Summer 2019

Improving the Sensitivity of the VRIN-r Scale on the MMPI-A-RF

Stefan E. LaTulip
Old Dominion University, slatu001@odu.edu

IMPROVING THE SENSITIVITY OF THE VRIN-r SCALE ON THE MMPI-A-RF

by

Stefan E. LaTulip
B.S. May 2010, Michigan State University
M.S. August 2013, Old Dominion University

A Dissertation Submitted to the Graduate Faculties of
Eastern Virginia Medical School
Norfolk State University
Old Dominion University
in Partial Fulfillment of the Requirements for the Degree of

DOCTOR OF PHILOSOPHY

CLINICAL PSYCHOLOGY

VIRGINIA CONSORTIUM PROGRAM IN CLINICAL PSYCHOLOGY
August 2019

Approved by:

Richard W. Handel (Director)

Scott M. Debb (Member)

Jennifer M. Flaherty (Member)

Serina A. Neumann (Member)

ABSTRACT

IMPROVING THE SENSITIVITY OF THE VRIN-r SCALE ON THEE MMPI-A-RF

Stefan E. LaTulip
Virginia Consortium Program in Clinical Psychology, 2019
Director: Dr. Richard Handel

When respondents to self-report measures fail to answer in accordance with item content it distorts the accuracy of obtained test scores, degrading the ability of clinicians to use results to make accurate diagnoses and recommendations. The Variable Response Inconsistency (VRIN) scale was created for the MMPI-2 and was later revised as the VRIN-r scale for the MMPI-2-RF and MMPI-A-RF in order to detect and invalidate protocols oversaturated with random responding. Analyses conducted by Pitta (2016) revealed that the VRIN-r scale for the MMPI-A-RF was not ideally sensitive to the detection of random responding in protocols. This study explored adding items drawn from another validity scale used to detect overreporting on the MMPI-A-RF (the F-r scale) to the VRIN-r scale to ascertain if the resultant hybrid scale was more sensitive to random responding. Using the MMPI-A-RF normative sample and an overreporting sample from the MMPI-A, analyses were conducted to identify the ideal number of F-r items to add to the VRIN-r scale that maximized sensitivity to random responding while maintaining specificity from detecting overreporting. F-r items were also added to another validity scale, the Combined Response Inconsistency (CRIN) scale, to ascertain if this hybrid scale was reliably sensitive to random responding. Analyses revealed that adding six F-r items to both the VRIN-r and CRIN scale greatly improved their sensitivity to random responding while maintaining adequate specificity from the detection of overreporting. Implications for these findings, limitations to the research design, and areas of future research are discussed.

This dissertation is dedicated to my mother and father. Without their constant support and investment, I never would have been able to complete my doctoral degree.

ACKNOWLEDGEMENTS

First, I want to express my gratitude to my research and clinical mentor Dr. Richard Handel. His advice and unwavering support instilled the confidence in me to pursue this dissertation topic and work on it until completion. Furthermore, his sense of humor, conviviality, and willingness to discuss anything greatly aided me through my graduate school career. I would also like to thank my committee members Drs. Scott Debb, Jennifer Flaherty, Serina Neumann, and Michael Stutts for their excellent suggestions and their ability to challenge me to stretch my research and clinical skills to the limits of my ability.

Finally, I want to express my sincerest thanks to my friends inside and outside of the graduate program, Dr. Willie McBride, Dr. Emily Oettinger, Dr. Amy Stamates, Jennika Jenkins, Callie Mansfield, and Caitlin Sorek, for being a constant, unconditional source of support for me throughout my time in graduate school.

TABLE OF CONTENTS

Page

LIST OF TABLES

Table Page

Table Page

Table Page

LIST OF FIGURES

Figure Page

CHAPTER I

INTRODUCTION

For clinicians and researchers to obtain valid information from self-report measures respondents need to answer items in accordance with each item's content. When respondents fail to respond to item content, it distorts the accuracy and validity of obtained scores. The use of techniques and scales within self-report measures to detect a respondent's failure to respond to test content in various ways has been incorporated into testing for several decades. At first, the bulk of the psychometric research and development of these scales were found in self-report cognitive testing (scholastic aptitude, vocabulary, mathematical reasoning, etc.), however this soon spread to self-report measures of personality and psychopathology (Tellegen, 1988). One of the first self-report, personality inventories to create scales to detect the several ways individuals can aberrantly respond to test items was the Minnesota Multiphasic Personality Inventory (MMPI) (Hathaway & McKinley, 1942), a measure designed to detect and describe problematic psychopathology symptoms and personality traits (Dahlstrom, Welsh, & Dahlstrom, 1972; Tellegen, 1988).

Over the evolution of the MMPI, multiple of these so called "Validity Scales" were added to the test, one of which, named the Variable Response Inconsistency (VRIN) scale, was used to detect random responding in protocols. It first appeared in the MMPI-2 (Butcher, Dahlstrom, Graham, Tellegen, & Kaemmer, 1989) and in the adolescent version of the MMPI, the MMPI-A (Butcher et al., 1992), and was retained and restructured as the VRIN-r scale in the newer MMPI-2-RF (Ben-Porath & Tellegen, 2008) and the MMPI-A-RF (Archer, Handel, Ben-Porath, & Tellegen, 2016). Recent research has delineated the sensitivity of the VRIN-r scale to detect random responding on the MMPI-2-RF (Handel, Tellegen, Ben-Porath, & Archer, 2010)

and the MMPI-A-RF (Pitta, 2016). Unfortunately, the sensitivity of the VRIN-r scale to simulated random responding on the newly created MMPI-A-RF was found to be significantly lower than the VRIN-r scale on the MMPI-2-RF, and it performed poorer at identifying protocols as invalid due to saturation with aberrant, random responding (Pitta, 2016). This study seeks to explore a procedure to improve the sensitivity of the VRIN-r scale in the detection of random responding on the MMPI-A-RF as to optimize the detection of invalid protocols oversaturated with random responses.

Traits, Aberrancy, and Content Non-Responsiveness

Tellegen (1988) defines a *trait* as a psychological construct that underlies a relatively stable behavior disposition, or a tendency to respond in certain ways in specific situations. A *personality trait* is a type of trait that underlies multiple behaviors that have several adaptational effects. In psychometrics, a *trait level* is a quantitative variable representing a single person's standing in a trait dimension. This trait level is an individual difference measure as it can be used to compare a single person's score to other individual's scores on that same dimension. A trait dimension can be conceptualized as signifying parametric or statistical differences among persons whose psychological structures, with respect to that trait in question, are basically the same. This specific type of trait that has a basic structure present in many persons who may vary considerably in trait level, has been called a *dimensional trait.* If it is thought to be common in most people, a dimensional trait is called *nomothetic* (as opposed to idiographic) (Tellegen, 1988). Self-report measures such as the MMPI capitalize on the idea of nomothetic dimensional traits as they seek to classify an individual who takes the assessment on several dimensions that are thought to be represented in all people. This not only allows comparisons between

respondents on a certain trait dimension, but it can allow for the clinician to ascertain if the trait level a person falls on likely predicts certain maladaptive behaviors and cognitions.

Consistency on a trait score can be defined as the degree of stability of scores for a given trait with a dimensional model. This requires scores representing the trait behave strictly as estimates of that trait and that scores obtained from others on this trait fall on one continuum. When model-based statistical tests show too little interindividual covariation, or too much intraindividual scatter then the data are considered too inconsistent and in violation of the model specified (Tellegen, 1988). There are multiple measures that have been created in the assessment literature that have been used to detect inconsistency. Most of these measures were originally used in the cognitive assessment literature, specifically for tests of school achievement, mathematics reasoning, vocabulary, or professional certification (Tellegen, 1988). Nearly all these early measures of inconsistency focused on, what Tellegen (1988) calls, aberrancy, or records so inconsistent with trait models that they are considered uninterruptable. These measures were created by first identifying what an "average" scoring record looked like for the test in question, and then making a scale that detected protocols that were so deviant from this "average" protocol they raised questions as to whether valid data was provided. In cognitive assessment, deviant protocols were usually caused by a responder misunderstanding the instructions, severe scoring errors, or cheating. The personality assessment literature slowly followed the cognitive literature in adding measures to detect aberrancy in self-report measures (Tellegen, 1988).

In self-report personality assessment, one type of aberrancy is caused when a respondent tries to minimize or maximize their symptoms or traits. Overreporting, feigning, or "faking bad" is a response set in individuals seeking to magnify or even feign their symptoms in order to

present themselves as very psychologically disturbed or hurt (Baer, Kroll, Rinaldo, & Ballenger, 1999). Overreporting is usually done so a person can obtain a goal such as admission to a psychiatric hospital, avoidance of responsibility for a crime, or to win compensation for psychological harm in personal injury litigation (Berry, Baer, & Harris, 1991). At the other extreme, underreporting, also known as defensiveness, "faking good," or socially desirable responding is a common response in individuals who are trying to present themselves in the best possible light, denying even minor faults that are common to most people. Like in overreporting, this usually occurs in high stakes personality assessments whereby exhibiting well-adjusted psychology will aid the person in obtaining a goal they want, e.g., job selection, child custody, or early release from hospital (Baer, Wetter, & Berry, 1992).

Another type of aberrancy response in personality assessment is called content non-responsiveness. This is when test-takers fail to respond to the content of items on self-report measures (Handel, Ben-Porath, Tellegen, & Archer, 2010). Content non-responsiveness has two types: fixed responding and random responding. Fixed responding is when a test-taker tends to endorse items in the same way repeatedly regardless of item content. In a true-false questionnaire this problem would manifest as endorsing test items with indiscriminate true (acquiescence) or indiscriminate false (counter-acquiescence) responses (Handel et al., 2010). Random responding, the focus of this study, occurs when a test-taker fails to pay attention to test content and indiscriminately marks down his or her answers aimlessly (Handel et al., 2010). Content non-responsiveness may be produced by test takers who do not understand instructions, have reactivity towards observation, have lack of motivation to cooperate with the testing, are disinterested or fatigued, have reading problems, or have cognitive impairment (Berry et al., 1992; Meier, 1994; Ben-Porath, 2008).

History of Scales Used to Detect the Types of Aberrancy on the MMPI

Overreporting. Multiple scales were developed for the MMPI and later the MMPI-2 and MMPI-A for detection of aberrancy. The original MMPI (Hathaway & McKinley, 1942) included the F (Infrequency) scale which was created to detect protocols that were saturated with items rarely endorsed in the normative sample. The scale worked by identifying protocols so unusual (compared to the normative sample) that it was doubtful that these protocols were measuring true personality and psychopathology data, and rather were the result of respondents overreporting response style The F scale was also included in the creation of the MMPI-2 (Butcher et al., 1989) and the MMPI-A (Butcher et al., 1992) to measure the same construct. Arbisi and Ben-Porath (1995) developed an additional scale on the MMPI-2 to better detect overreporting of severe psychopathology. They selected items in the MMPI-2 sample that were rarely selected in both the normative sample *and* in a sample of psychiatric patients. This scale was named the Fp (Infrequency Psychopathology) scale and was created because the original F scale sometimes indicated a person was overreporting when they were reporting true, severe psychopathology (Arbisi & Ben-Porath, 1995). The Fp scale was created to more finely discriminate between those who were overreporting and those who were reporting actual severe psychopathology. For the MMPI-A, the original F scale was composed of items infrequently endorsed by the normative MMPI-A (Butcher et al., 1992). The F and Fp scales were revised for the MMPI-2-RF as the F-r and Fp-r scales (Tellegen & Ben-Porath, 2008) and the F scale in the MMPI-A was revised as the F-r scale for the MMPI-A-RF (Archer et al., 2016), all continuing to measure the overreporting level of respondents to these instruments. The F-r scale for the MMPI-A-RF was also updated such that it was composed of items that were not only rarely endorsed by the normative sample, but also from rarely endorsed from the combined developmental sample

in outpatient, inpatient, correctional, and school setting (Archer et al., 2016). As such, the F-r scale for the new MMPI-A-RF functions similarly to the Fp-r scale on the MMPI-2-RF (Tellegen & Ben-Porath, 2008).

Underreporting. Two scales were also developed for the original MMPI to detect underreporting. The L (Lie) scale (Hathaway & McKinley, 1951) was created by identifying items that were rarely endorsed by the normative sample and included content that presented the test-taker as overly virtuous or extremely well-mannered. This scale was created to detect individuals who minimized their psychopathology and presented themselves in such a positive light that it inhibited the ability of scores to accurately represent the person's psychopathology and personality. The K (correction) scale (McKinley, Hathaway, & Meehl, 1948) was created by identifying items that tended to be endorsed by individuals who were hospitalized with psychological difficulties, yet yielded profiles that indicated no severe psychopathology. Thus, greater scores on this scale indicated a person that was highly defensive and minimizing their psychopathology, however this defensiveness was more subtle than the defensiveness detected by the L scale (McKinley, Hathaway, & Meehl, 1948). All the L items and all but one of the K items are keyed in the false direction. Thus, although neither scale was originally designed to detect content non-responsiveness, both scales are sensitive to counter-acquiescent responding as well (Handel et al. 2010). The L and K scales were carried over from the original MMPI to the MMPI-2 (Butcher et. al, 1989; Ben-Porath, 2008), and then were revised as the L-r and K-r scales and included in the MMPI-2-RF (Tellegen & Ben-Porath, 2008). L and K scales were also created to detect underreporting in protocols of the MMPI-A (Butcher et al., 1992; Graham, Archer, Tellegen, Ben-Porath, & Kaemmer, 2006), and similar to what was done in the MMPI-2-RF, were revised as L-r and K-r for the MMPI-A-RF (Archer et al., 2016).

Content non-responsiveness. Scales specifically designed to detect random and fixed responding on the MMPI started with Beuchly and Ball's (1952) Test-Retest (TR) index. The TR was composed of sixteen repeated items that were originally included to aid in machine scoring. If an individual answered one of these repeated items differently than they initially did then a raw score point would be added to the scale. Although TR was sensitive to random responding, Tellegen (1988) pointed out that it was not sensitive to fixed responding. Tellgen in 1982 developed scales for his Multidimensional Personality Questionnaire (MPQ) (Tellegen, 1982; Tellegen & Waller, 2008) designed to detect random and acquiescent/counter-acquiescent responding, and these scales were named the Variable Response Inconsistency Scale (VRIN) and the True Response Inconsistency (TRIN) scale, respectively. The two MPQ scales were used as the basis for the development of the same-named scales on the MMPI-2 (Butcher et al., 1989; Butcher, Graham, Ben-Porath, Tellegen, Dahlstrom, & Kaemmer, 2001) and on the MMPI-A (Butcher et al., 1989; Graham et al., 2006). The VRIN and TRIN were revised as the VRIN-r and TRIN-r scales and included in the MMPI-2-RF (Tellegen & Ben-Porath, 2008) and the MMPI-A-RF (Archer et al., 2016). Finally, the Combined Response Inconsistency scale (CRIN) was created by Archer et al. (2016) for the MMPI-A-RF as composite of both the VRIN-r and TRIN-r scales, and was designed to measure the overall content non-responsiveness (both fixed and random) in the protocol (Archer et al., 2016).

Creation of the VRIN, VRIN-r, and CRIN Scales for the MMPI-2, MMPI-2-RF, and MMPI-A-RF

The Variable Response Inconsistency scale was developed for the MMPI-2 to detect random responding (Butcher et al., 1989, 2001). It consists of 67 item pairs with either similar or opposite content. The VRIN scale for the MMPI-A consists of 50 paired items and was like the

VRIN scale on the MMPI-2 in that it includes pairs of items with similar and opposite content (Butcher et. al, 1992). Thus, test-takers that answered an item pair as FT, TF, FF, or TT would suggest inconsistent responding and add a raw score point to the VRIN scale.

Tellegen and Ben-Porath (2008) revamped the VRIN scale for the MMPI-2-RF, creating the VRIN-r scale for the test. The procedures they used were the same as the procedures used by Archer et al. (2016) to create the VRIN-r scale for the MMPI-A-RF. Each raw score unit for the VRIN-r scale is composed of a pair of items from the inventory, and the keyed direction is determined by the configural response to the two items. For the VRIN-r scale, these two items include similar content (unlike the previous VRIN scale used in the MMPI-2 and the MMPI-A where questions could be similar or opposites). The two items have four possible response patterns (TT, TF, FT, and FF), two of these responses (TF and FT) can suggest inconsistent responding and can be keyed as a *positive* score and coded "1." The two other responses (TT and FF) do not suggest inconsistent responding and are always keyed as a *negative* score and coded "0." The pair of items that are scored in a configural pattern were labeled *c-composites* by Tellegen and Ben-Porath (2008). Their criteria for identifying suitable c-composites to compose the VRIN-r scale was:

1. The two items within each c-composite had to be reliably, positively correlated, thus suggesting that they were related in content.
2. The content of each of the c-composites, when scored positively, were judged to be inconsistent by Tellegen and Ben-Porath (2008).
3. Each c-composite was required to have a low ratio of *observed frequencies* to *expected frequencies* to ensure that any keyed response pair was statistically improbable.

4. The c-composite was also required to have a *low degree of content saturation* as determined by a negligible or weak correlation with an additive composite consisting of the same two inventory items but now scored to measure the shared content component of the two items. An example helps for this explanation. Suppose the hypothetical item pair ("I am sad most of the time" and "I am blue most of the time") had been combined to create TF and FT c-composites for possible use as supposedly content-free random response indicators. Criterion 4 would have called for computing correlations of the TF and FT c-composites with a measure made up of the same two inventory items combined additively. The two-item scale would therefore assess the shared content of the two items, specifically, the respondent's feelings of dysphoria. Low correlations of the c-composites with this scale would have been required as empirical evidence that the c-composites were indeed relatively free of content and only assessing random responding.
5. The final criterion was designed to prevent the response to a single inventory item from exerting excessive influence on the total VRIN-r scale. It required that no inventory item belong to more than one of each of the four kinds of c-composites (TT, TF, FT, or FF). Thus, one inventory item could not be used multiple times in different c-composite to indicate random responding. This prevents one inconsistent response from the test-taker being used multiple times in c-composites with other similar items, thus inflating the VRIN-r total score.

The resulting VRIN-r scale for the MMPI-2-RF consisted of 53 c-composites (Tellegen & Ben-Porath, 2008), and the resulting VRIN-r scale on the MMPI-A-RF consisted of 27 c-composites (Archer et al., 2016). The VRIN-r score for both tests equals the number of scored inconsistences. These raw scores are converted to linear *T* scores; scores 65 and above suggest

some inconsistent responding, where scores 80 and above invalidate a protocol for interpretation on the MMPI-2-RF (Tellegen & Ben-Porath, 2008). On the MMPI-A-RF *T* scores of 75 and above invalidate a protocol (Archer et al., 2016).

As can be seen the VRIN-r scales for both the MMPI-2-RF and MMPI-A-RF are distinct from the old VRIN scales from the MMPI-2 and MMPI-A in that item pairs (c-composites) consist of only similar content items. Thus, only true-false and false-true responses can add a raw score point to the VRIN-r scale, whereas both similar and opposite item content was used on the VRIN scales such that all four possible responses (TF, FT, TT, and FF) could possibly indicate inconsistency and add a raw score point to the VRIN scale. Other differences are that the Tellegen and Ben-Porath (2008) conducted procedures to ensure VRIN-r items were relatively content independent and that no item appeared twice in a VRIN-r c-composite. These procedures, coupled with the general reduction of the item pool when creating the MMPI-2-RF from the MMPI-2 and the MMPI-A-RF from the MMPI-A, lowered the number of items within the VRIN-r scales as compared to the original VRIN scales. The MMPI-2's VRIN scale of 67 paired items was reduced to 53 pairs of items in the combined VRIN-r and F-r scale of the MMPI-2-RF, and the MMPI-A's VRIN scale of 50 paired items was dramatically reduced to 27 paired items in the MMPI-A-RF's VRIN-r scale.

As mentioned above, the VRIN-r scale for the MMPI-A-RF was smaller than the VRIN-r scale in the MMPI-2-RF. This was also the case for the TRIN-r scale in the MMPI-A-RF, as it was composed of thirteen c-composites compared to the MMPI-2-RF TRIN-r's 26 c-composites (Archer et al., 2016; Tellegen & Ben-Porath, 2008). In light of the relative brevity of both the TRIN-r and VRIN-r scale within the MMPI-A-RF, Archer and colleagues (2016) created a composite of both the VRIN-r and TRIN-r scale called the CRIN scale to measure overall

content non-responsiveness (both fixed and random). The creation of the scale was made to provide increased, incremental sensitivity to content non-responsiveness on the MMPI-A-RF to combat the possibility that the shorter VRIN-r and TRIN-r scales would not be as efficient at detecting content non-responsiveness as their longer MMPI-2-RF counterparts.

Efficiency of the VRIN and VRIN-r Scales

VRIN for the MMPI-2. Multiple studies have found that the VRIN scale is sensitive to random responding for the MMPI-2. Berry et al. (1991) instructed 180 college students to randomly respond to an MMPI-2 protocol at certain points in the test protocol and found that the VRIN scale was sensitive at detecting these students' response sets. Interestingly, this study found that the F and Fb scale (both infrequency scales described above) were better at detecting random responding when it occurred later in the test protocol. This was explained by the VRIN items being more concentrated in the beginning of the test protocol, whereas F and Fb items were more evenly distributed throughout the test (Berry et al., 1991). Another study conducted by Berry, Wetter, Baer, Larsen, Clark, and Monroe (1992) surveyed college students, community members, and police academy applicants across multiple experiments asking them if they randomly responded to the test, and if so, how many items they thought they randomly responded to. Not only did the study find that a significant number of participants admitted to answering some items randomly (29-60% of the sample depending on the experiment), but they found that participant's indication of the number of items they responded to randomly was reliably and positively correlated with VRIN scores. Cramer (1995) and Gallen and Berry (1996) attempted to test various formulas composing different validity scales of the MMPI-2 to better detect random responding. In both studies it was found that the VRIN scale reliably detected either simulated or respondent-instructed randomness in profiles. Finally, Charter and Lopez

(2003) utilized a Monte Carlo approach to generate 5,000 completely random responses for multiple validity scales including the VRIN. They found that the VRIN was sensitive to simulated random responding, however they argued that the cutoff for the VRIN scale was too high and did not optimally detect and eliminate random protocols. All of the aforementioned studies confirmed that the VRIN scale is sensitive to random responding, although there were findings that the VRIN was less sensitive to back protocol randomness due to item distribution (Berry et al., 1991) and there was an argument concerning the proper cutoff for the VRIN scale (Charter and Lopez, 2003).

VRIN for the MMPI-A. The VRIN scale for the MMPI-A has also been found to be sensitive to random responding. Baer, Ballenger, Berry, and Wetter (1997) used a community sample of 106 adolescents, and using procedures like Berry et al. 1991, instructed respondents to randomly respond at different points in the MMPI-A. They found that the VRIN scale was sensitive to random responding, but again, like what was found by Berry et al. (1991) in the MMPI-2, the F scale was better at detecting random responding in later portions of the test due to its items being more equally distributed throughout the test unlike the VRIN's items. Another study conducted by Baer, Kroll, Rinaldo, and Ballenger (1999) gave 24 nonclinical adolescents instructions to fill out an MMPI-A protocol to indicate severe psychological distress, and then had another 20 nonclinical adolescents fill out the MMPI-A protocol without access to the test booklet to simulate random responding. This was done to determine if the detection of random responding and overreporting could be discriminated by looking at the VRIN scale and the F scale. They found that the VRIN scale reliably detected only random responding, whereas the F scale could reliably detect both random responding and overreporting. Archer and Elkins (1999) examined 354 MMPI-A protocols obtained in a clinical setting and compared them to 354

MMPI-A protocols saturated with computer-simulated random responding. They found that the VRIN scale was sensitive to random responding and the cutoff value proposed in the MMPI-A manual (Butcher et al., 1989) was largely appropriate. Archer, Handel, Lynch, and Elkins (2002) examined five samples of MMPI-A protocols: one included MMPI-A protocols from 100 inpatient adolescents, another was composed of 100 totally random, computer generated protocols, and the final three were composed of varying degrees of computer generated randomness inserted into the back half the MMPI-A. They found that the VRIN was sensitive to random responding, however, like what was found in some previous research, it performed worse detecting randomness in the back of the MMPI-A profile and at detecting partially random protocols. Pinsoneault (2005) used protocols from 43 adolescent respondents to the MMPI-A and inserted varying degrees of computer-simulated random responses into their protocols. He found that VRIN was the most sensitive at detecting all random protocols, however it still failed to correctly identify 35% of all random protocols. All these studies indicate that, like the VRIN scale in the MMPI-2, the VRIN scale in the MMPI-A is sensitive to random responding, however it is not as sensitive to random responding situated in the back of the MMPI-A protocol (Baer et al., 1997; Archer et al., 2002), and to partial random responding (Archer et al., 2002).

