STABLE INCONSISTENCY: A STUDY OF RESPONSE INCONSISTENCY OVER TIME

By

James Nauert

Michael D. Biderman
Professor of Psychology
(Chair)

Christopher J. L. Cunningham
I-O Psychology Graduate Program Director
(Committee Member)

Katherine H. Rogers
Professor of Psychology
(Committee Member)
STABLE INCONSISTENCY: A STUDY OF RESPONSE INCONSISTENCY OVER TIME

By

James Nauert

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

The University of Tennessee at Chattanooga
Chattanooga, Tennessee

May 2017
ABSTRACT

Personality measures are currently a popular method for selection in the business world, despite issues such as poor predictive ability and the potential for output manipulation by participants. Another issue with personality testing that is often overlooked is that individuals sometimes respond differently on test items that otherwise measure the same traits. This phenomenon has been called Response Inconsistency (Reddock, Biderman, & Nguyen, 2011). The focus of this study is to attempt to show the phenomenon as a measurable trait that is stable over time. The study administered two different Big Five Inventories to participants taken from a local university. The two inventories were administered at different dates. Inconsistency measured as the mean standard deviation of responses within domains was computed and then correlated across the two time periods to determine reliability of the inconsistency measure. High correlations supported the hypothesis that response inconsistency is stable across time.
ACKNOWLEDGEMENTS

Special thanks to my thesis chair Dr. Biderman for providing me with everything I needed to make it through this trying time. Thanks to my committee members Dr. Cunningham and Dr. Rogers for advising me throughout the thesis process and laughing at my jokes. Additional thanks to Dr. O’Leary for assisting with political issues.
REFERENCES .................................................................................................................. 22

APPENDIX

A. IRB APPROVAL LETTER FOR RESEARCH ............................................................. 30

B. EXAMPLE BFI-2 INVENTORY ASSESSMENT ....................................................... 33

C. EXAMPLE HEXACO 60-ITEM QUESTIONNAIRE .................................................. 38

VITA .................................................................................................................................. 42
LIST OF TABLES

1. Mean Correlational Matrix Between Scale Scores ........................................26
2. Response Inconsistency Correlational Matrix .................................................27
3. Regression, BFI .................................................................................................28
4. Regression, HEXACO .......................................................................................29
LIST OF ABBREVIATIONS

AD-HD, Attention Deficit-Hyperactivity Disorder

BFI – 2, Big Five Inventory – 2

IER, Insufficient Effort Responding

IPIP, International Personality Item Pool

I/O, Industrial/Organizational
CHAPTER I
INTRODUCTION

Personality psychology seeks to understand people, and what makes them different from one another (Funder, 2001). It is one of the more far-reaching branches of psychological theory, with roots in both nature and nurture. Personality psychology has applications in many different fields of psychology, from developmental to social psychology. It can allow us to have a better understanding of those around us (McAdams, 1995) and helps us to make sense of them as people. Personality factors can even be used to predict important life events, such as mortality and divorce rate (Roberts, Kuncel, Shiner, Caspi, & Goldberg, 2007). Personality psychology is important to understand because it gives us insight into ourselves and the people we surround ourselves with.

A central theory of personality is known as trait theory. Allport and Odbert (1936) are often identified as some of the more prominent influences on the development of trait theory (Piedmont & Aycock, 2007). They created a list of words, or a taxonomy, that describe people. This was a vital step in personality research, as it gave researchers a means to communicate with each other using the same vocabulary (John, Naumann, & Soto, 2008). From there, researchers began trimming down the list of words by pairing similar terms into groups. As research progressed, these groups became larger, more encompassing, eventually becoming general categories of behaviors known as traits. Trait theory suggests that traits are measureable, fundamental elements of personality that reflect the most important ways people can differ. The
number of these traits, as well as what they are, has been the subject of debate. Currently, one of the most agreed upon trait frameworks of this sort is the five-factor model, also known as the Big Five (Digman, 1990; Gnambs, 2015; Karwowski & Lebuda, 2016).

The five-factor model of personality emphasizes five continuous trait dimensions, across which key aspects of personality can be measured (Goldberg, 1993). These traits are a person’s openness to experience, conscientiousness, extraversion, agreeableness, and neuroticism, sometimes referred to positively as emotional stability (P. T. Costa, Jr. & McCrae, 2009). A person can measure as high or low in each of these traits, and in theory each trait is roughly independent of each other. A similar trait theory is that of the six-factor HEXACO model of personality. As the name might suggest this model contains all of the five traits previously mentioned plus a sixth trait known alternatively as either Humility or Honesty (Lee & Ashton, 2004).

**Personality in Business**

A strong trend in business is toward the utilization of personality measures for selection of employees (Sackett & Lievens, 2008; Schmit & Ryan, 1993). Meta-analyses have shown that each of the five traits discussed in the previous section predict desirable behaviors in an organizational setting (Hurtz & Donovan, 2000; Parks-Leduc, Feldman, & Bardi, 2015). For example, conscientiousness and agreeableness are considered good general predictors of job performance, meaning that individuals who display higher levels of those traits will generally succeed at their job more often (Sackett & Walmsley, 2014; Schmidt & Hunter, 1998; Witt, Burke, Barrick, & Mount, 2002). In other words, more conscientious or agreeable people tend to perform their jobs better than less conscientious or agreeable people.
Other traits also have utility in predicting specific behaviors. Individuals who are open to new experiences have been shown to benefit more from training programs (Ziegler et al., 2014). Extraversion is seen as a useful trait mainly in jobs where employees are expected to be in regular contact with others, such as in sales occupations (Salgado, 1997). Finally, Emotional Stability is another trait that generally predicts good performance, however businesses are reluctant to select based on this trait, as it can bring up the uncomfortable topic of mental wellness (Sackett & Walmsley, 2014).

Personality traits can also be used to predict organizational behaviors that are harder to quantify. More extraverted people tend to show more commitment to an organization (Zhang, 2015). Personality tests can also identify which employees are more likely to hurt the company through counter-productive work behaviors. These behaviors can include stealing pens, coming in to work late and so on (Salgado, 2002). Additionally, personality factors can predict which employees will go above and beyond to help out their coworkers (Chiaburu, Oh, Berry, Li, & Gardner, 2011). Measuring personality can give us the ability to predict these behaviors over and above other factors external to the person, such as organizational culture. Because of these factors, personality testing has become a popular measure to include in selection batteries.

