalization study of the Teacher Efficacy Scale and related instruments. *Educational and Psychological Measurement, 61*, 404-420.
- Hogan, T. P., Benjamin, A., & Brezinski, K. L. (2000). Reliability methods: A note on the frequency of use of various types. *Educational and Psychological Measurement, 60*, 523-531.
- Horrigan, T. J., & Piazzzi, N. (1999). The Substance Abuse Subtle Screening Inventory minimizes the need for toxicology screening of prenatal patients. *Journal of Substance Abuse Treatment, 17*, 243-247.
- Hudson, S. L. (1997). Predicting substance abuse in a family counseling agency. Retrieved from Dissertations and Theses database. (AAT 9721045)
- Ingersoll, J. B. (2003). Predicting the underreporting of substance abuse symptoms in adolescent males in an outpatient substance abuse treatment program. Retrieved from Dissertations and Theses database. (AAT 3066661)
- Joseph, C. L., Rasmussen, J., Ganzini, L., & Atkinson, R. M. (1997). Outcome of nursing home care for residents with alcohol use disorders. *International Journal of Geriatric Psychiatry, 12*, 767-772.
- Kallen, M. (2005). Understanding reliability when using measurement instruments in the VA population. *Measurement Excellence and Training Resource Information Center (METRIC) Newsletter*.

- Kazdin, A. E. (Ed.). (1992). *Methodological issues & strategies in clinical research*. Washington, DC: American Psychological Association.
- Keith, T. Z. (2006). *Multiple regression and beyond*. (534 pp.). Boston: Allyn & Bacon.
- Keppel, G. (1991). *Design and analysis: A researcher's handbook* (3rd. ed.). Upper Saddle River, NJ: Prentice Hall, Inc.
- Kieffer, K. M. & Reese, R. J. (2002). A reliability generalization study of the Geriatric Depression Scale. *Educational and Psychological Measurement, 62*, 969-994.
- Kieffer, K. M., Cronin, C., & Fister, M. C. (2004). Exploring variability and sources of measurement error in Alcohol Expectancy Questionnaire reliability coefficients: A meta-analytic reliability generalization. *Journal of Studies on Alcohol, 65*, 663-671.
- Kieffer, K. M., Reese, R. J., & Thompson, B. (2001). Statistical techniques employed in *AERJ* and *JCP* articles from 1988 to 1997: A methodological review. *Journal of Experimental Education, 69*, 280-309.
- *Kikuchi, A., & Wada, K. (2000). The Japanese version of Substance Abuse Subtle Screening Inventory and its psychometric assessment. *Japanese Journal of Alcohol Studies & Drug Dependence, 35*, 388-401.
- Klein, S. M. (2000). Deception and substance abuse. Retrieved from Dissertations and Theses database. (AAT 9949559)
- Kuder, G. F., & Richardson, M. W. (1937). The theory of the estimation of test reliability. *Psychometrika, 2*, 151-160.
- Lane, G. G., White, A. E., & Henson, R. K. (2002). Expanding reliability generalization methods with KR-21 estimates: An RG study of the Coopersmith Self-Esteem Inventory. *Educational and Psychological Measurement, 62*, 685-711.