VRIN-r for the MMPI-2-RF. The MMPI-2-RF technical manual (Tellegen & Ben-Porath, 2008) presented some preliminary analyses that showed that the VRIN-r scale was sensitive to random responding (Handel, Ben-Porath, Tellegen, & Archer, 2007), and these same authors presented the complete results in an article in 2010. They used a computer simulation of random responding to insert varying percentages of random responses into a sample of originally valid 2,109 MMPI-2-RF protocols. The insertion was increased by increments of 10% until all protocols were composed of 100% randomly inserted responses. They then analyzed the average

VRIN-R T-score for the sample and obtained the percentage of the sample that yielded a VRIN-r T score that was 80 and above (suggesting an invalidated protocol) for each of the 10% increments. Results revealed that the VRIN-r scale was sensitive to random responding. For instance, at the 50% level of random response insertion, the mean VRIN-r *T* score for the sample was 81.5 (*SD* = 12.8) (the original, i.e. unmodified, sample had a mean *T* score of 49.5, *SD* = 9.5), and the VRIN-r scale identified 53.7% of the profiles as being invalid (*T* score ≥ 80). At the 70% level the mean *T* score was 90 (*SD* = 13.6), and the scale suggested invalid protocols for 77% of the sample. These results provide evidence that the VRIN-r scale on the MMPI-2-RF is sensitive to random responding and can efficiently detect problematic random responding even when profiles are only partially saturated with it.

VRIN-r on the MMPI-A-RF. Using the same procedure performed by Handel et al. (2010), Pitta (2016) inserted varying degrees of randomness into 1215 MMPI-A-RF protocols to test the efficiency of its VRIN-r scale. At 50% random insertion, the data produced an average VRIN-r *T* score of 64.3 (*SD* = 10.1) with 14.5% of protocols identified as invalid. At 100% random insertion, the average VRIN-r *T* score was 71.6 (*SD* = 10.7) with 36.6% of the protocols identified as invalid. This means that if an individual randomly responded to an entire MMPI-A-RF, the chances that protocol would be thrown out due to an elevated score on the VRIN-r would be about one in three. Conversely, this means that two out of three times this totally random protocol would fail to be thrown out by the VRIN-r scale and, barring failure of other validity indices, would go on to be interpreted in full. Pitta (2016) also analyzed the CRIN scale (a composite of both fixed and random responding) to see if it incrementally added sensitivity to the detection of random responding. The VRIN-r and CRIN scale together invalidated only 50.1% of protocols saturated with 100% random responding, thus the CRIN scale did provide

some modest incremental improvement in random response detection than just the VRIN alone. However, it seems that the VRIN-r scale's reliability at detecting random responding alone is much weaker than the same scale in the MMPI-2-RF, and even with the inclusion of CRIN, only about half the totally random protocols were correctly identified as invalid due to random responding.

Why is the VRIN-r Scale on the MMPI-A-RF Less Sensitive to Random Responding?

One problem is caused by VRIN-r scale being formed into c-composites. As mentioned above each VRIN-r raw score point is determined by a respondent answers to two items. If, for instance, we suppose a person is responding randomly to those two items, statistically, there is only, on average, a 50% chance that the respondent will answer the two items in the c-composite in such a way as to cause one raw score point being added to the VRIN-r scale. This means that, on average, 50% of the time the random responder will produce a c-composite that *fails* to detect that random responding has occurred even when the respondent is truly responding randomly. Thus, not only are two items needed for one raw score point on the VRIN-r, but the two items only detect random responding half the time, at best. This means there needs to be enough c-composites to provide a proper sample of the opportunities for randomly responding. As mentioned above the VRIN-r scale for the MMPI-2-RF is more sensitive to random responding than the VRIN-r scale for the MMPI-A-RF. The VRIN-r c-composites make up approximately 16% of the item pool for MMPI-2-RF, whereas this figure is 11% for the MMPI-A-RF. This approximate 5% reduction in composition could have greatly attenuated the sensitivity of the VRIN-r on the MMPI-A-RF to random responding.

Another factor is that the adolescent normative sample tends to endorse c-composites within the VRIN-r scale at a slightly higher rate than adults. In the MMPI-2-RF the normative

adult sample, on average, endorsed 6% of the VRIN-r items (Tellegen & Ben-Porath, 2008), whereas the adolescent sample for the MMPI-A-RF endorsed, on average, 10% of the VRIN-r items (Archer et al., 2016). The slightly higher rate of endorsement in the adolescent normative sample of VRIN-r items means that more items need to be endorsed for scores to go over the 75 *T* score cut off to identify problematic random responding in the MMPI-A-RF.

Another cause is due to the VRIN-r composition. As mentioned above, using c-composites can only detect random responding 50% of the time, at best. However, most c-composites within the VRIN-r item pool are not keyed in the reverse direction. Specifically, this means that if a c-composite has a keyed sequence of TF, one would assume that FT response would also be in the keyed direction, however most c-composites are keyed in one of these sequences. The reason that this occurred is because, although these two items in the c-composite are statistically related and were deemed to be of similar content, they failed to meet at least one of Tellegen and Ben-Porath's (2008) criteria (specifically one of criteria three, four, and five explained above) to be included as keyed in the reverse direction. This means that the detection rate of random responding for c-composites that are only include one of TF or FT as in the keyed direction falls to 25%. Of the 27 items in the VRIN-r item pools, only four c-composites are keyed in both the TF and FT direction, this means that are actually only 23 unique opportunities to detect random responding, and each opportunity can only detect true random responding a quarter of the time. The MMPI-2-RF's VRIN-r also has a similar percentage of its' total c-composites that are keyed in both directions: six c-composites out of the 53 being keyed in both directions. However, this means there is a much larger pool of unique opportunities to detect random responses on the VRIN-r in the MMPI-2-RF with 47, as compared to the 23 for the VRIN-r in the MMPI-A-RF. Although the rate of detection is still at 25% for these opportunities,

the significantly greater amount of unique opportunities to detect randomness likely explains why VRIN-r scale on the MMPI-2-RF is more sensitive to random responding than the VRIN-r scale for the MMPI-A-RF (Pitta, 2006).

Exacerbating this, the VRIN-r on the MMPI-A-RF is roughly half the length of both the original MMPI-A VRIN scale and the VRIN-r scale on the MMPI-2-RF. Thus, the 23 unique c-composites on the VRIN-r scale for the MMPI-A-RF may not be a large enough sample of opportunities to reliably detect random responding. The obvious solution to this would be to simply add more c-composites to the VRIN-r scale, but unfortunately all possible c-composites have been exhausted in the MMPI-A-RF's sample of items and creating new items would necessitate an entirely new standardization process for the test. Lowering the cutoff T score for the VRIN-r is another solution, but this may harm the tests specificity and produce many false positives which would invalidate protocols that were valid for interpretation.

Improving the VRIN-r Scale

One solution is to utilize some of the items of the MMPI-A-RF that are within the F-r scale, combining them with the VRIN-r scale. As was mentioned previously, the Infrequency (F) scale was first created in the original MMPI to detect protocols that were saturated with endorsements of items that were very rarely endorsed by the normative sample (Hathaway & McKinley, 1942). Thus, this scale was originally designed to detect protocols from individuals that were overreporting very rare symptoms, likely suggesting overreporting. The original MMPI-A also had an F scale designed for this same purpose (Butcher et al., 1992), and the MMPI-A-RF revised and included this scale as the Infrequency-Revised or F-r scale (Archer et al., 2016). The F-r scale is composed of items infrequently endorsed by the normative MMPI-A-RF sample and in the combined developmental sample of adolescents in outpatient, inpatient,

correctional, and school settings (Archer et al., 2016). The logic behind using items within the F-r scale to help detect random responding is that these items should be rarely endorsed by respondents; *however,* a random responder has a 50% chance of answering an F-r item in the keyed direction, which is much greater than what is to be expected in the general population of adolescents. As such, the F-r scale should be sensitive to not only overreporting, but to random responding as well. Indeed, a study by Pinsoneault (2005) found that the MMPI-A's F scale was sensitive to random responding, although not as sensitive as the VRIN scale. Similar results were found for the F-scale within the MMPI-2 (Berry et al., 1992; Cramer, 1995; Charter & Lopez, 2003). Although there have been no studies of the sensitivity of the F-r scale of the MMPI-A-RF to random responding, preliminary analysis by the authors of this paper revealed that inserting increasing degrees of simulated random responding in protocols resulted in increasing mean T scores for the F-r scale. These findings taken together suggest that supplementing the current VRIN-r scale of the MMPI-A-RF with items from the F-r scale is possibly a viable solution to improve detection of random responding and is worthy of further investigation.

The Importance of Improving the VRIN-r Scale

Improving the sensitivity of the VRIN-r scale is crucial to better allow clinicians and researchers to detect and invalidate what Tellegen (1988) called aberrant protocols which do not represent the true psychological symptoms of respondents. The MMPI-A-RF helps clinicians make decisions about the diagnosis, treatment, forensic placement, educational placement, etc. of adolescent respondents, and these decisions could be negatively impacted if protocols saturated with random responses are deemed valid for interpretation. Additionally, researchers may fail to eliminate protocols that are invalid from their samples, deforming reliability and validity estimates and greatly impacting the conclusions drawn analyses. A less obvious effect is that

many protocols may go on to be invalidated for the wrong reason. As discussed above, the VRIN-r scale of the MMPI-A-RF fails to detect about 2/3 of totally random protocols (Pitta, 2016), and the F-r scale has been tentatively shown to be sensitive to random responding. Thus, there is a good chance that some random protocols that are not invalidated by the VRIN-r scale will be invalidated by the F-r scale (or other validity scales). Clinicians will then (rightly) throw out the protocol but will believe that the protocol was invalidated not because of random responding, but, wrongly, because of over-reporting (or another reason). Although information will not be interpreted from the protocol, the clinician's conceptualization of the adolescent respondent will likely be affected by how a protocol was invalidated. Thus, it is also important to improve the sensitivity of the VRIN-r scale in order to communicate to clinicians the correct reason as to why a specific protocol was invalidated.

CHAPTER II

METHOD

Procedure

Items within the F-r scale of the MMPI-A-RF were reviewed as possible candidates for addition to the combined VRIN-r and F-r scale. A total of fifteen items were added to explore their effects. Efforts were made to select F-r items that were evenly distributed throughout the protocol, and not clustered in one half or one area of the test. This ensured that items can be located to better detect individuals who begin to randomly respond in the back half of the test.

Next, analyses were completed to identify how many items from the F-r scale should be added to the VRIN-r scale. The procedure attempted to find an optimal balance of improving the modified VRIN-r scale's sensitivity to random responding, while not adding too many items such that the modified VRIN-r scale will then become overly sensitive to over-reporting of psychopathology, somatic complaints, and/or cognitive complaints. In other words, there was an attempt to maintain the new hybrid scales specificity from detecting overreporting. This latter idea is important because, like was said above, it is important to communicate to clinicians *why* a protocol is invalidated. If the combined VRIN-r and F-r scale starts to become too indistinguishable from the F-r scale, it becomes impossible to tell if an elevation of the new scale is due to random responding or over-reporting.

To ascertain how the addition of F-r items improves the sensitivity of the VRIN-r scale a procedure similar to what Handel et al. (2010) and Pitta (2016) did to test the sensitivity of the VRIN-r scale in the MMPI-2-RF and the MMPI-A-RF, respectively, was conducted. This involved utilizing a subset of the normative sample of the MMPI-A-RF and using a computer simulation to insert varying degrees of random responses into these protocols. A composite

score was created that includes the original VRIN-r items combined with each additional F-r item that is added. This composite score for each protocol member was then converted into untruncated/unrounded *T* score units. *T* scores were calculated at 25%, 50%, 75%, and 100% increments of random insertion. Invalid protocols were indicated by *T* scores that are equal to or exceed 75, which is the cutoff set by the MMPI-A-RF manual (Archer et al., 2016). The percentage of cases in the sample that were at or above a *T* score of 75 for the new composite VRIN-r scale were collected at each level of random responding and item addition level. If the percentage of cases invalidated by the new composite is larger than the percentage of cases invalidated by the original VRIN-r scale then this would provide evidence that the combined VRIN-r and F-r hybrid scale is more efficient at detecting random responding than the original VRIN-r scale. Special attention was also paid to the possibility that, at some point, adding additional F-r items would not incrementally improve the sensitivity of the combined VRIN-r and F-r scale to a significant degree. Trend analyses were completed on the mean *T* score and percentage above cutoff data by item addition level to ascertain if this does occur.

To ensure that the combined VRIN-r and F-r scale was sufficiently specific from the F-r scale, the new scale was analyzed in a sample of adolescents who took the MMPI-A with the instructions to create protocols that were indicative of severe psychopathology (Stein, Graham, & Williams, 1995), which were then rescored into MMPI-A-RF protocols. A mean *T* score was calculated for each time an F-r item was added to the new scale. Attention was paid to how *T* score points for the combined VRIN-r and F-r composite elevated depending on the number of F-r items added to it. Particular attention was paid to the point at which the combined VRIN-r and F-r scale became reliably elevated within the overreporting sample, i.e., at what point did more than 10% of the cases have *T* scores on the combined VRIN-r and F-r composite that are at

or above the 75 point cutoff. This 10% level was selected because the research literature in symptom and performance validity often sets 90% as the minimum level of specificity for a validity scale in order to minimize the amount of false positive calls (Boone, Lu, & Wen, 2005; Greve, Ord, Curtin, Bianchini, & Brennan, 2008; Kim, Boone, Victor, Marion, Amano, Cottingham et al., 2010).

Correlations were also obtained between the F-r scale and newly created VRIN-r hybrid scale. These data were used as a comparison point for the VRIN-r/F-r and the F-r scale to understand how much statistical overlap in variance occurred between the two as F-r items were added to the hybrid scale. These procedures helped delineate the bound at which adding additional F-r items to the combined VRIN-r/F-r scale renders it too related to F-r scale such that it does not retain proper specificity from overreporting. Finally, once the proper number of items to add to the VRIN-r scale was identified, the VRIN-r hybrid scale and the F-r scale were correlated with the Restructured Clinical scales in the normative sample. The VRIN-r scale has been found to be less strongly related to the Restructured Clinical scales than the F-r scale (Archer et al., 2016). Thus, this procedure allowed for a comparison between correlations of the VRIN-r hybrid scale and the F-r scale to the Restructured Clinical scales. If the VRIN-r hybrid scale had correlations with the Restructured Clinical scales that more resemble the low correlations the original VRIN-r scale had with the Restructured Clinical scales, and did not resemble the greater correlations the F-r scale has with the Restructured Clinical scales, this would provide additional evidence that is sufficiently specific from the F-r scale and was not oversaturated with overreporting items.

Analyses completed by Pitta (2016) indicated that the CRIN scale was also sensitive to random responding. As such, the same procedures used to calibrate the combined VRIN-r and F-

r scale discussed above were used on the CRIN scale to create a combined CRIN and F-r scale. The combined CRIN and F-r hybrid scale was then compared to the combined VRIN-r and F-r scale to ascertain if this former scale is more sensitive to random responding, or how much the former scale adds to the incremental sensitivity to random responding over the combined VRIN-r and F-r hybrid scale.

Two item addition procedures were investigated. In the first procedure, F-r items were selected and added in ascending order starting with the least endorsed item in the *normative sample*, then moving to the next least endorsed item, and so forth. The logic behind this procedure was two-fold: First, it maximized the sensitivity of the new VRIN-r and CRIN hybrid scales to random responding, as items that are least endorsed should produce the lowest mean raw scores for the new hybrid scales with each additional item. As such, random responding that activates these items in the keyed direction will be more likely to deviate from this original raw score mean and produce *T* scores that are above the cutoff. Second, as these items are the least endorsed F-r items in the normative sample, they would best suggest aberrant responding from individuals who activate these items in the keyed direction. The second procedure added the F-r items to the new composite VRIN-r scale in ascending order that were the least endorsed in our *overreporting sample* (Stein et al., 1995). This procedure attempted to try to maximize the specificity of the VRIN-r and CRIN hybrid scales to best ensure it was detecting random responders and was less likely to elevate for individuals that have an over-reporting response style. Finally, using two item addition procedures provided the opportunity to compare them regarding their ability to improve sensitivity to detecting random responding while maintaining specificity from overreporting in the new VRIN-r and CRIN hybrid scales. The specific items

used from the MMPI-A-RF for each addition procedure along with the position they were added can be found in Table 1.

Table 1.

MMPI-A-RF F-r Items Selected by Each Addition Procedure.

Item Addition Position	MMPI-A-RF Item Selected for First Addition Procedure and Keyed Direction (Endorsement Rate Normative Sample)	Change in Endorsement Rate from Previous Item (Δ)	MMPI-A-RF Item Selected for Second Addition Procedure and Keyed Direction (Endorsement Rate Faking Bad Sample)	Change in Endorsement Rate from Previous Item (Δ)
1st	95T (5.20%)	-	3F (43.50%)	-
2nd	96F (6.50%)	1.30%	50T (52.90%)	8.40%
3rd	20T (7.70%)	1.20%	10F (60.10%)	7.20%
4th	27T (8.40%)	0.70%	63T (60.90%)	0.80%
5th	134T (9.60%)	1.20%	108T (61.60%)	0.70%
6th	122F (9.60%)	0.00%	96F (63.00%)	2.40%
7th	68T (10.10%)	0.50%	116T (63.80%)	0.80%
8th	81T (10.70%)	0.60%	180T (63.80%)	0.00%
9th	103T (11.00%)	0.30%	81T (64.50%)	0.70%
10th	148T (11.50%)	0.50%	103T (64.50%)	0.00%
11th	3F (11.60%)	0.10%	95T (65.20%)	0.70%
12th	63T (11.70%)	0.10%	122F (65.20%)	0.00%
13th	108T (12.00%)	0.30%	189T (66.70%)	1.50%
14th	180T (12.70%)	0.70%	43T (68.60%)	1.90%
15th	116T (13.00%)	0.30%	13T (69.60%)	1.00%

Participants

The MMPI-A-RF normative sample is composed of a subset of the original MMPI-A normative sample by Butcher et al. (1992) rescored as MMPI-A-RF protocols. This original sample was collected with a strategy to obtain a representative sample of the United States population based on ethnicity, geography, and rural vs. urban residents. Data were collected in junior high and high schools in California, Minnesota, New York, North Carolina, Ohio, Pennsylvania, Virginia, and Washington (Butcher et al., 1992). The MMPI-A-RF normative sample is consisting of 1610 adolescents, with half being boys and half being girls (n = 805). The average age for the boys in the sample was 15.54 (SD = 1.17), and the average age of the girls was 15.59 (SD = 1.18). The sample of boys was 76.5% Caucasian, 12.4% African American, and 11.1% other. For the girls, 76.1% were Caucasian, 12.3% were African American, and 11.6% were other.

The overreporting sample that was used to calibrate the specificity of the combined VRIN-r and F-r and the combined CRIN and F-r scales is Stein et al.'s (1995) non-clinical, overreporting sample of adolescents who took the MMPI-A. Stein et al.'s (1995) sample consisted of 58 male and 80 female high school students, with the average age of the boys being 15.5 (SD = 0.68) and the girls being 15.7 (SD = 0.58). Of the boys, 96.6% were white, with none indicating that they were Asian American, African American, Native American, or Hispanic, and 3.4% reporting that they were an ethnicity other than six previously mentioned. 87.5% of the girls were white, 2.5% were Asian American, 6.3% were African American, 1.3% Hispanic, with the remaining 2.4% indicating their ethnicity was not one of the six listed. The MMPI-A-RF uses a subset of the items of the MMPI-A (Archer et al., 2016), thus Stein et al.'s (1995) MMPI-A

protocols were converted into MMPI-A-RF protocols in order to be used in analyses to evaluate the hybrid scales.

Stein et al.'s (1995) over-reporting participants were instructed to respond to the MMPI-A in a way that would "give the impression that you have serious psychological problems and need hospital treatment where you can talk with a counselor, psychologist, or other doctor about your emotional problems" (p. 420). This sample also completed the MMPI-A under standard instructions. Results revealed that the mean *T* score for the F scale for girls that were asked to overreport was significantly greater than both mean *T* scores of the F scale obtained when these same girls completed the test under standard instructions ($t(80)= 15.68$, $p < .001$) and in a subsample of 80 different girls drawn from a clinical sample by Williams and Butcher (1989) who took it under standard instructions ($t(80) = 13.94$, $p < .001$). The mean T score of the F scale for the overreporting boys was also significantly greater than the *T* score obtained with the same boys under standard conditions ($t(58) = 10.64$, $p < .001$) and in 58 boys drawn from the clinical sample (Williams & Butcher, 1989) ($t(58) = 10.37$, $p < .001$). Conversely, the K scale mean which indicates defensiveness and minimization, was significantly lower in the overreporting girls and boys group compared to both their own scores under standard instructions (girls: $t = -6.01$, $p < .001$, boys: $t = -3.07$, $p < .003$) and compared to the girls ($t = -4.92$, $p < .001$) and boys ($t = -3.86$, $p < .001$) from Williams and Butcher's (1989) clinical sample. Finally, every clinical scale mean, except for the Masculine and Feminine interest scale in the boys, was significantly greater in the overreporting group when compared to both the same group under standard instructions and in the clinical sample (Williams & Butcher, 1989). Specific means and *t* statistics between the three samples can be fully reviewed in Stein et al.'s (1995) article. These data suggest that the sample being used in this study did significantly overreport their

psychological symptoms, as the F scale and most clinical scales for both girls and boys was significantly greater and the K scale significantly lower when compared to both their own MMPI-A scores when taking the test under standard instructions and with a separate clinical sample from Williams and Butcher (1989).

When the Stein et al. (1995) data were converted to MMPI-A-RF profiles the mean unrounded, untruncated *T* score for the F-r scale for the sample was 94.34 (*SD* = 29.41) which indicated general overreporting in the new F-r scale format, as it did in the original F scale. Furthermore, in general, there was no evidence of excessive fixed responding (TRIN-r *T* score *M* = 51.08, *SD* = 9.18), random responding (VRIN-r *T* score *M* = 45.84, *SD* = 9.47), or combined non-content responsiveness (CRIN *T* score *M* = 46.15, *SD* = 10.56). Nor was there any evidence of underreporting or defensiveness (L-r *T* score *M* = 49.59, *SD* = 11.98); K-r *T* score *M* = 42.05, *SD* = 9.85). Thus, when the overreporting data was converted to MMPI-A-RF protocols, the validity scales indicated that the sample only overreported, and did not tend to have any other aberrant response styles.

Measures

The Minnesota Multiphasic Personality Inventory-Adolescent-Restructured Form (MMPI-A-RF) is a 241-item test that is subset of the original items in the MMPI-A. The MMPI-A-RF is an instrument used to assess a broad array of psychopathology symptoms and personality traits in adolescents and is intended for use in psychological evaluations (Archer et al., 2016).

The Variable Response Inconsistency-Restructured (VRIN-r) scale of the MMPI-A-RF was created to detect random-responding to the protocol and was created by using the steps delineated by Tellegen and Ben-Porath (2008) mentioned above (Archer et al., 2016). It is

composed of 27 pairs of items. Alpha coefficients for VRIN-r scale in the normative sample were .45 for adolescent males and .37 for adolescent females. The low alpha coefficients for VRIN-r are expected as they were designed to be content free, further mostly cooperative individuals in the normative sample would likely not produce reliable and highly variant invalid responses (Archer et al., 2016). Archer et al. (2016) also noted that the VRIN-r variances are low, and their standard errors are small enough to warrant use in invalidating protocols at the suggested cutoffs.

The Combined Response Inconsistency (CRIN) scale for the MMPI-A-RF is a composite of both VRIN-r and TRIN-r scales and was made to capture both a random and fixed response style. Thus, it is a measure of content non-responsiveness, in general, within MMPI-A-RF protocols. The CRIN scale is composed of 40 pairs of items. Alpha coefficients for the CRIN scale in the normative sample were .60 for adolescent males and .52 for adolescent females (Archer et al., 2016). The less than ideal alpha coefficients can be explained by the fact that the CRIN scale is composed of the relatively content-free VRIN-r and TRIN-r scales. The greater alpha coefficients of the CRIN scale as compared to the VRIN-r scale are largely a product of the CRIN scale being longer than the VRIN-r scale. Like the VRIN-r scale, Archer et al. (2016) reported that the CRIN-r variances and standard errors are small enough to be reliable at invalidating protocols at the suggested cutoffs.

The Infrequent Reponses - Restructured, (F-r) scale of the MMPI-A-RF was created to detect endorsement of items that were indicative of severe psychopathology and were also very rarely endorsed in the normative sample, thus suggesting overreporting. Alpha coefficients for the F-r scale in the normative sample were .74 for boys and .71 for girls (Archer et al, 2016).

