The Effect of Divulging the Intent of the Conditional Reasoning Test of Aggression to Responses

Nathan Carter
Western Kentucky University
THE EFFECT OF DIVULGING THE INTENT OF THE CONDITIONAL
REASONING TEST OF AGGRESSION TO RESPONDENTS

A Thesis
Presented to
The Faculty of the Department of Psychology
Western Kentucky University
Bowling Green, Kentucky

In Partial Fulfillment
Of the Requirements for the Degree
Master of Arts

By
Nathan T. Carter

August, 2006
THE EFFECT OF DIVULGING THE INTENT OF THE CONDITIONAL REASONING TEST OF AGGRESSION TO RESPONDENTS

Date Recommended July 31, 2006

Director of Thesis

Date 8/1/2006

Dean of Graduate Studies and Research
Acknowledgements

Through the course of my life, there have been many people who have affected me, be it personally or academically. I would like to take this opportunity to thank those who have truly had an impact on me. Had it not been for the social and professional support of these persons, the completion of my master’s degree would not have been possible.

First, I would like to thank my family: To my mother, who never gave up on me, despite some early academic shortfalls; to my father, who always taught me that when you’re in doubt about something, look it up; and to my little brother, Kyle, for being a great friend and a loyal sibling.

Certain faculty members at Western Kentucky University are also due much credit for supporting and promoting my academic development: To Sam McFarland, thank you for being the first to help me find research opportunities and to really get me thinking; to Elizabeth Shoenfelt, thank you for all the applied and academic research opportunities in both my undergraduate and graduate careers; to Reagan Brown, thanks for being an incredible teacher and actually being fun to work with (“If loving regression is wrong… then I don’t want to be right”); and to John Bruni, Jr., who pushed me harder than anyone had as an undergraduate, and for being there for me as a graduate student.

Although professors at Western Kentucky University have played a very important part in my academic development, my fellow graduate students have made the past two years not only tolerable but enjoyable. In particular, I would like to thank Paula Starling for being a great friend and co-investigator, for pushing me to apply to Bowling Green State University’s doctoral program, despite the fact that I thought I wouldn’t be
accepted, and for getting me involved with activities and people that I believe were
instrumental to my acceptance to Bowling Green State University.

Finally, I would like to dedicate this paper to the memory of my grandfather,
Elmer T. Carter, who unfortunately is not with us today to see the accomplishments of
mine and my brother. Were he here I know he would be proud and that provides a great
source of satisfaction for me. He is and will always be missed.
# Table of Contents

Abstract.................................................................vi
Introduction...............................................................1
Method.................................................................14
Results...............................................................16
Discussion.............................................................18
References..........................................................20
Appendix A- Informed Consent Document.................................25
Appendix B – Procedural Script for Control Group Experimental Trials........27
Appendix C – Procedural Script for Experimental Group Experimental Trials...29
Appendix D – Demographics Form........................................31
Appendix E – Instructions Provided to Participants..........................33
THE EFFECT OF DIVULGING THE INTENT OF THE CONDITIONAL REASONING TEST OF AGGRESSION TO RESPONDENTS

Nathan T. Carter
August 1, 2006
39 Pages

Directed by: Reagan D. Brown, Elizabeth L. Shoenfelt, and Kathi Miner-Rubino

Department of Psychology
Western Kentucky University

Under the conditional reasoning (CR; James, 1998) system of personality measurement, items are constructed to have the appearance of a logic or reasoning test. This characteristic of CR tests is designed to combat response distortion by test takers in personality testing. The CR system was used to develop the Conditional Reasoning Test of Aggression (CRT-A; James & McIntyre, 2000), a tool used to screen employees whose scores indicate an aggressive disposition. An assumption of the CRT-A is that respondents are unaware of the construct being measured. This study examines the effect of informing respondents that the CRT-A measures aggression. A sample of 80 participants from a large southeastern university in the United States were randomly assigned to either the control (CRT-A standard administration) or experimental (respondents are aware the test is a personality measure of aggression) group. Results revealed no CRT-A score differences between groups, suggesting that the CRT-A is robust to violations of the assumption that test takers be under the impression that the test is a measure of reasoning. Descriptive statistics and effect size are reported. A discussion of the implications of the present findings and suggestions for future research is also provided.
CHAPTER 1