- *Laux, J. M., Salyers, K. M., & Kotova, E. (2005). A psychometric evaluation of the SASSI-3 in a college sample. *Journal of College Counseling, 8*, 41-51.
- *Lazowski, L. E., Miller, F. G., Boye, M. W., & Miller, G. A. (1998). Efficacy of the Substance Abuse Subtle Screening Inventory-3 (SASSI-3) in identifying substance dependence disorders in clinical settings. *Journal of Personality Assessment, 71*, 114-128.
- Leach, L. F., Henson, R. K., Odom, L. R., & Cagle, L. S. (2006). A reliability generalization study of the Self-Description Questionnaire. *Educational and Psychological Measurement, 66*, 285-304.
- Li, H., Rostenthal, R., & Rubin, D. B. (1996). Reliability of measurement in psychology: From Spearman-Brown to maximal reliability. *Psychological Methods, 1*, 98-107.
- Lincoln, A. K., Liebschutz, J. M., Chernoff, M., Nguyen, D., & Amaro, H. (2006). Brief screening for co-occurring disorders among women entering substance abuse treatment. *Substance Abuse Treatment, Prevention, and Policy, 1*.
- Lord, F. M., & Novick, M. R. (1968). *Statistical theory of mental test scores*. Reading, MA: Addison-Wesley.
- MacAndrew, C. (1965). The differentiation of male alcoholic outpatients from non-alcoholic psychiatric outpatients by *Ms* of the MMPI. *Quarterly Journal on Alcohol Studies, 26*, 238-246.
- Meier, S. T., & Davis, S. R. (1990). Trends in reporting psychometric properties of scales used in counseling psychology research. *Journal of Counseling Psychology, 37*, 113-115.
- McLellan, A. T., Luborsky, L., Woody, G. E. & O'Brien, C. P. (1980). An improved diagnostic evaluation instrument for substance abuse patients: The Addiction Severity Index. *Journal of Nervous and Mental Disease, 168*, 26-33.

McLellan, A. T., Kushner, H., Metzger, D., Peters, R., Smith, I. & Grissom, G., et al. (1992).

The fifth edition of the Addiction Severity Index. *Journal of Substance Abuse Treatment*, 9, 199-213.

*Makini, G. K., Jr., et al. (1996) Psychiatric symptoms of Hawaiian adolescents. *Cultural Diversity & Mental Health*, 2, 183-191.

Meier, S. T., & Davis, S. R. (1990). Trends in reporting psychometric properties of scales used in counseling psychology research. *Journal of Counseling Psychology*, 37, 113-115.

Miller, C. A., Shields, A. L., Campfield, D., Wallace, K. A., & Weiss, R. D. (in press). Substance use scales of the Minnesota Multiphasic Personality Inventory: An exploration of score reliability via meta-analysis. *Educational and Psychological Measurement*. Miller, F.G.,

Roberts, J., Brooks, M. K., & Lazowski, L. E. (1997). SASSI-3 users guide: A quick reference for administration and scoring. Bloomington, IN: Baugh Enterprises.

Miller, G. A. (1985). The Substance Abuse Subtle Screening Inventory manual. Bloomington, IN: Glenn A. Miller.

Miller, G. A. (1985). The Substance Abuse Subtle Screening Inventory-Adolescent (SASSI-A). Spencer, IN: Glenn A. Miller.

Miller, G. A. (1994). Supplement to the Substance Abuse Subtle Screening Inventory manual. Bloomington, IN: Glenn A. Miller.

*Myerholtz, L., & Rosenberg, H. (1998). Screening college students for alcohol problems: Psychometric assessment of the SASSI-2. *Journal of Studies on Alcohol*, 59, 439-446.

Nilsson, J. E., Schmidt, C. K., & Meek, W. D. (2002). Reliability generalization: An examination of the Career Decision-Making Self-Efficacy Scale. *Educational and Psychological Measurement*, 62, 647-658.