**Personality-Related Selection Issues**

Though benefits of understanding employee personality are obvious, personality testing may not be the best method to use for every business. Research has shown that the predictive validity of conscientiousness, widely considered the best general predictor of job performance from among the Big Five, is substantially less than that of General Mental Ability (Schmidt & Hunter, 1998). It is important to realize that businesses should use powerful predictors of job
performance - not only because businesses want to find the best possible candidates for their organization, but there is also an important legal consideration. Court cases such as McAdams (1995) have ruled that tests that have been proven to predict job performance are sufficient evidence against discriminatory intent in disparate impact cases. This is not to say that personality testing is somehow “bad”, but one issue is that other kinds of selection tests may give employers better predictive power.

Another issue in personality testing is that most tests utilize self-report measures (Widhiarso & Himam, 2015). Self-report data consist typically of surveys where individuals report their own feelings rather than getting data from an outside observer. While this provides insight into the subject’s internal psychological processes, there are two key assumptions that should be met before self-report measures are used. The first assumption is that the participants understand themselves well enough to be able to answer or respond accurately (Soto & Tackett, 2015). Given that personality is a complicated structure, it can be difficult for individuals to know themselves well enough to provide an accurate image. The second assumption is that participants will be honest with their responses (Widhiarso & Himam, 2015). To give an example, potential employees realize that how they do on a personality test will affect their chance of landing the job. They may then distort their responses to appear as more attractive candidates. This phenomenon is known as response distortion or “faking good” (Widhiarso & Himam, 2015).

Response Variability

Yet another issue that personality tests face is response variability, or inconsistency. Costa, McCrae, and Arenberg (1980) first introduced the idea that personality was not actually a
group of stable traits. They believed that personality is more variable at certain life stages, such as childhood. They found that some aspects of personality were stable across the lifespan, such as sociability, or general activity. On the other hand, other traits such as emotional stability could vary wildly across time. This variation, as expected, was more pronounced at younger ages. In other words, adults tend to have more stable personality characteristics and trait manifestations than children. What was not expected by the researchers was that as adults grew older, their personalities became more variable. In other words, variability of personality traits is more pronounced in youth and in old age. Though aging was the main issue they examined, Costa et al. argued that it was likely that other factors could influence personality variability.

Fleeson (2001) looked at personality variability and carried it further by suggesting that measurements of traits may not be measuring a set value, but rather the mean of a distribution. Fleeson applied a concept called density distributions into the process of measuring personality traits. This concept of density distributions introduced the dimension of variability into the process of personality testing. He argued that while personality can vary, responses on personality tests still provided a useful measurement. The scores we would typically associate with these personality traits were not static; rather they reflected a mean of behaviors in recent memory. These changes in behavior can be the result of internal processes, but are most visible when they are the result of an outside event.

As an example, imagine an office worker driving home from a job she loves. She is driving very carefully, being considerate of other drivers. Suddenly another driver swerves in front of her, almost running into her before speeding off. Because of this event, suddenly our driver is frustrated. The general situation of driving home has not changed, but because of this poor mood, she starts driving recklessly. According to Fleeson’s theory, the environment has
caused a change in her mental state, which then caused a change in her displayed level of conscientiousness. This serves as a good example for how the outside environment can affect a person, but not every event has to be so extreme. Events that cause change can be big or small, according to Fleeson, as can the changes they cause.

To explain density distributions, think of a snapshot of a person’s life. In a given year you can take many snapshots of this person’s life. These snapshots will rarely show the same thing twice, but if we take enough of them, we can plot out a general map of the person’s life. This map would reflect the variation personality within a person, and can provide an anchor with which researchers could guess where the person is at a given time.

Another variability-related issue is the idea that responses to personality items on questionnaires can vary. Schmidt, Le, and Ilies (2003) attributed some instability of trait measurements to measurement error and categorized that error into three different types. The first, random response error is error caused by variations in the participant’s mental state from moment to moment. According to Schmidt et al., this could be from changes in attention levels, mental efficiency, or even from distractions in the environment, essentially “noise”. Because of these momentary variations, participants can give a different response to the same item if it appears in two different places in the questionnaire.

Schmidt et al. (2003) next discussed a form of error they called transient error. Transient error is similar to what Fleeson (2001) called state changes. This error occurs between two testing periods, as events impact the participant’s mood, feelings, or mental processing ability. This error therefore affects constructs that are measured across a time gap. The changed mental processes then influence the participant’s responses, causing them to change their responses.
Normally these changes are small enough that the overall measurement is not heavily affected, however significant events can cause the overall trait measurement to increase or decrease.

Schmidt et al. (2003)’s final form of error is known as specific factor error. This form of error comes from participant’s unique reactions to elements of the assessment. In other words, the responses of participants are influenced by factors such as the specific wording of the questions and the interaction between the wording and the respondent’s ability to process that wording.

**Response Inconsistency**

Reddock et al. (2011) examined a phenomenon they called response inconsistency. They found that certain individuals gave different answers to similar test items that appeared at different points of a personality survey as might be expected by the Schmidt et al. (2003) concept of random response error. But, contrary to the assumption that random response error is completely random from one item to the next, they found that measures of inconsistent responding correlated positively between all five of the Big Five personality traits in a single test. Reddock et al. argued that it is unlikely that issues in the test content are to blame, as the test they used, the International Personality Item Pool Big Five Inventory (IPIP, BFI), has previously been shown to otherwise have good reliability (Ehrhart, Roesch, Ehrhart, & Kilian, 2008; Zheng et al., 2008). Biderman and Reddock (2012) continued this line of thinking by examining measures of response inconsistency across separate tests administered at the same time. They found that even across different personality tests, response inconsistency measures within persons were correlated. This means that if a participant’s answers were highly variable on one questionnaire, they would likely be variable on another given at the same time. The findings of
these studies indicate that certain individuals may be more prone than others to inconsistent responding, separate from any personality trait.