These alpha coefficients are respectable, and standard errors are small enough to be reliable at invalidating protocols at the suggested cutoffs (Archer et al., 2016).

Overview of Analyses

First, Mean T scores and percentage above the 75 T score cutoff data for the CRIN and VRIN-r hybrid scales by varying degrees of randomness in the normative sample were calculated. General trends in the data were explored by fitting trends lines. This allowed a general overview of how adding additional items tends to affect the Mean *T* and percentage above cutoff scores for each hybrid scale. Furthermore, it provided an opportunity to look at how these trends were affected by the varying levels of randomness. In this part of the analysis larger Mean *T* scores and greater percentages above the cut-off suggested that the scales were working more efficiently at identifying random responding.

Next, a comparison between the efficacy of the VRIN-r and CRIN hybrid scale at detecting random responding was completed by comparing the mean *T* score and percentage above cutoff data for the normative sample within each of the two item addition procedures. This was done in two ways. First, because fifteen items had been added and there were four levels of randomness (25%, 50%, 75%, and 100%), there were 60 cells of data that were produced for both mean T scores and percentage above cutoff in the analyses. These 60 cells were compared between the VRIN-r hybrid scale the CRIN scales to ascertain which scale had the greater value at the same level of item addition and percentage of randomness within the same item addition procedure (either the first or the second procedure). The scale that had the greatest amount of high values would indicate better efficiency at detecting random responding. The second analysis obtained the mean *T* scores and mean percentage above cutoff values by each level of randomness, collapsing across all fifteen items added. This produced four mean T scores and

four mean percentage above cutoff scores for both the VRIN-r and CRIN hybrid scale for both item addition procedures (a total of eight means for each scale). T-tests were then used to compare these means between the CRIN and VRIN-r hybrid scales. The scale that has the greatest amount of significantly greater means would indicate that it is performing best at detecting random responding.

Once this was completed, the same procedure mentioned in the previous paragraph was used to compare the efficacy of the item addition procedures used to create the VRIN-r hybrid scale and CRIN hybrid scale. A comparison was completed between the 60 cells of data within the VRIN-r hybrid scale and the CRIN scale by addition procedure type. Again, the addition procedure that yields the greatest number of cells with the highest scores would suggest it is the most efficient at detecting random responding. Means collapsing across all fifteen items were also obtained for each of the four random insertion levels. *t*-tests were used to compare the means produced within the VRIN-r and CRIN hybrid scale by addition procedure. The addition procedure that produced the largest amount of significantly greater means would indicate that it is performing best at detecting random responding.

Next, Mean *T* scores and percentage above the cutoff data for the VRIN-r and CRIN hybrid scales was obtained in Stein et al.'s (1995) overreporting sample. Again, trend lines were fitted to aid in exploring how adding additional items tended to affect the data in an overreporting sample. In this analyses, lower mean T scores and percentage above cut-off are ideal, as this suggests that the VRIN-r and CRIN hybrid scales were more specific from the F-r scale and are not oversaturated with F-r items.

Again, like what was done with the analyses within the normative sample, a comparison of the efficacy of the VRIN-r hybrid scale vs. the CRIN hybrid scale was conducted along with a

comparison of the addition procedures within the overreporting sample. Twenty cells of data are produced in these analyses which allowed direct comparison between the VRIN-r hybrid scale and CRIN hybrid scale within addition procedures and between addition procedures within the VRIN-r hybrid scale and CRIN scale themselves. The scale and the addition procedure that produced the lower value would suggest better specificity from the F-r scale. Means collapsing across all item addition levels will also be obtained and compared with *t*-tests. Again, the scale and addition procedure that produces the greater amount of significantly lower means would suggest the best specificity.

Next, the correlations between the VRIN-r hybrid and CRIN hybrid scale with the F-r scale were examined by item addition level. Special attention was paid to when the correlations become significant and large, as this possibly suggested the item level where both hybrid scales become oversaturated with F-r items are and are not significantly unique from the F-r scale.

Once these data have been analyzed, the ideal number of items to add to both the VRIN-r and CRIN hybrid scales was identified. This was be done by determining which item addition procedure was best at producing the highest sensitivity to random responding and the highest specificity from overreporting. Once the best item addition procedure was identified, the item level within this procedure where specificity of 90% (the rationale for this was discussed above) was located (i.e., had the percentage above cutoff score within the overreporting sample that was *at or below* 10%). This served as the upper bound of items to add, as items added above this produced unwanted, low specificity. Next, item addition levels that fell below this upper bound were reviewed in the normative sample for each hybrid scale that had the highest sensitivity to random responding (i.e., had the percentage above cutoff within the normative sample that was the *highest)*. Although the percentage above cutoff scores are not *exact* measures of sensitivity

and specificity they serve as practical proxy measures. These procedures identified the ideal number of items to add to each hybrid scale by maximizing sensitivity to randomness while maintaining an adequate level of specificity from overreporting.

Once the ideal number of items was determined for both the VRIN-r and CRIN hybrid scales, correlations between these new scales and the Restructured Clinical scales were obtained and compared to correlations that were produced between the original VRIN-r, CRIN, and F-r scales and the Restructured Clinical scales. This provided another opportunity to ascertain if the new VRIN-r and CRIN hybrid scales were significantly unique from the variance of the F-r scale. Ideally, the correlations between the new hybrid scales and the Restructured Clinical scales will be low and resemble the correlations the original CRIN and VRIN-r scales have with the Restructured Clinical scales. This would provide additional evidence that the new hybrid scales are sufficiently unique from the F-r scale; sensitive to random responding but not overly sensitive to overreporting.

CHAPTER III

RESULTS

Mean *T* scores by Simulated Randomness in Normative Sample

Mean *T* scores combined VRIN-r/F-r and CRIN/F-r Scales overall trend as items were added: first addition procedure. The mean T scores of the newly created VRIN-r/F-r and CRIN/F-r Scales by item addition level and percentage of inserted randomness for the first addition procedure in the normative sample can be seen in Tables 2 and 3 and graphically in Figures 1 and 2, respectively. Multiple trends can be discerned from these data. First, as the percentage of randomness was increased, mean *T* scores increased at every item level. This provides evidence that VRIN-r/F-r and CRIN/F-r scales are sensitive to higher degrees of random responding, and as F-r items are added. Second, there appeared to be a trend in both hybrid scales where the mean *T* score increases rapidly as initial items are added but then begins to increase more slowly as later items are added. This suggests a quadratic trend in the data, and data analyses presented below bore this out.

Trend analyses of the mean T scores for both hybrid scales revealed statistically significant quadratic trends for the 25% to 50% levels of random as item levels were increased; however, there did not appear to be a clinically significant changes in these scores. Specifically, looking at the 25% and 50% randomness level of the VRIN-r/F-r hybrid scale, the quadratic equations of $y = -0.0155x^2 + 0.3062x + 58.557$ and $y = -0.0298x^2 + 0.6462x + 65.187$ (in this equation and all future equations y = predicted Mean *T* score or percent above cutoff and x = the number of items added) were fit to the data (see Figure 1). The quadratic trends were both

Table 2.

Mean T Scores and Standard Deviations for Combined VRIN-r/F-r Scale in MMPI-A-RF Normative Sample for First Item Addition Procedure.

Percent Simulated Randomness	Number of F-r Items Added to VRIN-r Scale															
	0	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
25%	**58.40**	**58.67**	**59.20**	**59.49**	**59.78**	**59.86**	**59.93**	**59.96**	**59.92**	**59.57**	**60.14**	**59.99**	**59.88**	**60.10**	**59.83**	**59.74**
	10.04	*9.96*	*10.23*	*10.05*	*9.92*	*10.11*	*9.76*	*9.70*	*9.79*	*9.76*	*9.55*	*9.60*	*9.42*	*9.24*	*9.57*	*9.46*
50%	**64.98**	**65.40**	**66.38**	**67.32**	**67.79**	**67.91**	**67.85**	**68.37**	**68.36**	**68.34**	**68.37**	**68.60**	**68.17**	**68.89**	**68.51**	**68.36**
	10.10	*10.28*	*10.14*	*10.23*	*10.29*	10.12	*10.18*	*9.99*	*9.92*	*9.70*	*9.85*	9.58	*9.35*	*9.44*	*9.17*	*9.03*
75%	**68.70**	**70.47**	**72.06**	**72.81**	**73.29**	**74.27**	**74.20**	**74.61**	**75.28**	**74.68**	**74.86**	**75.20**	**75.13**	**75.64**	**76.07**	**75.89**
	10.17	*10.82*	*10.29*	*10.52*	*10.43*	*10.69*	*10.53*	*10.37*	*10.51*	*9.96*	*9.73*	*9.72*	*9.34*	*9.36*	*9.35*	*9.20*
100%	**70.98**	**73.02**	**74.74**	**75.69**	**77.40**	**78.27**	**79.11**	**79.39**	**80.23**	**80.11**	**80.54**	**80.70**	**81.12**	**81.77**	**81.44**	**81.37**
	10.70	10.91	*10.54*	*10.60*	*10.66*	*10.76*	*10.46*	*10.31*	*10.15*	*10.00*	*9.85*	*9.58*	*9.64*	*9.45*	*9.11*	*8.90*

M in bold and *SD* in italics. Total sample size 1610 for each cell.

Table 3.

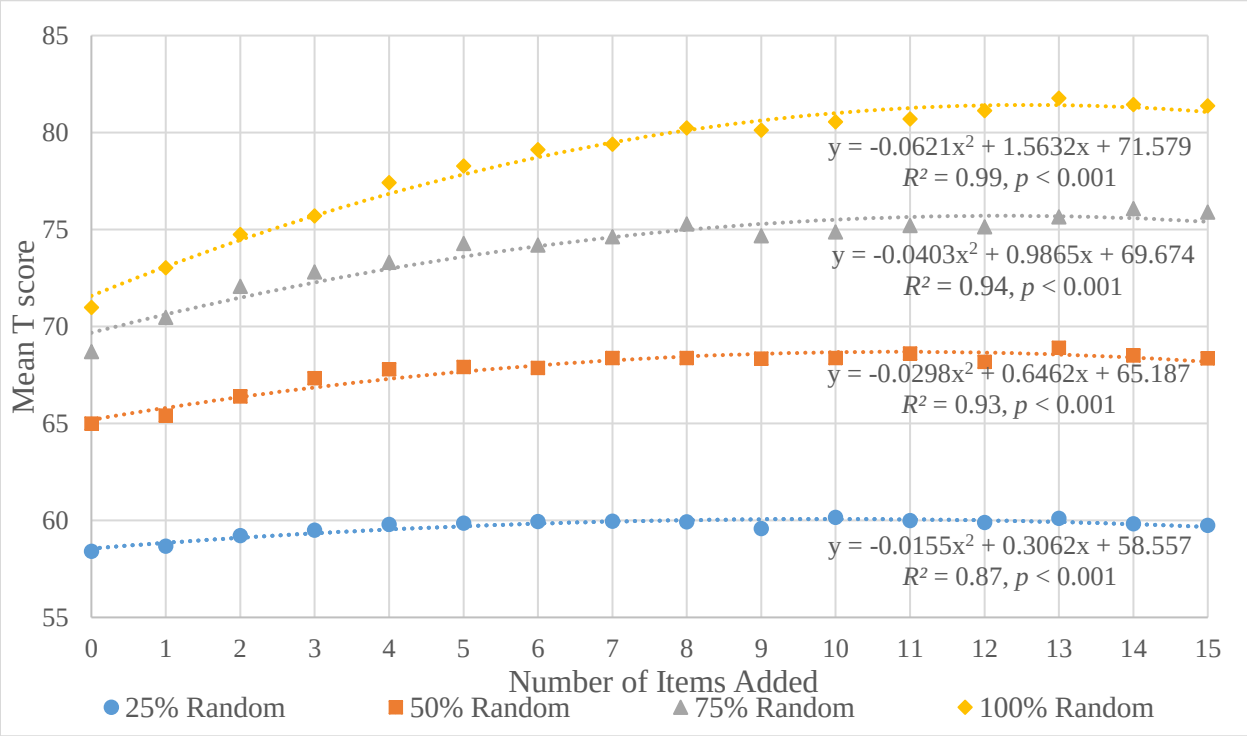
Mean T scores for Combined CRIN/F-r Scale in MMPI-A-RF Normative Sample for First Item Addition Procedure.

Percent Simulated Randomness	Number of F-r Items Added to CRIN Scale															
	0	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
25%	**58.67**	**58.78**	**59.42**	**59.42**	**59.58**	**59.73**	**59.76**	**59.65**	**59.82**	**59.43**	**59.81**	**59.67**	**59.74**	**59.81**	**59.73**	**59.67**
	9.56	*9.40*	*9.65*	*9.54*	*9.23*	*9.37*	*9.05*	*9.21*	*9.10*	*9.17*	*8.99*	*9.17*	*9.01*	*8.90*	*9.24*	*9.18*
50%	**65.27**	**65.76**	**66.44**	**66.88**	**67.36**	**67.51**	**67.47**	**67.94**	**67.96**	**67.79**	**68.05**	**68.19**	**67.74**	**68.28**	**67.98**	**67.89**
	9.28	*9.36*	*9.19*	*9.21*	*9.34*	*9.22*	*9.05*	*9.07*	*8.93*	*8.94*	*9.09*	*8.69*	*8.55*	*8.69*	*8.61*	*8.50*
75%	**69.21**	**70.80**	**71.80**	**72.24**	**72.85**	**73.32**	**73.53**	**73.54**	**74.34**	**73.72**	**73.98**	**74.40**	**74.12**	**74.74**	**75.15**	**75.06**
	9.20	*9.77*	*9.24*	*9.18*	*9.37*	*9.54*	*9.42*	9.20	*9.42*	*8.87*	*8.89*	*8.95*	*8.49*	*8.68*	*8.62*	*8.54*
100%	**71.90**	**73.48**	**74.84**	**75.44**	**76.60**	**77.34**	**77.87**	**78.03**	**78.93**	**78.63**	**79.16**	**79.25**	**79.54**	**80.20**	**80.04**	**80.06**
	9.66	*9.63*	*9.44*	*9.40*	*9.68*	*9.56*	*9.42*	*9.37*	*9.24*	*9.08*	*9.18*	*8.84*	*8.88*	*8.50*	*8.43*	*8.27*

M in bold and *SD* in italics. Total sample size 1610 for each cell.

Figure 1.

Mean T scores for combined VRIN-r/F-r scale in MMPI-A-RF normative sample for first item addition procedure.

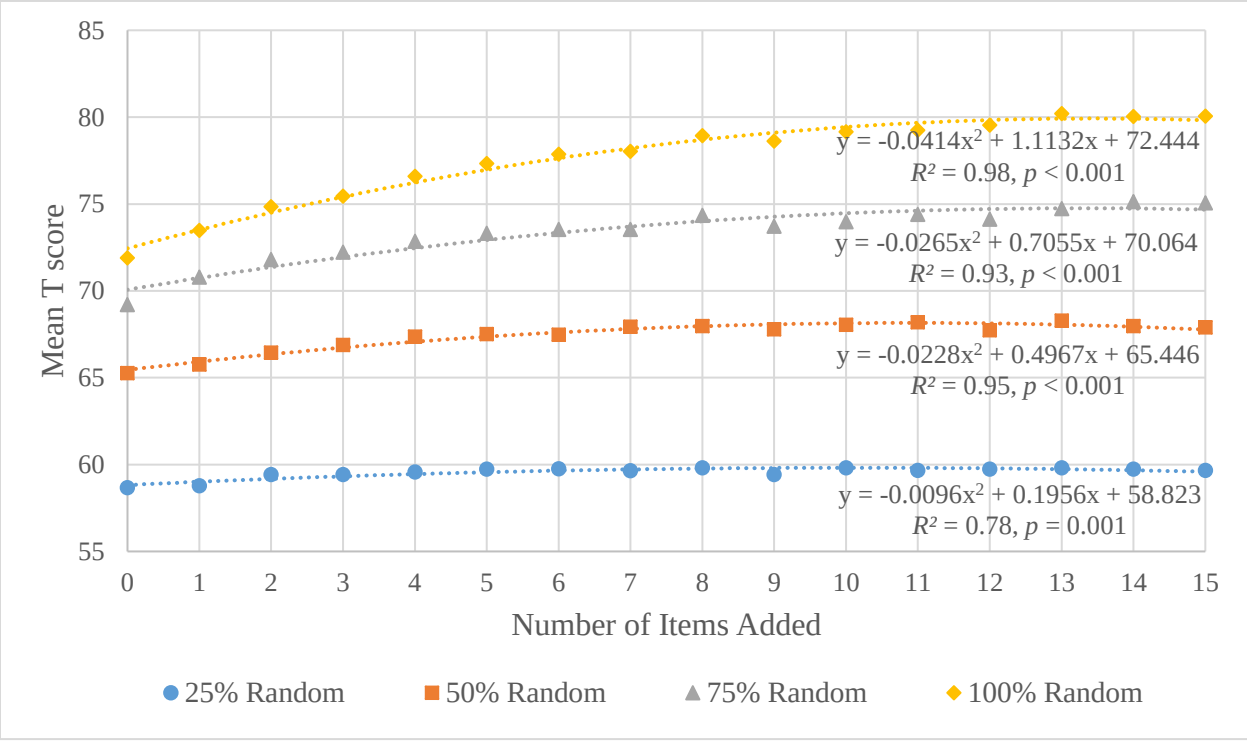


significant (both $p < 0.001$) and excellent fits for the data ($R^2 = 0.87$ and 0.93, respectively). Quadratic equations of $y = -0.0096x^2 + 0.1956x + 58.823$ and $y = -0.0228x^2 + 0.4967x + 65.446$ were fit to the data of Mean *T* scores by item addition in the 25% and 50% level of randomness on the CRIN hybrid scale, respectively (see Figure 2). These equations were also statistically significant fits ($p = 0.001$ and < 0.001, correspondingly) and excellent in their fit to the data ($R^2 = 0.75$ and 0.95, respectively). Although these trends were all significant for the 25% and 50% randomness levels for both hybrid scales, a difference of five *T* score points is considered "clinically significant" (Greene, 1987) on the MMPI-A-RF. Reviewing Tables 2 and 3, the Mean

T scores for the 25% and 50% randomness levels never become greater than five *T* score points from baseline in either hybrid scales. Thus, when randomness is kept at 50% or below, both hybrid scales do not seem to produce clinically significant changes in mean *T* scores when comparing them to baseline, suggesting they may not be ideally sensitive to low levels of random responding.

Figure 2.

Mean T scores for combined CRIN/F-r scale in MMPI-A-RF normative sample for first item addition procedure.



On the other hand, both the VRIN-r/F-r and CRIN/F-r scales showed more robust changes in mean *T* scores as items were added at the 75% and 100% randomness level. For the VRIN-r/F-r scale, the quadratic equations of $y = -0.0403x^2 + 0.9865x + 69.674$ and $y = -0.0621x^2 + 1.5632x + 71.579$ were fit to the 75% and 100% random data as item level was increased (see Figure 1). Both were statistically significant ($p < 0.001$) and very excellent fits ($R^2 = 0.94$ and 0.99, respectively). The CRIN hybrid scale produced quadratic fit equations of $y = -0.0265x^2 + 0.7055x + 70.064$ and $y = -0.0414x^2 + 1.1132x + 72.444$ for the 75% and 100% randomness, respectively, which were both significant ($p < 0.001$) and of ideal fit ($R^2 = 0.93$ and 0.98, correspondingly). Keeping in mind the five-point *T* score difference indicating clinical significance, the VRIN-r hybrid scale discerned a clinically significant change compared to baseline after five items were added at the 75% level, and after four items were added at the 100% level. For the CRIN-r hybrid scale this was after eight items and five items were added for the 75% and 100% levels, respectively. Reviewing all the quadratic equations for both hybrid scales for the first addition procedure found that, as randomness increases, the linear component of the equation is much larger at the 75% and 100% randomness levels compared to the 25% and 50% randomness levels, indicating a steeper increase in mean *T* scores at these latter levels of randomness as items are added. Taken together, these data suggest the hybrid scales are much more sensitive to detecting clinically significant changes in mean *T* scores as more items are added when random saturation of profiles is 75% or above.

Mean *T* scores combined VRIN-r/F-r and CRIN/F-r Scales overall trend as items are added: second addition procedure. Trends in Mean T score changes for the VRIN- and CRIN hybrid scales on the second addition procedure in the normative sample were very similar to those found in the first addition. These data can be viewed in Tables 4 and 5 and

Figures 3 and 4, respectively. First, again mean *T* scores at each level increased as the simulated randomness was increased, replicating the finding in the first addition procedure that the hybrid VRIN-r and CRIN scales were sensitive to increasing levels of random responding in the second addition procedure, as well as to increased sensitivity as F-r items were added. However, unlike the first addition procedure, here a linear trend to the data appeared to be the ideal fit, which was supported in analyses.

Figure 3.

Mean T scores for combined VRIN-r/F-r scale in MMPI-A-RF normative sample for second item addition procedure.

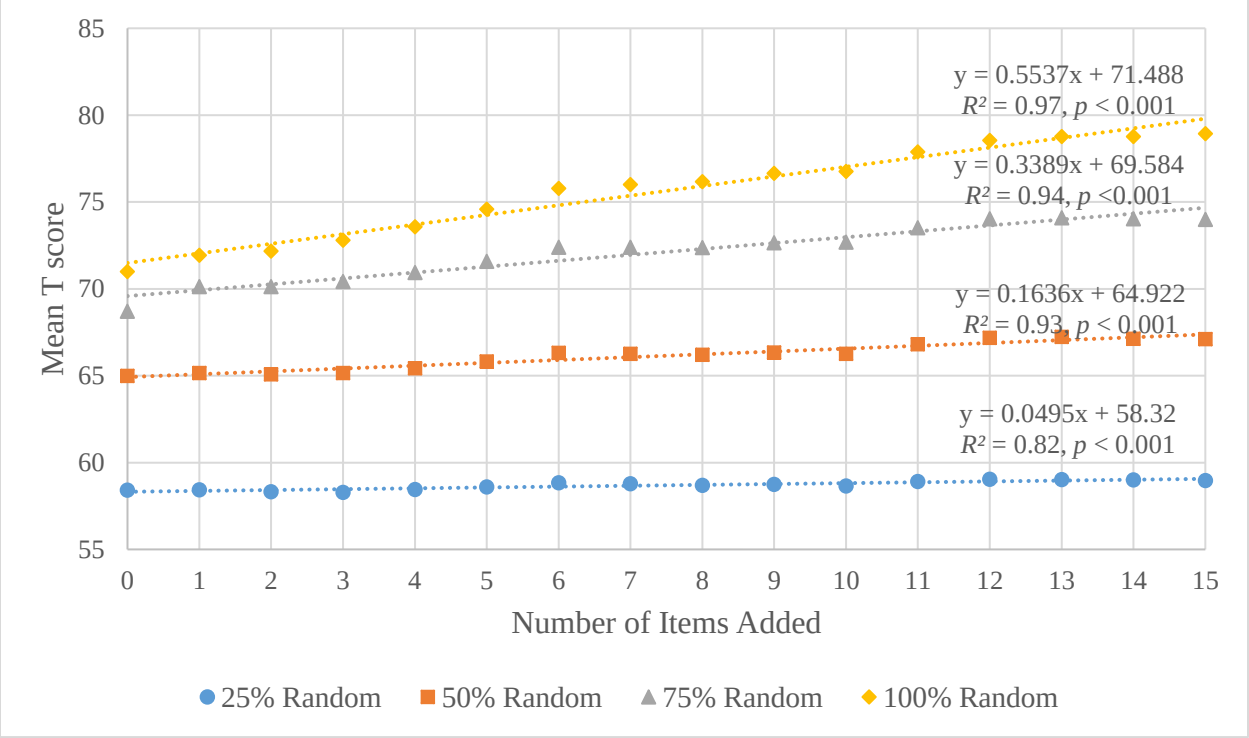


Table 4.

Mean T and scores for Combined VRIN-r/F-r Scale in MMPI-A-RF Normative Sample for Second Item Addition Procedure.