Introduction

In recent decades personality measurement has attracted the attention of industrial psychologists as a predictor of organizational variables such as job performance due to its low likelihood of nontrivial adverse impact against minority groups (Hough & Oswald, 2000). Unfortunately, research on personality and its relationship with many organizational variables, such as job performance and absenteeism, has shown modest results. Meta-analyses of hundreds of studies have failed to report validity coefficients in excess of .3 (Barrick & Mount, 1991; Ones, Viswesvaran, & Schmidt, 2003; Salgado, 2003; Schmitt, Gooding, Noe, & Kirsch, 1984; Tett, Douglas, & Rothstein, 1991). In contrast, a great deal of evidence exists to support the predictive validity of cognitive ability tests for many jobs (Outtz, 2002). However, tests of general cognitive ability have been found to show considerable differences based on racial group membership (Hunter & Hunter, 1984), whereas group differences in personality test scores are most often negligible.

As with personality testing, use of integrity testing has been associated with small score differences between various racial groups (Collins & Gleaves, 1998). Meta-analytic investigations have found integrity tests to be an effective predictor of job performance and absenteeism. Ones and Viswesveran’s (1998) meta-analytic investigation of integrity testing found a mean criterion related validity coefficient of .41 for the prediction of job performance. In 2003, Ones, Viswesveran, and Chocklingham reported an average
validity coefficient of .33 for personality based integrity tests for the prediction of absenteeism, but less impressive results for overt integrity tests ($r = .09$).

Self-report personality tests and personality based integrity tests are both susceptible to test taker response distortion. One can easily see the returns gained from the job applicant who fakes on a personality inventory or integrity test, namely an increased likelihood of success obtaining the job. A considerable amount of research (e.g., Furnham, 1997; McFarland, 2003; Rosse, Stecher, Levin, & Miller, 1998; Viswesveran & Ones, 1999) has been conducted on the propensity for applicants to fake tests of personality. In contrast, research on faking integrity tests has been somewhat sparse.

**Personality Tests and Faking**

Personality testing is one of many tools employers use to objectively determine the best candidate for the job. However, these tests have been shown to be susceptible to response distortion by candidates in an attempt to lower or raise their scores in order to maximize their potential to attain employment. For the purpose of the current study, a discussion of the literature on faking personality tests follows to familiarize the reader with recent research developments.

For those conditions in which respondents are told to fake, differences can be quite large, and dimensions of personality appear to be equally fakable (Viswesveran & Ones, 1999). A study conducted by Furnham (1997) indicated that test takers are capable of predicting their own scores on the NEO-FFI (Costa & McCrae, 1988) with significant accuracy. This finding suggests that participants are perceptive of the construct measured by the test and their standing on the construct. However, significant differences between
fake good (i.e., conditions in which respondents are instructed to respond in the desirable direction on a test) and control conditions were only present on the dimensions of agreeableness, conscientiousness, and neuroticism. These are the three dimensions of the five factor model of personality that have been found to be related to several counterproductive work behaviors (Salgado, 2002). One of them, conscientiousness, is the dimension found to be related to performance in the most occupations (Barrick & Mount, 1991).

Rosse et al. (1998) tested incumbents (those currently holding the job) and applicants (those attempting to get the job) and found that applicants scored consistently higher than the incumbents on the NEO-PI-R (Costa & McCrae, 1992), suggesting that those trying to get a job were attempting to distort their responses in the desirable fashion. The effect of these distortions could have a dramatic impact on test validity and utility. McFarland’s (2003) study compared test takers warned against faking by being told the test contained a social desirability scale to detect faking and a control group. The results indicated a significant difference between groups on NEO-FFI (Costa & McCrae, 1988) dimensions of neuroticism, extraversion, and conscientiousness such that the group warned against faking scored less desirably than the control group, suggesting that the warning deterred participant faking (McFarland).

Hogan, Hogan, and Roberts (1996) presented the concept that faking capability may be related to job performance because in the work environment one must identify and display the appropriate behaviors to succeed in the workplace. This proposal added to Ones, Viswesvaran, and Reiss’s (1996) suggestion of faking as a “red herring” due to its lack of attenuation of the correlation coefficient. Evidence against Hogan et al.'s position
was offered by Rosse, Levin, and Nowicki (1999) who reported that scores on a faking measure were negatively related to supervisory ratings of positive work behavior (e.g.,
responding to customers with warmth, listening carefully to customers) and positively
related to negative work behavior (e.g., lying about mistakes, exaggerating to customers).