Recent research has identified a similar phenomenon known as Insufficient Effort Responding, or IER (Huang, Bowling, Liu, & Li, 2015). IER occurs when a responder to a test does not put real effort into answering to the best of their ability. This can occur in two ways. The first form of IER is random responding, which is similar to response inconsistency (Biderman & Reddock, 2012). This occurs when participants answer randomly, as the name would suggest. This is one possible cause of inconsistent responding. The second method of IER involves participants responding consistently, but in ways that don’t make sense. This form of IER occurs when participants perhaps do not fully read, or do not fully care about their response (Huang et al., 2015). They will then give a response that does not hold up to scrutiny I.E. “I was born in the Republic of Texas before it joined the United States”. IER and response inconsistency are related according to Bowling et al. (2016), however the presence of one does not mean the presence of both.

Khan (2012) continued the research on response inconsistency by measuring individuals on two types of inconsistency. The first was “state inconsistency”, variability in states across time as proposed by Fleeson (2001) and perhaps what Schmidt et al. (2003) called transient error. The second was “response inconsistency” (Reddock et al., 2011). Khan (2012) then compared the two on several different factors, including the possibility that Attention Deficit-Hyperactivity Disorder (AD-HD) was a contributing factor to these types of error. Khan found that while measures of state inconsistency positively correlated with measures of AD-HD, response inconsistency from item to item within a scale displayed no relation to the disorder. Additionally, Khan found that response inconsistency was unrelated to measures of desirable
responding suggesting that response inconsistency is not the individual attempting to change their answers to suit their desires.

**Summary and Objectives of Current Study**

Though there has been considerable work examining transient error (Dwight, Wolf, & Golden, 2002; Fleeson, 2001; Saville et al., 2011), to-date there has been little work done in measuring response inconsistency. Even less research has explored response inconsistency as a stable trait within the individual. The limited research that does exist has provided a little evidence that inconsistency may in some ways be a unique trait of the individual. Also shown is that response inconsistency is stable from one domain of the big five to another (Reddock et al., 2011), as well as across multiple questionnaires during a single testing period (Biderman & Reddock, 2012).

To date, there has been no correlation of inconsistency measures from one time to the next. That is, it has yet to be determined if response inconsistency is stable within-person across multiple testing periods. To determine if this is the case, the current study involves the application of two separate personality inventories, with a several-week span of time between the two administrations. Though the previously mentioned studies involved only one time period, the expectation is that we will find high test-retest reliability, showing that response inconsistency is relatively stable across time.

\( H_1: \) Measures of respondent inconsistency obtained between the first and second testing periods will be significantly, positively correlated.
CHAPTER II

METHODOLOGY

Participants

The sample consisted of 233 undergraduate- and graduate-level college students from The University of Tennessee at Chattanooga. The participants joined through the university’s Sona research participation system. No incentives were offered by the researcher to encourage participation, but participating students may have been able to earn extra credit points in certain courses (at the discretion of separate course instructors).

Five individuals participated in the final survey without participating in the prescreen, and their data had to be discarded. If participants took the second study less than two weeks after taking the prescreen, their data were also discarded to ensure their consistency correlations were not inflated due to memory; this rule lead to the exclusion of 56 participants. Eight participants had a standard deviation of response that was zero across all domains. In other words, they gave the same answer for every response across the entire questionnaire. While it is possible that these are their honest responses, it is far more likely that these participants were not giving sufficient effort in their responses (Bowling et al., 2016). The removal of these participants brought the final sample pool to 181 individuals. Of the participants who provided demographic information ($N = 181$), the mean age of the participants was 20.31 ($SD = 4.911$). Of these remaining participants, 43 were male, while 138 were female. 88% of the sample self-identified
as non-Hispanic white, 8% self-identified as non-Hispanic black, 2% identified as a non-Hispanic mix of multiple races, and 2% did not want to provide racial information.

**Method**

The data for this study were gathered with two different personality inventories. Both inventories were applied through the university Sona system, which is entirely internet based. For the purposes of this study, inconsistency is operationalized as the standard deviation of responses within a given personality trait domain. The first test period (referred to hereon as the prescreen) used the HEXACO 60-item personality scale (Lee & Ashton, 2004). The HEXACO 60 item test measures the scales of the Big Five (Openness, Agreeableness, Conscientiousness, Neuroticism, and Extraversion), as well as a sixth scale that is meant to measure Honesty and Humility (See appendix C). This test period opened at the start of the semester in August, and did not close until the end of the semester in December. The second test period used the Big Five Inventory-2 (BFI2) 60-item questionnaire (Soto & John, 2016). The BFI is a measure of the Big Five, and contains 60 questions that ask the participant to view statements that reflect aspects of a personality characteristic, and rate how well those statements apply to them (See Appendix C). The test is divided into five sections, each twelve questions long and each measuring one of the Big Five traits. These questions were measured on a 7 point scale of agreement with a descriptive statement (I.E. “I am someone who has a forgiving nature”). The responses available range from 1 (Strongly Disagree) to 7 (Strongly Agree). Participants are instructed to answer questions openly and honestly, but otherwise are free to answer questions at their own pace.

A different secondary test was chosen to avoid any chance of carry-over effects during the second testing period through participants remembering their responses during the previous
testing period. A study measuring the comparative validity of several five factor and six factor personality measures found that the two questionnaires exhibited convergent validity (Thalmayer, Saucier, & Eigenhuis, 2011). In other words, trait scores on HEXACO questionnaires correlate with trait scores on Big Five questionnaires. For the research here, the HEXACO was administered as part of a Prescreen questionnaire given to all students participating in research. Similar to the BFI-2, participants read descriptive statements (I.E. “I feel I am an unpopular person”) and rates their agreement on a 7 point scale. This application also queries participants on demographic information, such as race, sex, and age. Participants are allowed to decline to provide this information if they do not feel comfortable providing it. Additionally, the prescreen assigns each participant an identification number through which their scores can be linked to other studies. No personally identifiable information was collected or stored for this project.

**Procedure**

The data collection period began at the start of the 2016 fall semester at UTC. Participants who signed up to participate in research projects first took the prescreen questionnaire, which included the HEXACO and demographic questions. The second collection period became available to participants in early October, and continued until the end of the school year in December. This time period was selected to provide enough time between the first and second applications of the test to reduce any possible effects caused by memory of the prescreen (Nunnally & Bernstein, 1994). October was chosen for the beginning of the second questionnaire to provide as long a time period as possible between the first and second time
periods while at the same time reducing the confounding impact of end of the year events, such as final examinations (Fleeson, 2001; Schmidt et al., 2003).

