12-1-2013

Experimental study examining observational and objective methods of assessing effort in an undergraduate sample

Aaron Gabriel Schrader

UNIVERSITY OF NORTHERN COLORADO

Greeley, Colorado

The Graduate School

AN EXPERIMENTAL STUDY EXAMINING OBSERVATIONAL AND OBJECTIVE METHODS OF ASSESSING EFFORT IN AN UNDERGRADUATE SAMPLE

A Dissertation Submitted in Partial Fulfillment

of the Requirements for the Degree of

Doctor of Philosophy

Aaron Gabriel Schrader

College of Educational and Behavioral Sciences
Department of School Psychology

December, 2013

ABSTRACT

Schrader, Aaron Gabriel. *An Experimental Study Examining Observational and Objective Methods of Assessing Effort in an Undergraduate Sample*. Published Doctor of Philosophy dissertation, University of Northern Colorado, 2013.

Psychological assessment is predicated on the assumption that the subject is putting forth maximum effort during testing. When the individual fails to do so, it is known as response bias. The assessment of response bias is an area of research that has garnered attention in psychology because of the implications it has on psychological evaluation validity. This study examined the assessment of effort and response bias through two means, observationally and objectively, with confederates who were instructed to put forth their best effort or suboptimal effort. While being video recorded, confederates were administered a commonly used cognitive task along with two well established tests of effort. The videos were then viewed by graduate students in psychology and the perceived effort by the undergraduate confederates was then rated using a Likert-type scale, a dichotomous rating scale, and a single open-ended question. Similarly to previous research, participants had difficulty in accurately detecting suboptimal effort. More specifically, results illustrated that the ratings had a high level of misclassification of suboptimal effort but accurately identified all confederates putting forth adequate effort. Discriminant analysis of the results showed a marginal increase in accuracy in the detection of suboptimal effort in comparison to chance probability. These results are discussed within the context of psychological evaluations within the school

setting, with a focus on the importance of validity of the data ascertained within evaluations as well as current training programs.

ACKNOWLEDGEMENTS

I would like to express my sincerest gratitude to many people who made this dissertation possible due to their support and guidance. To my committee chair, Dr. Thomas Dunn, you have provided more support and guidance throughout this process than I could have ever imagined. I can confidently say that I have only made it to this point of my career because of you. Thank you for all that you have done for me through all of these years. To my committee members, Dr. Achilles Bardos, Dr. Robyn Hess, and Dr. Stephen Wright, thank you. You have all persevered with me throughout this process and allowed me to achieve one of my most significant life goals. For this, I thank you.

To my wife, Kristin Schrader, thank you. You have continuously pushed me to strive to be a better partner, professional, and friend. You have always believed in me in spite of my own doubts and have always provided me the love and support I needed. Thank you for loving me.

To my friend and research assistant, Niraj Patrawala, thank you. Without your always willing to lend a helping hand spirit, this dissertation may not have been possible. Thank you for volunteering your time and effort to support me in the pursuit of one of my dreams.

To my parents, family, and close friends, thank you. You have provided me with more than I could ever adequately thank you. It is because of the love that you have

provided as well as the values and morals that you all have instilled in me that I am here at this point of my life with a dream achieved. Thank you!

TABLE OF CONTENTS

CHAPTER		
I.	INTRODUCTION ……...	1
	Methods of Assessing Effort Best Practices in School Psychology Purpose of the Study Research Questions Limitations Delimitations Definitions	
II.	REVIEW OF LITERATURE …..……...	19
	Response Bias Methods of Assessing Response Bias Literature in School Psychology	
III.	METHODS	34
	Participants Procedure Variables Instrumentation Research Hypotheses Data analysis	
IV.	RESULTS	47
	Phase Two Sample Description Accuracy Rates Fisher's Exact Test Discriminant Analysis Open-Ended Question Analysis Summary of the Statistical Findings	

V. DISCUSSION 61

Findings
Implications
Limitations
Future Directions
Conclusions

REFERENCES 72

APPENDIX A – INSTITUTIONAL REVIEW BOARD APPROVAL 85

APPENDIX B – INFORMED CONSENT FOR CONFEDERATES..... 87

APPENDIX C – INFORMED CONSENT FOR PARTICIPANTS 89

APPENDIX D – PARTICIPANT DEMOGRAPHIC INFORMATION FORM 91

APPENDIX E – OBSERVATIONAL RATING FORM 93

APPENDIX F – PHASE ONE STATISTICS 95

APPENDIX G – SUMMARY OF STUDY IN ARTICLE FORMAT 98

LIST OF TABLES

Table 1. Demographics for the Total Sample 48

Table 2. Accuracy of Dichotomous Ratings 49

Table 3. Contingency Table for Full Effort Condition 50

Table 4. Contingency Table for Suboptimal Effort Condition 51

Table 5. Cognitive Training Group Classification Summary using Linear Discriminant Function 54

Table 6. Cognitive Training Group Classification Summary using Quadratic Discriminant Function 54

Table 7. Cognitive Training Group Error Count Estimates using Linear and Quadratic Discriminant Functions 55

Table 8. Without Training Group Classification Summary using Linear Discriminant Function 56

Table 9. Without Training Group Classification Summary using Quadratic Discriminant Function 56

Table 10. Without Training Group Error Count Estimates using Linear and Quadratic Discriminant Function 57

Table 11. Demographics for Phase One Confederates 96

Table 12. Descriptive Statistics for Phase One Performance 97

CHAPTER I

INTRODUCTION

In a general sense, the purpose of psychological assessment is to quantify behavior (Kaplan & Saccuzzo, 2001). A quantitative description of behavior allows for comparison between groups through statistical means. Results from psychological assessment allow clinicians to screen individuals for potential acceptance into specific programs, classification, counseling, diagnosis, treatment, and determination of intervention efficacy (Aiken, 1994). Therefore, the psychometric properties of validity and reliability of the psychological tools utilized within the decision-making process are critical. Most professionals can agree on the obvious importance of psychometrically sound tools, however some view clinical judgment as an equally important strategy for assessing human behavior.

A debate in the field of psychology has arisen with respect to the systematic use of clinical judgment in comparison to empirical evidence to inform decision-making (Zeldow, 2009). Many professionals argue that clinical judgment alone is insufficient and prone to error and bias, among other confounds (Haynes & O'Brien, 2000). This is important to consider, especially since clinical judgment is often based upon internal mental processes that can be difficult to explicitly specify, whereas empirical or actuarial approaches are explicit in nature (Dawes, Faust, & Meehl, 1989). Although Dawes et al. (1989) state that a clinician may be more adept at recognizing facial expressions, actuarial methods provide a means for a more consistent and reliable predictions. These

researchers also noted that having the ability to observe in a unique manner does not equate to having the ability to predict in a unique manner based on the observations (Dawes et al., 1989). Humans cannot consistently apply optimal weights to variables which invariably make them more prone to error when compared to explicit equations (Grove & Meehl, 1996). The predisposition clinical judgment has towards error and bias makes the incorporation of empirical results pragmatically crucial and essential to best account for potential confounds.