*Integrity Tests and Faking*

Another psychological assessment tool used by employers to predict and select
the best employee for a job is the integrity test. Although the literature on integrity testing
and faking is not as abundant as the literature on personality testing and faking, a separate
discussion of recent investigations is in order to familiarize the reader with the topics.

Under instructions to fake integrity tests, participants can apparently be quite
successful in doing so. Brown and Cothern (2002) found that participants instructed to
fake good (in a realistic manner) scored significantly higher \( (d = 1.3) \) than participants in
the control condition. This finding is also supported by an earlier meta-analysis by
Alliger and Dwight (2000) concerning the sensitivity of integrity tests to faking. Faking
instructions produced a large effect size \( (d = .90) \), which increased when coaching
instructions were provided \( (d = 1.32) \).

Another milestone in integrity testing can be seen when considering the results of
a study that suggested a positive relationship between perceived invasiveness of an item
and the degree to which the item is resistant to faking (Alliger & Dwight, 2001). In other
words, the items that applicants viewed as being most invasive were also the items
applicants appear to be unable to fake. This is especially disturbing when considering that
perceived invasiveness of selection tests has been found to be positively related to
applicants’ perceived likelihood of applicant complaints concerning the fairness of a
selection procedure (Thibodeaux & Kudisch, 2003). These results suggest that tests containing faking resistant items are more likely to instigate complaints based on the fairness of the test.

The Conditional Reasoning Test of Aggression

From the review above, one can see that there are serious problems with self-report personality and integrity tests that confound validity studies concerned with self-report variables. Researchers have used projective techniques (e.g., Rorschach Ink Blot Test, Thematic Apperception Test) to combat intentional distortion by test takers for the purpose of measuring constructs with greater accuracy. The intent of projective measurement is to allow for a less contaminated measure of the construct. The logic supporting this position is that if test takers are not aware of the construct being measured, it will be difficult for them to fake in a strategic way (i.e., achieve the desired goal). In studies conducted by Bornstein, Rossner, Hill, and Stepanian (1994) it was found that when comparing a projective and a self-report measure of dependency, participants were able to identify the construct and fake their responses on the self-report measure. However, the projective measure was robust to the attempts of faking and identification of the construct. Furthermore, the authors reported that scores could be manipulated on the self-report measure by portraying the construct of dependency in a positive light prior to testing whereas scores on the projective measure were unaffected by this manipulation.

Although projective measurement methods help to deter faking in respondents there are several problems with their use in employment settings. First, results of validity studies of projective tests have been less than desirable. In a review by Reilly and Chao
the average validity coefficient was found to be only .18, far below that of other predictors. Miner (1977) found significant differences between race and gender groups for a projective sentence completion test such that African American males scored higher on the test than both Caucasian males and females in general. Reilly and Chao also indicated that the “poor face validity, the need for psychological interpretation, and multiplicity of scoring schemes make the application of projective tests difficult” in employment settings and conclude in their paper that “projectives appear to lack the basic requirements of sufficient reliability and validity” (p. 52).

In 1998, Lawrence R. James introduced, not just a suggestion for deterring respondents from distorting responses, but a whole new measurement system aimed at tapping into personality constructs indirectly to combat the problem of possible faking inherent in self-report measures of dispositional, latent traits. This new type of measurement approach is called conditional reasoning testing. In this measurement model, personality is assessed with items under the guise of a reasoning or logic test. As such, respondents are given the impression that the items have correct and incorrect answers, thus they do not think about desirable responses. Rather, test takers consider responses to be right and wrong, as with an ability test. Respondents are not aware that the solutions they select are used to indicate an underlying disposition. Following is a discussion of the conditional reasoning measurement system and the tests that have been developed using the system.

Overview of the Conditional Reasoning System of Measurement

The premise of the conditional reasoning measurement system (James, 1998) centers on the assumption that a person’s behavioral choice is governed by the way the
individual reasons in a particular situation. This reasoning process justifies the person’s behavior, allowing the individual to appear to himself as a rational being whose behavior is governed by logic. The reasoning processes used by individuals are referred to as Justification Mechanisms (JM). It is hypothesized that the different JM’s used by an individual should reflect tendencies related to the individual’s disposition. These JM’s result in reasoning that is termed *conditional*. For instance, someone who engages in aggressive behavior should have unique JM’s that enable the choice to behave aggressively to appear as the rational behavioral choice in that situation. Ideally, a conditional reasoning measurement system offers the faking resistant benefits of projective testing and the scoring benefits of self-report testing.