Upon signup, participants were instructed to read informed consent sheets informing them of possible risks, as well as detailing their right to stop participating at any time. After they read and signed the sheet, and provided non-identifying background information, they were provided with a link to take the BFI-2 test. Once participants finished the questionnaire, they were thanked for participating, and given a debriefing form detailing the purpose of the experiment, as well as means to contact the researcher with any questions. The survey is meant to be completed in a single sitting. There is no set time limit, but the participant cannot leave the survey and return later to finish it. If a participant is unable to finish the test before they leave, they are allowed to submit it early. Any participants who do this will receive credit if they have completed at least the first half of the questionnaire.
CHAPTER III
RESULTS

Convergent Validity

The mean time difference between testing periods was 45.5 days ($SD = 18.1$). Participant’s personality trait scores were computed and compared across both questionnaires to assess convergent validity between the measures of identical traits from the two questionnaires. Of the convergent validity correlations, all were moderately correlated $r(179) > .5, p < .001$ in absolute value. It is important to note that the correlation between the HEXACO’s Emotional Stability and the BFI’s Neuroticism is negative because the prescreen HEXACO application was reverse scored from the typical HEXACO emotionality scale. Essentially, the results summarized in Table 1 indicate that the two personality questionnaires used for the present study measured the same concepts. Large correlations, as we see here, show that the two questionnaires are similar enough that we can reliably compare the two sets of responses.

Inconsistency

The measures of response inconsistency within the two personality assessments were computed by obtaining standard deviations of a participant’s responses to items within each personality domain. The standard deviation has been determined to be an effective method of measuring item inconsistency, on par with more complicated analyses such as examining skewness and kurtosis (Saville et al., 2011). This procedure was performed for each domain for each questionnaire. The standard deviation of responses in each personality trait domain was
then correlated with its counterpart on the other questionnaire. Table 2 shows the correlation matrix between every domain. All correlations between Time 1 scale inconsistency values and Time 2 scale inconsistency values were significantly larger than 0 (p < .05). Though the correlations were relatively small, these results support hypothesis 1.

Next, mean inconsistency measures for each questionnaire were computed. Each questionnaire’s inconsistency measure was computed by averaging the domain standard deviations across the domains assessed by the questionnaire. For the prescreen results, this meant averaging the six HEXACO domains, while for the second set of results the five BFI items were averaged. The mean BFI Inconsistency Measure was 1.56 (SD = .35), while the mean HEXACO measure was 1.26 (SD = .34). We also computed the reliability of each mean inconsistency scale. The BFI Inconsistency scale had an alpha value of .723, while the HEXACO inconsistency scale had an alpha value of .804. The high alpha value provides more support that inconsistency is stable across time. The two main inconsistency measures were then correlated to assess stability of overall consistency between questionnaires. The correlation $r(179) = .506$ was moderately strong and significantly larger than 0 ($p < .001$).

**Trait Items’ Relation to Inconsistency**

To further establish the nomological net of inconsistency, analyses going beyond the original focus of the study were conducted. An overall inconsistency rating for each participant was computed by averaging their two questionnaire inconsistency scores. The trait scores from each questionnaire were then used as independent variables in a regression, while the overall inconsistency was used as the dependent variable. This would assess if any personality traits might be significant predictors of inconsistency. The results of the analysis were not consistent
(see table 3 and 4 for the regressions of the HEXACO and the BFI, specifically). The regression for the BFI questionnaire indicates that Neuroticism, Agreeableness and Openness are significant (p < .05). The regression for the HEXACO indicates that only Extroversion is significant (p < .05).
CHAPTER IV
DISCUSSION AND CONCLUSION

The purpose of this study was to assess whether response inconsistency is a stable construct across time. Though there has been some research on inconsistent responding, research on inconsistency as a stable, measurable trait of the individual is sparse. While previous studies confirm response inconsistency as a stable phenomenon across questionnaires within short time periods, no studies to date included a time factor (Biderman & Reddock, 2012). This study adds to the knowledge base of test-taking behavior by filling a gap in previous research.

The main findings of this study can be found in Table 2. The main hypothesis, that inconsistency measures for each trait would be significantly, positively correlated was supported. The convergent validity of overall inconsistency is similar to that of trait scores, \( r(179) = .510, p < .01 \). This correlation shows that whatever characteristic of the participant causes inconsistency is stable over a period of time, almost as stable as would be expected from a personality trait.

**Causes of Inconsistency**

Though the overall inconsistency measures correlated, we can currently only theorize about the cause of the correlation. Within each scale, it seems likely that some portion of the inconsistency we see is due to random response error (Schmidt et al., 2003). If inconsistency were due only to a completely random process, we would not expect to find correlations between the measures of inconsistency with the delay we used here. Since positive correlations were
found it could be that certain individuals are consistently more prone to random response error than others.

It is also possible that certain individuals have consistent individual differences of specific factor error. That is, some individuals are more prone to interpret items in unusual ways leading to idiosyncratic responses regardless of item content while others are less likely to give unusual interpretations to items. If the consistency of inconsistency is due to individual differences in specific factor error, the most likely explanation is that there are certain quirks used in personality questionnaires that some people find difficult to understand. This then causes inconsistency in responses as participants misunderstand what we are trying to assess. The consistency of inconsistency could also be a combination of random response and specific factor error. The two forms of error are not mutually exclusive, so it is probable that both forms have some degree of impact on individual inconsistency.

It is unlikely that transient error is involved in the inconsistency within only the BFI or only the HEXACO, as each questionnaire was given all at once. Where Transient error does come into play is in the stability of inconsistency across time periods. In the time delay used in this study, it is likely that the participants have changed at least somewhat and that transient error change may have changed the response inconsistency from Time 1 to Time 2, which could explain why the correlations of inconsistency values for each domain between Time 1 and Time 2 were relatively weak. With no changes due to transient error, random response error and specific factor error should relatively stay the same across the time delay.