The same argument regarding validity and reliability can also be applied to an individual's performance on standardized psychological tests. An individual's performance can be biased either knowingly or unknowingly (Lezak, Howieson, & Loring, 2004). This form of bias can make performance appear either higher or lower depending on the type of measure being administered. For example, scores on measures that require individuals to perform based on ability can appear lower when the individual provides a biased response style. Sattler (2008) described this type of bias when he stated that poor performance on cognitive measures may be accounted for by poor effort or motivation, among many other variables. On the other hand, measures utilizing self-report of behaviors or personality traits can be biased in response to make the results appear to be either higher or lower (McGrath, Mitchell, Kim, & Hough, 2010). Overall, these types of bias, which can affect scores either positively or negatively, are typically referred to within the literature as "response bias" (Bush et al., 2005; Heilbronner et al., 2009; McGrath et al., 2010). Response bias has been researched in its many forms because it can invalidate assessment results. Even more concerning, is that if psychologists are not aware of the potential for response bias during evaluations, there is

a potential to discredit clinicians. Consequences can be far-reaching, including inaccurate diagnosis and misappropriation of funds.

Effort is one component to closely examine in relation to individual performance and response bias. Similarly, effort can be influenced by any number of internal or external factors, which can include, but is not limited to: fatigue, pain, potential secondary gain, or disinterest in the task being presented (Hunt, Ferrara, Miller, & Macciocchi, 2007; Sattler, 2001). These factors may cause the amount of effort an individual exerts to increase or decrease. Ultimately, psychologists want individuals to put forth their best effort on a test so that the results reflect the "best estimate" of their ability, performance, or whatever construct is being measured. However, when the level of effort is influenced, usually in the negative direction, the individual is considered to have responded in a biased manner. As noted, this phenomenon makes accounting for response bias paramount because it can invalidate results and lead to misinformed decisions on the part of the professional.

Response bias is viewed as an umbrella term that encapsulates many different constructs, depending on the task at hand. It can include, but is not limited to: negative impression management, positive impression management, suboptimal effort, feigning, and malingering (Dunn, 2006; McGrath et al., 2010). Some researchers have adopted a viewpoint of response bias as being a continuum rather than simply an umbrella term. This continuum ranges from non-purposeful response bias to purposeful response bias, which would be associated with malingering (Dunn, 2006). Malingering is viewed as being a more sinister version of response bias, as it involves the intentional faking or

exaggerating of symptoms in order to gain access to an external incentive (Heubrock & Petermann, 1998).

Malingering is closely associated with the construct of effort because it can be viewed as exerting suboptimal effort purposefully within some specific performance tasks, to appear as though a loss of skills has occurred. However, it should be noted that on other psychological measures, the converse may also be true. That is, an individual who is actively malingering on some measures may be putting forth significant effort in "faking bad." It is important to keep this differentiation in mind when specifically considering performance on a designated measure.

Research on malingering, the most severe form of response bias, has illustrated that it is a commonly encountered phenomenon. According to survey data from 131 members of the American Board of Clinical Neuropsychology (ABCN), it was found that 29% of personal injury, 30% of disability, 19% of criminal, and 8% of medical cases involved probable malingering (Mittenberg, Patton, Canyock & Condit, 2002). These data estimates clearly illustrate a need for appropriate techniques for detecting suboptimal effort and inclusion of such techniques in training programs for applied psychology.

Both fields of clinical and school psychology rely heavily on assessment to inform decisions by identifying concerns, diagnosing disorders, and determining the effectiveness of interventions. Within the scope of school psychology, assessment is a significant portion of job responsibilities, as it directly pertains to particular educational diagnostic and associated exclusionary criteria (Kamphaus, Reynolds, & Imperato-McCammon, 1999). Among other unique traits, school psychologists hold an exclusive position within the educational setting because of their specialized training in the

administration and interpretation of different psychological assessments. The evaluations that school psychologists complete yield important data used within the special education identification process. According to a national survey from members of the National Association of School Psychologists (NASP), school psychologists completed, on average, 19 to 26 hours of assessment a week depending on their region in the United States (Hosp & Reschly, 2002). Although there is a current change in special education evaluation processes that places less emphasis on comprehensive assessment batteries, some school psychologists perceive the need for continued use of cognitive assessments to rule out intellectual disabilities (Machek & Nelson, 2010). Research related to the evaluation of response bias has not yet been conducted in relation to assessment practices within either the traditional discrepancy model or the more recent Response to Intervention model, although similar constructs have gained attention.

Understanding that assessment remains a dominant practice in school psychology, one must consider ethical obligations of appropriate evaluation techniques. According to the American Psychological Association (APA) ethical standards 9.01 and 9.02 (APA, 2010) and NASP ethical standard II.3.2 (NASP, 2010), psychologists are expected to use appropriate and responsible assessment procedures (Jacob & Hartshorne, 2007). Factors of whether the test subject may not be responding to test items in an authentic manner must be considered in an appropriate and responsible manner, just as a clinician would gather data on other aspects of cognitive or emotional functioning.

Although some might believe malingering to be specific to adult populations, research has also illustrated that children and adolescents are adept at feigning commonly encountered disorders (Faust, Hart, & Guilmette, 1988b; Faust, Hart, Guilmette, & Arkes,

1988c). More current research continues to examine the phenomena of pediatric malingering, as well as other forms of response bias within psychological evaluations. In a study which examined 193 pediatric patients who had sustained mild traumatic brain injuries, it was discovered that suboptimal effort had a base rate of 17% (Kirkwood & Kirk, 2010). Another study by Lu and Boone (2002) illustrated that children as young as age 9 were capable of feigning cognitive deficits. Because of this potential to exhibit response bias, many efforts have also been made to provide evidence for valid use of established tests of effort as well as embedded measures with pediatric populations (Blaskewitz, Merten, & Kathmann, 2008; Donders, 2005; Brooks, Sherman, & Krol, 2012; Gunn, Batchelor, & Jones, 2010; Kirkwood, Hargrave, & Kirk, 2011). It is important to keep in mind, however, that children who do not put forth maximum effort may be exhibiting response bias due to less sinister issues such as fatigue or disinterest. Nonetheless, establishing an efficient and accurate means of evaluating effort within pediatric psychological evaluations continues to be an important area of focus.

Methods of Assessing Effort

Effort that is exerted by the individual during a psychological evaluation can be assessed through many different methods. Specifically, within the fields of clinical, forensic, and neuropsychology, multiple methods of evaluating effort have been created and are used in practice regularly. Mittenberg et al. (2002) as well as Sharland and Gfeller (2007), reviewed the various methods used within psychological practice; they identified a number of different approaches to support diagnostic impression of response bias. These methods included severity of impairment which is inconsistent with condition, inconsistent performance on cognitive tests, performance scores below

empirical cutoffs on forced-choice tests, discrepancies among reported and observed behavior, unexplainable reported symptoms in an interview, performance scores below empirical cutoffs on other malingering tests, unexplainable changes in performance across multiple examinations, and performance scores above validity scale cutoffs on personality tests (Mittenberg et al., 2002; Sharland & Gfeller, 2007). The final methods noted varied slightly, as one simply identified below chance scores on forced-choice tests, while the other included scores below cutoff points on embedded measures of effort and scores on empirically derived discriminant function analyses. The most salient of methods are discussed below in greater depth.