- Nishimura, S. T., Hishinuma, E. S., Miyamoto, R. H., Goebert, D. A., Johnson, Y., Noelle, Y. C., et al. (2001). Prediction of DISC substance abuse and dependency for ethnically diverse adolescent. *Journal of Substance Abuse, 13*, 597-607.
- Nunnally, J. C., & Bernstein, I. H. (1994). *Psychometric theory* (3rd ed.). New York: McGraw-Hill.
- O'Rourke, N. (2004). Reliability generalization of responses by care providers to the center for epidemiologic studies-depression scale. *Educational and Psychological Measurement, 64*, 973-990.
- Pearson, B. S. (2000). Validation of the Substance Abuse Subtle Screening Inventory-3(SASSI-3) with the chronically mentally ill population. Retrieved from Dissertations and Theses database. (AAT 9953953).
- Piazza, N. J. (1996). Dual diagnosis and adolescent psychiatric inpatient. *Substance Use & Misuse, 31*, 215-223.
- Reese, R. J., Kieffer, K. M., & Briggs, B. K. (2002). A reliability generalization study of select measures of adult attachment style. *Educational and Psychological Measurement, 62*, 619-646.
- Rogers, R., Cashel, M. L., Johansen, J., & Sewell, K. W. (1997). Evaluation of adolescent offenders with substance abuse: Validation of the SASSI with conduct-disordered youth. *Criminal Justice and Behavior, 24*, 114-128.
- Rosenthal, R. (1979). The "file drawer problem" and tolerance for null results. *Psychological Bulletin, 86*, 638-641.

- Ross, M. E., Blackburn, M., & Forbes, S. (2005). Reliability generalization of the patterns of adaptive learning survey goal orientation scales. *Educational and Psychological Measurement, 65*, 451-464.
- Ryngala, D. J., Shields, A. L., & Caruso, J. C. (2005). Reliability generalization of the Revised Children's Manifest Anxiety Scale. *Educational and Psychological Measurement, 65*, 259-271.
- Saunders, J. B., Aasland, O. G., Babor, T. F., De la Fuente, J. R., & Grant, M. (1993). Development of the alcohol use disorders identification test (AUDIT): WHO collaborative project on early detection of persons with harmful alcohol consumption-II. *Addiction, 88*, 791-803.
- Schmidt, E. A. (2001). Pre- and post-test administration of the Adult Substance Abuse Subtle Screening Inventory-2 (SASSI-2): Detecting change. *Alcoholism Treatment Quarterly, 19*, 63-73.
- Selzer, M. L. (1971). The Michigan Alcohol Screening Test: The quest for a new diagnostic instrument. *American Journal of Psychiatry, 127*, 1653-1658.
- Shields, A. L., & Caruso, J.C. (2003). Reliability generalization of the Alcohol Use Disorders Identification Test. *Educational and Psychological Measurement, 63*, 404-413.
- Shields, A. L., & Caruso, J.C. (2004). A reliability induction and reliability generalization study of the CAGE questionnaire. *Educational and Psychological Measurement, 63*, 1-17.
- Shields, A. L., Campfield, D. C., Miller, C. A., Howell, R. T., Wallace, K., & Weiss, R. D. (in press). Score reliability of adolescent alcohol screening measures: A meta-analytic inquiry. *Journal of Child and Adolescent Substance Abuse*.
- Skinner, H. A. (1982). The Drug Abuse Screening Test. *Addictive Behaviors, 7*, 363-371.

- Spearman, C. (1910). Correlation calculated with faulty data. *British Journal of Psychology*, 3, 271-295.
- Strube, M. J. (2000). Reliability and generalizability theory. In L. G. Grimm & P. R. Yarnold (Eds.) *Reading and understanding MORE multivariate statistics* (pp. 23-66), Washington, DC: American Psychological Association.
- Swartz, J. A. (1998). Adapting and using the Substance Abuse Subtle Screening Inventory-2 with criminal offenders: Preliminary results. *Criminal Justice and Behavior*, 25, 344-365.
- Sweet, R. I., & Saules, K. K. (2003). Validity of the Substance Abuse Subtle Screening Inventory-Adolescent (SASSI-A). *Journal of Substance Abuse Treatment*, 24, 331-340.
- Teslak, A. G. (2000). The utility of the CAGE, MAST, DAST, and SASSI-3 in assessing substance use/misuse in a psychiatric population. Retrieved from Dissertations and Theses database. (AAT 9974837).
- Thompson, B. (1994). Guidelines for authors. *Educational and Psychological Measurement*, 54, 837-847. Thompson, B., & Snyder, P. A. (1998). Statistical significance and reliability analyses in recent JCD research articles. *Journal of Counseling and Development*, 76, 436-441.
- Vacha-Haase, T. (1998). Reliability generalization: Exploring variance in measurement error affecting score reliability across studies. *Educational and Psychological Measurement*, 58, 6-20.
- Vacha-Haase, T., Henson, R., & Caruso, J. C. (2002). Reliability generalization: Moving toward improved understanding and use of score reliability. *Educational and Psychological Measurement*, 62, 562-569.