To date, James and his colleagues are the only researchers with published investigations of the conditional reasoning measurement system (James, 1998; James, McIntyre, Glisson, Bowler & Mitchell, 2004; James et al., 2005). Published research describes the development of two conditional reasoning measures: achievement motivation (James, 1998) and aggression (James et al., 2005). However, the only commercially available test thus far is the Conditional Reasoning Test of Aggression (CRT-A; James & McIntyre, 2000). The following is a summation of the structure of items under the conditional reasoning framework, the conditional reasoning measure developed for the construct of aggression, and the available empirical analyses of the measure.

*Structuring Items that Reflect Justification Mechanisms*

The CRT-A was developed to measure the construct of aggression with two motive dispositions: the *motive to aggress* and the *motive to hold a favorable view of the*
The motive to aggress is identified as the desire to harm an individual or entity and the motive to hold a favorable view of the self is signified by desiring to view oneself as having done the socially or morally right thing. The degree to which these two motives oppose one another is determined by the respective strength of the individual motive (James et al., 2005).

The JM's identified as being needed to rationalize aggressive behavior are: a) hostile attribution bias, in which the person believes that others are also motivated by the prospect of doling out harm to others; b) potency bias, in which the person believes that all conflicts are competitions in which one is trying to exhibit dominance by winning; c) retribution bias, in which the person believes that paying people back is more important than mending relationships or coming to an agreement; d) the victimization by others bias, in which the person believes that due to all the wrong a person sees as being done to him, it only seems fit that this person inflicts wrong doing on others; e) the derogation of target bias, in which the person paints a picture of the person as being pure evil and therefore, in the person's mind, a more deserving recipient of aggressive acts; and f) the social discounting bias, in which the person views social norms and customs as disrupting the natural way of things, thus viewing acting out on aggressive impulses as an expression of their freedom (James et al., 2005).

Item stems were constructed in such a way that they allow the respondent to choose the most logical of the provided conclusions based on the statement contained in the stem. Of the four inferences available as options, two are logically valid. One valid response reflects a motive to aggress, and the other reflects the motive to hold a favorable view of the self (and thus, to not aggress). The CRT-A in its tested form included 22 of
these items. One point is given for the selection of JM’s that represent the underlying motive to aggress, and one point is deducted for the selection of the option reflecting the motive to view oneself as being in keeping with what is socially or morally right. No points are awarded for either of the logically invalid answers, described by James et al. as “meant to be clearly illogical and rejected by respondents” (2005, p. 77).

**Validity Evidence for the CRT-A**

The Conditional Reasoning Test of Aggression (CRT-A) is a 22-item measure of the aggression construct (James & McIntyre 2000). James et al. (2005) reported results from eleven empirical criterion related validation studies using different samples of persons in various positions, seven of which will be focused on here. Scores on this test were significantly related to a) nuclear facility operator absences ($r = .42$), b) restaurant employee tenure ($r = .34$), c) package handler absence ($r = .34$), d) supervisory ratings of temporary employees ($r = -.43$), e) witnessed theft in undergraduate students ($r = .64$), f) number of hard fouls in a college intramural basketball game ($r = .38$), and g) cheating on an Internet based simulation game ($r = .40$). All validity coefficients were uncorrected (i.e., no adjustments for reliability or range restriction).

Correlations with measures of critical intellectual skills as measured by American College Testing (ACT) scores showed no significant relationship with scores on the CRT-A in three different samples of undergraduate students ($r = -.06; r = -.05; r = -.08$; James et al., 2005). These correlations support the position that a personality test (regardless of measurement method) should have no relationship with intellectual skills and provides discriminant validity evidence for the CRT-A. Regarding the CRT-A’s relationship with gender, the CRT-A proved to be significantly related to gender in two
of seven samples, with correlations ranging from .20 to .22. Nonsignificant correlations did not exceed .08. James et al. stated that these findings indicate that males are slightly more likely to justify aggressive acts than women. In four different samples no relationships with race emerged as being significant, suggesting that race is not related to aggression test scores.