The first of these methods, severity of impairment that is inconsistent with condition, is determined when psychologists focus on cognitive functioning in relation to observed and expected scores (Iverson & Binder, 2000). Psychologists determine whether further investigation should be conducted based of their knowledge related to premorbid functioning, injury severity characteristics, and neuropsychological sequelae (Iverson & Binder, 2000). Studies have shown that specific to cognitive measures, patterns exist between actual and predicted intelligence quotients on tests of adult cognitive ability (Demakis et al., 2001). Predicted intelligence quotients in a study by Demakis et al. (2001) were based on regression equations from the Barona Index, the Best-3 method, and the Oklahoma Premorbid Intelligence Estimation. The patterns of discrepancies that emerged allowed the researchers to discriminate between insufficient effort and traumatic brain injury (Demakis et al., 2001). Similarly, psychologists can examine inconsistent performance across a battery relative to condition. This method is different from the previously described one in that it examines inconsistency across multiple tests in a

battery. Psychologists utilize this method by examining patterns of test performance on standard cognitive and neuropsychological measures that do not make sense neuropsychologically (Iverson & Binder, 2000). For example, if an individual performed well on spontaneous and cued recall on both immediate and delayed trials on a verbal memory task, it would be abnormal and unusual to obtain a low score on the delayed recognition portion of the same task, as it would be based on the previously well-learned information. These patterns have been identified through multiple studies which applied discriminant analysis procedures to test battery scores derived from performance of healthy individuals told to simulate impairment (Larrabee, 2007).

The second method of assessing response bias is the use of objective tests which utilize clever test designs to assess an individual's effort during a given task, rather than performance or ability levels (Blaskewitz, Merten, & Brockhaus, 2009). These tests provide a unique method of assessing response bias through objective means, which removes any potential observational limitations. Furthermore, these tests have been standardized and researched with diverse populations allowing for clinicians to draw comparisons and conclusions based on the individual's performance rather than the clinician's perception (Rees, Tombaugh, Gansler, & Moczynski, 1998; Green, Iverson, & Allen, 1999; Green & Flaro, 2003; MacAllister, Nakhutina, Bender, Karantzoulis, & Carlson, 2009). Generally speaking, these tests are designed to present tasks that appear difficult but in reality are easy even to individuals who have sustained brain damage. Everyone would be expected to do well on them, but because they appear more difficult, an individual who was not putting forth sufficient effort or attempting to deceive the

examiner might not attempt to complete the presented tasks. These tests are referred to as tests of malingering are also considered to be tests of effort.

Two categories of effort tests have been created and are referred to as forced-choice tests and non-forced-choice tests. Forced-choice tests utilize the binomial distribution to detect intentional poor performance. This can be determined through the use of chance probabilities to calculate whether an individual's performance falls below that of chance. If a person scores below chance, it can be determined that the individual intentionally performed poorly. Additionally, research has established empirically supported cutoff points to increase sensitivity and specificity due to the fact that patients involved with litigation often do not score below chance (Larrabee, 2007). The significance of these empirically supported cutoff points lie in their ability to improve detection rates beyond what is provided by the method of detection using only below chance performance.

Forced-choice tests are designed to provide test takers with two stimuli in which they are to choose the correct answer to which they have been previously exposed. Commercially available forced-choice measures include such measures as the Test of Memory Malingering (TOMM: Tombaugh, 1996) and the Word Memory Test (WMT: Green, 2005). The TOMM and WMT are two well-researched forced-choice measures of effort that have demonstrated validity and are also highly sensitive and specific in populations that have sustained various neurological traumas (Rees et al., 1998; Greiffenstein, Greve, Bianchini, & Baker, 2008; Bauer, O'Bryant, Lynch, McCaffrey, & Fisher; 2007). These measures have also been researched with diverse populations,

including children and individuals with various neurological conditions (Blaskewitz et al., 2008; Green, Flaro, & Courtney, 2009; MacAllister et al., 2009).

Non-forced-choice tests are another means of objectively assessing response bias. These tests are designed to be brief in nature, while forced-choice tests typically require additional administration time. A distinct benefit non-forced-choice tests is their brief nature. These non-forced-choice tests are designed to be independent of each other which allows for inclusion of multiple non-forced-choice tests throughout an evaluation. By providing multiple test results of response bias throughout an evaluation, psychologists can be confident of sustained levels of effort. Non-forced-choice tests include the Rey 15-Item Test, Dot Counting Test, b Test, and the Rey Word Recognition Test (Larrabee, 2007). However, these brief tests are limited in that they are relatively insensitive and nonspecific as compared to forced-choice tests (Whitney, Hook, Steiner, Shepard, & Callaway, 2008).

The third method of assessing for response bias involves noting any inconsistencies between reported and observed behaviors during the evaluation. Through this approach, a psychologist remains aware during the evaluation and observes any inconsistencies between presentation throughout an evaluation and impairments reported by the individual. For example, Iverson and Binder (2000) provided the example of an individual demonstrating profound difficulties with word finding on a Boston Naming Test but observing no difficulty with casual conversation during the interview.

The fourth method of evaluating response bias to be discussed is the use of validity scales on personality tests. Several personality tests have developed and integrated validity scales which can help provide indicators of response bias. Such tests

include the Minnesota Multiphasic Personality Inventory – Second Edition (Butcher et al., 2001), the Millon Clinical Multiaxial Inventory – Third Edition (Millon, Millon, Davis, & Grossman, 2009), and the Personality Assessment Inventory (Morey, 2007).

Best Practices in Assessment

As mentioned earlier, response bias and specifically effort are critically important factors when considering the validity and reliability of psychological assessment results. However, there is a notable lack of formal training or guidelines related to best practices of evaluating for response bias and effort during an evaluation. Classic studies from the field of neuropsychology have examined clinicians' ability to detect malingering or "faking bad" through informal means. The clinician participants were provided a brief background summary of the individual as well as the cognitive and neuropsychological test results. They were then asked to review the case and determine whether the "client" was malingering. Unfortunately, no clinicians were able to detect malingering beyond chance probability (Faust et al., 1988a, 1988b).

Research studies examining clinician's ability to detect malingering or suboptimal effort have come from the fields of clinical, forensic, and neuropsychology. There appears to be an absence of this type of work within the field of school psychology. This oversight is unfortunate as there is evidence to suggest even children engage in response bias. Alternatively, in the field of neuropsychology, consensus on the evaluation and diagnosis of malingering and response bias has been reached and a set of guidelines has been put forth by the American Academy of Clinical Neuropsychologists (Heilbronner et al., 2009). However, there are no formal best practice or professional guidelines designed with the school setting in mind, nor is this type of information typically included within

graduate training programs. Without any specific guidelines or training, school psychologists are then left to attempt to assess effort through unproven means such as observational methods. These discussion points provide a clear impetus for additional research and a call for attention to the assessment of response bias whenever a comprehensive evaluation is conducted with an individual, including in the field of school psychology. As a starting point, a study of the unproven means of assessing effort (i.e., observational methods) which school psychologists rely upon need to be examined.

Purpose of the Study

The purpose of this study was to examine the ability of observers, that is, graduate students in an applied field of psychology or education, to assess and evaluate levels of effort exerted by confederates during a task through informal observation, and then compare their ratings to standardized measures of response bias completed by confederates. Specifically, psychology or education graduate students with or without training in assessment were used in the study. These participants judged the effort put forth by undergraduate confederates on a common cognitive task who had been instructed to either put forth their best effort or exert suboptimal effort.

Although previous research has studied the ability of clinicians to detect suboptimal effort or malingering, this study contributes to the literature in a unique way. The selected research design represented a novel approach to studying this issue. Additionally, no studies have also compared those values to results from other graduate students' ratings as well as to results from two standardized tests of effort. Another unique aspect of the current study is that no other studies to date have evaluated the ability to detect suboptimal effort by school psychology graduate students from both

specialist and doctoral level programs. Including school psychology graduate students is this type of research is important because many of the common diagnoses encountered by school psychologists, such as learning disabilities and attention-deficit/hyperactivity disorder, can also be easily feigned (Sullivan, May, & Galbally, 2007).