An interesting finding was that not all of the correlations of standard deviations between the BFI and HEXACO were of equal strength. Given previous research, we would expect each measure of inconsistency across questionnaires to be relatively similar to one another (Reddock
et al., 2011). However, the within-trait (see bolded on table 2) inconsistency scores were typically stronger than between-trait inconsistency scores. It is possible that this is related to specific factor error (Schmidt et al., 2003). It is possible that specific factor errors are specific to each domain, so that individual differences in inconsistency due to specific factor errors will be slightly more highly correlated with individual differences in inconsistency the same domain than with individual differences in inconsistency in responses to other domains.

Limitations

There are several potential limitations for this study. The first was that the sample pool consisted of predominantly white, young adult college students. Taking a personality test to get extra credit for a class provides a different environment than if they were taking it as part of a selection battery. Furthermore, a greater limitation ties back to Costa et al. (1980). Recall that Costa et al. found that younger adults are less stable than older adults. It stands to reason that our sample, consisting of college-age students would show higher levels of response inconsistency than the general population. Further research should test if response inconsistency is as consistent in older populations. Additionally, there was a lack of experimental controls for the time factor. Though a hard limit of two weeks was set, participants took the survey at many different points during the semester. No controlled intervals were used. This was mostly a convenience issue, as there was no way to control when they took the prescreen, and therefore no easy way to control the time gap. Future research should utilize hard time controls, with different groups given the second test at different time intervals, for instance three different groups given the second questionnaire two weeks, one month, and two months after the first questionnaire. This can provide a better idea of how inconsistency can change over time.
Future Research

These findings point out several avenues for future research. Now that we have evidence that inconsistency is a stable trait of the individual, we should look more deeply for possible causes and relations to inconsistency. As Table 3 and 4 show, there are no consistent patterns relating general inconsistency to Big Five or HEXACO domains. While neuroticism, openness, and agreeableness scores on the BFI were significant predictors of inconsistency, their counterparts on the HEXACO were not. The same holds true for the extroversion scale of the HEXACO. Given that the two scales exhibit convergent validity, it is likely that the different pattern of results from the BFI vs the HEXACO is due to the relatively small sample size. A larger sample size would have likely provided more valid estimates of the Big Five and HEXACO traits and lead to smaller differences in the patterns of results shown in Tables 3 and 4. This, this may be the subject of future research.

Finally, we should take a closer look at why within-domain inconsistency (for example BFI Openness to HEXACO Openness) is stronger than across-domain inconsistency values (BFI Conscientiousness to HEXACO Agreeableness). If these findings are correct, and within-domain inconsistency tends to be stronger, why might that be? It could be related to specific factor error in that some individuals have more difficulty understanding, as an example, extroversion items while they have less difficulty understanding the other four factors. This would cause them to give more variable responses to extroversion scale questions, inflating the within-domain inconsistency scores. This phenomenon should be explored, as it could impact future personality questionnaire development.
I-O psychologists should also focus on the impact that inconsistency can have on the applied side of personality testing. Do people who answer inconsistently in personality tests also perform more inconsistently in their organization? Perhaps it could mean a person might be more variable in their academic or career interests? We have established that response inconsistency sticks with the individual across time, and now we must look into what value that can provide.
REFERENCES


Sackett, P. R., & Walmsley, P. T. (2014). Which personality attributes are most important in the workplace? *Perspectives on Psychological Science, 9*(5), 538.


Table 1 – Mean Correlational Matrix Between Scale Scores

<table>
<thead>
<tr>
<th>Trait Score Correlations</th>
<th>BFIE</th>
<th>BFIA</th>
<th>BFIC</th>
<th>BFIN</th>
<th>BFIO</th>
<th>hx</th>
<th>ha</th>
<th>hc</th>
<th>hs</th>
<th>ho</th>
<th>hh</th>
</tr>
</thead>
<tbody>
<tr>
<td>BFIE Pearson Correlation</td>
<td>1</td>
<td>.060</td>
<td>.225**</td>
<td>-.243**</td>
<td>.242**</td>
<td>.652**</td>
<td>-.126</td>
<td>.127</td>
<td>.078</td>
<td>.083</td>
<td>.059</td>
</tr>
<tr>
<td>BFIA Pearson Correlation</td>
<td>.060</td>
<td>1</td>
<td>.323**</td>
<td>-.253**</td>
<td>.289**</td>
<td>.162*</td>
<td>.585**</td>
<td>.191*</td>
<td>-.118</td>
<td>.125</td>
<td>.353**</td>
</tr>
<tr>
<td>BFIC Pearson Correlation</td>
<td>.225**</td>
<td>.323**</td>
<td>1</td>
<td>-.262**</td>
<td>.080</td>
<td>.105</td>
<td>.148*</td>
<td>.648**</td>
<td>.034</td>
<td>-.121</td>
<td>.321**</td>
</tr>
<tr>
<td>BFIN Pearson Correlation</td>
<td>-.243**</td>
<td>-.253**</td>
<td>-.262**</td>
<td>1</td>
<td>-.071</td>
<td>-.390**</td>
<td>-.237**</td>
<td>-.105</td>
<td>-.529**</td>
<td>-.067</td>
<td>-.045</td>
</tr>
<tr>
<td>BFIO Pearson Correlation</td>
<td>.242**</td>
<td>.289**</td>
<td>.080</td>
<td>-.071</td>
<td>1</td>
<td>.190*</td>
<td>.202*</td>
<td>.174*</td>
<td>.037</td>
<td>.762**</td>
<td>.117</td>
</tr>
<tr>
<td>hx Pearson Correlation</td>
<td>.652**</td>
<td>.162*</td>
<td>.105</td>
<td>-.390**</td>
<td>.190*</td>
<td>1</td>
<td>.062</td>
<td>.114</td>
<td>.083</td>
<td>.173*</td>
<td>.061</td>
</tr>
<tr>
<td>ha Pearson Correlation</td>
<td>-.126</td>
<td>.585**</td>
<td>.148*</td>
<td>-.237**</td>
<td>.202**</td>
<td>.062</td>
<td>1</td>
<td>.136</td>
<td>.087</td>
<td>.178*</td>
<td>.310**</td>
</tr>
<tr>
<td>hc Pearson Correlation</td>
<td>.127</td>
<td>.191*</td>
<td>.648**</td>
<td>-.105</td>
<td>.174*</td>
<td>.114</td>
<td>.136</td>
<td>1</td>
<td>.015</td>
<td>.082</td>
<td>.350**</td>
</tr>
<tr>
<td>hs Pearson Correlation</td>
<td>.078</td>
<td>-.118</td>
<td>.034</td>
<td>-.529**</td>
<td>.037</td>
<td>.083</td>
<td>.087</td>
<td>.015</td>
<td>1</td>
<td>.147*</td>
<td>-.084</td>
</tr>
<tr>
<td>ho Pearson Correlation</td>
<td>.083</td>
<td>.125</td>
<td>-.121</td>
<td>-.067</td>
<td>.762**</td>
<td>.173*</td>
<td>.178*</td>
<td>.082</td>
<td>.147*</td>
<td>1</td>
<td>-.024</td>
</tr>
<tr>
<td>hh Pearson Correlation</td>
<td>.059</td>
<td>.353**</td>
<td>.321**</td>
<td>-.045</td>
<td>.117</td>
<td>.061</td>
<td>.310**</td>
<td>.350**</td>
<td>-.084</td>
<td>-.024</td>
<td>1</td>
</tr>
</tbody>
</table>