- Vacha-Haase, T., Kogan, L. R., & Thompson, B. (2000). Sample compositions and variabilities in published studies versus those in test manuals: Validity of score reliability inductions. *Educational and Psychological Measurement, 60*, 509-522.
- Vacha-Haase, T., Ness, C., Nilsson, J., & Reetz, D. (1999). Practices regarding reporting of reliability coefficients: A review of three journals. *Journal of Experimental Education, 67*, 335-341.
- Vacha-Haase, T., Kogan, L. R., Tani, C. R., & Woodall, R. A. (2001). Reliability generalization: Exploring reliability variations on MMPI/MMPI-2 clinical scale scores. *Educational and Psychological Measurement, 61*, 45-59.
- Vacha-Haase, T., Tani, C. R., Kogan, L. R., Woodall, R. A., & Thompson, B. (2001). Reliability generalization: Exploring reliability variations on MMPI/MMPI-2 validity scale scores. *Assessment, 8*, 391-401.
- Viswesvaran, C., & Ones, D. S. (2000). Measurement error in “Big Five Factors” personality assessment: Reliability generalization across studies and measures. *Educational and Psychological Measurement, 60*, 224-235.
- Weed, N. C., Butcher, J. N., McKenna, T., & Ben-Porath, Y. S. (1992). New measures for assessing alcohol and drug abuse with the MMPI-2: The APS and the AAS. *Journal of Personality Assessment, 58*, 389-404.
- Whittington, D. (1998). How well do researchers report their measures? An evaluation of measurement in published educational research. *Educational and Psychological Measurement, 58*, 21-37.
- Wilkinson, L. & the American Psychological Association Task Force on Statistical Inference (1999). Statistical methods in psychology journals. *American Psychologist, 54*, 594-604.

- Willson, V. L. (1980). Research techniques in AERJ articles: 1969 to 1978. *Educational Researcher*, 9, 5-10.
- Yeh, L. B. (1993). A multiple-case study comparison of normal and substance-abusing adolescents and their families involving family assessment and videotaped dyadic interaction. Retrieved from Dissertations and Theses database. (AAT 9308006)
- Yin, P., & Fan, X. (2000). Assessing the reliability of Beck depression inventory scores: Reliability generalization across studies. *Educational and Psychological Measurement*, 60, 201-223.
- Youngstrom, E. A. & Green, K. W. (2003). Reliability generalization of self-report of emotions when using the Differential Emotions Scale. *Educational and Psychological Measurement*, 63, 279-295.
- *Young, T. M. (2005) The Substance Abuse Subtle Screening Inventory-Third Edition (SASSI-3) as an alcohol abuse screening tool for Asian Pacific American college students in Seattle. Ph.D. dissertation, Capella University, United States -- Minnesota.

* indicates study was used in meta-analysis

** indicates two independent sample

APPENDIXES

Table 1.

Scales Found on Differing Versions of the SASSI.