Research of this nature has important implications related to everyday practice of school psychology, as well as training programs in school psychology. Although the behaviors of undergraduate students were rated in the study, implications for practice within secondary education settings exist, such as obtaining testing accommodations for high stakes tests (e.g., SAT) or receiving post-school transitional services. Additionally, this study sought to provide additional data regarding observational assessment of effort in comparison to objective standardized assessment data. This study also evaluated current practice and training in the assessment of effort within psychological evaluations by graduate students in school psychology. The results of this study contribute to the literature of school psychology by providing a basis for additional research to be conducted specific to the evaluation of effort within psychological evaluations in the school setting.

Research Questions

The statistical procedure referred to as discriminant analysis allows researchers to use a number of variables for example, observations, to design a model which classifies individuals into one of two categories. Within the present study, the dependent variable referred to the full or suboptimal effort group membership of the undergraduate confederates. The independent variables included observational ratings provided by graduate students with training in assessment, an observational rating by graduate

students without training in assessment, and the undergraduate participant scores on the TOMM and WMT. Discriminant analysis represented the most appropriate means to address these questions as individuals were classified as either putting forth their best effort or exerting suboptimal effort a priori. The group membership (i.e., condition) allowed for a model to be created and evaluated using the categorical dependent variable (condition) and the independent variable (observational ratings). Additionally, the use of the jackknife method within discriminant analysis allowed for a systematic evaluation of classification models to occur. This method withheld an independent variable from the model and attempted to classify individuals based on the model created by the remaining independent variables. This process was repeated with each independent variable. The final result was an average of all classification models created to form a final model. The following research questions were researched using these methods:

Q1 Are observations by participants able to discriminate between full effort and suboptimal effort group membership conditions?

Q2 Does the addition of the observational rating scale improve the differential prediction of full effort or suboptimal effort group membership by participants?

Fisher's exact test allowed researchers to examine if there were significant differences between two groups of graduate student participants (i.e. those with training in assessment or those without training in assessment). Ratings on the observational rating scales were used as the basis for this comparison. These differences were examined as simulated response bias research has not typically incorporated participants with no training in assessment.

Q3 Does training in cognitive assessment aid in the ability to discriminate between groups using observational methods?

Lastly, using descriptive analysis, researchers were able to identify common categories or trends among responses. Within the current study, participants were asked to identify the salient features which informed their observational ratings in an open format. These categories were created through the use of information from an open ended question that was coded and analyzed for frequency among participants, a common technique endorsed by Merriam (1998). Significant responses were noted and made into a list where larger units of information were found and grouped into similar categories (Creswell, 2007). Using data from this question, the following research question was examined:

Q4 What salient features (for example, behaviors) informed the observational rating of effort by participants?

Limitations of the Study

Within the present study, there were limitations particularly related to the generalization of results. Firstly, graduate students who participated in this study were a part of a convenience sample from one university located in the western United States. These graduate students were enrolled in programs in school psychology, counseling psychology, counselor education and supervision, and special education programs. This convenience sampling method limits the extent to which results can be generalized.

Another limitation of the present study was the use of undergraduate students to simulate suboptimal conditions. A single exclusionary criterion was used in the selection process and required participants to have no prior history of concussion with loss of consciousness. The students received very limited coaching in how to exert suboptimal effort and as a result may not have followed directions appropriately throughout the entirety of the study. Conversely, students that were told to exert complete or optimal

effort may not have followed directions appropriately throughout the entirety of the study.

The final limitation of the present study also pertains to the use of undergraduate students in the generation of stimuli. Undergraduate students were used as confederates rather than children which limits the generalizability of the results. Although the inclusion of this age population in the study is applicable to school psychology as well as other professional psychology training programs, the lack of pediatric representation as confederates provides a limit to the generalizations that can be made in the observational ratings of children.

Delimitations of the Study

The TOMM and WMT were selected for this study based on the literature available which validates the use of these tests to measure effort. Both the TOMM and the WMT are sensitive and specific measures of effort commonly used in the detection of suboptimal effort (Rees et al., 1998; Gunn et al., 2010). Additionally, these tests utilize relatively easy recognition tasks which should not discriminate across cognitive ability levels within the subject population.

Undergraduate college students from a select part of the United States were utilized within the study and were not representative of the diversity within the United States population. Utilization of this subject population limits the ability to generalize the findings from the study.

A measure of perceived exertion was utilized within this study due to its Likert-type scale design which allowed for the psychology graduate students to quickly rate the perceived effort by individuals completing the cognitive task. Through this measure, the

researcher was able to compare the psychology graduate students subjective ratings based on informal observational methods of effort evaluation with objective quantitative findings from the TOMM and WMT.

Definitions

Effort: The level of exertion an individual utilizes during a given task. Effort can be described as mental, physical, or emotional exertion an individual puts forth during testing.

Maximum Effort: The level of exertion an individual utilizes during a given task in order to best complete the given task according to their relative ability.

Perceived Exertion: The level of effort an individual utilizes during a given task as perceived by the subject who engaged in the task. Perceived exertion is based on self-report using a Likert-type scale.

Response Bias: The phenomenon of inaccurate responding due to influence from other internal or external factors such as financial incentives, loss of interest in task, fatigue, and so forth. An example of response bias is malingering where an individual purposefully feigns a deficit in order for secondary gain.

Subjective Rating of Effort: The end result of a decision making process where a professional or clinician engages in a series of hypothesis testing based on descriptive data observed in a given task or situation. Subjective ratings are purely based on descriptive data which can include observations, conversations, or intuition.

Suboptimal Effort: The level of exertion utilized during a given task that is incomplete, meaning that an individual performed in a manner, whether purposeful or non-

purposeful, that is not a representative of his or her true ability. Suboptimal effort is also described in the literature as noncredible effort (Kirkwood et al., 2011), incomplete effort (Baker, Donders, & Thompson, 2000), or inadequate effort (Barrash, Suhr, & Manzel, 2004).

Confederates: The undergraduate students that were used within the current study to create the stimuli for the experiment. The stimuli were video recordings of the confederates completing a battery of tests, which were later rated by participants.

CHAPTER II

REVIEW OF LITERATURE

One of the most important tasks of an applied psychologist is clinical decision making based on data typically derived from assessment of individuals. In order to make the best decisions, both reliable and valid data are essential. This chapter reviews the research related to the evaluation of response bias in the field of applied psychology such as clinical neuropsychology, clinical psychology, and forensic psychology. However, for the purposes of this study, only literature that directly pertains to subject matter relevant to school psychology will be included. Finally, this chapter will expound on the literature and research within the field of school psychology that is related to effort and motivation.

Response Bias

Within the fields of applied psychology, a focus on response bias emerged beginning in the late 1980s (Morgan & Sweet, 2009). Response bias is a general or umbrella term that is multifarious and includes internal factors that affect or influence the way an individual responds to a given task or situation. One form of response bias, known as malingering, has gained significant attention due to the degree to which it has been involved in litigation. In a meta-analysis of 18 study groups and 2,353 subjects conducted by Binder and Rohling (1996), they computed prediction values indicating that financial incentives were associated with greater pathology. Their findings illustrated the need for inclusion of measures of motivation within psychiatric exams. However, other

forms of response bias exist even when there is no perceived financial gain. In fact, response bias has been theorized by some researchers to exist on a spectrum or continuum which ranges from non-purposeful bias such as suboptimal effort to purposeful exaggeration or feigning in order to gain access to external incentives (Rogers, 1984).