James et al.’s (2005) investigations of the CRT-A’s relationship to other self-report variables show a significant negative relationship to the NEO-PI-R’s Dutifulness scale (an indicator of low aggressiveness; $r = -.18$), a significant positive relationship with the Costa & McCrae Angry Hostility scale ($r = .26$) (1992), and a positive significant relationship to Buss and Perry’s (1992) Aggression Questionnaire ($r = .24$). These results suggest modest convergent construct validity evidence for the CRT-A. However, CRT-A scores were uncorrelated with the Personality Research Form’s (Jackson, 1967) Aggression, Dominance, and Impulsivity scales (James et al.). Inconsistent with previous findings is the report of an additional sample to assess the relationship between the NEO-PI-R’s Angry Hostility scale, which produced a nonsignificant correlation with the CRT-A ($r = .002$). In discussing these findings, James et al. noted that the reason for these weak findings may be explained by the fact that self-report measures and projective techniques measure “complimentary aspects of traits, motive, and need states” (p. 93). When the CRT-A, ACT, PRF Aggression, PRF Dominance, and PRF Impulsiveness were entered into a regression equation aimed at predicting not being truthful about extra credit in an undergraduate student sample, it was found that the CRT-A emerged as the only significant predictor and contributed 82% of the relative variance in the equation. The CRT-A was also entered into a regression
equation with the NEO-PI-R’s Angry Hostility and Dutifulness measures. All variables were significant predictors and the CRT-A appeared as the main contributor to prediction, subsuming 77% of the relative variance. When the CRT-A and the PRF Aggression questionnaire were both used as predictors of theft, the CRT-A again claimed 71% of the relative importance to prediction, and 74% when used with the NEO-PI-R Angry Hostility scale to predict hard fouls and fights in a basketball. The position of James et al. is that the CRT-A’s enhanced predictive validity is due to its resistance to faking. The other measures used in these studies were in a self-report format, where the CRT-A attempts to indirectly tap into the latent trait of aggression.

A simple perusal of any of the articles discussing the development of a test using the conditional reasoning measurement framework (James, 1998; James et al., 2004; James et al., 2005) makes obvious that developing such tests is hard work, built on theory and scientific evidence. However, it appears from some of the results described above that the returns in validity and prediction could be great. Additionally, the problem of adverse impact does not appear to be much of an issue. Most notably, from a prediction standpoint, the conditional reasoning measurement system thus far has consistently exceeded the $r = .3$ barrier often found in personality measurement. However, some limitations and opportunities for investigation exist in the available research.

Although the measurement system appears to be built to combat it, faking may still be a concern. Information about different types of tests occasionally leaks into the popular media (James, 1998). The effect of test takers knowing what is attempting to be measured is, as of yet, unreported in the literature. The conditional reasoning measurement system requires that the test taker be under the impression that the test is
measuring reasoning and not the personality construct it is designed to measure. As time passes, this requirement may become more and more difficult to achieve. As the CRT-A becomes more popular, its name will begin to circulate through not only the technical literature, but also popular literature. For instance, *Fortune* magazine writer Anne Fisher (2005) commented on “dozens of readers” (p. 34) who suggested colleagues use the CRT-A to screen out aggressive employees in response to an article on workplace violence. Also, in *Operations & Fulfillment* (2002), a business and retailing magazine, a title reading “It’s All the Rage” tells most of the story; of significance to the current argument is a section in which example items are given to readers to test themselves, complete with provided answers.

*The Present Study*

The ramifications of job applicant knowledge of the intent of the CRT-A on CRT-A test scores are presently unknown. James (1998) listed this issue as one of the research questions that should be addressed in future research concerning, not only the CRT-A, but the conditional reasoning measurement system as a whole. The current study addressed the question posed by James in regard to the effect of divulging to respondents the nature of the CRT-A’s measurement. To answer this question the present study examines the difference between the scores from two groups on the CRT-A. One group (the control group) will be administered the CRT-A under the standard testing procedure provided in the CRT-A test manual. The second group (the experimental group) will be administered the CRT-A under the same conditions as the control group with the exception that the experimental group was told that the test measures aggression. It is hypothesized that scores on the CRT-A will be lower (lower scores indicating a more
This desirable score) when respondents are told that the test measures aggression. This investigation is intended to serve as a response to James’ (1998) call for this research question to be addressed.
CHAPTER 2

Method

Participants

Data were collected from a large southeastern university. Participants for the study were college students attending undergraduate psychology classes. The sample consisted of 80 participants. Of the total sample, the majority (61%) were female. The sample consisted of Caucasians (88%), African-Americans (7%), and Asian/Pacific Islanders (4%). The mean age of the sample was 20.4 (SD = 2.32). When examined by condition it was found that the two groups were similar in terms of gender, race, and age composition. The mean age was 20.73 (SD = 1.89) and 20.08 (SD = 2.66) for the control and experimental groups, respectively. Race composition was the most disparate variable between the control group (5% African-American; 87% Caucasian; 7% Asian/Pacific Islander) and the experimental group (10% African American; 90% Caucasian). Gender composition was nearly equivalent for the control (62% female) and the experimental (60% female) groups.

