

AN EXAMINATION OF TWO METHODS OF MEASURING INCONSISTENCY

By
Rizwan Ahmed Khan

Approved:

Michael D. Biderman
Salem Carpet Professor
(Thesis Chair)

Dr. Brian O'Leary
Department Head, Associate Professor
(Committee Member)

Bart L. Weathington
UC Foundation Associate Professor
Sciences
(Committee Member)

Herbert Burhenn
Dean of the College of Arts and

A. Jerald Ainsworth
Dean of the Graduate School

AN EXAMINATION OF TWO METHODS OF MEASURING INCONSISTENCY

By

Rizwan Khan

A Thesis
Submitted to the Faculty of the
University of Tennessee at Chattanooga
In Partial Fulfillment of the Requirements
For the Degree of Master of Science
in Psychology

The University of Tennessee at Chattanooga
Chattanooga, TN

May 2012

ABSTRACT

Previous research has provided evidence for the notion that there are varying levels of inconsistency between individuals when responding to questionnaires with multiple response items. Specifically, there are individual differences in how consistently persons respond to items from the same dimension in a questionnaire (Reddock, Biderman & Nguyen, 2011). Currently, there is not a consensus on how inconsistency should be measured. In the present study inconsistency of responses to the IPIP Big Five questionnaire was measured. Two response formats permitting measurement of inconsistency were compared - a frequency-based format (FB) vs. a traditional Likert scale format. Furthermore, in an effort to study inconsistency in a broader context, the relationships of social desirability and ADHD to inconsistency were examined. The results provided no evidence for convergent validity between the two measures, discriminant validity for each measure, no evidence of a relationship between BIDR and inconsistency, but a positive relationship between FB based inconsistency and scores on the ADHD measure. Implications and limitations of the study are discussed.

DEDICATION

To my family, who have supported me in any endeavor that I have wished to pursue. To my friends, who were alongside me through this adventure.

ACKNOWLEDGEMENTS

The completion of this thesis could not have been done without the help of my advisor, Dr. Michael Biderman. His dedication to his students, patience and genuine interest in the topic at hand made for an enjoyable and thought-provoking experience. I could not have asked for a better thesis advisor! I would also like to thank Dr. Bart Weathington and Dr. Brian O'Leary for their insight, and ability to think about the big picture in regards to this research topic.

TABLE OF CONTENTS

DEDICATION	iv
ACKNOWLEDGMENTS	v
LIST OF TABLES	viii
LIST OF ABBREVIATIONS	ix
CHAPTER	
I. INTRODUCTION	1
The Stability of Inconsistent Responding	2
Convergent Validity	3
Discriminant Validity	4
Inconsistency in Relation to Socially Desirable Responding	6
Inconsistency in Relation to ADHD	6
II. METHOD	8
Participants	8
Measures	8
Big Five Scale	8
Inconsistency	9
SDR	10
ADHD	10
Procedure	11
III. RESULTS	12
IV. DISCUSSION	16
Study Limitations	19
Future Directions	19
REFERENCES	21

APPENDIX

A. A STANDARD IPIP BIG 5 ITEM IN LIKERT FORMAT24

B. A STANDARD IPIP BIG 5 ITEM IN FREQUENCY-BASED
FORMAT26

C. AN ITEM FROM THE BALANCED INVENTORY OF
SOCIALLY DESIRABLE RESPONDING30

D. AN ITEM FROM THE WENDER UTAH RATING SCALE.....32

E. A RELATIONSHIP BETWEEN POSITIVELY SKEWED
MEANS AND THEIR STANDARD DEVIATIONS34

F. A VISUAL DEPICTION SHOWING HOW THE FOUR
CONDITIONS IN THIS STUDY WERE BALANCED36

VITA38

LIST OF TABLES

1. Reliabilities of Dimensional Scale Scores and Inconsistency Measures	12
2. Correlations Between FB Scale Scores and Likert Scale Scores	13
3. Correlations Between FB Inconsistency score and Likert inconsistency Score	13
4. Correlations Between Inconsistency Measures and Dimensional Scale Scores	14
5. Correlations Between Inconsistency Measures and BIDR Scores	15
6. Correlations Between Inconsistency Measures and ADHD Scores	15

LIST OF ABBREVIATIONS

FB, Frequency-Based

FBSD, Frequency-Based Inconsistency

LBSD, Likert-Based Inconsistency

ADHD, Attention Deficient Hyperactivity Disorder

BIDR, Balanced Inventory of Desirable Responding

SDR, Socially Desirable Responding

CHAPTER I

INTRODUCTION

Extensive research has been conducted on what constitutes personality. Within the last century, many new theories have been offered which touch upon different aspects of personality. For example, Allport and Odbert (1936) established three main types of personality traits based on a lexical analysis. These include cardinal traits which are considered dominant through one's life, central traits which are more general in nature, and secondary traits which are situational or contextual. In essence, Allport and Odbert's personality trait theory proposed that some aspects of personality are temporary and other aspects of personality are more stable.

More recent research supports this notion, showing that some aspects of personality are more stable through one's life time than other aspects (Costa, McCrae, & Arenberg, 1980; Ferguson, 2010; Hampson, & Goldberg, 2006; Terracciano, McCrae, & Costa, 2010). On the other hand, specific personality traits, such as neuroticism, have been shown to be unstable or change over a period of time (Costa et al., 1980; Hampson & Goldberg, 2006; Hopwood et al., 2011; Vaidya, Gray, Haig, & Watson, 2002). Factors contributing to instability of personality traits causing them to vary include the natural maturation process or aging (Costa et al., 1980; Ferguson, 2010; Hampson & Goldberg 2006; Helson & Moane, 1987; Hopwood et al., 2011). Hence, a divide in personality research has been between consistency versus variability, and there is abundant research supporting both.

Recently there has been a resurgence of interest in the study of personality inconsistency

in responding (Biderman, & Reddock, 2012; Edwards & Woehr, 2007; Fleeson, 2001; Fleisher, Woehr, Edwards, & Cullen, 2011; Reddock, Biderman, & Nguyen, 2011). In view of this recent interest in stability vs. instability of personality traits, the purpose of this study is to investigate inconsistency or variability in responding to questionnaires and to determine the implications that may result from using two different measures of inconsistency.