Another perspective offered by McGrath et al. (2010) stated that within measures of performance, two primary forms of response bias can exist depending on the type of measure. It was noted that suboptimal effort exerted during neuropsychological measures is related to malingering while suboptimal effort on psychopathological measures (e.g., MMPI-2) is related to positive impression management (McGrath et al., 2010). In other words, exertion of poor effort on measures of innate ability results in depressed performance illustrating deficits whereas poor effort on measure of psychopathology or deviance results in depressed ratings illustrating better adjustment or fewer issues. Within the current study, the specific focus is given response bias rather than outright malingering. However, given that both malingering and suboptimal effort are similar constructs, select studies on malingering are reviewed since both are evaluated in the same manner.

To some degree, an individual's response to any given task or situation is affected by response bias. However, an examination of poor performance on tests is essential in order to determine if a deficit is truly present or if a form of response bias better accounts for this performance. When an individual demonstrates low performance on a task, there are any number of reasons why this might occur. For example, Sattler (2008) noted,

> "… when an examinee fails a test that requires repeating digits, the failure may be caused by difficulty in sequencing; limited short-term memory; lack of

> attention, motivation, auditory acuity, understanding of task demands, effort, or strategy usage; or peer-group pressure to do poorly" (p. 45).

Therefore, it may be necessary to administer additional specialized measures in order identify specific difficulties (Sattler, 2008). Furthermore, Bush et al. (2005) suggested that in order to ensure the validity of psychological measures as well as associated diagnoses and recommendations, one must assess the response validity of an individual's performance. These statements clearly illustrate the need for additional specific and sensitive measures in order to evaluate each component of an individual's performance, one of which includes effort.

The evaluation of effort is complicated and has many different aspects to consider. An individual may knowingly or unknowingly provide distorted or erroneous responses which are atypical when compared to that person's actual neuropsychological abilities (Lezak et al., 2004). Similarly, effort and honesty may vary within different points of an evaluation (Bush et al., 2005). A study which illustrates the tremendous effect that effort can have on performance was conducted by Green, Rohling, Lees-Haley, and Allen III (2001). In this study, a sample of 904 patients with neurological conditions, miscellaneous medical conditions, or psychiatric patients were administered a battery of neuropsychological measures. Results illustrated patients who demonstrated poor effort suppressed their overall mean score 4.5 times more than those with moderate to severe brain injury (Green et al., 2001). Brain injuries were categorized within this study using Glasgow Coma Scale scores, duration of post-traumatic amnesia, and loss of consciousness duration. This study also measured effort using two symptom validity tests, including the Word Memory Test, in order to differentiate between individuals exerting poor or adequate effort (Green et al., 2001). In a more recent study, Ord, Greve,

Bianchini, and Aguerrevere (2009) also found that performance on a neuropsychological test of executive functions was more significantly affected by effort during testing than severity of brain injury in their sample of 175 participants. Although no control group was used within this study, multiple exclusionary criteria were used to increase homogeneity of the sample and limit extraneous variables that could account for their performance. Kirkwood, Yeates, Randolph, and Kirk (2012) examined correlations between performance on symptom validity tests and ability-based tests. They found that performance on a symptom validity test explained 38% of the variance on the ability-based tests (Kirkwood et al., 2012).

Given that effort plays such an important role in test scores, researchers have also examined the prevalence of suboptimal effort within cognitive tasks. A related line of research has focused on how well as neuropsychologists' ability to recognize and detect abnormalities in performance. For example, Faust et al. (1988a) disseminated neuropsychological testing data to 600 psychologists. Of the respondents, approximately 80% were able to identify normal versus abnormal profiles. However, approximately 80% of the psychologist participants also made mistakes with their first hypothesis related to etiology (Faust et al., 1988a). It should be noted that although only minimal data were provided to participants, two control cases of typical, healthy individuals were included. The findings from this study suggest that diagnosis is prone to error and appropriate efforts should be made to increase the validity and sensitivity of diagnostic conclusions.

In addition to improving their own diagnostic skills, clinicians must also be able to identify deliberate attempts on the part of clients to present themselves in suboptimal

ways. Based on a survey of 131 members of the American Board of Clinical Neuropsychology (ABCN), Mittenberg et al. (2002) attempted to calculate estimated base rates of malingering and symptom exaggeration from a survey conducted. It was determined that adjusted base rates of probable malingering according to referral type occurred on 30.43% of personal injury cases, 32.73% of disability or worker's comp, 22.78% of criminal cases, and 8.11% of medical or psychiatric referrals (Mittenberg et al., 2002). No significant differences were found in relation to practice setting or geographic region. In addition to this finding, research from a survey of 188 National Academy of Neuropsychology (NAN) members found that these practitioners estimated that about 10% of patients probably gave suboptimal effort while 5% of patients definitely gave suboptimal effort (Sharland & Gfeller, 2007). These overall findings illustrate that malingering is a relatively common phenomenon faced by clinicians.

Methods of Assessing Response Bias

One of the earliest methods of assessing response bias is referred to as "symptom validity testing" and was first developed in the 1970s using the binomial theorem (Morgan & Sweet, 2009). These symptom validity measures presented individuals with a forced-choice measure with two choices with one response being correct. Overall, these symptom validity tests do not measure ability levels but rather effort that is put forth (Blaskewitz et al., 2009). The Test of Memory Malingering, the Word Memory Test, and the Validity Indicator Profile (Frederick, 1997) are examples of these types of symptom validity measures.

Another method of assessing response bias includes tests that utilize ceiling effects to enhance specificity and sensitivity in detection. These tests were designed to

appear difficult to unaware individuals but are in fact simple even for individuals who have sustained brain damage (Morgan & Sweet, 2009). This unique test design utilizing ceiling effects has provided the basis for current practices in malingering/symptom validity.

Research has shown that certain statistical formulas based on information collected from cognitive measures can provide another method of evaluation. For instance, within a study using the Wechsler Memory Scales – Third Edition (WMS-III), it was determined that an analysis of discrepant scores yielded a promising indicator of poor effort (Lange, Iverson, Sullivan, & Anderson, 2006). The groups within this study were formed using assessment data from previous neuropsychological evaluations which included tests of effort. Individuals who put forth adequate effort as evidenced by their results on tests of effort formed the comparison group. These individuals in the comparison group cannot be considered controls as they sustained head injuries of varying severity and were personal injury litigants (Lange et al., 2006). In a similar study using the Wechsler Adult Intelligence Scales – Third Edition (WAIS-III), a statistical equation incorporating the digit symbol age-corrected scaled scores plus accuracy and time scores from an experimental timed recognition trial added to the WAIS-III demonstrated 80% sensitivity and 88.7% specificity in the indication of response bias (Kim et al., 2010). This study included 82 participants who met criteria for probable malingered neurocognitive dysfunction according to multiple effort indicators and 89 participants who put forth adequate effort and were not seeking disability benefits to comprise the comparison group (Kim et al., 2010).