**. Correlation is significant at the 0.01 level (2-tailed).
*. Correlation is significant at the 0.05 level (2-tailed).

Models convergent validity across prescreen HEXACO with BFI. Bolded scores indicate convergent validity between trait domains between questionnaires.
Table 2 – Response Inconsistency Correlational Matrix

<table>
<thead>
<tr>
<th>Inconsistency Correlations</th>
<th>sdBFIE</th>
<th>sdBFIA</th>
<th>sdBFIC</th>
<th>sdBFIN</th>
<th>sdBFIO</th>
<th>sdhx</th>
<th>sdha</th>
<th>sdhc</th>
<th>sdhs</th>
<th>sdho</th>
<th>sdhh</th>
</tr>
</thead>
<tbody>
<tr>
<td>sdBFIE Pearson Correlation</td>
<td>1</td>
<td>.249**</td>
<td>.355**</td>
<td>.357**</td>
<td>.247**</td>
<td>.316**</td>
<td>.215**</td>
<td>.236**</td>
<td>.209**</td>
<td>.262**</td>
<td>.155**</td>
</tr>
<tr>
<td>sdBFIA Pearson Correlation</td>
<td>.249**</td>
<td>1</td>
<td>.430**</td>
<td>.352**</td>
<td>.405**</td>
<td>.209**</td>
<td>.221**</td>
<td>.120</td>
<td>.104</td>
<td>.266**</td>
<td>.099</td>
</tr>
<tr>
<td>sdBFIC Pearson Correlation</td>
<td>.355**</td>
<td>.430**</td>
<td>1</td>
<td>.334**</td>
<td>.313**</td>
<td>.245**</td>
<td>.177</td>
<td>.333**</td>
<td>.162</td>
<td>.261**</td>
<td>.186</td>
</tr>
<tr>
<td>sdBFIN Pearson Correlation</td>
<td>.357**</td>
<td>.352**</td>
<td>.334**</td>
<td>1</td>
<td>.387**</td>
<td>.386**</td>
<td>.225**</td>
<td>.233**</td>
<td>.185</td>
<td>.320**</td>
<td>.206**</td>
</tr>
<tr>
<td>sdBFIO Pearson Correlation</td>
<td>.247**</td>
<td>.405**</td>
<td>.313**</td>
<td>.387**</td>
<td>1</td>
<td>.355**</td>
<td>.200**</td>
<td>.165</td>
<td>.129</td>
<td>.372**</td>
<td>.182</td>
</tr>
<tr>
<td>sdhx Pearson Correlation</td>
<td>.316**</td>
<td>.209**</td>
<td>.245**</td>
<td>.386**</td>
<td>.355**</td>
<td>1</td>
<td>.418</td>
<td>.343</td>
<td>.359</td>
<td>.448</td>
<td>.438**</td>
</tr>
<tr>
<td>sdha Pearson Correlation</td>
<td>.215**</td>
<td>.221**</td>
<td>.177**</td>
<td>.225**</td>
<td>.200**</td>
<td>.418**</td>
<td>1</td>
<td>.540</td>
<td>.436</td>
<td>.524</td>
<td>.445**</td>
</tr>
<tr>
<td>sdhc Pearson Correlation</td>
<td>.236**</td>
<td>.120</td>
<td>.333**</td>
<td>.233**</td>
<td>.165</td>
<td>.343**</td>
<td>.540**</td>
<td>1</td>
<td>.315</td>
<td>.403</td>
<td>.411**</td>
</tr>
<tr>
<td>sdhs Pearson Correlation</td>
<td>.209**</td>
<td>.104</td>
<td>.162**</td>
<td>.185**</td>
<td>.129</td>
<td>.359**</td>
<td>.436**</td>
<td>.315</td>
<td>1</td>
<td>.263</td>
<td>.399**</td>
</tr>
<tr>
<td>sdho Pearson Correlation</td>
<td>.262**</td>
<td>.266**</td>
<td>.261**</td>
<td>.320**</td>
<td>.372**</td>
<td>.448**</td>
<td>.524**</td>
<td>.403</td>
<td>.263</td>
<td>1</td>
<td>.383**</td>
</tr>
<tr>
<td>sdhh Pearson Correlation</td>
<td>.155**</td>
<td>.099</td>
<td>.186**</td>
<td>.206**</td>
<td>.182</td>
<td>.438**</td>
<td>.445**</td>
<td>.411</td>
<td>.399</td>
<td>.383</td>
<td>1</td>
</tr>
</tbody>
</table>

**. Correlation is significant at the 0.01 level (2-tailed).
*. Correlation is significant at the 0.05 level (2-tailed).

Inconsistency correlations through correlation of standard deviations between BFI-2 and HEXACO. Bolded correlations indicate within-trait inconsistency correlations.
### Table 3 – Regression, BFI