The Stability of Inconsistent Responding

Much of the research concerning inconsistency involves variability in responses to personality questionnaires (Biderman & Reddock, 2012; Britt, 1993; Fleeson, 2001; Fleisher et al., 2011; Reddock, et al., 2011). Keeping this context in mind, Biderman and Reddock (2012) defined inconsistency as, “the tendency for respondents to give different responses to items for which identical or nearly identical responses would seem to be appropriate,” (p.647).

Inconsistency in this context has largely been attributed to differences in item wording, but rarely to inconsistency as a personality characteristic. However, Reddock et al. (2011) found support for the latter notion; some individuals exhibited inconsistency across all dimensions of a single IPIP Big Five questionnaire. Because the IPIP Big Five has been shown to be a highly reliable scale, this suggests that variability is due to a personal characteristic, rather than it being due to the scale; otherwise variability would not be exhibited in every dimension, rather it would be found more in some than others. Furthermore, Biderman and Reddock (2012) found that persons inconsistent in one questionnaire were more likely to be inconsistent in another questionnaire. Additional support for the stability of inconsistent responding has been provided by Fleeson (2001), who conducted a study where participants reported their behavior and emotions using a Big 5 type of scale, 5 times per day for about 2-3 weeks. The results indicated that across time, inconsistent individuals remained inconsistent. Therefore, research on inconsistency in this

context has established the notion that variability can manifest itself as a characteristic of the individual. The purpose of this study was to compare two methods of measuring inconsistency in responses.

Convergent Validity

According to Saville et al. (2011), there are not a sufficient number of studies comparing different measures of inconsistency. In the present study, two methods of measuring inconsistency were compared to establish convergent validity; one measure of inconsistency is based on a typical Likert scale response and has been used by Reddock et al. (2011) and Biderman and Reddock (2012), while the other is based on a frequency-based (FB) response format (Appendix B), which has been used by Edwards and Woehr, (2007) and Fleisher et al., (2011). Although Fleisher et al. (2011) assessed and found convergent validity of scale scores computed using Likert and FB methods, they did not compare measures of inconsistency computed using the FB scale with the Likert type scale, making the assessment of convergent validity of the two measures of inconsistency a primary objective in this study. Because both measures are believed to measure inconsistency by their respective users, they should correlate with each other - to exhibit convergent validity. In an effort to replicate the convergent validity found in Fleisher et al. (2011), this study correlates the scale scores (i.e., level) of each format as well. Therefore:

H1: The scale scores computed using the FB and Likert response formats will be positively correlated, exhibiting convergent validity of scale scores.

H2: The measures of inconsistency will be positively correlated, exhibiting convergent validity of inconsistency scores.

Discriminant Validity

Another purpose of this study is to further ensure that the construct of inconsistency is indeed a unique and separate construct from already established constructs by establishing discriminant validity. This was done by correlating inconsistency scores obtained using both measures with the dimensions of the Big Five. More specifically, I hypothesize that the scores on each dimension will have a non-existent to low correlation with the measures of inconsistency. If inconsistency is truly a separate construct from these dimensions, then a person exhibiting inconsistency should do so, regardless of the level of each dimension they possess. In addition to the independence of inconsistency from the Big Five traits, the *measures* of inconsistency should also be theoretically independent of the measures of the traits. For example, if an individual possesses extraversion to a high degree, this should not necessarily result in high consistency merely because they scored high on those items, resulting in low variation, which would lead to labeling the person as consistent. Rather, all individuals, regardless of the level of extraversion they may possess (i.e., low versus high) should be able to exhibit varying levels of inconsistency.

Previous research has indicated when using FB method, there may be correlations between the measures of the traits and the measures of inconsistency (Fleisher et al., 2011). More Specifically, Fleisher et al. found a correlation of .77 between inconsistency and level of agreeableness, a correlation of .68 between inconsistency and conscientiousness, and a correlation of .22 between inconsistency and stability. On the other hand, a similar correlation between inconsistency and levels of the Big Five using a Likert type scale in Reddock et al. 2011 resulted in low correlations ranging from -.17 to .07. This, it appears that the two ways of measuring inconsistency may have different discriminant validities with trait levels.

Previous research has suggested a way to reduce spurious correlations between trait level and inconsistency using the Likert format. According to Baird, Le, and Lucas (2006), correlations between level and inconsistency measures based on Likert scale responses may be due to extreme values of mean responses, where participants either tend to answer very high or very low in response to multiple items. Since responders who tend to answer either high or low will inherently have low variance due to the closeness of their responses to the end of the response scale, compared to moderate responders who can have either high or low variance, this may result in a pseudo relationship between the standard deviation and means. In other words, a correlation between variability of responses and mean dimension may exist not due to a genuine relationship but because the distributions of means towards one end of the scale will inherently result in low variation. This is depicted in Appendix E. Baird et al. (2006) proposed modified measures to reduce the spurious correlations. However, Reddock et al. (2011) found that the modifications were not needed. Currently, no research has been conducted on modifying the FB measure of inconsistency to reduce the correlation with trait scores. Therefore:

H3: The inconsistency measures based on Likert Scale responses will be less highly related to mean scores than will the measures using FB responses.

Inconsistency in Relation to Socially Desirable Responding

Socially desirable responding (SDR) refers to individuals who alter their responses to a questionnaire to portray themselves positively (Paulhus, 2002). Paulhus (1984) defined two types of SDR, self-deception and impression management. Self-deception refers to an individual who truly believes in the positive light in which they portray themselves, whereas impression management refers to an individual who consciously attempts to portray themselves positively. The construct of SDR should exhibit a low to medium strength correlation with inconsistency.

Specifically, I predict persons high on SDR should be less inconsistent, so there should be a negative correlation between SDR and inconsistency. That is persons high in SDR would respond in a socially desirable manner, which will result in similar responses for each item within a dimension. Therefore, I hypothesize that responding desirably influences variability in responding.