Procedure

Participants were randomly assigned to one of two groups, experimental (n = 40) or control (n = 40). At the beginning of the testing session, participants were informed of their rights as a participant in the study and asked to sign the informed consent document provided in Appendix A. The control group was administered the CRT-A (James & McIntyre, 2000) following exactly the directions provided in the test manual published by The Psychological Corporation. Participants in both groups were also told to take the test as if they were taking the test as part of an application process for a job they wish to
get. The experimental, or *divulgence*, group was additionally told that, “This is a personality test that measures aggression.” The previous statement was the only difference in treatment for the two groups (see Appendixes B and C for procedural scripts). After completing the CRT-A, participants completed a demographics form (see Appendix D), providing information on their age, race, and gender.
CHAPTER 3

Results

The CRT-A was scored according to the rules prescribed in the accompanying manual (James & McIntyre, 2000). Aggressive responses were assigned a score of 1. Although the studies reporting the development of the scales (James et al., 2005) discussed in previous sections of this report assigned a score of -1 for nonaggressive responses, the manual states that all responses other than aggressive responses be assigned a score of 0 (James & McIntyre). The aggressive responses were summed to result in the total CRT-A score. The test manual also states that illogical responses should be counted to determine the validity of the test administration for each test taker. If a test taker has five or more illogical responses the test is considered invalid and not scored. No test takers in the present sample met the criteria for an invalid test. The mean illogical response score for the total sample was low ($M = .40, SD = .70$). No significant differences were found, $t(78) = .43, p > .05$, between the control ($M = .40, SD = .81$) and experimental ($M = .47, SD = .71$) conditions in terms of illogical responses.

The hypothesis of this study was investigated with a one-tailed independent samples $t$-test to determine if CRT-A scores were lower for the experimental (divulgence of the intent of measurement) group as compared to the control (no divulgence of intent of measurement) group. Examination of the data revealed that the $t$-test’s assumptions of normality and homogeneity of variance were tenable. The analysis revealed a nonsignificant difference, $t(78) = -.62, p > .05$, between the experimental ($M = 4.05, SD = 2.13$) and control ($M = 4.35, SD = 2.13$) conditions. Effect size calculation revealed a small effect size ($d = .14$).
In the faking literature, there is a line of thinking that test faking should not be examined at the scale level but rather at the item level. The rationale behind this position is that faking occurs at the item level, one item at a time. Put simply, respondents do not fake scales, they fake items (Zickar & Robie, 1999). Although the sample size for the current study precludes more sophisticated analytic techniques (e.g., confirmatory factor analysis), mean differences were analyzed at the item level. One-tailed independent samples t-tests were run for each of the items with the hypothesis that the experimental group would have significantly lower means than the control group on each item. Bonferroni corrections were applied to minimize Type I Error. The results for the item level analysis revealed no significant differences for any of the items. Without a Bonferroni correction, only one item (Item 10) displayed a significant difference the hypothesized direction.

It is also notable that reliability was considerably low for the control ($\alpha = .31$, $n = 40$), experimental ($\alpha = .366$, $n = 40$), and overall sample ($\alpha = .33$, $n = 80$). These low reliability estimates can be attributed to the small sample sizes (by correlational research standards) for each group (Schmidt, Hunter, & Urry, 1976).
CHAPTER 4

Discussion

The results of the current study suggest that the CRT-A (James & McIntyre, 2000) is robust to violations of the CR measurement system’s assumption that test takers believe they are taking a test of reasoning at both the test and item level. This finding is a very important consideration especially as the CRT-A begins to gain popularity in the professional and public media and more job applicants are aware of its format and its measurement purpose. In addition, these results attest to the stability of the CR measurement system in general. Although these results are encouraging, there is still much work to be done in examining the measurement properties of the CRT-A and CR measures in general.

Also troubling are the low reliability estimates provided in the previous section of this paper. This brings into question whether this reliability estimate is merely an artifact of the present study or a legitimate concern for the CRT-A’s consistency of measurement. Given that James et al. (2005) found that reliability of the CRT-A was .76 using the same method for estimating reliability employed in the current study (coefficient alpha) but used a much larger sample size ($n = 1,603$), it is likely that the results found in the present study were a product of small sample size.