#### BFI Regression Coefficients

<table>
<thead>
<tr>
<th>Model</th>
<th>Unstandardized Coefficients</th>
<th>Standardized Coefficients</th>
<th>Correlations</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>B</td>
<td>Std. Error</td>
<td>Beta</td>
</tr>
<tr>
<td>1 (Constant)</td>
<td>.962</td>
<td>.249</td>
<td></td>
</tr>
<tr>
<td>MeanBFIE</td>
<td>.017</td>
<td>.023</td>
<td>.056</td>
</tr>
<tr>
<td>MeanBFIA</td>
<td>.068</td>
<td>.032</td>
<td>.166</td>
</tr>
<tr>
<td>MeanBFIC</td>
<td>-.005</td>
<td>.029</td>
<td>-.015</td>
</tr>
<tr>
<td>MeanBFIN</td>
<td>-.055</td>
<td>.022</td>
<td>-.193</td>
</tr>
<tr>
<td>MeanBFIO</td>
<td>.053</td>
<td>.025</td>
<td>.162</td>
</tr>
</tbody>
</table>

a. Dependent Variable: OVERALLINCONSISTENCY

Regression of BFI Factors onto Overall Inconsistency
Table 4 – Regression, HEXACO

<table>
<thead>
<tr>
<th>Model</th>
<th>B</th>
<th>Std. Error</th>
<th>Beta</th>
<th>t</th>
<th>Sig.</th>
<th>Zero-order</th>
<th>Partial</th>
<th>Part</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>(.999)</td>
<td>.213</td>
<td></td>
<td></td>
<td>4.694</td>
<td>.000</td>
<td></td>
<td></td>
</tr>
<tr>
<td>hh</td>
<td>.007</td>
<td>.027</td>
<td>.021</td>
<td>.255</td>
<td>.799</td>
<td>.067</td>
<td>.019</td>
<td>.019</td>
</tr>
<tr>
<td>hs</td>
<td>-.019</td>
<td>.027</td>
<td>-.055</td>
<td>-.727</td>
<td>.468</td>
<td>-.028</td>
<td>-.055</td>
<td>-.054</td>
</tr>
<tr>
<td>hx</td>
<td>.051</td>
<td>.025</td>
<td>.154</td>
<td>2.037</td>
<td>.043</td>
<td>.163</td>
<td>.153</td>
<td>.151</td>
</tr>
<tr>
<td>ha</td>
<td>.044</td>
<td>.029</td>
<td>.123</td>
<td>1.538</td>
<td>.126</td>
<td>.139</td>
<td>.116</td>
<td>.114</td>
</tr>
<tr>
<td>hc</td>
<td>-.006</td>
<td>.029</td>
<td>-.015</td>
<td>-.194</td>
<td>.847</td>
<td>.029</td>
<td>-.015</td>
<td>-.014</td>
</tr>
<tr>
<td>ho</td>
<td>.011</td>
<td>.022</td>
<td>.038</td>
<td>.495</td>
<td>.621</td>
<td>.077</td>
<td>.038</td>
<td>.037</td>
</tr>
</tbody>
</table>

a. Dependent Variable: OVERALLINCONSISTENCY

Regression of HEXACO Factors onto Overall Inconsistency
APPENDIX A

IRB APPROVAL LETTER FOR RESEARCH
MEMORANDUM

TO: James Nauert IRB # 16-124

Dr. Michael Biderman

FROM: Lindsay Pardue, Director of Research Integrity

Dr. Amy Doolittle, IRB Committee Chair

DATE: 10/4/2016

SUBJECT: IRB #16-124: Stable Inconsistency: A Study of Personality Responses over Time

The IRB Committee Chair has reviewed and approved your application and assigned you the IRB number listed above. You must include the following approval statement on research materials seen by participants and used in research reports:

The Institutional Review Board of the University of Tennessee at Chattanooga (FWA00004149) has approved this research project # 16-124.

Annual Renewal. All approved research is subject to UTC IRB review, at least once a year. Please visit our website (http://www.utc.edu/research-integrity/institutional-review-board/forms.php) for the Form B (continuation / change / completion form) that you will need to complete and submit if your project remains active and UTC IRB approval needs to be renewed for another year. Unless your research moves in a new direction or participants have experienced adverse reactions, then renewal is not a major hurdle. You as Principal Investigator are responsible for turning in the Form B on time (2 weeks before one year from now), and for determining whether any changes will affect the current status of the project. When you complete your research, the same change/completion form should be completed indicating project termination. This will allow UTC’s Office of Research Integrity to close your project file.

Please remember to contact the IRB immediately and submit a new project proposal for review if significant changes occur in your research design or in any instruments used in conducting the
study. You should also contact the IRB immediately if you encounter any adverse effects during your project that pose a risk to your subjects.

For any additional information, please consult our web page http://www.utc.edu/irb or email instrb@utc.edu.

Best wishes for a successful research project.
APPENDIX B

EXAMPLE BIG FIVE INVENTORY – 2 ASSESSMENT
The following statements are to be rated on the degree to which they agree that the statement accurately reflects themselves. The scale is graded from 1 to 7. A rating of 1 indicates that the participant barely, of at all, feels the statement accurately reflects themselves. A rating of 7 indicates that the participant feels the statement reflects themselves highly accurately.

**Example Item:**
I am the life of the party.

\[
1 \quad 2 \quad 3 \quad 4 \quad 5 \quad 6 \quad 7
\]

(Strongly Disagree) (Strongly Agree)

**Extraversion**

*Sociability*

1. Is outgoing, sociable.
46. Is talkative.
r16. Tends to be quiet.
31r. Is sometimes shy, introverted.

*Assertiveness*

6. Has an assertive personality.
21. Is dominant, acts as a leader.
36r. Finds it hard to influence people.
51r. Prefers to have others take charge.

*Energy*

41. Is full of energy.
56. Shows a lot of enthusiasm.
11r. Rarely feels excited or eager.
26r. Is less active than other people.

*Agreeableness*

*Compassion*

2. Is compassionate, has a soft heart.
32. Is helpful and unselfish with others.
17r. Feels little sympathy for others.
47r. Can be cold and uncaring.
**Politeness**

7. Is respectful, treats others with respect.

52. Is polite, courteous to others.

22r. Starts arguments with others.

37r. Is sometimes rude to others.

**Trust**

27. Has a forgiving nature.

57. Assumes the best about people.

12r. Tends to find fault with others.

42r. Is suspicious of others' intentions.