H4: Inconsistency will be negatively related to social desirability

Inconsistency in Relation to ADHD

The final purpose of this study is to determine whether a relationship exists between inconsistency in responding and ADHD. ADHD is characterized by varying levels of attentive issues and/or hyperactivity among children which may lead into adulthood. For an adult to be diagnosed with ADHD, not only must the symptoms of ADHD be present currently, but it must be established that those symptoms were present during childhood (American Psychiatric Association, 2000). Self reports are used to retrieve information about the symptoms of ADHD during childhood as well as reports by individuals close to the person to corroborate this information when diagnosing adults with ADHD (Barkley, Murphy, & Fischer, 2008; Goldstein & Ellison, 2002; McGough & Barkley, 2004; Murphy & Gordon, 2006). This method clearly poses a problem in its validity as there may be numerous reasons for an incorrect description to be given. For example, recalling information that could be over 20 years old may be inaccurate, bias plays a role in any self report measure, and the intentions of the individual may not be truthful (e.g., trying to receive medication) (Knouse, Bagwell, Barkley, & Murphy, 2005; Booksh, Pella, Singh, & Gouvier, 2010; Harrison, Edwards, & Parker, 2007; Murphy & Gordon, 2006). Furthermore, the individual may not be aware of their behavior, resulting in the ADHD going unnoticed during childhood and subsequently during adulthood (Barkley, 1997; Wender,

1995). Hence, the usefulness of establishing an additional measure for ADHD to increase validity is apparent.

Current ADHD research involves analyzing reaction times during tasks to determine the degree to which inconsistency occurs in the reaction times, indicating attention deficiency.

Research indicates that individuals who possess ADHD are more likely to be inconsistent in their reaction times (Adams, Roberts, Milich, and Fillmore, 2011; Castellanos & Tannock, 2002).

Based on this research, the purpose of this part of the study is to overcome the barriers of using self-report by determining whether response patterns can be used to predict attention deficiencies. Therefore:

H5: There will be a positive correlation between inconsistency and scores on a measure of ADHD tendencies.

CHAPTER II

METHOD

Participants

Participants were 175 undergraduates at a medium-sized southeastern university, participating for course credit. Forty-three were male. Mean age was 18.83 (SD=2.04). Percentage of Whites was 78.4, Black/African-America was 12.0, Hispanic was 3.0, Asian was 2.6, and Other was 4.2.

Measures

Big Five Scale

The big five IPIP is a 100 item questionnaire, available at (www.ipip.ori.org). It is available in the Original 50 item questionnaire and as a 100 item questionnaire, which consists of the Original 50 items and an additional Other 50 items. Each participant completed two response formats of the IPIP Big Five. Items from the original 50-item sample questionnaire were administered in one response format. Items from the 100-item scale that were not in the 50-item sample questionnaire were administered in the other format. Format one consisted of a 7 point Likert scale ranging from 1= “completely inaccurate” to 7= “completely accurate” (Appendix A). Format two was based on a FB scale of the Big Five IPIP (Appendix B). More specifically, each item has three categories for responses (e.g., very inaccurate, neither inaccurate nor accurate, very accurate). For each category, the participant assigned a percentage value representing the frequency with which the behavior indicated by the item described the participant, with the

restriction that the percentages sum to 100% across the three categories. For example, when answering the following item, “I am the life of the party,” the participant may answer the following: very inaccurate 20% of the time, neither inaccurate nor accurate 50% of the time, and very accurate 30% of the time. Then, the percentages were combined to make one score for each item by assigning the following weights to each response category: very inaccurate =.01, neither inaccurate nor accurate = .04, very accurate=.07) so the minimum score and maximum score for each item would range from 1 to 7 ((20x.01) + (50x.04) + (30x.07) =4.3 for the example)). This computation converts the FB score into a Likert scale score.

Inconsistency

Inconsistency was operationalized for the Likert scale response by computing the standard deviation of responses to items within a dimension, and averaging those standard deviations across dimensions (Baumeister, 1991; Baumeister & Tice, 1988; Biderman and Reddock, 2012; Britt 1993; Fleeson, 2001; Reddock, et al., 2011). This inconsistency score is called Likert between-item standard deviation or LBSD.

Inconsistency was operationalized for the FB scale, as introduced by Kane (1986), using methods outlined in Fleisher et al. (2011), as shown in Appendix B. The standard deviation of each item was calculated based on the assigned percentages for each response category for that item, and then the mean of the standard deviations for each of the ten items within a dimension was computed to determine dimensional inconsistency. Finally, the mean of the standard deviations from the five dimensions was computed to obtain an overall measure of inconsistency. This inconsistency score is called Frequency-based within-item standard deviation or FBWSD. Note that one Likert based inconsistency measure and one Frequency based inconsistency measure was computed for each person.

SDR

Individuals were also given the 40-item balanced inventory of desirable responding, or BIDR (Appendix C), developed by Paulhus (1991). Participants responded on a scale of 1= “Not True” to 7= “Very True”. The first 20 items of the BIDR measure the construct of self-deception, while the last 20 items measure impression management. In accordance to Paulhus (1991), one BIDR score was calculated for each dimension, resulting in two scores. However, the way in which each dimension was scored here differed from the way suggested by Paulhus (1991). For the purpose of this study, the two dimensions were scored using the typical procedure for summated scales by finding the mean of the item responses within the dimension.

ADHD

Individuals also completed the 25-item Wender Utah Rating Scale, or WURS (Appendix D), developed by Ward, Wender, and Reihmerr (1993). The questionnaire consists of a 5 point scale ranging from 0= “Not at all or very slightly” to 4= “Very Much”. Although a cut off sum score of 36 has been used to be considered as potentially possessing ADHD, such a cutoff score was not employed here. Instead, correlations involving the mean score were analyzed. This is because the purpose of this study was to examine any relationship that may exist between ADHD and inconsistency, and not to diagnose a participant with ADHD.