Percent Simulated Randomness	Number of Fr-r Items Added to VRIN-r Scale															
	0	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
25%	**58.40**	**58.42**	**58.31**	**58.28**	**58.44**	**58.59**	**58.84**	**58.78**	**58.68**	**58.74**	**58.65**	**58.90**	**59.04**	**59.02**	**58.99**	**58.97**
	10.04	*9.99*	*9.85*	*9.83*	*9.78*	*9.75*	*9.69*	*9.59*	*9.49*	*9.43*	*9.31*	*9.32*	*9.38*	*9.39*	*9.39*	*9.33*
50%	**64.98**	**65.15**	**65.07**	**65.16**	**65.42**	**65.82**	**66.31**	**66.26**	**66.20**	**66.33**	**66.26**	**66.81**	**67.17**	**67.23**	**67.12**	**67.10**
	10.10	*9.92*	*9.64*	*9.58*	*9.50*	*9.50*	*9.34*	*9.20*	*8.99*	*8.49*	*8.78*	*8.78*	*8.86*	*8.88*	*8.83*	*8.75*
75%	**68.70**	**70.12**	**70.13**	**70.42**	**70.93**	**71.59**	**72.40**	**72.38**	**72.38**	**72.65**	**72.69**	**73.51**	**74.02**	**74.08**	**74.03**	**73.99**
	10.17	*10.21*	*9.73*	*9.66*	*9.53*	*9.61*	*9.44*	*9.30*	*9.09*	*9.08*	*8.87*	*8.91*	*8.90*	*8.88*	*8.83*	*8.64*
100%	**70.98**	**71.93**	**72.17**	**72.79**	**73.56**	**74.59**	**75.78**	**76.00**	**76.17**	**76.65**	**76.76**	**77.89**	**78.55**	**78.77**	**78.77**	**78.93**
	10.70	*10.46*	*10.12*	*9.93*	*9.78*	*9.83*	*9.61*	*9.39*	*9.09*	*8.88*	*8.70*	*8.80*	*8.78*	*8.79*	*8.68*	*8.52*

M in bold and *SD* in italics. Total sample size 1610 for each cell.

Table 5.

Mean T scores for Combined CRIN/F-r Scale in MMPI-A-RF Normative Sample for Second Item Addition Procedure.

Percent Simulated Randomness	Number of Fr-r Items Added to CRIN Scale															
	0	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
25%	**58.67**	**58.66**	**58.57**	**58.55**	**58.66**	**58.79**	**58.96**	**58.93**	**58.84**	**58.89**	**58.80**	**59.00**	**59.11**	**59.12**	**59.09**	**59.08**
	9.56	*9.51*	*9.42*	*9.41*	*9.36*	*9.35*	*9.30*	*9.25*	*9.19*	*9.13*	*9.09*	*9.10*	*9.13*	*9.15*	*9.14*	*9.09*
50%	**65.27**	**65.31**	**65.26**	**65.33**	**65.53**	**65.86**	**66.21**	**66.19**	**66.14**	**66.27**	**66.17**	**66.63**	**66.92**	**67.01**	**66.93**	**66.94**
	9.28	*9.11*	*8.90*	*8.87*	*8.81*	*8.83*	*8.70*	*8.58*	*8.42*	*8.38*	*8.35*	*8.35*	*8.3t9*	*8.37*	*8.29*	*8.24*
75%	**69.21**	**70.46**	**70.46**	**70.68**	**71.07**	**71.60**	**72.19**	**72.21**	**72.20**	**72.45**	**72.43**	**73.11**	**73.53**	**73.64**	**73.61**	**73.61**
	9.20	*9.32*	*8.99*	*8.93*	*8.83*	*8.88*	*8.74*	*8.67*	*8.48*	*8.48*	*8.38*	*8.41*	*8.37*	*8.36*	*8.29*	*8.14*
100%	**71.90**	**72.61**	**72.77**	**73.22**	**73.81**	**74.62**	**75.49**	**75.70**	**75.82**	**76.24**	**76.28**	**77.21**	**77.75**	**78.01**	**78.02**	**78.21**
	9.66	*9.30*	*9.04*	*8.89*	*8.82*	*8.84*	*8.68*	*8.52*	*8.28*	8.14	*8.08*	*8.15*	*8.13*	*8.13*	*8.05*	*7.93*

M in bold and *SD* in italics. Total sample size 1610 for each cell.

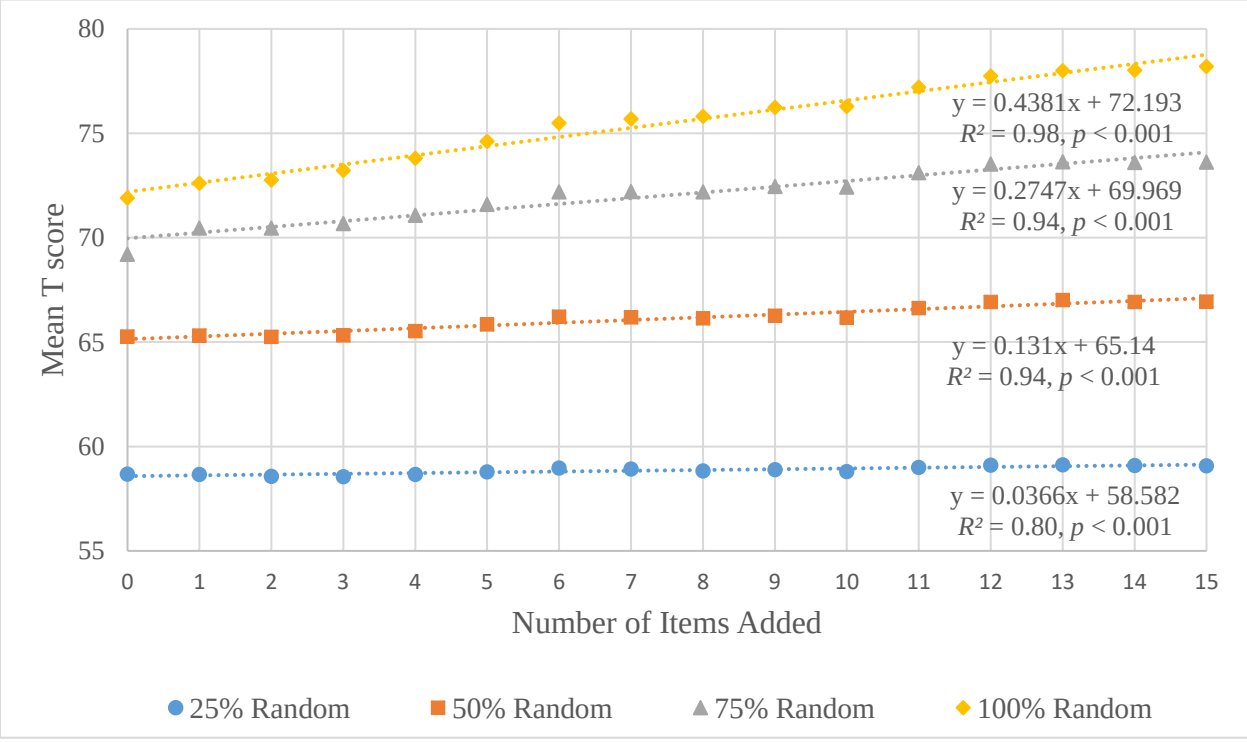
The VRIN-r and CRIN hybrid scales mean *T* scores at the 25% and 50% simulated had a slight, positive linear trend; however, they changed very little as items were added. Reviewing the VRIN-r hybrid scale, the linear equations of y = 0.0495x + 58.32 and y = 0.1636x + 64.922 fit the 25% and 50% randomness levels as items were increased significantly (both $p < 0.001$) and with excellent fit ($R^2 = 0.82$ and 0.93, correspondingly; see Figure 3). Likewise, for the CRIN scale, the linear equations for the 25% and 50% randomness levels of y = 0.0366x + 58.582 and y = 0.131x + 65.14 were both significant (both $p < 0.001$) and great fits ($R^2 = 0.80$ and 0.94, correspondingly; see Figure 4). Although all these equations were significant linear trends, the slopes on all four are miniscule, suggesting a very small increase in Mean *T* scores as items are added. Indeed, none of the mean *T* scores increased more than five points above baseline throughout the entire item addition process for both hybrid scales for the 25% and 50% randomness levels, suggesting both scales failed to detect any clinically meaningful changes in randomness as items were added.

More robust increases in mean *T* scores were detected by item addition in the VRIN-r and CRIN hybrid scales at the 75% and 100% level of randomness. The best fitting linear equations for the VRIN-r hybrid scale at the 75% and 100% randomness levels were y = 0.3389x + 69.584 and y = 0.5537x + 71.488, respectively. Both equations were statistically significant ($p < 0.001$) and very good fits ($R^2 = 0.94$ and 0.97, correspondingly; see Figure 3). For the CRIN/F-r scale, the equations of y = 0.2747x + 69.969 and y = 0.4381x + 72.193 fit significantly ($p < 0.001$) and excellently ($R^2 = 0.94$ and 0.97) for the 75% and 100% randomness levels, correspondingly (see Figure 4). At the 75% randomness level, the VRIN-r hybrid scale produced a mean *T* score clinically significant from baseline after twelve items were added; however, the CRIN hybrid scale failed to produce a mean *T* score that reached clinical significance from baseline during the

entire item addition procedure at this randomness level. At the 100% randomness level, clinical significance from baseline was reached after the seventh item was added for the VRIN-r/F-r scale and after the eleventh item was added for the CRIN/F-r hybrid. Like what was found in the first addition procedure, the linear slope component to the trend of Mean *T* score by item addition increased as the percentage of randomness increased. Taken together, these data suggest, just like what was found in the first addition procedure, the hybrid scales in the second addition procedure are much more sensitive to detecting clinically significant changes in Mean *T* scores in responding that is saturated with randomness of 75% or above.

Figure 4.

Mean T scores for combined CRIN/F-r scale in MMPI-A-RF normative sample for second item addition procedure.



Comparing VRIN/F-r and CRIN/F-r scales by mean *T* scores.

Within the first addition procedure. The mean *T* scores for the VRIN/F-r and CRIN/F-r data for the first addition procedure can be observed in Tables 2 and 3. Fifteen items were added by four levels of simulated randomness (i.e., 25%, 50%, 75%, and 100%) meaning there is a total of 60 cells of data within each of the two tables. Of these 60 cells, the VRIN/F-r scale produced a greater mean *T* score in 42/60 cells when compared to the corresponding data cell in the CRIN/F-r scale for the same item addition and simulated randomness level. This left 8/60 cells where the CRIN/F-r scale had a greater mean *T* score when compared to the corresponding VRIN-r/F-r data cell. These eight cells were concentrated in the first two item addition levels, after that, the VRIN-r/F-r scale consistently produced the greater mean *T* score at every additional item level and percentage of simulated randomness. These findings tentatively suggest that the VRIN-r/F-r scale is a bit more sensitive to random responding when three or more items have been added in the first addition procedure compared to the CRIN/F-r scale; however, statistically this did not hold.

The t-test comparisons of mean *T* scores between the VRIN-r and CRIN hybrid scales within the first addition procedure by each level of randomness can be observed in Table 6. Table 6 also contains the means, standard deviations, and confidence intervals for these data. Although, the VRIN-r hybrid scale consistently produces the greater mean *T* score compared to the CRIN hybrid scale for each percentage of randomness, these differences did not reach statistical significance. These data indicate that, although the VRIN-r/F-r scale produces consistently greater observed mean T scores than the CRIN/F-r scale, these differences are small and are not statistically different from naught; suggesting that the two hybrid scales are, in general, equally sensitive to random responding.

Within the second addition procedure. Turning to the second addition procedure, a few different findings were obtained relative to the ones discussed in the previous section. These mean *T* score data can be located in Tables 4 and 5. Again, looking at the 60 cells produced across item addition and simulated randomness levels, there was a 50% split as to which scale had the greatest mean *T* score by item level and randomness level with each scale capturing the greater mean *T* score in 30/60 of the cells. Here the CRIN/F-r scale captured the greatest mean *T* score across all levels of randomness for the first five items added. After that, the VRIN-r/F-r scale obtained the greatest mean *T* score for item addition levels of six and above, but only for the 50%, 75%, and 100% randomness level. procedure. The *t*-test statistics for the comparison of the mean *T* scores for the VRIN-r hybrid scale by addition procedure can be observed in Table 8. Mean differences reached statistical significance when collapsing across item addition level and looking within every level of randomness. Specifically, at the 25% randomness level, there was a significant difference in the mean *T* scores for the first addition procedure ($M = 59.74$, $SD = 0.38$) and the second ($M = 58.71$, $SD = 0.26$); $t(28) = 8.63$, $p < 0.001$. For the 50% randomness level, similarly the first addition procedure ($M = 67.91$, $SD = 0.92$) produced a statistically greater mean *T* score than the second procedure ($M = 66.23$, $SD = 0.77$); $t(28) = 5.43$, $p < 0.001$. Likewise, at the 75% and 100% randomness level, the first addition procedure's mean *T* scores ($M = 74.29$, $SD = 1.54$; $M = 78.99$, $SD = 2.67$, respectively) were significantly greater than the second addition procedure T scores ($M = 72.35$, $SD = 1.44$; $M = 75.95$, $SD = 2.46$, correspondingly) $t(28) = 3.56$, $p < 0.00$; $t(28) = 3.24$, $p = 0.003$. Effects sizes for these obtained differences at the 25%, 50%, 75%, and 100% randomness levels were large to huge (Cohen's d = 3.16, 1.98, 1.30, and 1.18, respectively; Cohen, 1988; Sawilowsky, 2009). This indicated that, collapsing across item addition level, the first addition procedure for the VRIN-r hybrid scale

Table 6.

Results of t-tests of Mean T scores Between the VRIN-r/F-r and CRIN/F-r Scales on the First Addition Procedure by Percentage of Simulated Randomness in the Normative Sample.

	Scale						95% CI for Mean Difference			Cohen's *d*
	VRIN-r/F-r			CRIN/F-r						
	M	*SD*	*n*	*M*	*SD*	*n*		*t*	*df*	
25% Random	59.74	0.38	15	59.60	0.27	15	-0.11, 0.38	1.13	28	0.42
50% Random	67.91	0.92	15	67.55	0.70	15	-0.25, 0.97	1.20	28	0.44
75% Random	74.29	1.54	15	73.57	1.23	15	-0.32, 1.77	1.42	28	0.52
100% Random	78.99	2.67	15	77.96	2.06	15	-0.75, 2.82	1.19	28	0.43

Table 7.

Results of t-tests of Mean T scores Between the VRIN-r/F-r and CRIN/F-r Scales on the Second Addition Procedure by Percentage of Simulated Randomness in the Normative Sample.

	Scale						95% CI for Mean Difference			Cohen's *d*
	VRIN-r/F-r			CRIN/F-r						
	M	*SD*	*n*	*M*	*SD*	*n*		*t*	*df*	
25% Random	58.71	0.26	15	58.87	0.20	15	-0.33, 0.01	-1.92	28	0.69
50% Random	66.23	0.77	15	66.18	0.62	15	-0.48, 0.57	0.19	28	0.07
75% Random	72.35	1.44	15	72.22	1.16	15	-0.84, 1.12	0.29	28	0.10
100% Random	75.95	2.46	15	75.72	1.95	15	-1.42, 1.90	0.29	28	0.10

was more sensitive to random responding, when compared to the second addition procedure, as it consistently produced greater mean T scores.

Within the CRIN/F-r scale. The results for the differences between the CRIN/F-r scale's mean *T* scores by addition procedure are very similar to what was found for the VRIN/F-r scale. Tables 3 and 4 contain the CRIN/F-r scales mean T scores by item addition procedure. Looking at the 60 cells, the CRIN/F- r scale composed of the first item addition procedure had the greatest mean T score at all item addition and simulated randomness levels compared to the CRIN hybrid scale created by the second item addition procedure. The *t*-test results comparing mean *T* scores for the CRIN hybrid scales found in Table 9 indicated that these differences were significant for each randomness level, collapsing across item level. The first addition procedure created CRIN hybrid scale had mean T scores for the 25% (M = 59.60, SD = 0.27), 50% (M = 67.55, SD = 0.70), 75% (M = 73.57, SD = 1.23) and 100% (M = 77.96, SD = 2.06) randomness levels that were significantly greater than the second addition procedure produced CRIN hybrid scale at the same corresponding randomness level (M = 58.87, SD = 0.20; M = 66.18, SD = 0.62; M = 72.22, SD = 1.16; M = 75.72, SD = 1.95, respectively). The corresponding *t*-tests for these four contrasts by randomness level were $t(28) = 8.57 (p < 0.001)$, 5.66 ($p < 0.001$), 3.12 ($p = 0.004$), and 3.07 ($p = 0.005$). The respective Cohen's *d'*s of 3.58, 2.07, 1.14, and 1.10 suggest large to huge effect sizes for each of these contrasts. These data all suggested that, like what was discovered within the VRIN-r/F-r scale, the first addition procedure consistently produced greater mean *T* scores for the CRIN hybrid scale than the second addition procedure create CRIN hybrid scale, which provides evidence that the first addition procedure creates a CRIN hybrid scale that is more sensitive to random responding.

Table 8.

Results of t-tests of Mean T scores Between the VRIN-r/F-r Scales by Addition Procedure and Percentage of Simulated Randomness in the Normative Sample.

	Scale						95% CI for Mean Difference			Cohen's *d*
	VRIN-r/F-r, First Addition Procedure			VRIN-r/F-r, Second Addition Procedure						
	M	*SD*	*n*	*M*	*SD*	*n*		*t*	*df*	
25% Random	59.74	0.38	15	58.71	0.26	15	0.78, 1.27	8.63***	28	3.16
50% Random	67.91	0.92	15	66.23	0.77	15	1.05, 2.32	5.43***	28	1.98
75% Random	74.29	1.54	15	72.35	1.44	15	0.83, 3.06	3.56***	28	1.30
100% Random	78.99	2.67	15	75.95	2.46	15	1.12, 4.96	3.24**	28	1.18

** $p < 0.01$, *** $p < 0.001$

Percentage Above Cutoff by Simulated Randomness in Normative Sample

Notable trend in data. The percentage above cutoff for the VRIN-r and CRIN hybrid scales for the first item addition procedure within the normative sample can be seen in Tables 10 and 11, respectively and for the second item addition procedure within the normative sample in Tables 12 and 13, respectively. Graphically, the percentage above cutoff data for the VRIN-r hybrid scale for the first and second addition procedure can be seen in Figures 5 and 6, correspondingly, and in the CRIN hybrid scale by addition procedure in Figures 7 and 8. Before reviewing the specific findings, it is best to clarify one of the interesting overall trends in the data. There is a distinct pattern for the percentage above cutoff data for the VRIN-r and CRIN hybrid scales as item levels are increased which can best be described as sinusoidal. This has to do with the how the *T* score that is just below the 75 cutoff fluctuates with each additional item. The *T* score point level just below the 75-cutoff hovered between 70 and 75 at every level of the item addition process. When this *T* score point was closer to 70, percentages above the cutoff tended to be higher as the next highest *T* score point tended to just barely be over 75 (allowing for more respondents to be captured above 75). On the other hand, when the T score point below 75 was just barely below this point (e.g., at 74) percentages above the cutoff tended to be lower (allowing less respondents to be captured above 75; they just snuck under the 75-point level).

An example is helpful to illustrate looking at the VRIN-r hybrid scale percentage above cutoff for the first addition procedure in Table 10, at the 100% random level of insertion for item addition level three and four there was a decrease in percentage above cutoff from 61.61% to 52.61%. The T score just below the cutoff at the three-item level was 70.37 and the next T score level was 75.17, barely above the 75-point cutoff. When another item was added (the four-item level), the T score just below the cutoff was 74.15, just barely under the 75-point cutoff,

Table 9.

Results of t-tests of Mean T scores Between the CRIN/F-r scales by Addition Procedure and Percentage of Simulated Randomness in the Normative Sample.

	Scale						95% CI for Mean Difference			Cohen's *d*
	CRIN/F-r, First Addition Procedure			CRIN/F-r, Second Addition Procedure						
	M	*SD*	*n*	*M*	*SD*	*n*		*t*	*df*	
25% Random	59.60	0.27	15	58.87	0.20	15	0.56, 0.91	8.57***	28	3.07
50% Random	67.55	0.70	15	66.18	0.62	15	0.87, 1.86	5.66***	28	2.07
75% Random	73.57	1.23	15	72.22	1.16	15	0.46, 2.25	3.12**	28	1.13
100% Random	77.96	2.06	15	75.72	1.95	15	0.75, 3.74	3.07**	28	1.12

** $p < 0.01$, *** $p < 0.00$

Table 10.

Percentage of VRIN-r Scores Above Cutoff in MMPI-A-RF Normative Sample for First Item Addition Procedure.

Percent Simulated Randomness	Number of Fr-r Items Added to VRIN-r Scale															
	0	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
25%	**6.00**	**6.52**	**9.19**	**9.75**	**5.40**	**7.20**	**8.20**	**5.22**	**6.58**	**8.82**	**5.90**	**7.20**	**8.32**	**5.78**	**7.64**	**9.38**
	96	*105*	*148*	*157*	*87*	*116*	*132*	*84*	*106*	*142*	*95*	*116*	*134*	*93*	*123*	*151*
50%	**15.78**	**18.51**	**24.41**	**28.45**	**18.82**	**23.54**	**28.20**	**20.75**	**23.60**	**29.38**	**21.43**	**25.90**	**28.14**	**22.67**	**24.97**	**29.13**
	254	*298*	*393*	*458*	*303*	*379*	*454*	*334*	*380*	*473*	*345*	*417*	*453*	*365*	*402*	*469*
75%	**24.10**	**34.53**	**43.79**	**50.37**	**35.84**	**45.53**	**50.50**	**42.24**	**50.00**	**53.98**	**44.29**	**51.18**	**56.71**	**49.19**	**56.15**	**60.99**
	388	*556*	*705*	*811*	*577*	*733*	*813*	*680*	*805*	*869*	*713*	*824*	*913*	*792*	*904*	*982*
100%	**32.67**	**44.04**	**54.78**	**61.61**	**52.61**	**62.17**	**68.20**	**59.88**	**68.57**	**74.47**	**67.52**	**73.98**	**78.70**	**73.48**	**77.89**	**82.80**
	526	*709*	*882*	*992*	*847*	*1001*	*1098*	*964*	*1104*	*1199*	*1087*	*1191*	*1267*	*1183*	*1254*	*1333*

Percentage in bold, *n* in italics. Total sample size 1610 for each cell.

Table 11.

Percentage of CRIN Scores Above Cutoff in MMPI-A-RF Normative Sample for First Item Addition Procedure.

Percent Simulated Randomness	Number of Fr-r Items Added to CRIN Scale															
	0	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
25%	**7.00**	**7.14**	**9.32**	**5.59**	**5.65**	**7.70**	**7.89**	**6.27**	**6.34**	**5.28**	**6.02**	**7.39**	**5.09**	**5.71**	**7.70**	**6.21**
	113	*115*	*150*	*90*	*91*	*124*	*127*	*101*	*102*	*85*	*97*	*119*	*82*	*92*	*124*	*100*
50%	**17.39**	**20.19**	**24.29**	**16.77**	**20.00**	**22.86**	**25.90**	**21.24**	**22.73**	**18.07**	**22.05**	**24.35**	**17.64**	**21.30**	**23.91**	**18.45**
	280	*325*	*391*	*270*	*322*	*368*	*417*	*342*	*366*	*291*	*355*	*392*	*284*	*343*	*385*	*297*
75%	**29.32**	**36.89**	**43.60**	**35.96**	**39.88**	**45.90**	**51.30**	**41.80**	**48.07**	**40.43**	**46.40**	**50.50**	**41.61**	**49.19**	**54.78**	**48.45**
	472	*594*	*702*	*579*	*642*	*739*	*826*	*673*	*774*	*651*	*747*	*813*	*670*	*792*	*882*	*780*
100%	**39.75**	**48.26**	**57.52**	**48.07**	**55.84**	**62.42**	**67.70**	**60.06**	**69.01**	**60.50**	**67.76**	**72.61**	**64.72**	**73.23**	**76.34**	**70.37**
	640	*777*	*926*	*774*	*899*	*1005*	*1090*	*967*	*1111*	*974*	*1091*	*1169*	*1042*	*1179*	*1229*	*1133*

Percentage in bold, *n* in italics. Total sample size 1610 for each cell.

Table 12.

Percentage of VRIN-r Scores Above Cutoff in MMPI-A-RF Normative Sample for Second Item Addition Procedure.