Scale	Version of the SASSI				
	SASSI	SASSI 2	SASSI 3	SASSI-A	SASSI-A2
RPSA (FVA)	X	X	X	X	X
RPSD (FVOD)	X	X	X	X	X
OAT	X	X	X	X	X
SAT	X	X	X	X	X
DEF1	X	X	X	X	X
DEF2	X				
ALD	X				
FAM	X	X	X		X
SYM			X		X
SAM (replaced DEF2)		X	X	X	X
COR		X	X	X	X
RAP		X	X		
FRISK					X
ATT					X
VAL					X
SCS					X

RPS-Risk Prediction Scales (renamed Face Valid Alcohol and Face Valid Other drug), OAT-Obvious Attributes, SAT-Subtle Attributes, DEF1-Defensiveness 1, DEF2-Defensiveness 2

(replaced by SAM), ALD-Alcohol/Other drug, FAM-Family Pattern, SYM-Symptoms, SAM-Supplemental Addiction Measure, COR-Tendency to involvement with Correctional system, RAP-Random Answering Pattern, FRISK-Family/Friends Risk, ATT-Attitudes toward substance use, VAL-Validity Check, and SCS-Secondary Classification Scale.

Table 2.

Reliability Reporting Across SASSI Subscales.

	<i>N</i>	<i>Accurate Reliability Reported</i>	<i>%</i>	<i>Reliability Induction by Report</i>	<i>%</i>	<i>Reliability Induction by Omission</i>	<i>%</i>
Scale							
Scale	48	3	6.2	8	16.7	37	77.1
FVA	48	11	22.9	3	6.2	34	71
FVOD	48	11	22.9	3	6.2	34	71
SYM	48	6	12.5	2	4.2	40	83.3
OAT	48	11	22.9	3	6.2	34	71
SAT	48	11	22.9	3	6.2	34	71
DEF	48	11	22.9	3	6.2	34	71
SAM	48	11	22.9	3	6.2	34	71
FAM	48	11	22.9	3	6.2	34	71
COR	48	11	22.9	3	6.2	34	71
Total	48	13	27.1	8	16.7	27	56.2

Table 3.

Descriptive Statistics for Reliability Coefficients of SASSI Total Scale and Subscale Scores.

Variable	N	M	Mdn	SD	Range	*Percentage (U)	**Percentage (T)
Reliability (α)							
Scale	3	.87	.93	.116	.74-.95	23%	6.25%
FVA	11	.88	.89	.039	.78-.93	85%	23%
FVOD	11	.92	.93	.052	.78-.98	85%	23%
SYM	6	.65	.75	.256	.14-.81	46%	12.5%
OAT	11	.59	.63	.144	.24-.76	85%	23%
SAT	11	.23	.25	.163	.01-.49	85%	23%
DEF	11	.52	.53	.129	.29-.66	85%	23%
SAM	11	.36	.32	.109	.22-.57	85%	23%
FAM	10	.26	.28	.075	.15-.36	77%	20.8%
COR	10	.63	.65	.106	.47-.78	77%	20.8%

N = Number of studies reporting internal consistency reliability estimates among all studies employing the SASSI.

*Percentage (U) of studies reporting internal consistency reliability estimates among the studies reporting reliability. Descriptives are based on these studies.

**Percentage (T) of studies reporting any reliability estimates among all studies employing the SASSI.

Scale = Full scale score, FVA=Face Valid Alcohol, FVOD=Face Valid Other Drug, Sym=Symptoms, OAT=Obvious Attributes, SAT=Subtle Attributes, DEF=Defensiveness, SAM=Supplemental Addiction Measure, FAM=Family History, & COR=Tendency to Involvement in Correctional Setting.

Table 4.

Descriptive Statistics for FVA Reliability Coefficients and Predictor Variables.

Variable	N	M	Mdn	SD	Range	r.	p-value
Score reliability (α)	11	.88	.89	.039	.78 - .93		
SD of FVA scores	7	6.66	7	1.82	4.38 - 10.10	.171	.714
Average age	11	27.72	31	5.07	19 - 32.48	.761	.006
Proportion male	11	.63	.72	.21	.21 - .81	.118	.729
Proportion non-Caucasian	11	.37	.33	.36	.02 - 1.0	-.134	.712
Sample type (%clinical)	13	.69	--	.48	.0 - 1.0	.113	.713

Note. Sample type was coded as clinical (1) and nonclinical (0).