Although research has shown that CRT-A scores do not correlate with scores on the ACT, an examination of the relationship between cognitive ability and the number of invalid responses provided by a test taker might reveal possible measurement contamination in or limits to the use of the CRT-A. When James et al. (2005) examined the relationship between what they term *critical intellectual skills*, American College
Testing (ACT) scores were used as the measure of intellectual skills. However, the ACT is more of an academic achievement test and less of a measure of pure cognitive ability. It could be argued that the ACT reflects a respondent’s intelligence but it is more likely to reflect crystallized intelligence, which is dependent on learning. It would be more interesting to see the relationship between the CRT-A and measures of fluid intelligence (e.g., Raven’s Standard Progressive Matrices; Raven, Court, & Raven, 1993), which is independent of past experience (Cattell, 1987). This issue is especially important considering that five or more invalid responses results in an invalid test.

In addition, the application of more sophisticated statistical and psychometric techniques at the item and scale level have been shown to shed interesting results concerning the effects of faking on the measurement properties of tests, such as confirmatory factor analysis (see Schmit & Ryan, 1993), item response theory (IRT; see Zickar & Robie, 1999) and mixed IRT models such as the combination of latent class analysis and item response theory used by Zickar, Gibby, and Robie (2004). Finally, the development of new CR measures for a variety of personality constructs and the psychometric evaluation of those measures would provide greater insight into the utility of the CR framework of measurement established by James (1998).
References


Psychological Assessment Resources.


reactions, the likelihood of complaints, and organizational attractiveness. *Journal of Business and Psychology, 18*, 247-257.


APPENDIX A

Informed Consent Document
INFORMED CONSENT DOCUMENT

You consent to serve as a subject in the research investigation entitled: Study of a New Test for Job Applicants. The nature and general purpose of the study have been explained and the attached statement has been read to you by Nathan T. Carter from the Psychology Department.

“This study is meant to examine a new test for selecting job applicants. You will be asked to complete a 25-item test and a short demographics sheet. You will NOT be required to provide any identifying information on any of the forms. The researcher will NOT be able to identify participants. All test scores will be kept confidential. Questions about the study may be answered by using the contact information provided on your copy of this document.”

You understand the purpose of this research is to investigate a new test used for hiring job applicants, and that the research procedures involve completing a 25-item test as well as a demographics sheet.

The potential benefits for participants are the participation in psychological research and an opportunity to see how that research is conducted, as well as either filling the requirement for research participation in your course or gaining extra credit. There are no known risks to participants in this study.

You understand that your participation is voluntary and that all information is confidential and your identity will not be revealed; you are free to withdraw consent and to discontinue participation in the project at any time; any questions you may have about the project will be answered by the researcher named below or by the Human Protections Administrator for WKU, Phillip E. Myers, telephone 270-745-4652.

Western Kentucky University and the investigator named below have responsibility for ensuring that participants in research projects conducted under institutional auspices are safeguarded from injury or harm resulting from such participation.

_________________________________________  ________________________________
Participant's Signature                    Researcher's Signature

For further information regarding the HSRB, contact the Office of Sponsored Programs:

Director -- Dr. Phil Myers, Human Protections Administrator
(270) 745-4652; Fax (270) 745-4211; E-mail: phillip.myers@wku.edu

Compliance Manager — Mr. Sean Rubino
(270) 745-2129; Fax (270) 745-4211; E-mail: sean.rubino@wku.edu

For further information regarding the research you have participated in please contact either:

Nathan T. Carter – Principal Researcher
(270) 993-9887 ; E-mail : carternt@yahoo.com

Dr. Reagan Brown – Research Sponsor
(270) 745-6939 ; E-mail : reagan.brown@wku.edu
APPENDIX B

Procedural Script for Control Group Experimental Trials
PROCEDURAL SCRIPT (Control Condition)

Hello and thank you for participating in this study. My name is Nathan Carter. This study is meant to examine the characteristics of a new test for employee selection. Your full attention and cooperation will be needed for the study to produce useful results. You have been informed of your rights as a participant, agreed to participate, and are now ready to begin the study.

I will now pass out the tests, directions, and pencils. Please keep your tests face down until I tell you to start.

(Pass out the tests FACE DOWN, a pencil and the directions FACE UP on top of the tests)

I will now read the directions. Please do not open your test until I tell you to start. You may look at the direction sheet while I read aloud.