Percent Simulated Randomness	Number of Fr-r Items Added to VRIN-r Scale															
	0	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
25%	**6.00**	**7.89**	**5.53**	**6.71**	**8.51**	**5.09**	**6.21**	**7.45**	**5.53**	**6.34**	**4.72**	**5.28**	**6.58**	**7.89**	**6.02**	**6.96**
	96	*127*	*89*	*108*	*137*	*82*	*100*	*120*	*89*	*102*	*76*	*85*	*106*	*127*	*97*	*112*
50%	**15.78**	**20.81**	**12.92**	**16.71**	**21.18**	**13.66**	**16.34**	**20.99**	**14.35**	**17.64**	**12.11**	**14.97**	**18.63**	**22.98**	**16.71**	**20.50**
	254	*335*	*208*	*269*	*341*	*220*	*263*	*338*	*231*	*284*	*195*	*241*	*300*	*370*	*269*	*330*
75%	**24.10**	**36.46**	**28.01**	**34.22**	**41.30**	**32.61**	**38.32**	**44.78**	**36.46**	**41.74**	**34.35**	**39.07**	**45.71**	**51.06**	**42.92**	**48.45**
	388	*587*	*451*	*551*	*665*	*525*	*617*	*721*	*587*	*672*	*553*	*629*	*736*	*822*	*691*	*780*
100%	**32.67**	**43.66**	**34.84**	**43.66**	**51.55**	**43.73**	**51.55**	**59.69**	**51.43**	**58.76**	**51.74**	**58.32**	**65.16**	**71.43**	**64.72**	**70.99**
	526	*703*	*561*	*703*	*830*	*704*	*830*	*961*	*828*	*946*	*833*	*939*	*1049*	*1150*	*1042*	*1143*

Percentage in bold, *n* in italics. Total sample size 1610 for each cell.

Table 13.

Percentage of CRIN Scores Above Cutoff in MMPI-A-RF Normative Sample for Second Item Addition Procedure.

Percent Simulated Randomness	Number of Fr-r Items Added to CRIN Scale															
	0	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
25%	**7.00**	**4.91**	**6.02**	**7.27**	**5.34**	**6.15**	**7.02**	**5.40**	**6.83**	**7.52**	**5.65**	**6.27**	**7.14**	**5.47**	**6.34**	**7.52**
	113	*79*	*97*	*117*	*86*	*99*	*113*	*87*	*110*	*121*	*91*	*101*	*115*	*88*	*102*	*121*
50%	**17.39**	**11.12**	**13.79**	**16.96**	**12.36**	**14.53**	**17.39**	**12.10**	**15.78**	**18.26**	**13.79**	**16.02**	**19.13**	**14.91**	**17.14**	**20.62**
	280	*179*	*222*	*273*	*199*	*234*	*280*	*195*	*254*	*294*	*222*	*258*	*308*	*240*	*276*	*332*
75%	**29.32**	**27.33**	**33.35**	**38.01**	**31.37**	**36.21**	**41.24**	**34.10**	**40.06**	**43.98**	**37.83**	**41.74**	**46.77**	**40.19**	**44.84**	**49.07**
	472	*440*	*537*	*612*	*505*	*583*	*664*	*549*	*645*	*708*	*609*	*672*	*753*	*647*	*722*	*790*
100%	**39.75**	**34.41**	**40.68**	**48.76**	**41.24**	**48.26**	**55.78**	**49.01**	**55.03**	**61.30**	**54.66**	**60.99**	**66.40**	**59.63**	**64.78**	**69.88**
	640	*554*	*655*	*785*	*664*	*777*	*898*	*789*	*886*	*987*	*880*	*982*	*1069*	*960*	*1043*	*1125*

Percentage in bold, *n* in italics. Total sample size 1610 for each cell.

and the next highest T score was 78.87. As each item is added this pattern repeated: The T score just below the cutoff starts just under 75 (producing lower percentages) and as more items are added this slowly drops to approach 70 (producing higher percentages) before it resets back near the 75-point cutoff. This is what produced the oscillating, sinusoidal trend in the data, the fluctuating *T* score just below the cutoff is directly linked to how many respondents fall above and below the cutoff.

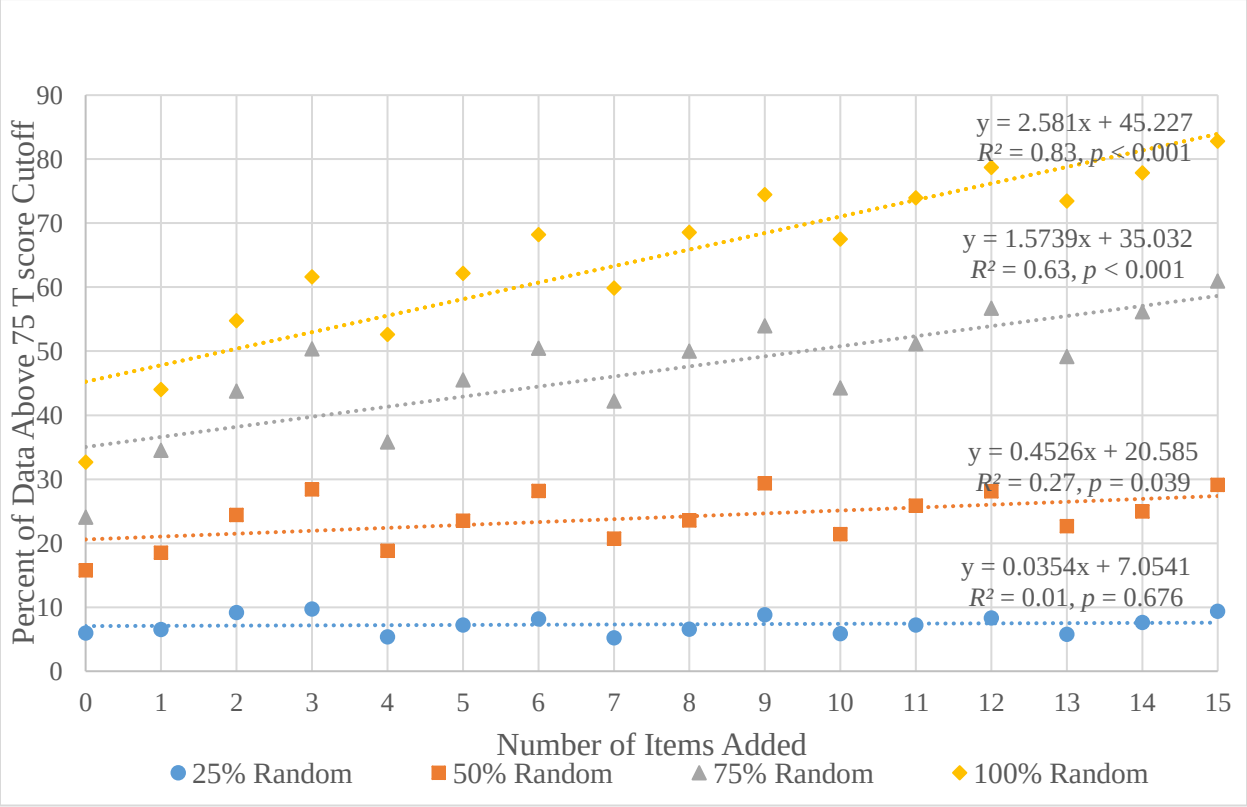
Percentage above cutoff VRIN-r/F-r and CRIN/F-r Scales, overall trend as items are added: First addition procedure. The sinusoidal pattern of the data was discussed above, but other trends interact with this sinusoidal trend as item levels are increased. Most of the sinusoidal patterns appear to be increasing as items are added, so a linear trend line was fit to the percentage above cutoff data for each of the hybrid scales at each level of randomness to help better identify this overall rising pattern. Just like what was found with Mean T scores, there is a difference in trends between the 25 and 50% randomness levels and the 75 and 100% randomness levels.

Looking at the 25% randomness level for the VRIN-r hybrid scale, the linear equation y = 0.0354x + 7.0541 explained the overall linear trend; however, it was nonsignificant ($p = 0.676$) and fit the data poorly ($R^2 = .01$; see Figure 5). At 25% randomness for the CRIN hybrid scale, the linear equation y = -0.0774x + 7.2254 best fit the trend, but again was non-significant ($p = 0.227$) and was a poor fit ($R^2 = 0.10$; see Figure 6). At 50% randomness within the CRIN hybrid scale, the linear equation y = 0.0643x + 20.589 was fit to the data, but was again non-significant ($p = 0.691$), and was even a worse fit than what was found at 25% randomness ($R^2 = 0.10$; see Figure 6). On the other hand, at 50% randomness for the VRIN-r hybrid scale the linear equation y = 0.4526x + 20.585 was significant ($p = 0.039$) and was an okay fit ($R^2 = 0.27$; see Figure 5).

Taken together these data suggest that the VRIN-r and CRIN do not seem to reliably detect more respondents over the cutoff as items are added for the 25% and 50% randomness levels. Only at the 50% randomness level for the VRIN-r hybrid scale was there a significant linear trend, but the slope for this equation was modest, and indeed, the greatest change from baseline cutoff detection was a small, but respectable, 15% throughout the entire item addition procedure.

Figure 5.

Percent of scores above cutoff for VRIN-r/F-r scale in MMPI-A-RF normative sample for first item addition procedure.



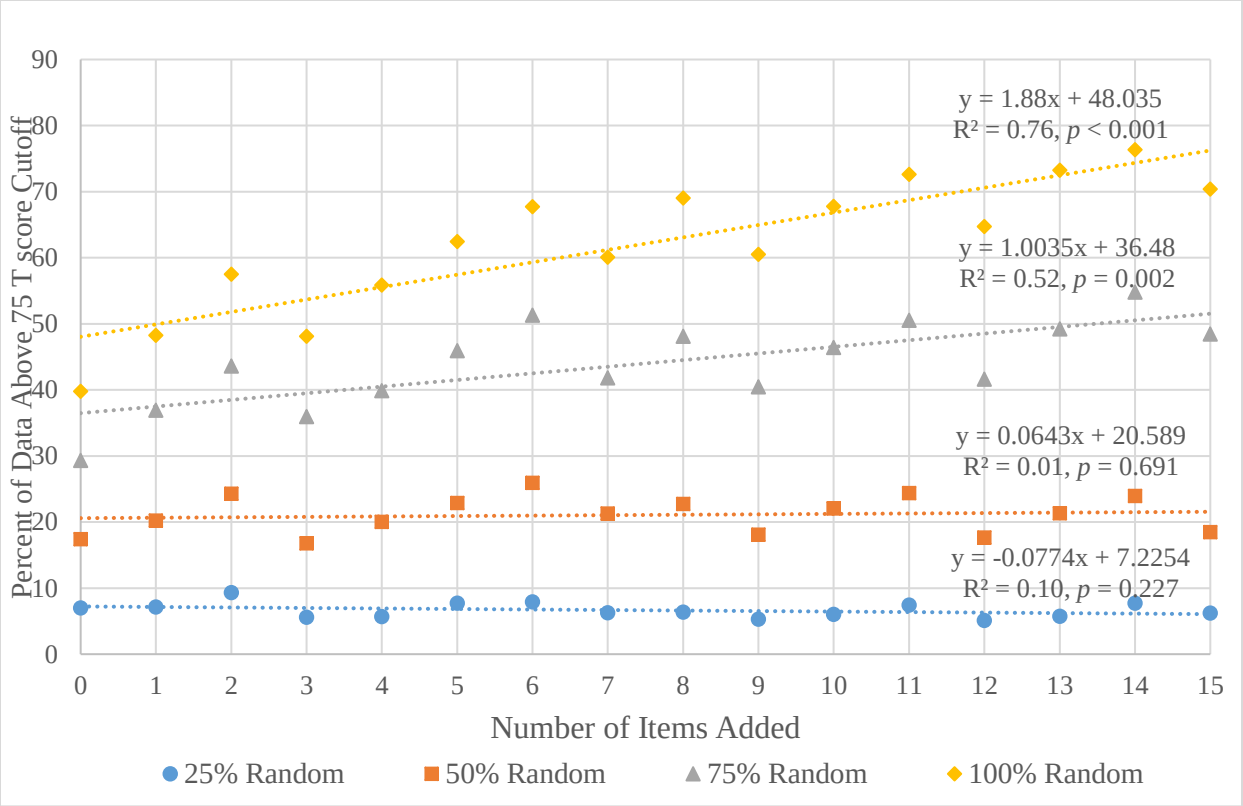
On the other hand, looking at the 75% and 100% randomness level for both hybrid scales there is a clearer upward trend to the sinusoidal pattern as item level increases. At randomness levels 75% and 100% for the VRIN-r/F-r scale linear trend equations of y = 1.5739x + 35.032 and y = 2.581x + 45.227 were produced (respectively) which were both significant ($p < 0.001$) and excellent fits for the data ($R^2 = 0.63$ and 0.83, correspondingly; see Figure 5). Likewise, at the 75% and 100% randomness levels for the CRIN/F-r scale, the corresponding linear trend equations of y = 1.0035x + 36.48 and y = 1.88x + 48.035 (correspondingly) were both significant ($p = 0.002$ and < 0.001, respectively) and an excellent fit to the data ($R^2 = 0.52$ and 0.76, correspondingly; see Figure 6). These data suggest that at 75% randomness and above, both hybrid scales are generally much more sensitive to detecting responding over the *T* score cutoff as items are added, especially when comparing these scales to their own respective performance as items are added at lower levels of randomness (i.e., 50% and below).

Overall, these trends of percentage above the cut-off by item addition matched the trends of mean T scores by randomness level for the VRIN-r and CRIN hybrid scales. That is, there is not much change at randomness levels at and below 50%; however, there is a much stronger positive trend at randomness levels at and above 75%. This is logical because percentage above cutoff is directly related to the mean T score: as the mean T score for the VRIN-r and CRIN-r hybrid scales rises with each additional item, the percentage of scores above the 75-point cutoff mathematically must rise, as well.

Percentage above cutoff VRIN-r/F-r and CRIN/F-r Scales, overall trend as items are added: Second addition procedure. Like the first addition procedure, percentage above cut-off remained relatively stable at the 25 and 50% random insertion level for both the CRIN and VRIN-r hybrid scales for the VRIN-r hybrid scale at 25% and 50% randomness linear

Figure 6.

Percent of scores above cutoff for CRIN/F-r scale in MMPI-A-RF normative sample for first item addition procedure.



trend equations of y = -0.0163x + 6.5412 and y = 0.1346x + 16.257 were produced, respectively, both were non-significant (p = 0.795 and 0.470, correspondingly) and were poor fits to the data (R^2 = 0.01 and 0.04, correspondingly; see Figure 7). Analogously, the CRIN hybrid scale linear trend equations of y = 0.0382x + 6.0786 and y = 0.2748x + 13.645 for the 25% and 50% randomness levels (respectively) were non-significant (p = 0.423 and 0.053, correspondingly). The initial linear trend for the 25% level was a poor fit (R^2 = 0.05); however, the second linear equation, despite being non-significant, was a reasonable fit for the data (R^2 = 0.24; see Figure

8). Again, like what was found for the first addition procedure, both hybrid scales exhibited negligible to quite modest increased sensitivity to detecting percentages of respondents above the cutoff at profiles saturated with 50% randomness or below.

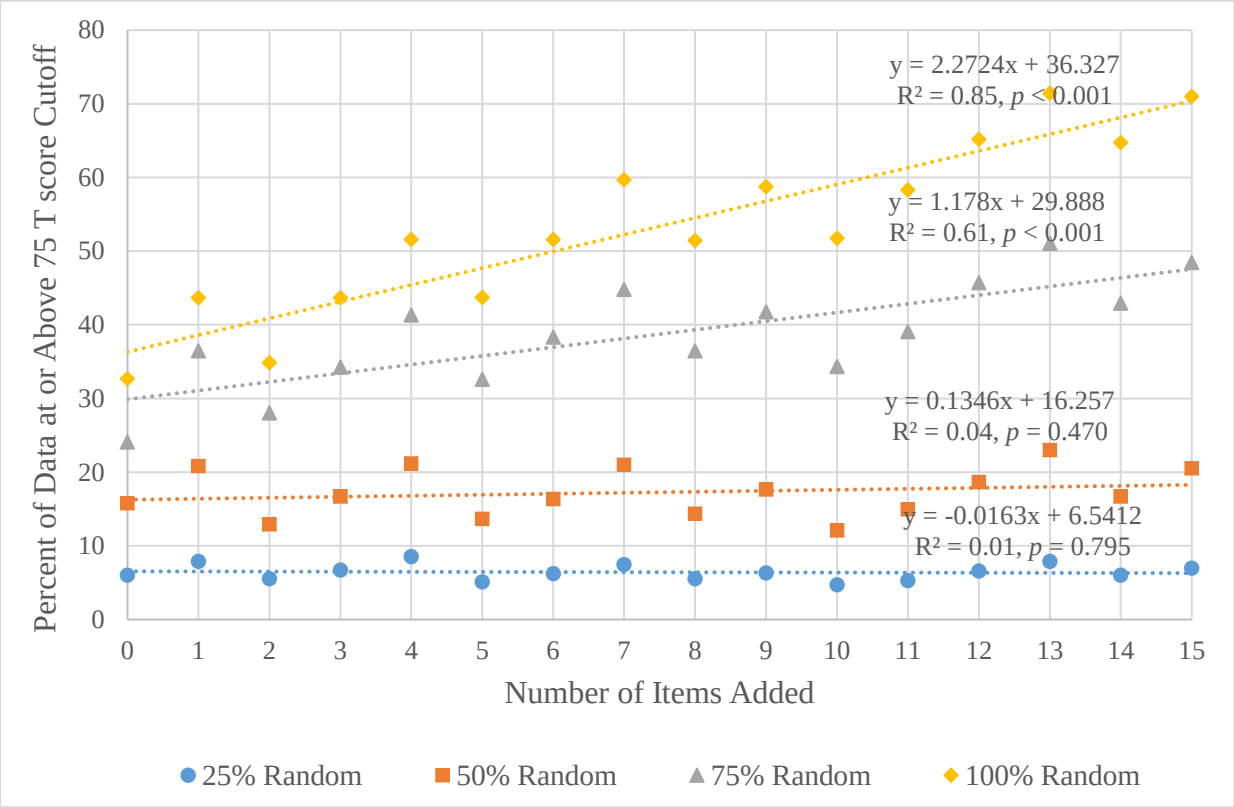
A more pronounced upward trend in percentage above cut-off scores is noticeable at the 75 and 100% level of random insertion for both VRIN-r and CRIN hybrid scales as items are added. At the 75% and 100% randomness level for the VRIN-r hybrid scale, linear trend equations of y = 1.178x + 29.888 and y = 2.2724x + 36.327 were fit to the data, correspondingly, and both were found significant ($p < 0.001$) and an excellent fit to the data ($R^2 = 0.61$ and 0.85, correspondingly; see Figure 7). Correspondingly, these same randomness levels within the CRIN/F-r scale were fit with the equations y = 1.1365x + 29.939 and y = 2.0687x + 37.645, respectively. Both were significant ($p < 0.001$) and great fits for the data ($R^2 = 0.75$ and 0.88, correspondingly; see Figure 8). Similar to what was found in the first item addition procedure, these equations suggest that a stronger linear trend is produced in percentage above cutoff scores as items are added for both the VRIN-r and CRIN hybrid scales at the 75% and above randomness levels, especially when compared to the trends at and below 50%.

Comparing the VRIN-r/F-r and CRIN/F-r Scales for percentage above cutoff.

Within the first addition procedure. Comparing the VRIN-r and CRIN hybrid scales percentage above cutoff values for both addition procedures were somewhat complicated due to the sinusoidal trends associated with both scales as items were added. The percentage above cutoff values for the VRIN-r and CRIN hybrid scales can be observed in tables 10 and 11, correspondingly. For the first addition procedure, the VRIN-r/F-r scale began to consistently have higher percentage values after the eleventh added item for each level of randomness.

Figure 7.

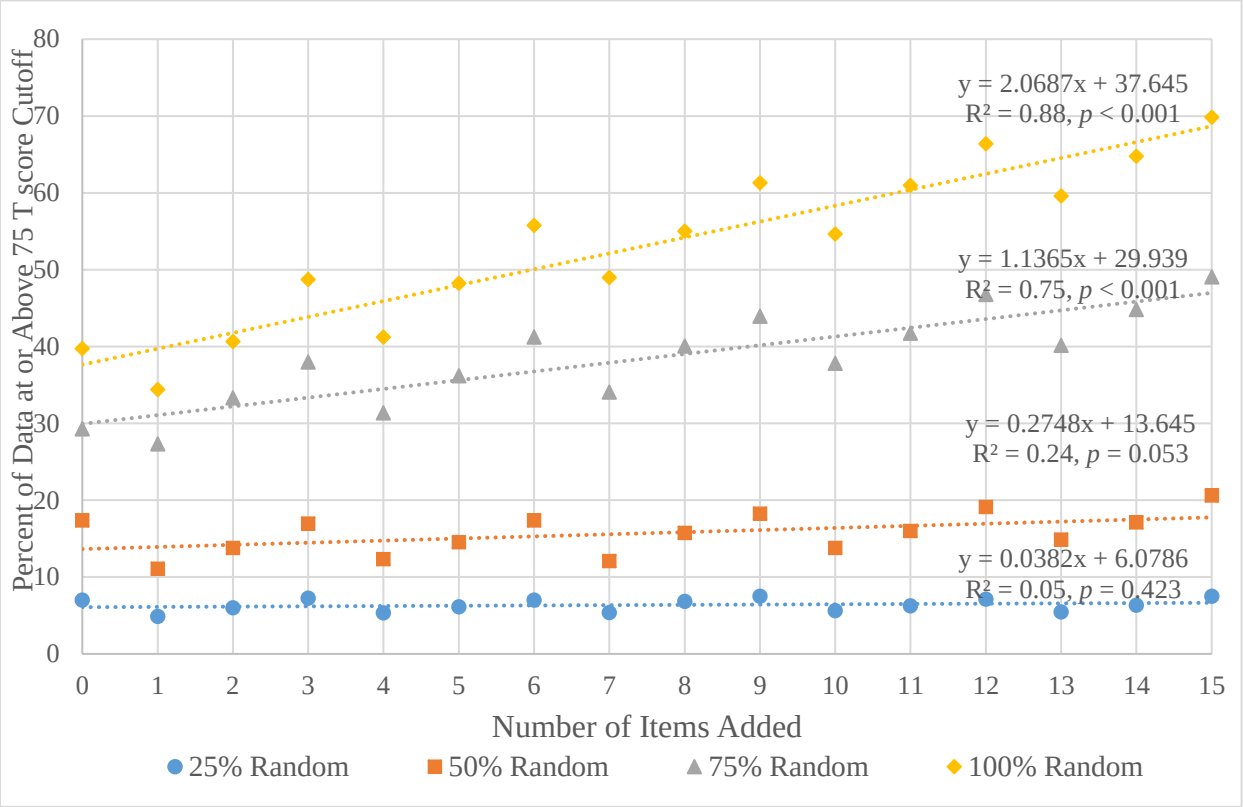
Percent of scores above cutoff for VRIN-r/F-r scale in MMPI-A-RF normative sample for second item addition procedure.



Specifically looking at the item addition levels of 11-15, 17/20 cells of the VRIN-r hybrid percentages above cutoff were higher than their corresponding CRIN hybrid percentages. There were only two occasions when the CRIN hybrid scale exhibited higher percentage above-cutoff than its corresponding VRIN-r hybrid scale value, both were confined to the 25% level of randomness (there was one occasion when the VRIN-r and CRIN hybrid scales had exactly the same percentage value). Below the eleven-item addition level, while there were clearly differences in percentage scores between the VRIN-r/F-r scale and CRIN/F-r scale at

Figure 8.

Percent of scores above cutoff for CRIN/F-r scale in MMPI-A-RF normative sample for second item addition procedure.



individual item levels, there did not appear to be a consistent trend of one scale producing consistently higher (or lower) percentages as items were added. Specifically looking at the 40 cells within item levels one through 10, the VRIN-r hybrid scale produced higher percentages in 18/40 cells when compared to the corresponding CRIN hybrid scale cell at the same item level and percentage of randomness. The CRIN hybrid scale produced higher percentages in the remaining 22/40 cells when compared to their corresponding cells in the VRIN-r hybrid scale. This suggested the CRIN hybrid scale did produce slightly more data points between item

addition level one and ten where percentage above the cutoff was higher than the corresponding VRIN hybrid scale value; however, this difference was quite small. What seemed to happen is that the VRIN-r and CRIN hybrid scales traded sinusoidal peaks in percentage above cutoff values as items were added. This explains why a clear trend is difficult to discern, and why looking within the 40 cells between item addition level one and ten yielded a near 50% split between either the VRIN-r/F-r and CRIN/F-r owning the highest percentage above cutoff values when compared to its counterpart value.

The *t*-test comparisons of the percentage above cutoff scores between the VRIN-r and CRIN hybrid scales within the first addition procedure by each level of randomness can be observed in Table 14. These data suggested that at the 50% randomness level, within the first addition procedure, the VRIN-r hybrid scale ($M = 24.53$, $SD = 3.65$) produced a statistically greater mean percentage above cutoff score than the CRIN hybrid scale ($M = 21.32$, $SD = 2.76$); $t(28) = 2.72$, $p = 0.011$. The effect size of this difference was large ($d = 0.99$). On the other hand, although the VRIN-r hybrid scale's percent above cutoff mean was consistently larger than the CRIN hybrid scale's mean at the three other randomness levels, these differences were not statistically significant. These data indicated that, although the VRIN-r/F-r scale produces consistently greater observed percentage above cutoff scores than the CRIN/F-r scales for the first item addition procedure, especially when item addition numbers are greater than 10, only at the 50% randomness level were these differences reliably different from 0. Overall, this suggests that for most of the data the VRIN-r and CRIN hybrid scales were equally sensitive to detecting random responding within the first addition procedure.