Procedure

The two response formats of the Big 5 scale were given to participants. To control for sequence effects that could have resulted if all participants were given the same order of formats, the order of presentation of the two formats and two versions of the Big Five questionnaire were counterbalanced. Specifically, one-fourth of the respondents responded to the Original 50 IPIP

items using the Likert response format and then to the Other 50 IPIP items using the FB response format. A second fourth responded to the Original 50 IPIP items using the FB response format and then to the Other 50 IPIP items using the Likert response format. A third quarter of the respondents responded to the Other 50 IPIP items using the Likert response format and then to the Original 50 IPIP items using the FB response format. The final fourth of the respondents responded to the Other IPIP items using the FB response format and then to the Original 50 IPIP items using the Likert response format. This setup is shown in Appendix F. Inconsistencies were computed using the respective procedures described above for each format. The questionnaire measuring socially desirable responding was administered, followed by the questionnaire measuring ADHD.

For the analyses to follow, data were collapsed across order of presentation of scales. All participants who responded using the FB format to the Original IPIP scale and using the Likert format to the Other IPIP scale, regardless of order of presentation of the two scales were treated as one group. This group had a total of 86 participants. Similarly, all participants who responded using the FB format to the Other IPIP scale and the Likert format to the Original scale were treated as a second group. This group had a total of 89 participants. Parallel analyses were conducted on the two groups.

CHAPTER III

RESULTS

Table 1 presents reliabilities of the Likert Scales, the FB scales, and the inconsistency measures. As can be seen from inspection of the table, all reliabilities with the exception of the reliability of LBSD from the original questionnaire were at least marginally acceptable.

Table 1

Reliabilities of Dimensional Scale Scores and Inconsistency Measures.

	E	A	C	S	O	Inconsistency
Likert	.834 .885	.664 .716	.815 .728	.717 .907	.638 .821	.693 .533
Frequency	.752 .745	.666 .638	.768 .739	.780 .809	.745 .729	.869 .915

Note. First number =FB Original vs. Likert Other; Second number= Likert Original vs. FB Other;
E= Extraversion; A= Agreeableness; C= Conscientiousness; S= Stability; O= Openness;

Table 2 presents the correlations between the computed FB dimensional scale scores with the Likert dimensional scale scores. Inspection of this table indicates high, positive correlations between the FB scale scores and Likert scale scores measuring the same dimension, showing support for hypothesis 1.

Table 2

Correlations Between FB Scale Scores and Likert Scale Scores.

Variable	EL	AL	CL	SL	OL
EFB	.662^c .741^c	.307 ^b .024	.013 .267 ^a	-.052 .356 ^b	.248 ^a .143
AFB	.297 ^b .219 ^a	.660^c .554^c	.375 ^c .252 ^a	.050 .221 ^a	.256 ^a .146
CFB	.109 .080	.234 ^a .147	.719^c .765^c	.149 -.003	.261 ^a .082
SFB	.026 .172	.094 -.056	.027 .058	.643^c .717^c	.083 .201
OFB	.364 ^b .129	.170 .178	.146 .142	.046 .350 ^b	.663^c .745^c

Note. Convergent validities are in boldface; First number =FB Original vs. Likert Other; Second number= Likert Original vs. FB Other; ^a p<.05; ^b p<.01; ^c p<.001; E= Extraversion; A= Agreeableness; C= Conscientiousness; S= Stability; O= Openness; FB=Frequency-based scale format; L= Likert scale format.

Table 3 presents the correlations between the inconsistency scores using the FB response format and the inconsistency scores using the Likert response format. Inspection of this table indicates that the correlation between the two inconsistency formats was not significant for either group. This suggests that hypothesis 2 was not supported.

Table 3

Correlations Between FB Inconsistency Score and Likert Inconsistency Score

Variable	LBSD
FBWSD	-.145 -.074

Note: First number =FB Original vs. Likert Other, Second number= Likert Original vs. FB Other; FBWSD=Frequency-based inconsistency score; LBSD= Likert inconsistency score.

Table 4 presents the correlations between the Likert dimensional scale scores and the Likert inconsistency score and the correlations between the FB dimensional scale scores and FB inconsistency scores. With the Likert scale format, the mean discriminant validity was .057, with only one dimensional scale score correlated significantly with the Likert inconsistency measure,

exhibiting discriminant validity. For the FB format, the mean discriminant validity coefficient was more than five times as large at -.303, with eight FB scale scores significantly correlated with the FB inconsistency measure, indicating general a lack of discriminant validity. These results show support for hypothesis 3 only for the Likert based inconsistency measure.

Table 4

Correlations Between Inconsistency Measures and Dimensional Scale Scores

Variable	E	A	C	S	O
LBSD	.007 .205	-.032 .170	.224 ^a -0.037	-.102 .063	-.032 .103
FBWSD	-.075 -.282 ^b	-.443 ^c -448 ^c	-.505 ^c -374 ^c	.105 -.305 ^b	-.313 ^b -386 ^c

Note: First number =FB Original vs. Likert Other, Second number= Likert Original vs. FB Other; ^a p<.05; ^b p<.01; ^c p<.001; LBSD= Likert Inconsistency Score; FBWSD= Frequency-based inconsistency score E=Extraversion; A=Agreeableness; C=Conscientiousness; S=Stability; O=Openness.

The first row shows the correlations between the Likert dimensional scale scores and Likert inconsistency score. The second row shows the correlations between the FB dimensional scale scores and FB inconsistency score.

Table 5 presents the correlations of the Likert scale inconsistency score and FB inconsistency score with the scores on the BIDR. Only one negatively significant relationship was found between Likert inconsistency and BIDR, at -.236 (p < .05). The other significant relationship was positive (r=.238, p < .05). For the FB format, the only significant correlation was positive (r = .279, p < .01). In general, Hypothesis 4 was not supported.

Table 5

Correlations Between Inconsistency Measures and BIDR Scores

Variable	BIDRSD	BIDRIM
LBSD	.079 .238 ^a	.190 -.236 ^a
FBWSD	-.041 -.019	.158 .279 ^b

Note: First number =FB Original vs. Likert Other, Second number= Likert Original vs. FB Other; ^a p<.05; ^b p<.01; ^c p<.001; FBWSD= Frequency-based inconsistency score; LBSD= Likert inconsistency score; BIDRSD= Mean score of self-deception items on BIDR; BIDRIM= Mean score of impression management items on BIDR.