**Conscientiousness**

**Orderliness**

18. Is systematic, likes to keep things in order.

33. Keeps things neat and tidy.

3r. Tends to be disorganized.

48r. Leaves a mess, doesn't clean up.

**Industriousness**

38. Is efficient, gets things done.

53. Is persistent, works until the task is finished.

8r. Tends to be lazy.

23r. Has difficulty getting started on tasks.

**Reliability**

13. Is dependable, steady.

43. Is reliable, can always be counted on.

28r. Can be somewhat careless.

58r. Sometimes behaves irresponsibly.

**Neuroticism**

**Anxiety**
19. Can be tense.
34. Worries a lot.
4r. Is relaxed, handles stress well.
49r. Rarely feels anxious or afraid.

*Depression*

39. Often feels sad.
54. Tends to feel depressed, blue.
9r. Stays optimistic after experiencing a setback.
24r. Feels secure, comfortable with self.

*Volutility*

14. Is moody, has up and down mood swings.
59. Is temperamental, gets emotional easily.
29r. Is emotionally stable, not easily upset.
44r. Keeps their emotions under control.

*Openness to Experience*

*Aesthetic Sensitivity*

20. Is fascinated by art, music, or literature.
35. Values art and beauty.
5r. Has few artistic interests.
50r. Thinks poetry and plays are boring.

*Intellect*

10. Is curious about many different things.
40. Is complex, a deep thinker.
25r. Avoids intellectual, philosophical discussions.
55r. Has little interest in abstract ideas.
Imagination

15. Is inventive, finds clever ways to do things.
60. Is original, comes up with new ideas.
30r. Has little creativity.
45r. Has difficulty imagining things.
APPENDIX C

EXAMPLE HEXACO 60-ITEM QUESTIONNAIRE
DIRECTIONS

On the following pages you will find a series of statements about you. Please read each statement and decide how much you agree or disagree with that statement. Then write your response in the space next to the statement using the following scale:

7 = Strongly Agree
6 = Agree
5 = Somewhat Agree
4 = Neutral (neither agree nor disagree)
3 = Somewhat Disagree
2 = Disagree
1 = Strongly Disagree

Please answer every statement, even if you are not completely sure of your response.
1 I would be quite bored by a visit to an art gallery.
2 I plan ahead and organize things, to avoid scrambling at the last minute.
3 I rarely hold a grudge, even against people who have badly wronged me.
4 I feel reasonably satisfied with myself overall.
5 I would feel afraid if I had to travel in bad weather conditions.
6 I wouldn’t use flattery to get a raise or promotion at work, even if I thought it would succeed.
7 I’m interested in learning about the history and politics of other countries.
8 I often push myself very hard when trying to achieve a goal.
9 People sometimes tell me that I am too critical of others.
10 I rarely express my opinions in group meetings.
11 I sometimes can’t help worrying about little things.
12 If I knew that I could never get caught, I would be willing to steal a million dollars.
13 I would enjoy creating a work of art, such as a novel, a song, or a painting.
14 When working on something, I don’t pay much attention to small details.
15 People sometimes tell me that I’m too stubborn.
16 I prefer jobs that involve active social interaction to those that involve working alone.
17 When I suffer from a painful experience, I need someone to make me feel comfortable.
18 Having a lot of money is not especially important to me.
19 I think that paying attention to radical ideas is a waste of time.
20 I make decisions based on the feeling of the moment rather than on careful thought.
21 People think of me as someone who has a quick temper.
22 On most days, I feel cheerful and optimistic.
23 I feel like crying when I see other people crying.
24 I think that I am entitled to more respect than the average person is.
25 If I had the opportunity, I would like to attend a classical music concert.
26 When working, I sometimes have difficulties due to being disorganized.
27 My attitude toward people who have treated me badly is “forgive and forget”.
28 I feel that I am an unpopular person.
29 When it comes to physical danger, I am very fearful.
30 If I want something from someone, I will laugh at that person's worst jokes.

Continued...
I’ve never really enjoyed looking through an encyclopedia.
I do only the minimum amount of work needed to get by.
I tend to be lenient in judging other people.
In social situations, I’m usually the one who makes the first move.
I worry a lot less than most people do.
I would never accept a bribe, even if it were very large.
People have often told me that I have a good imagination.
I always try to be accurate in my work, even at the expense of time.
I am usually quite flexible in my opinions when people disagree with me.
The first thing that I always do in a new place is to make friends.
I can handle difficult situations without needing emotional support from anyone else.
I would get a lot of pleasure from owning expensive luxury goods.
I like people who have unconventional views.
I make a lot of mistakes because I don’t think before I act.
Most people tend to get angry more quickly than I do.
Most people are more upbeat and dynamic than I generally am.
I feel strong emotions when someone close to me is going away for a long time.
I want people to know that I am an important person of high status.
I don’t think of myself as the artistic or creative type.
People often call me a perfectionist.
Even when people make a lot of mistakes, I rarely say anything negative.
I sometimes feel that I am a worthless person.
Even in an emergency I wouldn’t feel like panicking.
I wouldn’t pretend to like someone just to get that person to do favors for me.
I find it boring to discuss philosophy.
I prefer to do whatever comes to mind, rather than stick to a plan.
When people tell me that I’m wrong, my first reaction is to argue with them.
When I’m in a group of people, I’m often the one who speaks on behalf of the group.
I remain unemotional even in situations where most people get very sentimental.
I’d be tempted to use counterfeit money, if I were sure I could get away with it.
VITA

James Alexander Nauert was born in San Antonio, TX, to the parents of James J. and Suzanne Nauert. He is the oldest of five children, with two brothers and two sisters. He was homeschooled through the elementary, middle and high schools until the age of 16. Afterward he attended the Lone Star community college system in a program providing dual high school and college credit. After a few false starts to his major field of study he became interested in psychology. After graduating high school, he transferred to Sam Houston State University. As he completed course credit, he became interested in how psychology can be applied to the average worker. He completed the Bachelors of Science degree in December 2014 in general psychology, after which James applied to the University of Tennessee at Chattanooga in the Industrial/Organizational Psychology Program. James is currently working toward his Masters of Science in Industrial/Organizational Psychology, and plans to work in the consulting field post-graduation.