Table 14.

Results of t-tests of Mean Percentage Above Cutoff Scores Between the VRIN-r/F-r and CRIN/F-r Scales on the First Addition Procedure by Percentage of Simulated Randomness in the Normative Sample.

	Scale						95% CI for Mean Difference			Cohen's *d*
	VRIN-r/F-r			CRIN/F-r						
	M	*SD*	*n*	*M*	*SD*	*n*		*t*	*df*	
25% Random	7.41	1.50	15	6.62	1.19	15	-0.22, 1.80	1.59	28	0.58
50% Random	24.53	3.65	15	21.32	2.76	15	0.79, 5.63	2.72*	28	0.99
75% Random	48.35	7.43	15	44.98	5.51	15	-1.52, 8.26	1.41	28	0.52
100% Random	66.71	10.83	15	63.63	8.64	15	-4.24, 10.43	0.86	28	0.36

* $p < 0.05$

Within the second addition procedure. For the second addition procedure, a different trend emerged when analyzing the difference between percentage above cutoff scores on the CRIN and VRIN-r hybrid scales (seen in Tables 13 and 12, respectively). Here it appeared that the CRIN/F-r scale consistently outperformed (had higher percentage above cutoff values) the VRIN-r/F-r scale. Looking at the 60 cells of data between item addition level one and fifteen, the CRIN-r hybrid scale captured the higher percentage cutoff value in 43 of the cells. Indeed, the VRIN-r/F-r scale only had higher percentage above cutoff values when compared to the corresponding CRIN/F-r values, at item levels one, four, seven, thirteen, and at the 100% randomness level of fifteen. Here again, the sinusoidal trend of the data is the key, as at these item levels were when the VRIN-r/F-r scale obtained a local acme. The CRIN/F-r scales local acmes were out of phase with the VRIN-r/F-r scale trend, thus producing the relatively rarer occurrence of the VRIN-r/F-r scale capturing the highest percentage above cutoff value.

The t-test statistical comparisons between mean T scores of the VRIN-r and CRIN hybrid scales for the second addition procedure can be found in Table 15. These data point to a different picture then discussed in the previous paragraph. When collapsing across item addition level and looking within the percentage of randomness levels for the second addition procedure, the mean percent above cutoff score for the VRIN-r hybrid scale was always greater than the CRIN hybrid scale mean score at all four levels of randomness. This suggests that, although the CRIN hybrid scale had the greater percent above cutoff score in most of the corresponding data cells, when the VRIN-r hybrid scale had the greater percent score it was to a much larger degree. However, *t*-statistics indicated that these observed differences between the mean percent above cutoff scores for the VRIN-r and CRIN hybrid scales never reached statistical significance. This provides

evidence that, statistically, the VRIN-r and CRIN hybrid scales were equally sensitive to detecting random responding when created by the second item addition procedure.

Comparing addition procedures for percentage above cutoff.

Within the VRIN-r/F-r Scale. Analyzing the percentage above cutoff values by addition procedure yielded results that strongly suggest that the first addition procedure was more likely to produce higher percentage above cutoff values, and thus higher sensitivity to random responding. Specifically when looking at the data cells between item addition level one and fifteen by all random insertion levels, the VRIN-r/F-r scale (Table 10) for the first procedure produced a higher percentage above the cutoff in 46/60 cells when compared to the corresponding values in the VRIN-r hybrid scale for the second addition procedure (Table 12). The VRIN hybrid scale for the second addition procedure captured the higher score in only 14/60 cells with these incidents occurring at item addition level one, four, seven, and thirteen.

The *t*-tests comparing the mean percent above cutoff scores for the first and second addition procedures within the VRIN-r/F-r scale can be found in Table 16. The VRIN-r/F-r mean percent above cutoff collapsing across item addition levels for the first addition procedure was consistently greater than the means for the second addition procedure at all four levels of simulated randomness, and they reached statistical significance at the 50%, 75%, and 100% randomness levels. Specifically, the VRIN-r hybrid scale created by the first addition procedure at the 50% (M = 24.53, SD = 3.65), 75% (M = 48.35, SD = 7.43), and 100% (M = 66.71, SD = 10.83) randomness levels produced significant *t*-statistics when compared with the VRIN-r hybrid scale created by the second addition procedure (M = 17.37, SD = 3.39; M = 39.07, SD = 6.29; M = 54.75, SD = 10.71, correspondingly) of $t(28) = 5.57$ ($p < 0.001$), 3.45 ($p = 0.002$), and 3.04 ($p = 0.005$), respectively. Effect sizes for these three significant *t*-statistics were large (d =

Table 15.

Results of t-tests of Mean Percentage Above Cutoff Scores Between the VRIN-r/F-r and CRIN/F-r Scales on the Second Addition Procedure by Percentage of Simulated Randomness in the Normative Sample.

	Scale						95% CI for Mean Difference			Cohen's *d*
	VRIN-r/F-r			CRIN/F-r						
	M	*SD*	*n*	*M*	*SD*	*n*		*t*	*df*	
25% Random	6.45	1.13	15	6.32	0.86	15	-0.63, 0.87	0.34	28	0.13
50% Random	17.37	3.39	15	15.59	2.72	15	-0.52, 4.07	1.58	28	0.58
75% Random	39.70	6.29	15	39.07	5.95	15	-3.95, 5.20	0.28	28	0.10
100% Random	54.75	10.71	15	54.05	10.25	15	-7.15, 8.54	0.18	28	0.07

2.03, 1.26, and 1.11, respectively). These data indicated that, in general, just like what was found in the mean *T* score contrasts, the first addition procedure seemed to be more sensitive to detecting random responders over the *T* score cutoff compared to the second addition procedure within the VRIN-r hybrid scale.

Within the CRIN/F-r Scale. Very similar results were found when comparing the CRIN hybrid scales by addition procedure. The CRIN hybrid scale for the first addition procedure captured the higher percentage above cutoff value in 46/60 cells between item addition levels one through fifteen. The CRIN/F-r scale for the second addition procedure captured the higher value in 14/60 cells, which only occurred at item level three, eight, nine, and fourteen. This provides tentative evidence that, just like what was found with the VRIN-r/F-r scale, the CRIN/F-r scale created by the first addition procedure performed better at detecting random responders above the T score cutoff than the CRIN/F-r scale produced by the second addition procedure.

The *t*-test findings bore this out and were also very similar to VRIN hybrid analyses. These statistics can be observed in Table 17. Again, the CRIN/F-r scale created by the first addition procedure produced a greater mean percent above cutoff score at every level of simulated randomness when compared to the CRIN/F-r scale made by the second addition procedure. Just like the VRIN-r/F-r analyses, these differences reached statistical significance at the 50%, 75%, and 100% randomness level, but not at the 25% randomness level. The CRIN/F-r scale mean percent above cutoff score for the first addition procedure at the 50% (M = 21.32, SD = 2.76), 75% (M = 44.98, SD = 5.51), and 100% (M = 63.63, SD = 8.64) randomness insertion level, when contrasted with the respective statistics for the CRIN/F-r scale produced by the second addition procedure (M = 15.59, SD = 2.72; M = 39.07, SD = 5.95; M = 54.05, SD = 10.25,

respectively) produced specific significant *t*-statistics of $t(28) = 5.72$ ($p < 0.001$), 2.83 ($p = 0.009$, and 2.77 ($p = 0.01$), correspondingly. Cohen's *d* statistics of 2.09, 1.03, and 1.01 for each respective contrast indicated large effect sizes for these differences. Overall, just like what was found for the VRIN-r/F-r scale, statistical analyses suggest that the first addition procedure creates a CRIN/F-r scale that is more sensitive to detecting random responders over the T score cutoff, especially at randomness levels of 50% and above.

Mean *T* scores by Simulated Randomness in the Faking Bad Sample

Mean *T* score trends as items are added. Mean T scores for the VRIN-r and CRIN hybrid scales for both addition procedures in the faking bad sample can be seen in Tables 18 and 19 and Figures 9 and 10. A linear trend appeared to be the best fit for these data, and indeed analyses supported this. For the first and second addition procedure within the VRIN-r hybrid linear equations of $y = 1.6499x + 50.644$ and $y = 2.0385x + 47.364$ were obtained with both being significant ($p < 0.001$) and near perfect fits for the data ($R^2 = 0.94$ and 0.96, correspondingly; see Figure 9). Similar results were found for the first and second addition procedures within the CRIN hybrid scale (see Figure 10): the respective equations of $y = 1.3452x + 49.124$ and $y = 1.6813x + 46.865$ were produced, which were both significant ($p < 0.001$) and excellent fits ($R^2 = 0.96$ and 0.94, correspondingly; see Figure 10). These data suggest that the Mean *T* scores for VRIN-r and CRIN hybrid scales reliably increased as items were added within the faking bad sample, which provided evidence that these hybrid scales are sensitive to overreporting and necessitate precautions for maintaining specificity from this style of reporting.

Before results of the analyses comparing mean *T* score and percent above cutoff scores within the faking bad sample are reported below, the reader is reminded that, unlike the previous

Table 16.

Results of t-tests of Mean Percentage Above Cutoff Scores Between the VRIN-r/F-r Scales by Addition Procedure and Percentage of Simulated Randomness in the Normative Sample.

	Scale						95% CI for Mean Difference			Cohen's *d*
	VRIN-r/F-r First Addition Procedure			VRIN-r/F-r Second Addition Procedure						
	M	*SD*	*n*	*M*	*SD*	*n*		*t*	*df*	
25% Random	7.41	1.50	15	6.45	1.13	15	-0.03, 1.95	1.98	28	0.72
50% Random	24.53	3.65	15	17.37	3.39	15	4.53, 9.79	5.57***	28	2.03
75% Random	48.35	7.43	15	39.70	6.29	15	3.51, 13.80	3.45**	28	1.26
100% Random	66.71	10.83	15	54.75	10.71	15	3.91, 20.02	3.04**	28	1.11

** $p < 0.01$, *** $p < 0.00$

Table 17.

Results of t-tests of Mean Percentage Above Cutoff Scores Between the CRIN/F-r Scales by Addition Procedure and Percentage of Simulated Randomness in the Normative Sample.

	Scale						95% CI for Mean Difference			Cohen's *d*
	CRIN/F-r First Addition Procedure			CRIN/F-r Second Addition Procedure						
	M	*SD*	*n*	*M*	*SD*	*n*		*t*	*df*	
25% Random	6.62	1.19	15	6.32	0.86	15	-0.48, 1.07	0.78	28	0.29
50% Random	21.32	2.76	15	15.59	2.72	15	3.67, 7.77	5.72***	28	2.09
75% Random	44.98	5.51	15	39.07	5.95	15	1.62, 10.20	2.82**	28	1.03
100% Random	63.63	8.64	15	54.05	10.25	15	2.49, 16.66	2.77**	28	1.01

** $p < 0.01$, *** $p < 0.001$

analyses within the normative sample, *lower* mean *T* scores and percent above cutoff percentages are preferred in these analyses. In the normative sample higher scores indicated greater sensitivity to random responding, whereas in the faking bad sample greater scores indicated problematic sensitivity to overreporting. Lower values in the faking bad analyses suggest that the VRIN-r and CRIN hybrid scales are more specific in their detection of overreporting.

Comparing the VRIN-r/F-r and CRIN/F-r Scales for mean *T* score. The VRIN-r/F-r and CRIN/F-r mean T scores for the first addition procedure in the faking bad sample can be observed in Tables 18 and 19, respectively. At every level of item addition, the VRIN-r hybrid scale produced a higher mean *T* score when compared to the corresponding CRIN hybrid mean *T* score. This tentatively suggests that the CRIN hybrid scale is more specific in detection of overreporting than the VRIN hybrid scale within the first addition procedure. The *t*-statistic for this contrast, however, was not significant (see Table 20), which indicated, that although the CRIN hybrid scale's observed mean *T* scores suggested it was more specific in its detection of overreporting than the VRIN hybrid scale, these differences were not statistically different from zero. The VRIN/F-r and CRIN/F-r mean T scores for the second addition procedure in the faking bad sample can also be seen in Tables 18 and 19, respectively. Except for the first item addition level, at every other item addition level the VRIN-r/F-r scale had a higher mean *T* score than the corresponding CRIN/F-r mean. This again, tentatively suggested that the CRIN/F-r scale for the second addition procedure was more specific from detecting overreporting than the second addition produced VRIN-r hybrid scale. However, the *t*-statistic for this contrast (see Table 20), was again not significant, suggesting these observed differences were not statistically different from naught.

Table 18.

Mean T scores *for Combined VRIN-r/F-r Scale in Faking Bad Sample.*

Item Addition Configuration Type	Number of Fr-r Items Added to VRIN-r Scale														
	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
First	**48.94**	**51.75**	**54.71**	**57.47**	**61.91**	**62.00**	**63.95**	**65.65**	**66.96**	**68.64**	**69.06**	**70.18**	**71.39**	**72.14**	**72.89**
	9.02	*8.80*	*9.15*	*9.85*	*10.90*	*10.83*	*11.46*	*12.17*	*12.77*	*13.38*	*13.64*	*14.44*	*15.35*	*15.73*	*16.33*
Second	**45.01**	**51.58**	**54.36**	**56.78**	**55.51**	**61.65**	**63.41**	**65.10**	**66.92**	**68.49**	**70.50**	**72.31**	**73.49**	**74.46**	**75.52**
	5.61	*8.30*	*8.18*	*8.71*	*8.93*	*10.11*	*10.74*	*11.25*	*12.00*	*12.49*	*13.44*	*14.18*	*14.84*	*15.27*	*15.77*

M in bold and *SD* in italics. Total sample size 138 for each cell.

Table 19.

Mean T scores *for Combined F-r/CRIN Scale in Faking Bad Sample.*

Item Addition Configuration Type	Number of Fr-r Items Added to CRIN Scale														
	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
First	**48.34**	**50.36**	**52.50**	**54.54**	**58.11**	**58.07**	**59.63**	**61.04**	**62.17**	**63.61**	**64.07**	**65.08**	**66.15**	**66.92**	**67.69**
	7.93	*9.88*	*9.97*	*10.22*	*10.52*	*10.47*	*10.71*	*11.04*	*11.34*	*11.73*	*11.88*	*12.44*	*13.11*	*13.40*	*13.88*
Second	**45.38**	**50.30**	**52.40**	**54.25**	**53.36**	**58.03**	**59.51**	**60.95**	**66.92**	**63.84**	**65.49**	**67.01**	**68.13**	**69.07**	**70.10**
	7.14	*9.56*	*9.35*	*9.54*	*9.39*	*10.11*	*10.39*	*10.61*	*11.08*	*11.36*	*11.99*	*12.52*	*13.04*	*13.36*	*13.77*

M in bold and *SD* in italic. Total sample size 138 for each cell.

Figure 9.

Mean T scores for the VRIN-r/F-r scale in faking bad sample.

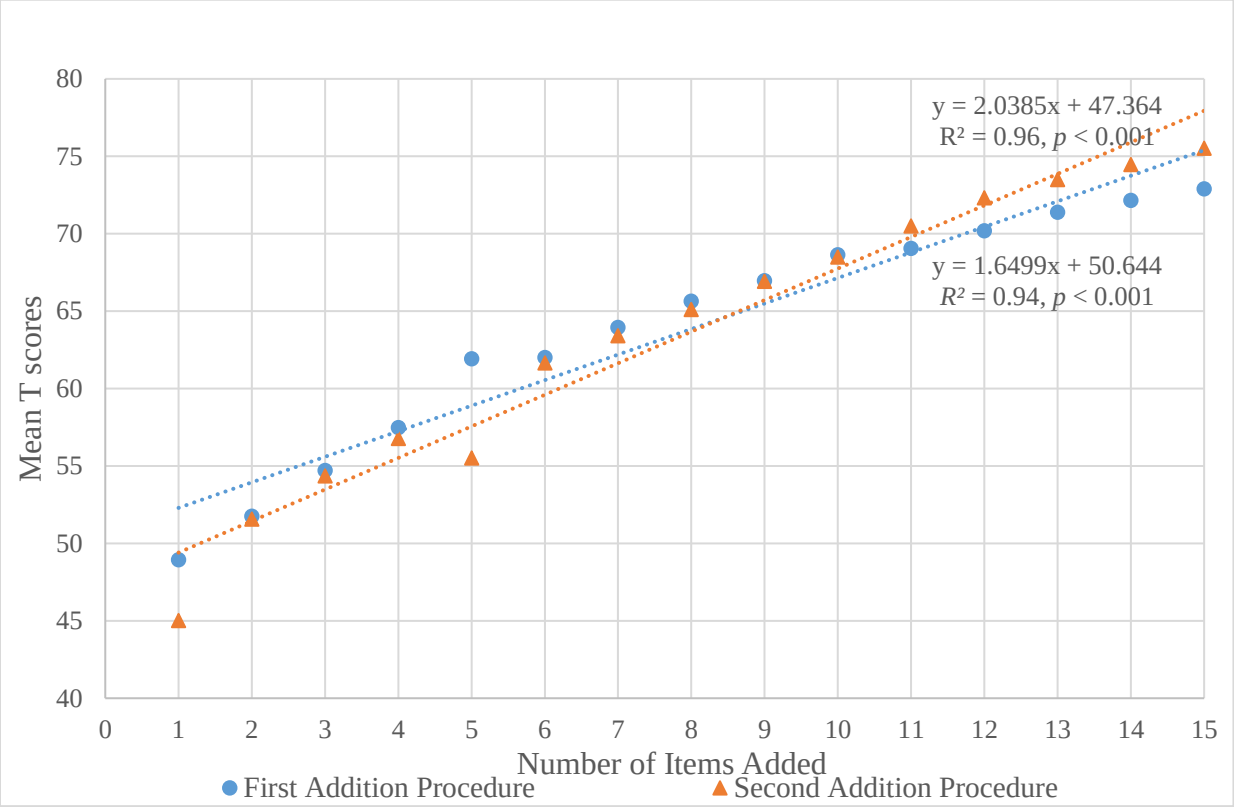


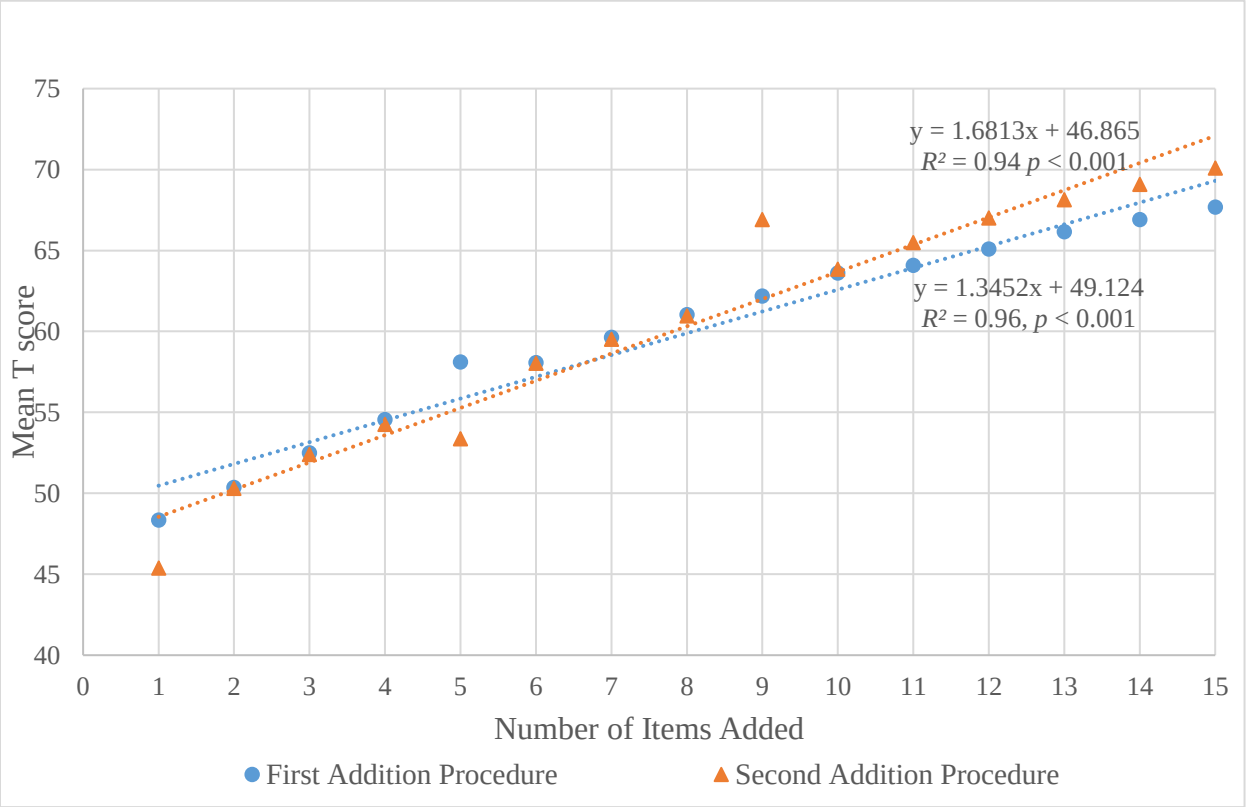
Table 20.

Results of t-tests of Mean T scores Between the VRIN-r/F-r and CRIN/F-r Scales for the First and Second Addition Procedure in the Faking Bad Sample.

	Scale						95% CI for Mean Difference			Cohen's *d*
	VRIN-r/F-r			CRIN-r/F-r						
	M	*SD*	*n*	*M*	*SD*	*n*		*t*	*df*	
First Addition Procedure	63.84	7.61	15	59.89	6.14	15	-1.21, 9.13	1.57	28	0.57
Second Addition Procedure	63.67	9.30	15	60.32	7.74	15	-3.04, 9.75	1.08	28	0.39

Figure 10.

Mean T scores for the CRIN/F-r scale in the faking bad sample.



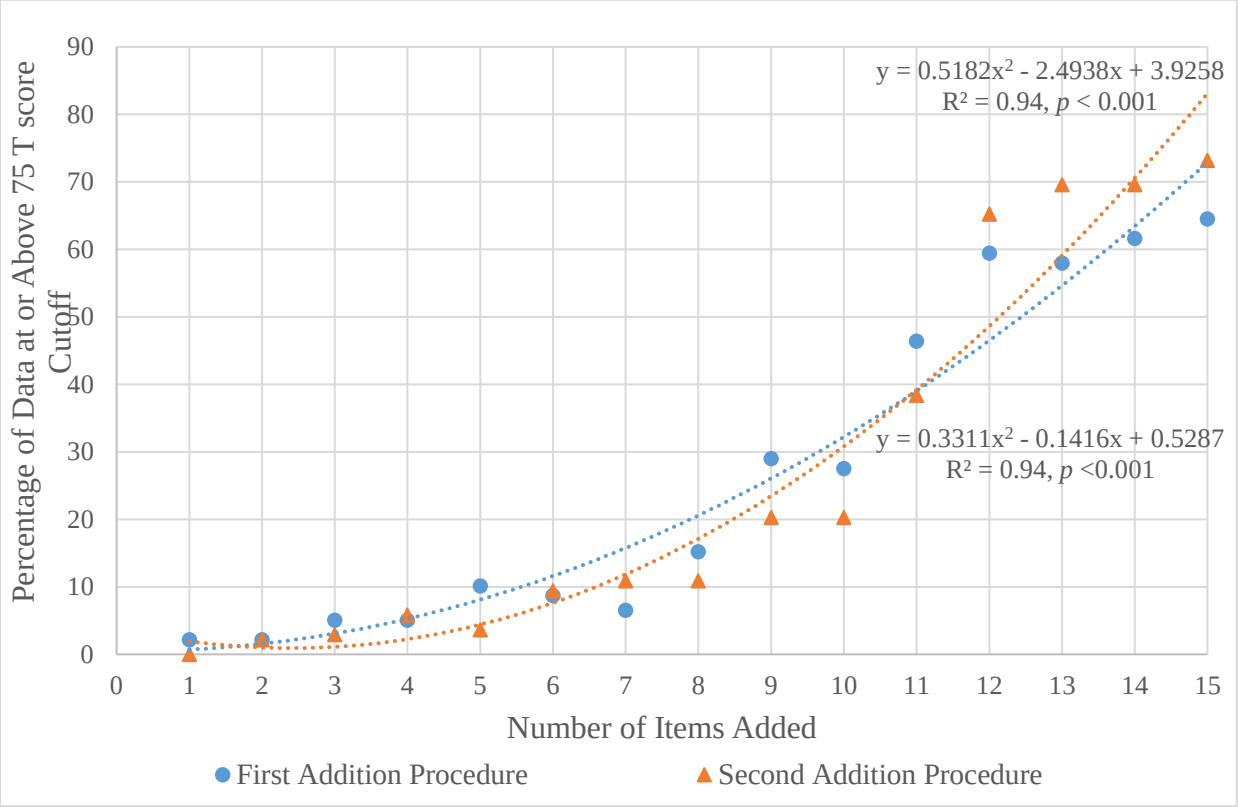
Comparing addition procedure for mean *T* scores. The first addition procedure produced higher mean *T* scores compared to the second addition procedure within the VRIN-r hybrid scale in the sample from item addition level one to item addition level ten (see Table 18). From this point on the second addition procedure began producing consistently higher mean *T* scores than the first addition procedure (item addition level eleven through fifteen) within the VRIN-r hybrid scale. A similar pattern was revealed in the CRIN hybrid scale (see Table 19) within the sample. Initially, the first addition procedure consistently produced higher mean *T* scores from item addition level one to level nine compared to the second addition procedure.