Current practices of malingering detection include analysis of consistency of information, performance on psychological and neurocognitive tests, symptom validity tests, and forced-choice tests (Bush et al., 2005; Heilbronner et al., 2009). In the previously mentioned survey study of NAN members (Sharland & Gfeller, 2007), participants were also asked to describe their practices related to assessment of effort. Results indicated that 57% of neuropsychologists reported using measures of effort within their evaluations and that the most frequently used measures of effort and response bias by the participants included the Test of Memory Malingering (TOMM), Minnesota Multiphasic Personality Inventory – Second Edition (MMPI-2) Frequency - Defensiveness ratio, MMPI-2 Fake Bad Scale, Rey 15-item test, and the California Verbal Learning Test (CVLT) (Sharland & Gfeller, 2007). Additionally, the neuropsychologist participants rated the most accurate measures of effort to include the Validity Indicator Profile (VIP), Word Memory Test (WMT), Victoria Symptom Validity Test, and Computerized Assessment of Response Bias (Sharland & Gfeller, 2007). Finally, it should be noted that 52% of neuropsychologists reported that they never or rarely provided warning that effort indicators would be used with their clients.

Many methods for evaluation response bias have been developed, validated, and are routinely used within some fields of psychology, however, other fields of applied psychology have not yet fully utilized these methods to their full potential. One such field is school psychology. The following section presents an overview of the application of similar methods for evaluating bias within the field of school psychology.

School-based Assessment

On a daily basis, school psychologists within educational settings must use assessment data to make decisions about test accommodations and qualification for special education. Additionally, school psychologists also provide assessment data which is used to make decisions regarding other types of support programs such as vocational rehabilitation. Because of this, it is important to consider the ways that response bias might affect student performance in the school setting. Unfortunately, in the field of school psychology there seems to be a lack of awareness as well as research pertaining to the assessment of effort and response bias during school based evaluations for special education services. Furthermore, there is a lack of awareness in regards to the effects response bias might have on different aspects of a child's performance. Educational research efforts have tended to focus on other forms of effort, such as achievement motivation as related to academic success and school completion.

Test taking behaviors have been researched in relation to low stakes tests in education (Abdelfattah, 2010). Abdelfattah (2010) defined low stakes tests as large scale national examinations used in all levels of education that have little to no direct consequence for the individual student but rather have direct effects on school personnel or other stakeholders. In this study, 797 participants completed a self-report scale of motivation following completion of an achievement test. Results from this study illustrated that individuals who reported high levels of motivation performed better on achievement tests regardless of content (Abdelfattah, 2010). Similarly, self-report ratings have been used as a measure of test taking effort in order to examine the effects of motivation on test performance (Wolf & Smith, 1995). In a study of 158 college students,

participants' ratings of self-reported motivation indicated that individuals behaved differently according to the testing situation as well as whether the test had a direct consequence to an examinee (Wolf & Smith, 1995). Participants were randomly assigned to testing conditions and dependent measures were presented in a counterbalanced manner to prevent ordering effects. Results from the same study also illustrated that performance levels were nearly the same whether the participant was in a high motivation and high anxiety or in a low motivation and low anxiety condition (Wolf & Smith, 1995). The findings of this large scale study suggest the important negative effects that low motivation as well as heightened levels of anxiety can have on test performance.

Several studies from adjacent fields of applied psychology provide a rationale for further study of response bias in the field of school psychology. One such study examined the performance of 66 college students on attention-deficit/hyperactivity disorder (ADHD) and learning disorder (LD) assessments that were being evaluated in order to receive services while in college. Additional services while in college can be viewed as secondary gain which allowed researchers to administer participants the Word Memory Test in order to assess response bias. Based on their findings, Sullivan et al. (2007) asserted that poor effort was present to a significant degree in the ADHD and LD evaluations with college students in comparison to the control group (Sullivan et al., 2007). Interestingly, Sullivan et al. (2007) also concluded that self-report indices should not be used to confirm a diagnosis if they have no validity scales embedded within them as this makes them vulnerable to the feigning of symptoms.

A similar study by Harrison, Edwards, and Parker (2008) evaluated 125 college students' ability to feign dyslexia based on their performance on common

psychoeducational assessments. These students did not have dyslexia and were randomly assigned to honest or feigning groups. Harrison et al (2008) concluded that students who were instructed to act as though they had dyslexia were able to produce scores as low or lower than students with true deficits (Harrison et al., 2008).

More recently, a study by Sollman, Ranseen, and Berry (2010) placed 73 college students in experimental conditions including one in which students were told to feign ADHD, and were given information on ADHD that is readily available on the internet. Results from this study showed that college students could provide ADHD consistent profiles across ADHD specific rating scales that did not include validity indicators/scales (Sollman et al., 2010). Interestingly, this study also used an examiner who was blind to the experimental condition to help reduce any examiner effects that may have biased results. Comparable findings were found in a study with 70 college student participants. This particular study asked a group of randomly assigned students to feign ADHD symptoms while comparing their results to students without ADHD and those diagnosed with ADHD. Findings illustrated that ADHD symptoms were easily feigned and profiles were indistinguishable from those students with ADHD using archival data sets (Harrison, Edwards, & Parker, 2007). The findings from these three studies seem to suggest that it was not unusual for college age populations to demonstrate response bias and that they were able to easily produce convincing profiles on ADHD specific measures consistent with an ADHD diagnosis.

Within pediatric populations, studies have also attempted to evaluate whether children and adolescents can feign believable deficits in performance as measured on psychological assessments. One such study examined children's ability to fake believable

deficits on neuropsychological testing. A sample of 42 clinical neuropsychologists was given assessment results and a brief background history. These practitioners were considered to be well seasoned with over 8,000 hours of professional service. Two main conclusions were made based on the results. First, clinicians were unable to detect feigned deficits and secondly, children can easily fake or malinger with minimal instruction (Faust et al., 1988b). It should be noted that the participants only reviewed results and did not observe or test the child themselves. A control condition was also utilized in this study where participants were only provided a brief history. Another interesting finding of this study was that level of training or experience did not improve the accuracy in detection (Faust et al., 1988b). A similar two part study first examined the ability of 113 neuropsychologists to detect feigned deficits in adolescent populations. The second part of the study examined the ability of another 125 neuropsychologists to detect feigned deficits in adolescent populations while also given information on base rates of malingering. Findings from this study illustrated that neuropsychologists were not able to detect malingering better than chance probability (Faust et al., 1988c). However, this study used a survey method and did not include direct interaction with any of the adolescents. Both of these studies suggest that malingering is more difficult to detect than practitioners currently believe as both children and adolescents with minimal training can "fake bad" effectively. However, because of the indirect methodology, it is possible that the practitioners may have been able to detect poor effort through direct observation and interaction with the children and adolescents.

Response bias in high school students has also been examined. In a study by Hunt et al. (2007), researchers administered two brief tests of effort within a brief

neuropsychological evaluation to 199 high school athletes during baseline testing and discovered that 11% of students exerted poor effort. The student participants were kept blind to the purpose of the study. At baseline testing, the researchers discovered that 11% of students exerted poor effort. Interestingly, due to the fact that only brief measures of effort were used and are less sensitive as they can only detect substantially poor effort, one in ten high school athletes still failed (Hunt et al., 2007). The results of these studies provide sufficient evidence that response bias should represent a concern within the field of school psychology. Important decisions are based on psychological assessment and school-age children are able to effectively feign difficulties. Similar findings have resulted when using effort measures with typically developing children (Blaskewitz et al., 2008) as well as children with other disabilities or medical conditions (MacAllister et al., 2009). Despite this growing body of research, formal training on measures of effort are not included in school psychology training programs.