Table 6 presents the correlations of the Likert inconsistency score and FB inconsistency score with the ADHD score. The results show that there were no significant relationships between Likert inconsistency score and ADHD. However, the FB inconsistency score did positively correlate with ADHD, with a mean correlation of +.267. The Likert format correlations gave no support for Hypothesis 5, while the Frequency based correlations did.

Table 6

Correlations Between Inconsistency Measures and ADHD Scores

Variable	ADHDM
LBSD	.006 -.012
FBWSD	.248 ^a .285 ^b

Note: First number =FB Original vs. Likert Other, Second number= Likert Original vs. FB Other; ^a p<.05; ^b p<.01; ^c p<.001; FBWSD= Frequency-based inconsistency score; LBSD= Likert inconsistency score.

CHAPTER IV

DISCUSSION

The purpose of this study was to compare two methods of measuring inconsistency to establish convergent validity. Although there is evidence for the construct validity of each method, no study to date had compared methods for measuring inconsistency in questionnaire responding. Another purpose of this research was to show evidence for a relationship of inconsistency with social desirability and ADHD. Therefore, this study replicates previous research findings and adds to the literature of inconsistency in responding to questionnaires.

First, this study found support for the notion that scales created from the FB format of responding and Likert scales have convergent validity - the scale scores for each format had strong positive correlations with each other. This is consistent with the results found in Fleisher et al. (2011). It should be noted here that the strong positive correlations shown in Table 1 were correlations across response format and across questionnaires. This suggests that the FB format measures personality to a degree similar to the Likert scale.

The next step was to find support for the notion that the FB inconsistency scores would exhibit convergent validity with the Likert inconsistency scores. Because the scale scores were found to have convergent validity, and both format types are thought to measure inconsistency in responding, one would expect evidence for convergent validity in inconsistency between the two formats. As shown in table 2, the results do not support this hypothesis. This suggests that there is a certain point at which the two formats cease to have a strong relationship with each other

(i.e., the scale scores had convergent validity, but the inconsistency scores did not). This suggests that the two formats may measure different kinds of inconsistency. It could be that the Likert response format measures inconsistency in responding to items, while the FB response format measures inconsistency in the personality dimensions themselves. This is because the FB response format is asking the participant to report the degree to which the behavior was displayed in the last six months suggesting that the FB response format is measuring personality inconsistency rather than response inconsistency. It is important to note that this study is the first to our knowledge to show that the two measures of inconsistency may actually be measuring different kinds of inconsistency.

The purpose of this study was also to establish discriminant validity for the two measures. This was been done by correlating the inconsistency score of each format to their respective dimensional scale scores. In the past, results have suggested that there is little to no relationship when correlating Likert dimensional scale scores to the Likert inconsistency score (Reddock, et al., 2011). However, when correlating FB dimensional scale scores with the FB inconsistency score, positive relationships have been found (Fleisher, et al., 2011). Therefore, in an effort to replicate these results in the same study, discriminant validity was computed for each measure. The results supported previous findings. More specifically, when correlating the FB inconsistency score with dimensions of the FB scale scores, eight significant negative relationships emerged. Only one positive relationship emerged when correlating the Likert inconsistency score with the dimensions of the Likert scale scores. It should be noted that the power to detect a population correlation of .3 was .84 for these analyses, based on a sample size of 90 per correlation. So the failure to find significant correlations was probably not due to small sample sizes. However, one possibility for a lack of significant correlations for the Likert

inconsistency measure could be due to the low reliability of the measure (Table 1). These findings further provide support for the notion that these two measures of inconsistency are measuring different kinds of inconsistency. Otherwise one would expect the two measures to have similar discriminant validities with their respective scale scores. Instead, the FB based inconsistency measure exhibited a lack of discriminant validity, whereas the Likert based inconsistency measure exhibited discriminant validity.

The final purpose of this study was to expand the literature on inconsistent responding by examining the relationship between inconsistency and other constructs thought to influence the degree to which how inconsistently a person responds. The first relationship examined was between inconsistency and social desirability. It was hypothesized that individuals who respond in a socially desirable manner would make their responses more consistent, and therefore there would be a negative relationship. This hypothesis was given only slight support by the fact that one significantly negative relationship was found between the Likert inconsistency score and impression management ($r = -.236, p < .05$), as shown in table 5. Since two positive correlations with BIDR scores were also found, the overall pattern of results gives little support to this hypothesis. At the present time, we have no explanation for the lack of significant findings, other than it could be due to the low number of participants in the study.

The second part of expanding the literature on inconsistent responding was to examine the relationship between ADHD and inconsistency. Since ADHD is characterized by inattentiveness and impulsivity, it was hypothesized that individuals with ADHD will have a greater inconsistency score. There were no significantly positive relationships found between Likert inconsistency and ADHD. However, significantly positive relationships were found between the FB inconsistency score and ADHD. This provides further support for the notion that

FB inconsistency may be measuring personality inconsistency because ADHD is also characterized by impulsiveness, which may result in inconsistent behavior.

Study Limitations

One limitation of this study was the low power, particularly for the correlation between the BIDR scores and inconsistency and the correlation between ADHD and inconsistency. For example, if it was expected that the correlation between BIDR and inconsistency would have a population r to equal about .2, then to achieve a power of .80, 280 participants would have been needed. In this case, 175 participants were used, achieving a power of .47.

Future Directions

Keeping this in mind, it would be interesting to see future research replicate this study with more participants to see if the larger samples would lead to more statistically significant relationships of social desirability and ADHD to inconsistency. More importantly, the correlations between the two measures of inconsistency were non-existent, suggesting that these two measures of inconsistency measure different kinds of inconsistency- response inconsistency measured by Likert inconsistency and personality inconsistency measured by FB inconsistency. As such, this study should be replicated to examine the stability of these results.