Table 5.

Descriptive Statistics for FOD Reliability Coefficients and Predictor Variables.

Variable	N	M	Mdn	SD	Range	r.	p-value
Score reliability (α)	12	.92	.93	.05	.78 -.98		
SD of FOD scores	6	5.6	5.95	2.06	2.41 - 8.33	.581	.226
Average age	10	27.69	31	5.35	19 - 32.48	.255	.477
Proportion male	10	.67	.72	.17	.26 - .81	.319	.370
Proportion non-Caucasian	9	.37	.33	.38	.02 - 1	-.310	.416
Sample type (%clinical)	12	.75	--	.45	.0 - 1.0	.333	.290

Note. Sample type was coded as clinical (1) and nonclinical (0).

Table 6.

Descriptive Statistics for SYM Reliability Coefficients and Predictor Variables.

Variable	N	M	Mdn	SD	Range	r.	p-value
Score reliability (α)	7	.67	.76	.24	.14 - .81		
SD of SYM scores	4	2.24	2.22	.36	1.92 - 2.6	.182	.818
Average age	6	27.47	28.2	5.45	20 - 32.48	-.403	.428
Proportion male	6	.63	.68	.20	.26 - .8	-.179	.735
Proportion non-Caucasian	5	.49	.33	.48	.05 - 1	-.154	.805
Sample type (%clinical)	7	.57	--	.53	.0 - 1.0	-.302	.510

Note. Sample type was coded as clinical (1) and nonclinical (0).

Table 7.

Descriptive Statistics for OAT Reliability Coefficients and Predictor Variables.

Variable	N	M	Mdn	SD	Range	r.	p-value
Score reliability (α)	12	.60	.63	.14	.24 - .76		
SD of OAT scores	6	2.29	2.13	.64	1.41 - 3.14	-.329	.524
Average age	10	27.68	31	5.35	19 - 32.48	.155	.668
Proportion male	10	.67	.72	.17	.26 - .81	.291	.415
Proportion non-Caucasian	9	.37	.33	.38	.02-1.0	.000	.999
Sample type (%clinical)	12	.75	--	.45	.0 - 1.0	.310	.326

Note. Sample type was coded as clinical (1) and nonclinical (0).

Table 8.

Descriptive Statistics for SAT Reliability Coefficients and Predictor Variables.

Variable	N	M	Mdn	SD	Range	r.	p-value
Score reliability (α)	12	.21	.25	.16	.01 - .49		
SD of SAT scores	6	1.19	1.15	.26	.9 - 1.64	.877	.022
Average age	10	27.68	31	5.35	19 - 32.48	-.274	.444
Proportion male	10	.67	.72	.17	.26 - .81	.120	.742
Proportion non-Caucasian	9	.37	.33	.38	.02 - 1.0	.128	.743
Sample type (%clinical)	12	.75	--	.45	.0 - 1.0	.131	.684

Note. Sample type was coded as clinical (1) and nonclinical (0).

Table 9.

Descriptive Statistics for DEF Reliability Coefficients and Predictor Variables.

Variable	N	M	Mdn	SD	Range	r.	p-value
Score reliability (α)	12	.53	.55	.13	.29 - .66		
SD of DEF scores	6	2.16	2.17	.14	1.99 - 2.32	.089	.867
Average age	10	27.68	31	5.35	19 - 32.48	.242	.501
Proportion male	10	.67	.72	.17	.26 - .81	.236	.511
Proportion non-Caucasian	9	.37	.33	.38	.02 - 1.0	-.159	.683
Sample type (%clinical)	12	.75	--	.45	.0 - 1.0	.098	.761

Note. Sample type was coded as clinical (1) and nonclinical (0).

Table 10.

Descriptive Statistics for SAM Reliability Coefficients and Predictor Variables.