“For each question, identify the one answer that is the most logical based on the information presented. Sometimes this will require you to cut through answers that look logical to get the most genuine or ‘real’ answer. Circle your answers on the [Scan-tron sheet].” Be sure to mark heavily in the provided bubbles. If you make a mistake or need to change an answer, be sure to erase your mark COMPLETELY.

On the sheet there is an example. I will now read this example aloud.

(Read aloud from Example)

While taking this test it is important that you imagine that you are applying for a job that you wish to have and that this test is a part of the application process. THIS IS A TEST OF REASONING.

Thank you again for your participation. You have 25 minutes to answer all 25 questions. If you finish early, please turn your test over and sit quietly while the others finish. When everyone is done taking the test I will take them up individually and give you another very short sheet to fill out. I will then take those up and record your name so that you may receive research participation credit.

Are there any questions?

You may begin NOW.

(Start Timer) (**NOTIFY THEM WHEN THEY HAVE 5 MINUTES LEFT***)

Your time is up. Please set down your pencils and bring your tests to the front of the room. Lay your test face down on my desk.

Thank you for your participation in this study. Have a good day.
APPENDIX C

Procedural Script for Experimental Group Experimental Trials
PROCEDURAL SCRIPT (Experimental Condition)
Hello and thank you for participating in this study. My name is Nathan Carter. This study is meant to examine the characteristics of a PERSONALITY TEST OF AGGRESSION. Your full attention and cooperation will be needed for the study to produce useful results. You have been informed of your rights as a participant, agreed to participate, and are now ready to begin the study.

I will now pass out the tests, directions, and pencils. Please keep your tests face down until I tell you to start.

(Pass out the tests FACE DOWN, a #2 pencil and the directions FACE UP on top of the tests)

I will now read the directions. Please do not open your test until I tell you to start. You may look at the direction sheet while I read aloud.

“For each question, identify the one answer that is the most logical based on the information presented. Sometimes this will require you to cut through answers that look logical to get the most genuine or ‘real’ answer. Circle your answers on the [Scan-tron sheet].” Be sure to mark heavily in the provided bubbles. If you make a mistake or need to change an answer, be sure to erase your mark COMPLETELY.

On the sheet there is an example. I will now read this example aloud.

(Read aloud from Example)

While taking this test it is important that you imagine that you are applying for a job that you wish to have and that this test is a part of the application process. THIS IS A PERSONALITY TEST THAT MEASURES AGGRESSION.

Thank you again for your participation. You have 25 minutes to answer all 25 questions. If you finish early, please turn your test over and sit quietly while the others finish. When everyone is done taking the test I will take them up individually and give you another very short sheet to fill out. I will then take those up and record your name so that you may receive research participation credit.

Are there any questions?
You may begin NOW.

(Start Timer) (**NOTIFY THEM WHEN THEY HAVE 5 MINUTES LEFT***)

Your time is up. Please set down your pencils and bring your tests to the front of the room. Lay your test face down on my desk.

Thank you for your participation in this study. Have a good day.
APPENDIX D

Demographics Form
Thank you for participating in this study. Please pay close attention to each question provided and answer the question as honestly as possible by either placing a mark in the provided space. Some items require you to write your answer in the blank next to the questions. Please take care in filling out this form.

Age (in years): _____

Gender (Please Check One):
[ ] Male
[ ] Female

Race (Please Check One):
[ ] African-American
[ ] Caucasian (non-Hispanic)
[ ] Asian/Pacific Islander
[ ] Middle Eastern/West Asian
[ ] Native American
[ ] Other ___________________________
APPENDIX E

Instructions Provided to Participants
Instructions: For each question, identify the one answer that is the most logical based on the information presented. Sometimes this will require you to cut through answers that look logical to get to the most genuine or “real” answer. Bubble in your answers on the Scantron sheet provided. If you need to change an answer, be sure to erase completely.

Example
Feeling like he had finally recovered from the flu, Tom talked his wife into going out for dinner. They both ordered the flounder and fully enjoyed their meals. Later that evening, Tom developed an upset stomach.

Which of the following is the most logical explanation for Tom’s upset stomach?

a. The flounder was spoiled.
b. Tom had not fully recovered from the flu.
c. They sat in the no-smoking section of the restaurant.
d. His wife had cheesecake for dessert.

Explanation
Answer b is the most logical. Answers a, c, and d involve other people, but Tom was the only one who got sick.

You have 25 minutes to answer all 25 questions.