Future research should also focus on confirming or disconfirming the notion that FB inconsistency measures personality. One method of examining this notion might be to administer the Big Five at two different time periods, and compute the standard deviation of the scale scores between the two time periods; larger standard deviations would suggest greater inconsistency in personality between the two time periods. Then, the correlation of the standard deviation of the scale scores with the FB inconsistency measure could be computed. One would

expect a positive correlation- those high in FB inconsistency will also have a larger standard deviation score between the two time periods.

Another focus of this study was to expand the nomological network of inconsistency as a construct by examining the relationship of inconsistency with socially desirable responding and ADHD. Future research should also further expand this network by examining the relationship of inconsistency with other constructs that would be thought to influence inconsistency. Finally, the relationship between ADHD and inconsistency should be examined again using a different measure of ADHD. It would be interesting to see if these results are replicated- that is whether there continues to be a relationship of ADHD with FB inconsistency, but not Likert inconsistency.

REFERENCES

- Adams, Z. W., Roberts, W. M., Milich, R., Fillmore, M. T. (2011). Does response variability predict distractibility among adults with attention-deficit/hyperactivity disorder? *Psychological Assessment*, 23(2), 427-436
- Allport, G.W. & Odbert, H.S. (1936). Trait-names: A psycho-lexical study. *Psychological Monographs*, 47(211).
- American Psychiatric Association. (2000). *Diagnostic and statistical manual of mental disorders* (4th ed.). Washington DC: Author.
- Baird, B. M., Le, K., & Lucas, R. E. (2006). On the nature of intra-individual personality variability: Reliability, validity, and associations with well-being. *Journal of Personality and Social Psychology*, 90(3), 512-527.
- Barkley, R. A. (1997). *ADHD and the nature of self-control*. New York, NY: Guilford Press.
- Barkley, R. A., Murphy, K. R., & Fischer, M. (2008). *ADHD in adults: What the science says*. New York, NY: Guilford Press.
- Baumeister, R.F. (1991). On the stability of variability: retest reliability of metatraits. *Personality and Social Psychology Bulletin*, 17, 633-639.
- Baumeister, R.F., & Tice, D.M. (1988). Metatraits. *Journal of Personality*, 56, 571-598.
- Biderman, M.D., & Reddock, C. M. (2012). The relationship of scale reliability and validity to respondent inconsistency. *Personality and Individual Differences*, 52, 647-651.
- Booksh, R. L., Pella, R. D., Singh, A. N., & Gouvier, W. D. (2010). Ability of college students to simulate ADHD on objective measures of attention. *Journal of Attention Disorders*, 13, 325-338.
- Britt, T. (1993). Metatraits: Evidence relevant to the validity of the construct and its implications. *Journal of Personality and Social Psychology*, 65, 554-562.
- Castellanos, F. X., & Tannock, R. (2002). Neuroscience of attention deficit/ hyperactivity disorder: The search for endophenotypes. *Nature Reviews Neuroscience*, 3, 617-628.

- Costa, P. T., McCrae, R. R., & Arenberg, D. (1980). Enduring dispositions in adult males. *Journal of Personality and Social Psychology*, 38(5), 793-800.
- Edwards, B. D., & Woehr, D. J. (2007). An examination and evaluation of frequency-based personality measurement. *Personality and Individual Differences*, 43, 803-814.
- Ferguson, C. J. (2010). A meta-analysis of normal and disordered personality across the life span. *Journal of Personality and Social Psychology*, 98(4), 659-667.
- Fleeson, W. (2001). Toward a structure- and process-integrated view of personality: traits as density distributions of states. *Journal of personality and social Psychology*, 80, 1011-1027.
- Fleisher, M. S., Woehr, D. J., Edwards, B. D., & Cullen, K. L. (2011). Assessing within-person personality variability via frequency estimation: more evidence for a new measurement approach. *Journal of Research in Personality*.
- Goldberg, L. R. (1999). A broad-bandwidth, public domain, personality inventory measuring the lower-level facets of several five-factor models. In I. Mervielde, I. Deary, F. De Fruyt, & F. Ostendorf (Eds.), *Personality Psychology in Europe*, Vol. 7 (pp. 7-28). Tilburg, The Netherlands: Tilburg University Press.
- Goldstein, S., & Ellison, A. T. (2002). *Clinician's guide to adult ADHD: Assessment and intervention*. Boston, MA: Academic Press.
- Hampson, S. E., & Goldberg, L. R. (2006). A first large cohort study of personality trait stability over the 40 years between elementary school and midlife. *Journal of Personality and Social Psychology*, 91(4), 763-779.
- Harrison, A. G., Edwards, M. J., & Parker, K. C. H. (2007). Identifying students faking ADHD: Preliminary findings and strategies for detection. *Archives of Clinical Neuropsychology*, 22, 577-588.
- Helson, R., & Moane, G. (1987). Personality change in women from college to midlife. *Journal of Personality and Social Psychology*, 53(1), 176-186.
- Hopwood, C. J., Donnellan, M., Blonigen, D. M., Krueger, R. F., McGue, M., Iacono, W. G., & Burt, S. (2011). Genetic and environmental influences on personality trait stability and growth during the transition to adulthood: A three-wave longitudinal study. *Journal of Personality and Social Psychology*, 100(3), 545-556.
- Kane, J. S. (1986). Performance distribution assessment. In R. A. Berk (Ed.). *Performance assessment: methods and applications* (pp. 237-273). Baltimore, MD: Johns Hopkins University Press.