From the tenth item added to the fifteenth, the second addition procedure began to yield higher mean *T* scores than the first addition procedure. The *t*-tests of the contrasts between addition procedures collapsing across all item addition levels for the VRIN-r/F-r and CRIN/F-r scale can be observed in Tables 21 and 22. Both *t*-statistics were not significant, which suggests that there was no reliable difference between addition procedures at helping or harming specificity to overreporting within both hybrid scales.

Percentage above cutoff overall trend as items are added. Percentage above cutoff values for the VRIN-r and CRIN hybrid scales for both addition procedures within the faking bad sample can be found in Tables 23 and 24, and Figures 11 and 12, respectively. These data appear to be a quadratic function for both addition procedures within both hybrid scales, and indeed, analyses supported this. The VRIN-r/F-r scale percent above cutoff by item addition number equation of best fit for the first item addition procedure was $y = 0.3311x^2 - 0.1416x + 0.5287$ and $y = 0.5182x^2 - 2.4938x + 3.9258$ for the second addition procedure. Both equations were significant ($p < .001$) and excellent fits for these data (both $R^2 = 0.94$; see Figure 11). The equations for the CRIN/F-r scale above cutoff values in the faking bad sample for the first and second addition procedures were $y = 0.1618x^2 - 0.902x + 6.3497$ and $y = 0.225x^2 - 1.2856x + 5.2636$, correspondingly. Like the VRIN-r/F-r equations, these were both significant ($p < .001$) and near perfect fits ($R^2 = 0.96$ and 0.94, respectively; see Figure 12). Like the findings for the mean *T* score trends in the faking bad sample, these equations evidenced that, as items are added, both the VRIN-r/F-r and CRIN/F-r scale are sensitive to overreporting. Indeed, the significant quadratic equation suggested an exponential increase in profiles above the cutoff as items are added, again indicating that procedures to minimize the hybrid scales specificity from overreporting were necessary.

Figure 11.

Percent of scores above cutoff for VRIN-r/F-r scale in MMPI-A-RF faking bad sample.



Comparing VRIN-r/F-r and CRIN Scales by percentage above cutoff. Percentage above cutoff values for the VRIN-r and CRIN hybrid scales for the first addition procedure within the faking bad sample can be found in Tables 23 and 24. For the first The contrast *t*-test between the VRIN-r/F-r and CRIN/F-r scales in the first addition procedure for the faking bad sample can be seen in Table 25.

The mean VRIN-r/F-r percent above cutoff score (M = 26.76 SD = 24.44) was significantly greater than the CRIN/F-r mean percent above cutoff score (M = 12.25 SD = 8.22);

t(17.13) = 2.14, *p* = 0.047. Due to the homogeneity of variance assumption being violated between these two groups, the *df* of the *t*-statistic was adjusted.

Figure 12.

Percent of scores above cutoff for CRIN/F-r scale in MMPI-A-RF faking bad sample.

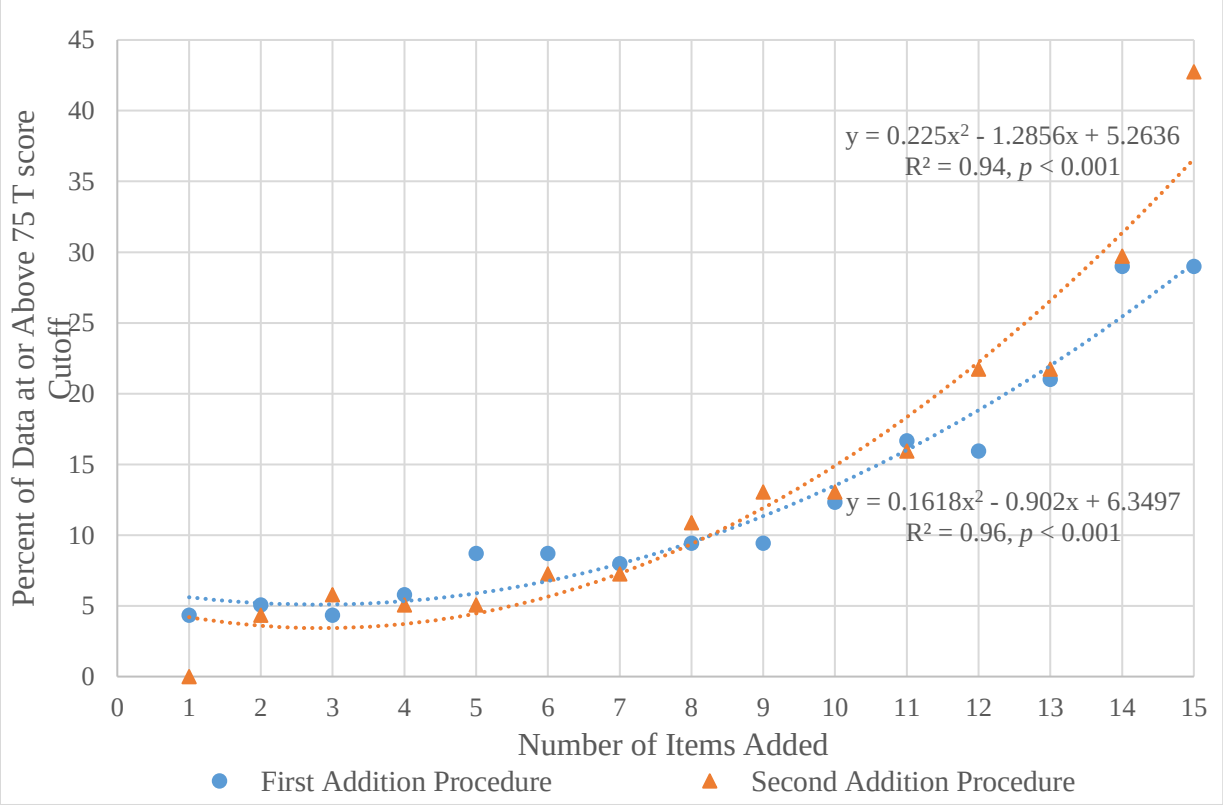


Table 21.

Results of t-tests of Mean T scores *Between the VRIN-r/F-r Scales for the First and Second Addition Procedure in the Faking Bad Sample.*

Scale						95% CI for Mean Difference			Cohen's *d*
VRIN-r/F-r First Addition Procedure			VRIN-r/F-r Second Addition Procedure						
M	*SD*	*n*	*M*	*SD*	*n*		*t*	*df*	
63.84	7.61	15	63.67	9.30	15	-6.18, 6.52	0.06	28	0.02

Table 22.

Results of t-tests of Mean T scores *Between the CRIN/F-r Scales for the First and Second Addition Procedure in the Faking Bad Sample.*

Scale						95% CI for Mean Difference			Cohen's *d*
CRIN/F-r First Addition Procedure			CRIN/F-r Second Addition Procedure						
M	*SD*	*n*	*M*	*SD*	*n*		*t*	*df*	
59.89	6.14	15	60.32	7.74	15	-5.66, 4.80	-0.17	28	0.06

Furthermore, it was not appropriate to calculate a Cohen's *d* for these data as the pooled variance between the two groups was not an appropriate reflection of the variance of the two groups. The significant *t*-statistic suggested that, overall, the CRIN/F-r scale is significantly more specific from overreporting causing more respondents to go over the *T* score cutoff than the VRIN-r/F-r scale. Percentage above cutoff values for the VRIN-r and CRIN hybrid scales for the second

addition procedure within the faking bad sample can also be found in Tables 23 and 24. For the second addition procedure the VRIN-r/F-r scale had the greatest percentage above cutoff score for 10/15 cells, while the CRIN-r/F-r scale captured the greatest percentage above cutoff score for 3/15 cells, while for 2/15 cells both the VRIN-r/F-r and CRIN/F-r scale had the exact same percentage above cutoff scores. The *t*-test statistics for mean comparison between these two scales can be found in Table 25. Here although, the VRIN-r/F-r scale did have a greater mean percentage above cutoff score when compared to the CRIN/F-r scale, it was not statistically significant. These data indicate that, for the second addition procedure, the observed CRIN/F-r percent above cutoff data suggested this scale is more specific than the VRIN-r/F-r scale at detecting overreporting; however, the differences observed between the two scales were not reliably different from zero when collapsing across item addition level.

Comparing item addition procedures by percentage above cutoff. For the VRIN hybrid scale (see Table 23), the second addition procedure began reliably producing higher percentages above cutoff compared to the first addition procedure after twelve items were added. Before this item level, the first addition procedure had the greatest percentage above cutoff value in 7/11 item levels, with the second addition procedure only having the greater percentage value in 3/11 item levels (item addition level two produced the same percent above cutoff value for both addition procedures). However, the *t*-test contrast between mean percentage above cutoff scores between the two addition procedures within the VRIN-r hybrid scale collapsing across item level (see Table 26) was not significant. This indicated that, although there are some observed slight variations between the two addition procedures within VRIN-r hybrid scale for the percent above cutoff scale, statistically these differences are not different form zero.

Table 23.

Percentage of Scores Above VRIN-r Cutoff Score in Faking Bad Sample.

Item Addition Configuration Type	Number of Fr-r Items Added to VRIN-r Scale														
	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
First	**2.17**	**2.17**	**5.07**	**5.07**	**10.14**	**8.70**	**6.52**	**15.22**	**28.99**	**27.54**	**46.38**	**59.42**	**57.97**	**61.59**	**64.49**
	3	*3*	*7*	*7*	*14*	*12*	*9*	*21*	*40*	*38*	*64*	*82*	*80*	*85*	*89*
Second	**0.00**	**2.17**	**2.90**	**5.80**	**3.62**	**9.42**	**10.87**	**10.87**	**20.29**	**20.29**	**38.41**	**65.22**	**69.57**	**69.57**	**73.19**
	0	*3*	*4*	*8*	*5*	*13*	*15*	*15*	*28*	*28*	*53*	*90*	*96*	*96*	*101*

Percentages in bold *n* in italics. Total sample size 138 for each cell.

Table 24.

Percentage of Scores Above CRIN Cutoff Score in Faking Bad Sample.

Item Addition Configuration Type	Number of Fr-r Items Added to CRIN Scale														
	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
First	**4.35**	**5.07**	**4.35**	**5.80**	**8.70**	**8.70**	**7.97**	**9.42**	**9.42**	**12.32**	**16.67**	**15.94**	**21.01**	**28.99**	**28.99**
	6	*7*	*6*	*8*	*12*	*12*	*11*	*13*	*13*	*17*	*23*	*22*	*29*	*40*	*40*
Second	**0.00**	**4.35**	**5.80**	**5.07**	**5.07**	**7.25**	**7.25**	**10.87**	**13.04**	**13.04**	**15.94**	**21.74**	**21.74**	**29.71**	**42.75**
	0	*6*	*8*	*7*	*7*	*10*	*10*	*15*	*18*	*18*	*22*	*30*	*30*	*41*	*59*

Percentages in bold *n* in italics. Total sample size 138 for each cell.

Table 25.

Results of t-tests of Percentage Above Cutoff Scores Between the VRIN-r/F-r and CRIN/F-r Scales for the First and Second Addition Procedure in the Faking Bad Sample.

	Scale						95% CI for Mean Difference		
	VRIN-r/F-r			CRIN-r/F-r					
	M	*SD*	*n*	*M*	*SD*	*n*		*t*	*df*
First Addition Procedure	26.76	24.44	15	12.51	8.22	15	0.61, 27.89	2.14*	17.13†
Second Addition Procedure	26.81	28.27	15	13.57	11.38	15	-3.27, 29.74	1.68	18.42†

* $p < 0.05$, † Homogeneity of Variance Assumption Violated Between Groups, *t* Statistic Not Calculated with Pooled Variance; Cohen's *d* not calculated

Table 26.

Results of t-tests of Percentage Above Cutoff Scores Between the VRIN-r/F-r Scales for the First and Second Addition Procedure in the Faking Bad Sample.

Scale						95% CI for Mean Difference			Cohen's *d*
VRIN-r/F-r First Addition Procedure			VRIN-r/F-r Second Addition Procedure						
M	*SD*	*n*	*M*	*SD*	*n*		*t*	*df*	
26.76	24.44	15	26.81	28.27	15	-19.82, 19.71	-0.01	28	< 0.01

This suggested that item addition procedure does not impact the specificity from detecting overreporting within the VRIN-r hybrid scale. Observing the CRIN hybrid scale (see Table 24), almost the exact same patterns were noted. The second addition procedure began reliably producing higher percentages than the first addition procedure after twelve items were added. Below this level, the first addition procedure obtained the greatest percentage value for 7/11 item levels, with the first addition procedure having the lower percentage value for the remaining 4/11 item levels. However, the *t*-statistics (see Table 27) indicate that these differences, collapsing across item level, did not reach statistical significance. Taken together, these data indicated that the CRIN hybrid scale, just like the VRIN hybrid scale, did not seem to have any differences in specificity from overreporting when comparing the two item addition procedures, collapsing across item addition level.

Table 27.

Results of t-tests of Percentage Above Cutoff the VRIN-r/F-r Scales for the First and Second Addition Procedure in the Faking Bad Sample

Scale						95% CI for Mean Difference			Cohen's *d*
CRIN/F-r First Addition Procedure			CRIN/F-r Second Addition Procedure						
M	*SD*	*n*	*M*	*SD*	*n*		*t*	*df*	
12.51	8.22	15	13.57	11.38	15	-8.49, 6.36	-0.29	28	0.11

Correlations between the hybrid scales and F-r overall trend as items are added.

Correlations between the VRIN-r/F-r and CRIN/F-r scales and the F-r sale by addition procedure within the faking bad sample can be found in Tables 28 and 29, respectively. Correlations between the both hybrid scales and the F-r scale within the faking bad sample generally increased linearly as items were added. It took the addition of only a few items for the correlations between the VRIN-r and CRIN hybrid scales with the F-r scale to become statistically significant. Looking at the VRIN-r hybrid scale for the first addition procedure, the correlation between the hybrid scale and the F-r scale first became significant at item addition level three [$r(138) = 0.30$, $p < .01$]. For the second addition procedure, the first statistically significant correlation also occurred at item addition level three [$r(138) = 0.25$, $p < .01$]. For the CRIN hybrid scale, the first statistically significant correlation with the F-r scale using both the first and second addition procedure occurred at item addition level four [$r(138) = 0.26$, $p < .01$; $r(138) = 0.23$, $p < .01$, respectively]. Using Cohen's (1988) suggestion that correlations at and above .50 suggest "large" effect sizes. The first and second addition procedure within the VRIN-r hybrid scale started to produce large effect sized correlations at and after item addition level five [$r(138) = 0.67$, $p < .01$); $r(138) = 0.51$, $p < .01$, respectively]. For the CRIN hybrid scale both first and second addition procedures yielded large effect size correlations at and above item addition level seven [$r(138) = 0.53$, $p < .01$); $r(138) = 0.50$, $p < .01$, respectively]. Indeed, correlations became extremely high between the hybrid scales and the F-r scale as item addition levels approached fifteen. For instance, the correlation obtained between the VRIN/F-r and F-r scale for the first and second addition procedures was $r(138) = 0.92$, $p < .01$ and $r(138) = 0.88$, $p < .01$, correspondingly, at item addition level fifteen. They stood at $r(138) = 0.84$, $p < 0.01$ and $r(138) = 0.79$, $p < 0.01$for the first and second addition procedures, respectively, at the fifteen

Table 28.

Correlations Between VRIN-r/F-r and F-r Scale in Faking Bad Sample.

Addition Configuration Type	Number of Fr-r Items Added to VRIN-r Scale														
	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
First	-0.09	0.12	0.30*	0.45*	0.67*	0.65*	0.72*	0.78*	0.81*	0.84*	0.86*	0.88*	0.90*	0.91*	0.92*
Second	-0.11	0.10	0.25*	0.41*	0.51*	0.61*	0.69*	0.74*	0.77*	0.80*	0.82*	0.83*	0.85*	0.87*	0.88*

*$p < 0.001$, $n = 138$ for each cell.

Table 29.

Correlations Between CRIN/F-r Scale and F-r Scale in Faking Bad Sample.

Addition Configuration Type	Number of Fr-r Items Added to CRIN Scale														
	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
First	-0.11	0.02	0.14	0.26*	0.47*	0.45*	0.53*	0.60*	0.66*	0.70*	0.73*	0.77*	0.80*	0.82*	0.84*
Second	-0.12	0.01	0.10	0.23*	0.32*	0.41*	0.50*	0.56*	0.61*	0.65*	0.69*	0.72*	0.75*	0.77*	0.79*

*$p < 0.001$, $n = 138$ for each cell.

item level for the CRIN hybrid scale. These data evidenced that it only takes a few added items to produce significant correlations between the two hybrid scales and the F-r scale in the faking bad sample. These significant correlations become large after only a few more items are added and become extremely large once item addition levels reach the fifteen-item level.

Comparing the VRIN-r/F-r and CRIN/F-r scales by correlation to F-r scale. Correlations by item addition level for both addition procedures were consistently less in the CRIN hybrid scale compared to the VRIN-r hybrid scale. This indicates that the CRIN/F-r scales were more statistically unrelated to the F-r scale than the VRIN-r/F-r scale was to the F-r scale in the faking bad sample for both addition procedures. This likely is explained by the original CRIN scale including more items than the initial VRIN-r scale, as it is composed of both VRIN-r and TRIN-r items. The addition of the TRIN items likely added an additional source of variance that tended to attenuate the correlations between the CRIN hybrid scale and the F-r scale when compared to the correlations that were produced between the VRIN-r hybrid scale and F-r scale.

Comparing item addition procedures by correlation to F-r scale. Correlations between both hybrid scales and the F-r scale were consistently lower for the second addition procedure compared to the first item addition procedure at every item level. This suggests that the second item procedure produced VRIN/F-r and CRIN/F-r scales that were slightly more statistically unrelated from the F-r scale in the faking bad sample compared to the first item addition procedure. This is likely explained by the second addition procedure being composed of the least endorsed F-r items in the faking bad sample, which likely resulted in it being much more statistically unique from the F-r scale for both hybrid scales when compared to the first addition procedure.

Selecting the Proper Number of Items to Add to the VRIN-r/F-r and CRIN/F-r Scales

Taking the preliminary analyses revealed above into account, these data suggested that, although there was observed differences, in general, there was no statistically significant differences between VRIN-r/F-r and CRIN/F-r scales in sensitivity in detecting random responders in the normative sample (as indicated by mean *T* and percent above cutoff scores). Although correlations suggested that CRIN/F-r was less correlated to the F-r scale than the VRIN-r/F-r scale in the faking bad sample, in practice this did not result in statistically significant differences in mean *T* and percentage above cutoff scores within the same data. This evidenced that the two scales provided statistically similar sensitivity from overreporting. As such, it was decided that both the VRIN-r and CRIN hybrid scales would be investigated as to the number of ideal items to add.

On the other hand, the preliminary analyses did reveal reliable statistical differences between the overall efficacy of the first and second item addition procedure within the hybrid scales. Specifically, the first item addition procedure consistently yielded higher sensitivity (as indicated by mean *T* and percent above the cutoff scores) to simulated random responding in the normative sample. This was especially pronounced once randomness insertion levels reached 75% and above. There was observed differences between item addition procedures within the faking bad sample and correlations suggested the second addition procedure produced hybrid scales that were more statistically unrelated to the F-r scale than the first addition procedure's resultant scales. However, in the analyses, none of these differences were statistically significant collapsing across item addition level. Overall, there did not appear to be any reliable differences between specificity in detecting overreporting between the two addition procedures within the hybrid scales. Thus, because the first addition procedure produced reliably better sensitivity to

simulated random responding and was not statistically different in specificity from the second addition procedure, it was the addition procedure used when selecting the proper number of items to add to the VRIN-r and CRIN hybrid scales.

To ensure that the VRIN and CRIN hybrid scales are sufficiently specific from detecting overreporting, the number of items added to both scales needed to produce percent above cutoff values for the faking bad sample that were at or under 10%. With this criterion in place, the goal was to ascertain the item addition level for both hybrid scales that has the highest percent above cutoff in the normative sample (suggesting maximized sensitivity to random responding) while simultaneously having a percent above cutoff level at or below 10% in the faking bad sample to maintain proper sensitivity. The item levels that meet the latter criteria can be found in Table 23 for the VRIN-r/F-r scale and Table 24 for the CRIN/F-r scale. For the first addition procedure, item addition levels one through four along with item level six and seven were all under 10% for the VRIN-r/F-r scale, and item levels one through ten were all under 10% for the CRIN/F-r scale. Remembering those item levels, the percentage above the cutoff data for the first addition procedure in the normative sample can be seen in Tables 10 and 11 for the VRIN-r and CRIN hybrid scales, respectively. For the VRIN-r data, item addition level six produced the highest percent above cutoff values for the 75% and 100% randomness level, whereas item addition level three produced the highest levels at the 25% and 50% level. The six-item level was selected as the ideal level because it had better sensitivity at the higher randomness levels, particularly at the 100% level, and because it's sensitivity levels at the 25% and 50% randomness level were very similar to item level three values. Looking at the ten items that met the specificity criterion for the CRIN hybrid scale, item level six produced the greatest percent above cutoff value for simulated randomness levels of 25%, 50%, and 75%, but item addition level eight produced the

greatest percentage value at the 100% level. The six item addition level was selected as the ideal number of items to add to the CRIN hybrid scale because it had three of four of the greatest percentage above cutoff values by simulated randomness level, and also only had a negligibly different value at the 100% randomness level compared to item level eight (67.70% vs. 69.01%, respectively). Furthermore, selecting a lower item addition level means that the new CRIN hybrid scale will be more statistically unrelated (by correlation) from the F-r scale. Taken together, the addition of six items using the first addition procedure in both the VRIN-r/F-r and CRIN/F-r scales created the ideal levels of both sensitivity to random responding and specificity from overreporting.

Adding the six F-r items to both the VRIN-r and CRIN scales also resulted in a dramatic improvement in the sensitivity to random responding compared to baseline. Looking at Table 10, the original VRIN-r scale detected 6.00%, 15.78%, 24.10%, and 32.67% of protocols being over the 75 *T* score cutoff for simulated randomness at 25%, 50%, 75%, and 100%. With the six F-r items added, the percentage of protocols detected to be over the cutoff increased to 8.20%, 28.20%, 50.50%, and 68.20% at each respective simulated randomness level. This means that adding the six F-r items to the VRIN-r scale approximately doubled the scales ability to detect protocols over the cutoff for randomness at 50% and above. Observing Table 11, the original CRIN scale detected 7.00%, 17.39%, 29.32%, and 39.75% of protocols being above the 75 *T* score cutoff for the simulated randomness levels of 25%, 50%, 75%, and 100%, correspondingly. When the six F-r items were added to the CRIN scale, these detection levels improved to 7.89%, 25.90%, 51.30%, 67.70% for each respective level of simulated random responding. Again, like what was seen in the VRIN-r scale, this represented a dramatic increase in sensitivity. Furthermore, comparing VRIN-r and CRIN hybrid scales with the six items added revealed that

both have very similar abilities in detecting protocols above the 75 *T* score cutoff for each level of simulated randomness.