Variable	N	M	Mdn	SD	Range	r.	p-value
Score reliability (α)	12	.35	.32	.11	.22 - .57		
SD of SAM scores	6	1.99	2.05	.32	1.41 - 2.35	-.294	.572
Average age	10	27.68	31	5.34	19 - 32.48	-.09	.804
Proportion male	10	.67	.72	.17	.26 - .81.	.022	.951
Proportion non-Caucasian	9	.37	.33	.38	.02 - 1.0	.275	.474
Sample type (%clinical)	12	.75	--	.45	.0 - 1.0	-.248	.437

Note. Sample type was coded as clinical (1) and nonclinical (0).

Table 11.

Descriptive Statistics for FAM Reliability Coefficients and Predictor Variables.

Variable	N	M	Mdn	SD	Range	r.	p-value
Score reliability (α)	11	.26	.27	.08	.15 - .36		
SD of FAM scores	6	1.82	1.82	.12	1.7 - 2.0	.892	.017
Average age	9	27.22	31	5.46	19 - 32.48	-.254	.509
Proportion male	9	.67	.72	.18	.26 - .81.	-.532	.140
Proportion non-Caucasian	9	.37	.33	.38	.02 - 1.0	.683	.043
Sample type (%clinical)	11	.73	--	.47	.0 - 1.0	-.579	.062

Note. Sample type was coded as clinical (1) and nonclinical (0).

Table 12.

Descriptive Statistics for COR Reliability Coefficients and Predictor Variables.

Variable	N	M	Mdn	SD	Range	r.	p-value
Score reliability (α)	11	.62	.65	.11	.47 - .78		
SD of COR scores	6	2.64	2.54	.37	2.3 – 3.23	.654	.159
Average age	9	27.22	31	5.46	19 - 32.48	-.342	.367
Proportion male	9	.67	.72	.18	.26 - .81	-.164	.674
Proportion non-Caucasian	9	.37	.33	.38	.02 - 1.0	.76	.017
Sample type (%clinical)	11	.73	--	.47	.0 - 1.0	-.463	.151

Note. Sample type was coded as clinical (1) and nonclinical (0).

VITA

JOSHUA A. WOODSON

- Personal Data: Date of Birth: April 30, 1981
Place of Birth: Greeneville, Tennessee
Marital Status: Single
- Education: Public Schools, Cocke County, Tennessee
B.A. Psychology, East Tennessee State University, Johnson City, Tennessee 2005
B.S. Human Development and Learning, East Tennessee State University, Johnson City, Tennessee 2005
M.A. Clinical Psychology, East Tennessee State University, Johnson City, Tennessee 2008
- Professional Experience: Primary Instructor, East Tennessee State University; Johnson City, Tennessee, 2006-2007
Graduate Assistant, East Tennessee State University, College of Arts and Sciences, 2005-2008.
- Publications: Miller, C.S., Woodson, J.A., Howell, R.T., & Shields, A.L. (in press). Assessing the reliability of scores produced by the Substance Abuse Subtle Screening Inventory (SASSI). *Substance Use and Misuse*, 44.
- Presentations: **Woodson, J.A.**, Shields, A.L. and Miller, C.S. (2006). Assessing the reliability of scores produced by the substance abuse subtle screening inventory (SASSI). The 40th annual Convention for the Association for Behavioral and Cognitive Therapies. Chicago, IL.

Shields, A.L, **Woodson, J.A.**, and Weiss, R.D. (2006). Reliability of scores produced by the Addiction Severity Index: A meta-analytic integration. The 27th Annual Scientific Meeting of the Research Society on Alcoholism. June. Baltimore, MD.

Woodson, J. A. & Dixon, W. E. (2005). Examining spontaneous labeling and vocabulary size as predictors of the success generalization of novel objects. *Appalachian Student Research Forum*, Johnson City, TN.
- Honors and Awards: Cum Laude
Dean's List