- Knouse, L. E., Bagwell, C. L., Barkley, R. A., & Murphy, K. R. (2005). Accuracy of self-evaluation in adults with attention-deficit hyperactivity disorder. *Journal of Attention Disorders, 8*, 221–234.
- McGough, J. J., & Barkley, R. A. (2004). Diagnostic controversies in adult ADHD. *American Journal of Psychiatry, 161*, 1948–1956.
- Murphy, K. R., & Gordon, M. (2006). Assessment of adults with ADHD. In R. A. Barkley (Ed.), *Attention deficit hyperactivity disorder: A handbook for diagnosis and treatment* (3rd ed., pp. 425–452). New York, NY: Guilford Press.
- Paulhus, D. L. (2002). Socially desirable responding: the evolution of a construct. In H. I. Braun, D. N. Jackson, & D. E. Wiley (Eds.), *The role of constructs in psychological and educational measurement* (pp.46-69). Mahwah NJ: Lawrence Erlbaum.
- Paulhus D. L. (1991). Measurement and control of response bias. *Measures of personality and social psychological attitudes* [e-book]. San Diego, CA US: Academic Press; 17-59. Available from: PsycINFO, Ipswich, MA. Accessed September 17, 2011.
- Paulhus, D. L. (1984). Two-component models of socially desirable responding. *Journal of Personality and Social Psychology, 46*(3), 598-609.
- Reddock, C. M, Biderman, M. D., & Nguyen, N. T. (2011). Increasing the validity of personality questionnaires. *International Journal of Selection and Assessment, 19*, 119-131.
- Saville, C., Pawling R., Trullinger, M., Daley, D., Intriligator, J., Klein, C. (2011). On the stability of instability: Optimizing the reliability of intra-subject variability of reaction times. *Personality and individual differences, 51*, 148-153.
- Terracciano, A., McCrae, R. R., & Costa, P. r. (2010). Intra-individual change in personality stability and age. *Journal of Research in Personality, 44*(1), 31-37.
- Ward, M. F., Wender, P. H., & Reimherr, F. W. (1993). The Wender Utah Rating Scale: An aid in the retrospective diagnosis of childhood attention deficit hyperactivity disorder. *The American Journal of Psychiatry, 150*(6), 885-890.
- Wender, P. (1995). *Attention-deficit hyperactivity disorder in adults*. New York, NY: Oxford University Press.
- Vaidya, J. G., Gray, E. K., Haig, J., & Watson, D. (2002). On the temporal stability of personality: Evidence for differential stability and the role of life experiences. *Journal of Personality and Social Psychology, 83*(6), 1469-1484.

APPENDIX A

A STANDARD IPIP BIG 5 ITEM IN LIKERT FORMAT

Goldberg, L. R. (1999). A broad-bandwidth, public domain, personality inventory measuring the lower-level facets of several five-factor models. In I. Mervielde, I. Deary, F. De Fruyt, & F. Ostendorf (Eds.), *Personality Psychology in Europe*, Vol. 7 (pp. 7-28). Tilburg, The Netherlands: Tilburg University Press.

Directions: Describe yourself as you generally are now, not as you wish to be in the future. Describe yourself as you honestly see yourself, in relation to other people you know of the same sex as you are, and roughly your same age. So that you can describe yourself in an honest manner, your responses will be kept in absolute confidence. Indicate for each statement whether it is 1. Completely Inaccurate, 2. Fairly Inaccurate, 3. Slightly Inaccurate, 4. Neither Inaccurate nor Accurate, 5. Slightly Accurate, 6. Fairly Accurate, or 7. Completely Accurate as a description of you.

	Completely Inaccurate	Fairly Inaccurate	Slightly Inaccurate	Neither Inaccurate nor Accurate	Slightly Accurate	Fairly Accurate	Completely Accurate
I am the life of the party							

Computing Likert inconsistency:

The IPIP Big Five measures five dimensions: Extraversion, Agreeableness, Conscientiousness, Stability, and Imagination. Each dimension is measured by 10 items on the questionnaire. The example above is an item that measures Extraversion. To compute inconsistency, the standard deviation of each dimension is computed using the 10 items to create a dimensional score. Then, the mean of those dimensional scores are computed to create an overall inconsistency score.

APPENDIX B

A STANDARD IPIP BIG 5 ITEM IN FREQUENCY-BASED FORMAT WITH
MEDIUM, HIGH, AND LOW INCONSISTENCY

Three examples showing the calculation of standard deviation for frequency based items. An example of medium inconsistency, high inconsistency and low inconsistency are given for the same item.

Directions: Of all the opportunities you've had to display this behavior in the past 6 months, think of how frequently this statement was descriptive of your actual behavior at each of the three levels. If you feel that "I am the life of the party" was very descriptive of your behavior 50% of the time, somewhat descriptive of your behavior 35% of the time, and not at all descriptive of your behavior 15% of the time, then your response would look like:

Medium Inconsistency

	% very inaccurate	% neither inaccurate nor accurate	% very accurate
I am the life of the party.	25	70	5

Computed "level" formula:

$$\text{Level} = ((\%VI \times 1) + (\%N \times 3) + (\%VA \times 5)) / 100$$

Computed "level" for this item would be: 3.4= 25*.01 + 70*.04 + 5*.07

Computed variability formula:

$$\text{Variability} = \sqrt{(((1 - \text{Level})^2 \times \%VI) + ((4 - \text{Level})^2 \times \%N) + (((7 - \text{Level})^2 \times \%VA)) / 100)}$$

Computed variability for this item would be:

$$1.53 = \sqrt{(((1 - 3.4)^2 \times 25) + ((4 - 3.4)^2 \times 70) + ((7 - 3.4)^2 \times 5))/100}$$

High Inconsistency

	% very inaccurate	% neither inaccurate nor accurate	% very accurate
I am the life of the party.	50	0	50

Computed "level" formula:

$$\text{Level} = ((\% \text{VI} \times 1) + (\% \text{N} \times 3) + (\% \text{VA} \times 5)) / 100$$

Computed "level" for this item would be: $4.00 = 50 \times .01 + 0 \times .04 + 50 \times .07$

Computed variability formula:

$$\text{Variability} = \sqrt{(((1 - \text{Level})^2 \times \% \text{VI}) + ((4 - \text{Level})^2 \times \% \text{N}) + (((7 - \text{Level})^2 \times \% \text{VA})) / 100}$$

Computed variability for this item would be:

$$3.00 = \sqrt{(((1 - 4.00)^2 \times 50) + ((4 - 4.00)^2 \times 0) + ((7 - 4.00)^2 \times 50)) / 100}$$