Correlations Between the VRIN-r/F-r and CRIN/F-r Scales and the Restructured Clinical Scales

The correlations for the unrounded, untruncated T scores in the normative sample for the original VRIN-r, original CRIN, F-r, VRIN-r/F-r, and CRIN/F-r scales with the Restructured Clinical (RC) scales can be seen in Table 30. Ideally, the newly created VRIN-r/F-r and CRIN/F-r scale's correlations with the RC scales should more resemble the original VRIN-r and CRIN scale's correlations and not the F-r scales correlations with the RC scales, as this would indicate more statistical unrelatedness from the F-r scale, providing more evidence of specificity of the new hybrid scales from detection of overreporting. Reviewing Table 30, the VRIN-r/F-r and CRIN/F-r scale's correlations with the RC scales were generally more alike the original VRIN-r and CRIN correlations than the F-r correlations. Two notable exceptions were the correlations observed with Low Positive Emotions (RC4) and Antisocial Behavior (RC6). Here, the original VRIN-r and CRIN scales had correlations that were very similar or even higher than the F-r scales correlations to the two RC scales. This meant that the resultant VRIN-r/F-r and CRIN/F-r correlations tended to be closer and even higher than the F-r correlations. Otherwise, correlations to the seven other Restructured Clinical scales suggest that the two new hybrid scales appear to resemble the original scales indicating adequate statistical uniqueness from the F-r scales.

Table 30.

Correlations Between Original VRIN-r, Original CRIN, New VRIN-r + 6 Items, New CRIN + 6 Items, and F-r Scales with the Restructured Clinical Scales.

	Demoralization (RCd)	Somatic Complaints (RC1)	Low Positive Emotions (RC2)	Cynicism (RC3)	Antisocial Behavior (RC4)	Ideas of Persecution (RC6)	Dysf. Negative Emotions (RC7)	Aberrant Experiences (RC8)	Hypomanic Activation (RC9)
VRIN-r Original	0.18***	0.28***	0.37***	0.10***	0.45***	0.30***	0.11***	0.38***	0.02
VRIN-r +6 Items	0.25***	0.39***	0.41***	0.14***	0.47***	0.39***	0.15***	0.47***	0.04
CRIN Original	0.19***	0.31***	0.44***	0.10***	0.50***	0.35***	0.10***	0.42***	0.01
CRIN +6 Items	0.24***	0.38***	0.46***	0.14***	0.51***	0.41***	0.14***	0.49***	0.02
F-r	0.37***	0.58***	0.42***	0.21***	0.53***	0.53***	0.31***	0.68***	0.11***

***$p < 0.001$, $n = 1610$ for each cell.

CHAPTER IV

DISCUSSION

Adding six F-r items to both the VRIN-r and CRIN scales, using an item addition procedure that focused on F-r items that were the most infrequently responded to in the keyed direction within normative sample, yielded a dramatic increase in both scales sensitivity to simulated random responding in the normative sample. Furthermore, descriptive and correlational analyses within the normative sample and the faking bad sample indicated that the two newly created hybrid scales were specific enough from the F-r scale such that they are not overly sensitive to overreporting. This means they can both be effectively used as scales that uniquely detect non-content responding.

Taken altogether, both newly created scales should greatly aid the clinical user in effectively identifying and invalidating "aberrant" protocols (Tellegen, 1988) that are oversaturated with random responding. Clinical users can also be reasonably confident that protocols that are invalidated by the VRIN-r/F-r and CRIN/F-r scales are solely due to non-content responding and not overreporting. All of this should help the clinician in interpreting MMPI-A-RF protocols that truly represent that respondents psychological and personality functioning, and to be able to tell the difference between protocols that were invalidated due to content non-responsiveness and overreporting. In turn, this will help the MMPI-A-RF user to make the proper diagnosis and treatment recommendations.

In process of exploring the feasibility of adding F-r items to the VRIN and CRIN scales and determining the ideal number of items to add, several notable findings were revealed in the analyses. These are delineated and explained below:

Trend Analyses

In most of the data there was a clear linear trend in the Mean T score and percent above cutoff data for both hybrid scales within both the normative and faking bad sample as items were increased. This indicated that the addition of F-r items did improve the sensitivity of both hybrid scales in detecting simulated randomness in the normative sample. Furthermore, it indicated that both hybrid scales were also sensitive to overreporting in the faking bad sample, and thus proper analyses were needed to ensure that the new hybrid scales were adequately specific to just detect random responding and not overreporting. Some notable findings in the trend analysis data are discussed and explained below:

First, analyses revealed the Mean T scores and percent above cutoff values for both the VRIN-r/F-r and CRIN/F-r scales created by both addition procedures generally remained unchanged at the 25% and 50% simulated randomness levels as items were added in the normative sample. The quadratic and linear trends that were fitted to the data showed very minimal to modest increases at these two levels of randomness, as well, for the Mean T score and percent above cutoff data. Additionally, these two statistics tended to be relatively low at baseline for both the original VRIN-r and CRIN scales. This suggests that, first, the VRIN-r and CRIN scales are not very sensitive to random responding that saturates less than 75% of an MMPI-A-RF protocol, and second, adding F-r items does not seem to remedy this insensitivity to an appreciable extent.

Another notable finding in the trend analysis data was the quadratic trends identified in both hybrid scales for the Mean T score for the first addition procedure in the normative sample and with the percent above cutoff data for both addition procedures in the faking bad sample. The initial finding is likely explained by the first addition procedures focus on adding the least

endorsed items in the normative sample. The first finding is likely explained by the fact that the initial items added within the first addition procedure tended to have significantly lower endorsement rates than the middle items that were added (see Table 1). This meant that the simulated randomness tended to produce more deviant scores from the baseline for these more rarely endorsed items, resulting in sharper increases in the Mean T score as items were initially added. As more items were added though, the marginal endorsement rates between items lessened, which resulted in a flattening of the Mean T score slope line for both hybrid scales. This initial sharp increase and flattening of the slope explains why a negative quadratic trend was fitted. The second addition procedure was tailored to the least endorsed items in the faking bad sample, which meant that this same pattern was not reproduced in the second addition procedure for Mean T scores, rather a linear trend was fitted.

The second quadratic finding in the percent above cutoff data for both hybrid scales and both addition procedures in the faking bad sample was likely explained by the frequency distribution of the F-r data scores in this sample. The faking responders in this data produced a "J" curve in the F-r T score data, which meant, that as each additional F-r item was added to both the VRIN-r and CRIN hybrid scales, more and more faking responders were going to score in the keyed direction in both addition procedures and score above the cutoff. As such, the percent above cutoff data resembles the original distribution curve of the F-r data in the faking bad sample. Additionally, the second addition procedure focused on maximizing specificity by adding, sequentially, the lowest endorsed F-r items to both hybrid scale, which maximized this "J" curve more in these data for both hybrid scales.

VRIN-r/F-r vs. CRIN/F-r

One of the aims of the study was to investigate whether there was a difference in sensitivity to simulated randomness and specificity from overreporting between the VRIN-r/F-r and CRIN/F-r scales. Although there were some observed differences between the performance of these two scales, most statistical comparisons collapsing across item addition level revealed that both scale's sensitivity and specificity were similar. As such, analyses were conducted to add the proper F-r items to both scales. It was perhaps expected that the CRIN scale may outperform the VRIN-r scale in sensitivity to random responding because the CRIN scale, because of its inclusion of the TRIN-r c-composites, is longer. However, the inclusion of the TRIN-r items may have served as a break on the sensitivity of the CRIN scale. As was discussed in the introduction, 100% random responding on a MMPI-A-RF protocol has a 50/50 equal chance to activate both c-composites and F-r items in both the keyed and the non-keyed configuration. As such, although the inclusion of the TRIN-r c-composites lengthened the CRIN scale, it is likely that the simulated random responding was equally likely to trip these TRIN-r c-composites in a non-keyed (less acquiescent and counter-acquiescent) direction those reducing Mean T scores and suppressing the amount of protocols over the cutoff.

First Addition Procedure vs. Second Addition Procedure

A second aim of this study was to investigate whether there were differences in the sensitivity to random responding in the normative sample and specificity from overreporting in the faking bad sample for both hybrid scales if two different item addition procedures were used. Analyses did reveal differences. Specifically, the first addition procedure generally produced statistically greater mean *T* scores and percentages above cutoff, collapsing across item addition level, in both hybrid scales within the normative sample at all four levels of simulated

randomness. This suggest that the first addition procedure produced hybrid scales that were more sensitive to simulated random responding in the normative sample compared to the second addition procedure. Indeed, maximizing sensitivity is exactly why the first addition procedure was created, and statistical analyses bore than out. On the other hand, statistical comparisons of mean *T* scores and percent above cutoff, collapsing across item level, for both hybrid scales in the faking bad sample revealed no significant differences. This suggested that, even though the second addition procedure was created to maximize specificity from overreporting, the analyses suggested no statistical differences in specificity from overreporting between the two addition procedures in both hybrid scales. Thus, due to the first addition procedures superior sensitivity production, it was utilized to create the new VRIN-r/F-r and CRIN/F-r scales.

Limitations and Future Research

One of the limitations of the study was the nature of the simulated randomness inserted into the normative sample that was used in testing the sensitivity of the hybrid scales. A unique data set was produced for each of the four levels of randomness, at every item addition level, for both hybrid scales, and for both addition procedures; however, this means that, due to the nature of randomness, that some of these unique data sets could have been extreme causing the data to not be prototypically random and biasing some of the individual sensitivity statistics for certain data cells in the analysis. This bias was likely lessened due to this study sampling across multiple levels of randomness and across many item additions levels; however, it is important that this study be replicated to ensure that the sensitivity statistics obtained for the VRIN-r and CRIN hybrid scales are stable. It may also be feasible to use bootstrapping statistics to produce confidence intervals for the mean *T* scores and percent above cutoff values at each level of item

addition and randomness in future studies to help ensure that observed values are not extreme due to sampling error.

An additional drawback of the current study was the post-hoc nature of fitting trend lines to the data. The primary aim of this study was to test the feasibility of adding F-r items to both the VRIN-r and CRIN scale to improve their sensitivity to random responding, and as such, no hypotheses were made about the trends of the mean *T* score and percent above cutoff data as F-r items were added. Trend lines were fitted to the data to help the reader understand patterns, and then possible explanations of why these trend lines fit were provided. Now that these trends have been observed, it is crucial to conduct a replication study to ascertain if they are reproducible. It may also be fruitful to dig deeper into the relationship between F-r item endorsement rates in the samples and their relationship to the trends of the Mean *T* and percent above cutoff scores as items are added. This would not only aid in better understanding how F-r items affect the efficacy of the VRIN-r and CRIN scales in the MMPI-A-RF, but the statistical concepts revealed could aid future researchers in designing other non-content response scales in other self-report measures.

Another limitation of this study was the use of simulated fakers to aid in the specificity analyses for the hybrid scales and the use of simulated randomness only within the normative sample. Boone (2013) outlines some of the drawbacks of using simulated fakers to represent actual overreporting in symptom validity studies. She points out that there is questionable generalizability of findings from simulated feigners to data from actual overreporting test takers sampled in the clinical milieu. Indeed, studies using samples of simulated feigners tend to observe higher sensitivity measures in symptom validity measures then are observed in studies using feigners drawn from clinical samples (Boone, 2013). Boone additionally explains that there

is enough evidence to suggest there are differences in simulated overreporting samples vs. clinical overreporting samples that often in forensic settings plaintiff attorneys successfully attack symptom validity measures that were solely validated on simulated fakers. Additionally, simulated feigners may not actually feign in intended ways in experimental conditions (Lindstrom, Coleman, Thomassin, Southall, & Lindstrom, 2011). Furthermore, non-content responding was investigated only within the normative sample and by inserting simulated randomness.

Taken together, there is some question as to whether the findings obtained in this study will generalize to clinical populations and to respondents in those settings that are actual feigners and random responders. The ideal solution to this is to replicate the findings found here in a population of actual feigners and non-content responders drawn from clinical populations. Investigating the effects of simulated randomness in MMPI-A-RF protocols drawn from clinical populations would also be helpful, as it could determine if simulated randomness affects the VRIN-r and CRIN scales differently within the normative sample and clinical samples. All these possible areas of future research would do much in helping determine the generalizability of this study's findings in clinical populations.

Another area of future research is the investigation of how the original VRIN-r and CRIN scales are sensitive to differing levels of randomness. One of the main findings of the analyses in this study was that the original VRIN-r and CRIN scales were not particularly sensitive to simulated randomness at and below 50%, and that adding F-r items did not greatly improve the sensitivity at these randomness levels. Analyzing the sensitivity of the original VRIN-r and CRIN scale at differing levels of randomness was beyond the scope of this current study, and thus only four levels of randomness were utilized in analyses. Future studies may seek to test the

two scales with a more complete range of randomness to understand at what levels these two scales begin to reliably detect random responding. This would aid clinicians in understanding the relationship between the portion of randomness in the MMPI-A-RF protocol and VRIN-r and CRIN scores, which may better help them detect non-content responders. The other main finding from this study is that the sequence in which F-r items were added to the hybrid scales significantly affected their sensitivity to random responding (but their specificity from overreporting). Future studies may seek to observe how different item addition procedures using the F-r items affect the mean *T* and percentage above cutoff scores of the hybrid scales.

Finally, it is important to remind the reader that the CRIN/F-r scale produced, although sufficiently unique from detection of overreporting, is not just a measure of random responding. Because the original CRIN scale incorporates items from the TRIN-r scale, it is also a measure of acquiescent and counter-acquiescent responding. Only the VRIN-r/F-r scale that was produced is solely a measure of random responding due to its unique c-composite structure (Tellegan & Ben-Porath, 2008). Thus, clinicians need to be cautious when interpreting CRIN/F-r elevations as indications of random responding, as it could be this scale is elevated solely due to acquiescent or counter-acquiescent responding. Checking the TRIN-r and the VRIN-r/F-r scales in the protocol should help the clinician to determine what type of non-content responding is elevating the CRIN/F-r scale, helping the clinician to draw the proper conclusions about an invalidated protocol.

REFERENCES

Arbisi, P. A. & Ben-Porath, Y. S. (1995). An MMPI-2 infrequent response scale use with psychopathological populations: The Infrequency-Psychopathology Scale F(p). *Psychological Assessment, 7,* 424-431.

Archer, R. P. & Elkins, D. E. (1999). Identification of random responding on the MMPI-A. *Journal of Personality Assessment, 73,* 407-421.

Archer, R. P, Handel, R. W., Ben-Porath, Y.S., & Tellegen, A. (2016). *The Minnesota multiphasic personality inventory – adolescent – restructured form (MMPI-A-RF).* Minneapolis, MN: University of Minnesota Press.

Archer, R. P., Handel, R. W., Lynch, K. D., & Elkins, D. E. (2002). MMPI-A validity scale uses and limitations in detecting varying levels of random responding. *Journal of Personality Assessment, 78,* 417-431.

Baer, R. A., Berry, D. T. R., & Wetter, M. W. (1997). Detection of random responding on the MMPI-A. *Journal of Personality Assessment, 68,* 139-151.

Baer, R. A., Kroll, L.S., Rinaldo, J, & Ballenger J. (1999). Detecting and discrimination between random responding and overreporting on the MMPI-A. *Journal of Personality Assessment, 72,* 308-320.

Baer, R. A., Wetter, M. W., & Berry, D. T. R. (1992). Detection of underreporting on the MMPI: A meta-analysis. *Clinical Psychology Review, 12,* 509-525.

Ben-Porath, Y.S. *Interpreting the MMPI-2-RF*. Minneapolis, MN: University of Minnesota Press.

Ben-Porath, Y.S. & Tellegen, A. *The Minnesota multiphasic personality inventory – 2 restructured form (MMPI-2-RF).* Minneapolis, MN: University of Minnesota Press.

Berry, D. T. R., Wetter, M. W., Baer, R. A., Larsen, L., Clark, C., & Monroe, K. (1992). MMPI-2 random response indices: Validation using self-report methodology. *Psychological Assessment, 4,* 340-345.

Berry, D. T. R., Wetter, M. W., Baer, R. A., Widiger, T. A., Sumpter, J. C., Reynolds, S. K, & Hallam, R. A. (1991). Detection of random responding on the MMPI-2: Utility of F, Back F, and VRIN scales. *Psychological Assessment, 3,* 418-423.

Boone, K. B. (2013). *Clinical practice of forensic neuropsychology: An evidence-based approach.* New York: Guilford Press.

Boone, K. B., Lu, P, & Wen, J. (2005). Comparison of various RAVLT scores in the detection of noncredible memory performance. *Archives of Clinical Neuropsychology, 20,* 301-319.

Buechly, R., Ball, H. (1952). A new test for "validity" for the group MMPI. *Journal of Consulting Psychology, 16,* 299-301.

Butcher, J. N., Graham, J. R., Ben-Porath, Y. S., Tellegen, A, Dahlstrom, W. G., & Kaemmer, B. (1989). *The Minnesota multiphasic personality inventory – 2 (MMPI-2).* Minneapolis, MN: Univeristy of Minnesota Press.

Butcher, J. N., Graham, J. R., Ben-Porath, Y. S., Tellegen, A, Dahlstrom, W. G., & Kaemmer, B. (2001). *The Minnesota multiphasic personality inventory – 2 (MMPI-2) revised.* Minneapolis, MN: Univeristy of Minnesota Press.

Butcher, J. N., Williams, C. L., Graham, J. R., Archer, R. P., Tellegen, A., Ben-Porath, Y. S., Kaemmer, B. (1992). *The Minnesota multiphasic personality inventory – adolescent (MMPI-A).* Minneapolis, MN: University of Minnesota Press.

Charter, R. A. & Lopez, M. N. (2003). MMPI-2: Confidence intervals for random responding to the F, F back, and VRIN scales. *Journal of Clinical Psychology, 59,* 985-990.

Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (2nd ed.). Hillsdale, NJ: Lawrence Earlbaum Associates.

Cramer, K. M. (1995). Comparing three new MMPI-2 randomness indices in a novel procedure for random profile derivation. *Journal of Personality Assessment, 65,* 514-520.

Dalhstrom, W. G., Welsh, G. S., & Dalhstrom, L. E. (1972). *An MMPI handbook volume 1: Clinical interpretation* (rev ed). Minneapolis, MN: University of Minnesota Press.

Gallen, R. T. & Berry, D. T. R. (1996). Detection of random responding in MMPI-2 protocols. A*ssessment, 3,* 171-178.

Graham, J. R., Archer, R. P., Tellegen, A., Ben-Porath, Y. S., & Kaemmer, B. (2006). *The Minnesota multiphasic personality inventory – adolescent (MMPI-A) supplemental manual.* Minnesota, MN: University of Minnesota Press.

Greene, R. L, (1987). Ethnicity and MMPI performance: A review. *Journal of Consulting and Clinical Psychology, 55*¸497-512.

Greve, K. W., Ord, J., Curtis, K. L., Bianchini, K. J., & Brennan, A. (2008). Detecting malingering in traumatic brain injury and chronic pain: A comparison of three forced choice symptom validity tests. *The Clinical Neuropsychologist, 22,* 896-918.

Handel, R. W., Ben-Porath, Y. S., Tellegen, A., & Archer, R. P. (April, 2007). *Psychometric functioning of the MMPI-2-RF VRIN-r and TRIN-r scales with varying degrees of*

randomness, acquiescence, and non-acquiescence. Paper presented at the 42nd annual meeting on Recent Developments in MMPI-2 and MMPI-A Research, Fort Lauderdale, FL.

Handel, R. W., Ben-Porath, Y. S., Tellegen, A., & Archer, R. P. (2010). Psychometric functioning of the MMPI-2-RF VRIN-r and TRIN-r scales with varying degrees of randomness, acquiescence, and counter-acquiescence. *Psychological Assessment, 22,* 87-95.

Hathaway, S. R. & McKinley, J. C. (1942). Th*e Minnesota multiphasic personality inventory.* Minneapolis, MN: University of Minnesota Press.

Hathaway, S. R., McKinley, J. C. (*1951*). *The Minnesota Multiphasic Personality Inventory Manual (Revised)*. New York: Psychological Corp.

Kim M. S., Boone, K. B., Victor, T., Marion, S. D., Amano, S., Cottingham, M. E., Ziegler, E. A., & Zeller, M. A. (2010). The Warrington Recognition Memory Test for words as a measure of response bias: Total score and response time cutoffs developed on "real world" credible and noncredible subjects. *Archives of Clinical Neuropsychology, 25,* 60-70.

Lindstrom W., Coleman, C., Thomassin, K., Southal, C. M., & Lindstrom, J. H. (2011). Simulated dyslexia in post-secondary students: Description and detection using embedded validity indicators. *The Clinical Neuropsychologist, 25,* 302-322.

McKinley, J. C., Hathaway, S. R., & Meehl, P. E. (1948). The Minnesota multiphasic personality inventory: IV. The K Scale. *Journal of Consulting Psychology, 12,* 20-31.

Meier S. T. (1994). *The chronic crisis in psychological measurement and assessment: A historical survey*. San Diego, CA: Academic Press.

Pinsoneault, T. B. (2005). Detecting random, partially random, and nonrandom Minnesota multiphasic personality inventory – adolescent protocols. *Psychological Assessment, 17*, 476-480.

Pitta, A. C. (2016). *Psychometric functioning of the MMPI-A Restructured Form VRIN-r, TRIN-r, CRIN, and Cannot Say scales with varying degrees of randomness, acquiescence, counter-acquiescence, and omitted items.* (Unpublished doctoral dissertation). Virginia Consortium Program in Clinical Psychology, Norfolk, VA.

Rogers, R. (Ed.). (1997). *Clinical assessment of malingering and deception* (2nd ed.). New York: Guilford Press.

Sawilowsky, S. (2009). New effect size rules of thumb. *Journal of Modern Applied Statistical Methods, 8*, 467-474.

Stein, L. A. R., Graham, J. R., & Williams, C. L. Detecting fake bad MMPI-A profiles. *Journal of Personality Assessment, 65*, 415-427.

Tellegen, A. (1982). Brief manual of the Multidimensional Personality Questionnaire. Unpublished manuscript.

Tellegen, A. (1988). Analysis of consistency in personality assessment. *Journal of Personality, 56*, 621-663.

Tellegen, A. & Waller, N.G. (2008) Exploring personality through test construction: Development of the Multidimensional Personality Questionnaire, In G.J. Boyle, G. Matthews, and D.H. Saklofske (eds.), *Handbook of personality theory and*

testing: Vol.II. Personality measurement and assessment (pp.261-292). Thousand Oaks, CA: Sage.

Williams, C. L., & Butcher, J. N. (1989). An MMPI study of adolescents: I. Empirical validity of the standard scales. *Psychological Assessment: A Journal of Consulting and Clinical Psychology, 1*, 251-259

VITA

Stefan E. LaTulip

Virginia Consortium Program in Clinical Psychology
700 Park Avenue/MCAR - 410
Norfolk, Virginia 23504

EDUCATION

2013 - 2019 **Virginia Consortium Program in Clinical Psychology**
Ph.D., Clinical Psychology, expected December 2019

2011 - 2013 **Old Dominion University**
M.S., Experimental Psychology

2006 - 2010 **Michigan State University**
B.S., Psychology, Specialization in Cognitive Science

CLINICAL EXPERIENCE

2019 - Present **Thomson Memory Center**
Neuropsychology Post-Doctorate Fellow

2018 - 2019 **Tennessee Valley Healthcare System**
Pre-Doctoral Psychology Intern

PUBLICATIONS & PRESENTATIONS

LaTulip, S. E. (2019, June). Matrix Reasoning on the WASI-II as a performance validity indicator. Paper presented at the research conference for the Tennessee Valley Healthcare System's psychology internship research day, Nashville, TN.

McBride W. F., **LaTulip S. E.,** Handel R., Brokenbourgh, M., Arbisi, P., & Wygant, D. (2016, May). *Examining the measurement invariance of the Minnesota Multiple Personality Inventory-2 restructured form externalizing specific problems scales in African American and Caucasian men*. Paper presented at the annual MMPI Convention, North Hollywood Beach, FL.

LaTulip, S. E., Handel, R., & Archer, R. (2015, September). *Psychometric functioning ofthe MMPI-A Restructured Form (MMPI-A-RF) Specific Problems (SP) and Personality Psychopathology-Five (PSY-5) Scales with varying degrees of simulated Random responding, acquiescence, and counter-acquiescence*. Paper presented at the Virginia Consortium Program's research day, Norfolk, VA.

LaTulip, S. E., Winstead, B., Lewis, R., & Scerbo, M. (2014, August). *Depression, social anxiety, and attachment as predictors of use and perception of quality of communication via social networking sites and instant messaging*. Poster presented at the annual American Psychological Association convention, Washington, D.C.

LaTulip, S. E., Handel, R., & Archer, R. (2014, April). *Psychometric functioning of the MMPI-A Restructured Form (MMPI-A-RF) Specific Problems (SP) and Personality Psychopathology-Five (PSY-5) Scales with varying degrees of simulated random responding, acquiescence, and counter-acquiescence*. Poster presented at the annual MMPI convention, Scottsdale, AZ.