Literature in School Psychology

There are a variety of reasons that an individual might put forth low effort or actively attempt to appear "different" on an assessment. One reason that has been discussed within the field of school psychology is that of "rapport" because of its association with effort. That is, if a student and examiner have developed a good rapport, the student is likely to put forth his or her best effort on an assessment (Teglasi & Freeman, 1983). One such article examined the common issues of beginning school psychology graduate student testers when building rapport. It was similarly stated within this article as with other literature that poor performance on psychological tests can be due to extraneous factors with lack of motivation or desire to appear unmotivated being

among them (Teglasi & Freeman, 1983). Understanding when it is appropriate to test limits as well as understanding how to appropriately probe for answers when poor effort is suspected, is particularly difficult for beginning testers (Teglasi & Freeman, 1983). In an earlier work, it was also noted that verbal encouragement should be used frequently in order to ensure that a child is putting forth their best effort (Witmer, Bornstein, & Dunham, 1971). Additionally, a practitioner may conclude that lack of motivation was present if a child improved performance following a request for the individual to put forth his or her best effort (Sattler, 2001). Within school psychology, informal means of assessing effort and motivation during psychological evaluations tends to be the primary method used. Efforts to quantify testing behaviors have resulted in behavioral checklists (e.g., McConaughy, 2005) which are commercially available. However, these checklists are still dependent on the observation of the examiner. While there is a great amount of utility for such measures, more objective measures based on the client's performance should be used in conjunction with these observational methods to assess client effort.

One such effort to measure client effort directly is referred to as "Can't do/Won't do assessment" or performance/skill deficit assessment. Recently, this approach has gained significant attention. This area of research focuses on the evaluation of skill deficit versus performance deficit in relation to academic achievement (VanDerHeyden & Witt, 2008). Can't do/Won't do assessment relies on a procedure which utilizes curriculum based measures being given in a repeated fashion. For students who perform poorly, incentives are provided and explicitly linked to higher performance on subsequent administrations. Duhon et al. (2004) hypothesized that if a poorly performing student improves his or her performance due to rewards then a performance deficit is present.

Conversely, if a poorly performing student does not improve in the presence of rewards then a skill deficit is assumed. This type of assessment research has provided a more quantifiable method for referral for special education identification rather than traditional teacher referral methods (VanDerHeyden, Witt, & Naquin, 2003). Within this literature, a procedure for improving validity of referrals for special education identification has been created.

Overall, the field of school psychology has not focused on assessing response bias or effort within psychological evaluations conducted in the schools. While there has been some attention toward building rapport and encouraging students to present their best effort through encouragement and incentives, the assessment of effort does not appear to be an important area of study within the field of school psychology. When it is addressed, a major assumption appears to be that motivation can be monitored by informal means instead of formal assessment techniques. This assumption is a critical threat to the validity of such findings in any research claiming to control for effort without using objective and statistically sound measures of effort. In a study examining performance characteristics of 121 college students on a psychoeducational test, it was noted that the motivation of the participants is a potential liability but still came to the conclusion that motivation was not a general concern based on the fact that examiners monitored individual's performance to ensure optimal effort (Laurent, 1997).

Overall, research from adjacent fields of applied psychology has provided a well-established basis for the assessment and study of effort within psychological evaluations. However, despite this rich research foundation, reliance on informal and subjective (i.e. observational) methods of assessing effort during psychological evaluations remains the status quo within the practice of school psychology. School psychology researchers have examined similar constructs such as achievement motivation but have not yet focused on response bias in psychoeducational evaluations. This study bridges the apparent gap

between fields of applied psychology specific to assessing effort during psychological evaluations.

CHAPTER III

METHODOLOGY

Participants

The participants selected for this study were collected through a convenience sample at a public university in the rocky mountain region. A total of 30 graduate level students who were enrolled in nationally accredited school psychology, counseling psychology, counselor education and supervision, and special education programs volunteered for this study. Participants were divided into two groups of 15 graduate students were used within this study and were composed of either psychology graduate students who had completed coursework and formal training in cognitive assessment or those that had not yet been exposed to coursework in cognitive assessment. The group of graduate students that had not yet been exposed to coursework in cognitive assessment served as the control group; those who had completed training in cognitive assessment served as the experimental group. These groups were formed as studies which have examined response bias detection often include different levels of seasoned professionals. However, research has not yet examined the ability of individuals without specific assessment training in their ability to detect performance that simulates intentional poor performance.

The graduate students in school psychology were from both specialist (Ed.S.) and doctoral (Ph.D.) programs. School psychology graduate students in the specialist program

were also solicited in this study because school psychologists with Ed.S. degrees are able to practice within the school setting without a doctoral degree. Graduate students in counseling psychology, counselor education and supervision, clinical mental health counseling, and special education programs were either in master's or doctoral degree programs and were included in the participant sample as they represented similar levels of overall education and were in service oriented professions. Graduate student participants were provided monetary compensation for their participation.

Instruments

Experimental stimuli were generated with the help of nine undergraduate students. The confederates were obtained through an undergraduate subject pool and received research credit in their course. Care was taken to assure that they were free from neurological disability as each confederate was asked if they had sustained any concussions or brain injuries. Confederates that endorsed a positive history of neurological disability were excluded from the study. Their performances on the cognitive task (see Block Design section) were videotaped and served as the stimulus materials that were viewed and rated by graduate student participants. These undergraduate confederates were comprised of two groups, one which included full effort and the other, suboptimal effort. Confederates within the full effort group were asked to try their best on all tasks presented. The confederates within the suboptimal effort condition were asked to purposefully deceive the examiner by not putting forth their best effort. Both groups were administered the Test of Memory Malingering (TOMM: Tombaugh, 1996), the Block Design subtest from the Wechsler Abbreviated Scales of Intelligence (WASI: Wechsler, 1999), and the Word Memory Test (WMT: Green, 2005).

Basic demographic information was also collected at this time and included gender, age, and history of head injury. These college students were deemed to be an appropriate sample for school psychology students to observe and rate as they similar in age to students in a secondary setting.

Block Design

Block Design is a subtest from the Wechsler Abbreviated Scale of Intelligence (WASI: Wechsler, 1999) which is an abbreviated form of the Wechsler Adult Intelligence Scale – Third Edition (WAIS-III: Wechsler, 1997). This subtest requires an individual to recreate visual patterns using colored blocks. Block Design is designed to assess an individual's nonverbal reasoning, visual spatial abilities, and visual construction skills. According to the WASI manual, the reliability of the Block Design subtest is .76 with established concurrent validity with the WAIS-III. This subtest was administered in a standardized manner to each confederate. T-scores were calculated based on the individual's age and associated norms included within the WASI manual.

Test of Memory Malingering

The Test of Memory Malingering (TOMM: Tombaugh, 1996) is a forced-choice measure of effort. It consists of a series of memory tasks which require an individual to identify between two visual stimuli which stimulus was previously presented. The stimuli within this test are black and white drawings of objects. Although the test consists of multiple trials, the first two trials are considered the standard administration with an optional retention trial that can be administered 15 minutes after the second trial. Each trial consists of 50 pictures which are visually exposed to the participant for three seconds each. After viewing all 50 stimuli, the individual is required to point to the picture that

was previously viewed. Utilizing a forced-choice test paradigm, the TOMM provides an objective means to assess for malingering and suboptimal effort. Scores below the recommended cutoff of 90% correct on the second trial are deemed to be evidence of suboptimal effort or possible malingering behaviors.