Low Inconsistency

	% very inaccurate	% neither inaccurate nor accurate	% very accurate
I am the life of the party.	0	100	0

Computed "level" formula:

$$\text{Level} = ((\%VI \times 1) + (\%N \times 4) + (\%VA \times 7)) / 100$$

Computed "level" for this item would be: $4.00 = 0 \times .01 + 100 \times .04 + 0 \times .07$

Computed variability formula:

$$\text{Variability} = \sqrt{(((1 - \text{Level})^2 \times \%VI) + ((4 - \text{Level})^2 \times \%N) + (((7 - \text{Level})^2 \times \%VA)) / 100}$$

Computed variability for this item would be:

$$0 = \sqrt{(((1 - 4.00)^2 \times 0) + ((4 - 4.00)^2 \times 100) + ((7 - 4.9)^2 \times 0)) / 100}$$

APPENDIX C

AN ITEM FROM THE BALANCED INVENTORY OF SOCIALLY DESIRABLE
RESPONDING (BIDR)

Paulhus D. Measurement and control of response bias. *Measures of personality and social psychological attitudes* [e-book]. San Diego, CA US: Academic Press; 1991:17-59. Available from: PsycINFO, Ipswich, MA. Accessed September 17, 2011.

Directions: Below is a set of statements about how people may describe themselves. Place a check mark in the bubble that represents how accurately the statement describes you.

Not True			Somewhat True			Very True
1	2	3	4	5	6	7

_____ 1. I never drop litter in the street.

APPENDIX D

AN ITEM FROM THE WENDER UTAH RATING SCALE (WURS)

Ward, M. F., Wender, P. H., & Reimherr, F. W. (1993). The Wender Utah Rating Scale: An aid in the retrospective diagnosis of childhood attention deficit hyperactivity disorder. *The American Journal of Psychiatry*, 150(6), 885-890.

Directions: Finally, we're interested in how you were as a child. Although it's been many years since you were a child, try to remember how it was and respond to the following items.

Not at All or Very Slightly	Mildly	Moderately	Quite a Bit	Very Much
--------------------------------	--------	------------	-------------	-----------

As a child, I was or (had)

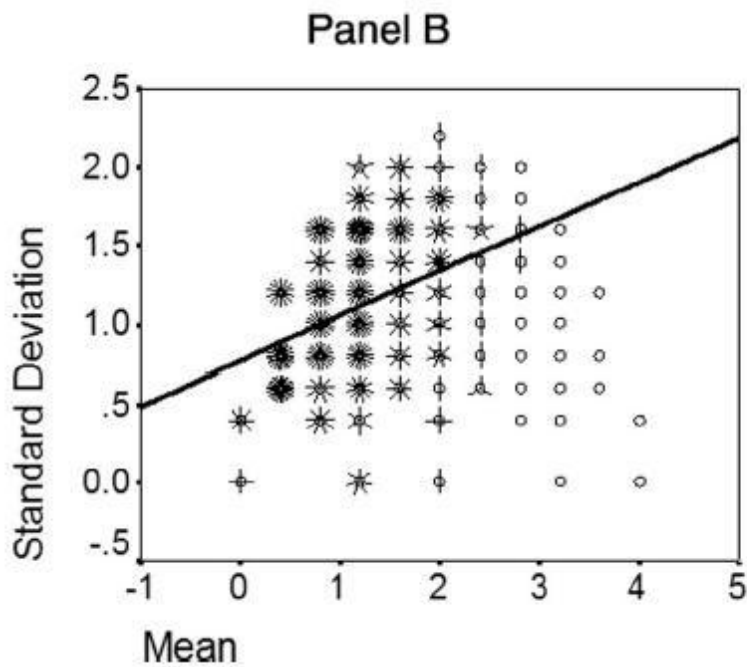
___ 1. Concentration problems, easily distracted.

APPENDIX E

A RELATIONSHIP BETWEEN POSITIVELY SKEWED MEANS AND THEIR
STANDARD DEVIATIONS

Baird, B. M., Le, K., & Lucas, R. E. (2006). On the nature of intra-individual personality variability: Reliability, validity, and associations with well-being. *Journal of Personality and Social Psychology*, 90(3), 512-527.

A graph depicting the relationship between positively skewed means and their standard deviations. The more spokes around a circle, indicate a larger mean. From Baird, Le, and Lucas (2006).



APPENDIX F

A VISIUAL DEPICTION SHOWING HOW THE FOUR CONDITIONS IN THIS STUDY
WERE BALANCED

One-fourth of the respondents responded to the Original 50 IPIP items using the Likert response format and then to the Other 50 IPIP items using the FB response format. A second fourth responded to the Original 50 IPIP items using the FB response format and then to the Other 50 IPIP items using the Likert response format. A third quarter of the respondents responded to the Other 50 IPIP items using the Likert response format and then to the Original 50 IPIP items using the FB response format. The final fourth of the respondents responded to the Other IPIP items using the FB response format and then to the Original 50 IPIP items using the Likert response format

Frequency Based Original 50 IPIP Items	Likert Other 50 IPIP Items	Condition = 1
Likert Other 50 IPIP Items	Frequency Based Original 50 IPIP Items	Condition = 2
Frequency Based Other 50 IPIP Items	Likert Original 50 IPIP Items	Condition = 3
Likert Original 50 IPIP Items	Frequency Based Other 50 IPIP Items	Condition = 4

VITA

Rizwan Ahmed Khan is from Ellicott City, Maryland, where he has resided most of his life. He is the youngest of four siblings, two older brothers and one older sister. He graduated from the University of Maryland, College Park in December 2009, with a Bachelor of Science in Psychology. During his undergraduate career, he taught experimental and statistical software as a teaching assistant and was involved with research regarding gender roles as a research assistant. Rizwan began his graduate studies in August 2010 at the University of Tennessee at Chattanooga. He has taught two semesters of Statistics in Psychology Laboratory, was the Director of Communications for the Society of Human Resources UTC chapter, and has offered pro-bono consulting services. Rizwan graduates in May 2012 with a Master of Science in Psychology: Industrial – Organizational Psychology.