Validation of the TOMM has been provided through many research studies. In a series of studies, it was determined that the TOMM had high face validity as it was perceived by participants to be a test of cognitive ability rather than malingering (Rees et al., 1998). Studies have also illustrated validity of use of the TOMM with populations with various psychogenic and neurological disorders (Cragar, Berry, Fakhoury, Cibula, & Schmitt, 2006) as well as use within forensic psychiatric settings (Weinbor, Orr, Woods, Conover, & Feix, 2003). Promising research has also illustrated potential utility of the TOMM in the detection of suboptimal effort in pediatric populations (Donders, 2005; Constantinou & McCaffrey, 2003; MacAllister et al., 2009). Concurrent validity specific to measures of cognitive effort has also been found with the TOMM and other measures of effort (Farkas, Rosenfeld, Robbins, & van Gorp, 2006). In relation to reliability, no test retest statistics are reported within the TOMM manual.

Word Memory Test

The Word Memory Test (WMT: Green, 2005) measures the effort an individual exerts in order to learn a list of related words that are commonly used. The test includes multiple trials as well as a recognition trial which can also be utilized to evaluate effort (Green, 2005). Performance on any of the trials which falls below the recommended cutoff point of 82.5% correct is considered to be a “Clear fail” and evidence of suboptimal effort or possible malingering behaviors. Scores between 83% and 90%

correct are deemed to be within the "Caution" range while scores above 90% are considered a "Clear pass." This test has been validated for assessing effort most commonly among populations that have sustained mild, moderate, or severe brain injuries (Green, 2005; Flaro, Green, & Robertson, 2007). Validation of the WMT has been shown in research in the detection of malingering and suboptimal effort at a high level of specificity and sensitivity to that of detection rates from the TOMM (Greffenstein et al., 2008; Bauer et al., 2007). Promising research has also demonstrated the utility of the WMT in pediatric populations as well (Green & Flaro, 2003).

Observational Rating of Effort

The observational rating of effort form is comprised of two items and is included in Appendix E. This form was created by the researcher. The first item is comprised of a 9 point ratio scale which asks the participant to rate the effort they perceived the confederate to be putting forth on the Block Design subtest. On this scale, the one point was defined as no effort or "did not participate" while the 9 point was defined as optimal or complete effort. The second item on the observational rating of effort required the graduate student to choose whether the individual put forth optimal effort or suboptimal effort in a dichotomous manner. The purpose of this second item was to have the graduate student provide an overall rating similar to how effort is reported within psychological reports.

Self-Report Rating

The self-report rating is a measure that is similar in design to the previously described observational rating of effort except that it required confederates to rate their own ability to follow the directions specific to the condition they were randomly placed

into on the same 9 point ordinal scale. This self-report rating question was also created by the researcher. On this scale, the one point was defined as "did not follow directions" and the 9 point was defined as "completely followed the directions." This rating was used in this study as an additional check to ensure that directions were followed appropriately.

Video Stimuli

The video stimuli created within the study were the end results from the administrations of Block Design, TOMM, WMT, and the self-report rating to each confederate. It should be noted that only the Block Design administration portion was video recorded. Demographic information of the confederates as well as descriptive statistics of their respective performances on the selected measures can be found in Appendix F.

Procedure

First Phase

Prior to beginning the study, permission was obtained from the university Institutional Review Board (IRB) (see Appendix A). The procedures of this study consisted of two phases. The first phase involved the generation of the stimuli material required for the second stage. After providing consent (see Appendix xx), a group of nine undergraduate confederate participants were randomly placed in one of two experimental conditions (full effort or suboptimal effort) by the primary researcher. The full effort condition included a set of instructions which stated:

> Today I would like you to try your best on all of the measures you are given. If you are unsure of an answer, please feel free to take guesses. It is important that you try your best on all the items.

The suboptimal effort condition included a set of instructions for the confederate participant which stated:

> Today I would like you to pretend that you are trying very hard when completing all of the measures, but, in reality, you are not performing to the best of your ability. It is important that you still complete the tasks you are given. Some tasks may be easy but some may be more difficult. It is important that the examiner thinks that you are trying your best, but once again please do not perform well on the measures.

Confederate participants were then asked by the lead researcher if they had any questions and then instructed to not divulge anything that was said to them. The lead researcher then stepped out of the room and sent in the examiner. A doctoral student in school psychology who was trained in the administration of the test battery served as the examiner in the study. The examiner was blind to which condition the confederate participants were assigned and the overall purpose of the study to reduce potential biases. This examiner was instructed to administer the test battery according to the formal instructions provided in each test manual.

The examiner then administered the series of tasks in a standardized manner in the following order: administration of the WMT, followed by the Block Design subtest, and finishing with the TOMM. The administration of the Block Design subtest was the only portion of the evaluation that was video recorded for viewing by the graduate student participants. Lastly, the confederate participants were also asked to complete the self-report measure to rate their ability during the study to follow the instructions of the condition to which they had been assigned. Total testing time for this sequence was approximately 60 minutes. All nine video recording sessions took place in a designated research room in the psychology department at a public university. Acknowledgment of the use of video recording equipment in the study was included as part of the informed

consent (see Appendix B). The primary researcher of this study operated the recording equipment.

Following completion of the video stimuli, the primary researcher reviewed the scores on the tests of effort and selected a total of eight videos that met appropriate criteria for the adequate effort and suboptimal effort conditions based on the scores on both the WMT and TOMM. Of these eight videos, four were selected for the full effort condition and four were selected for the suboptimal effort condition. One videotaped administration was excluded as the undergraduate confederate who had been assigned to the suboptimal effort condition, performed at a level deemed inappropriate for the condition according to the scores on both the WMT and TOMM. This undergraduate confederate earned scores that were in the full effort range when they were instructed to put forth suboptimal effort.

Demographic data regarding the confederates who participated within the first phase of this study are shown in Table 12 (Appendix F). The age range for the included confederates ranged from 18-20 years of age and included four females (50%) and four males (50%) with the mean age of 18.875 and a standard deviation of 0.64 years.

The performance of the confederates on the Block Design subtest, the TOMM, and the WMT within their respective conditions of either full or suboptimal effort as well as their adherence to directions self-report rating are shown in Appendix F Table 13. As mentioned previously, scores on the second trial of the TOMM below 90% correct are deemed to be within a range of suboptimal effort or the indication of possible malingering behaviors. Scores on any of the WMT subscales that are below 82.5%

correct are considered failing and indicative of suboptimal effort or possible malingering behaviors.

After examination of these scores, trends in the data suggested that the full effort confederate group scores were higher in every measure than the suboptimal effort confederate group. Each confederate within the full effort group scored within acceptable ranges across the measures of effort. Conversely, each confederate within the suboptimal group scored within the fail range on at least one area of the WMT. The adherence to directions self-report ratings, which provided an indication of the confederates were also higher within the full effort group in comparison to the suboptimal group.

Second Phase

The second procedure used within this study relates to the graduate student participation. All 30 graduate student participants were first given consent forms and asked to provide demographic information about themselves which included gender, age, graduate degrees currently held, and their graduate program. The participants were also asked to indicate whether they had received graduate level training in cognitive assessment or the assessment of response bias. Following the collection of the demographic information, the graduate students were provided instructions for the study. These instructions stated that they would view a series video recordings in which individuals were administered a cognitive task that required planning and nonverbal problem solving. The graduate students were also informed that they would be asked to rate the effort each individual in the video appeared to put forth. Participants then viewed each of the eight video recordings in a consistent order either individually or in a small group in the presence of the lead researcher. The participants were then asked to

complete the two questions related to the ratings of effort immediately following the completion of the video. Each videotaped administration yielded a total of 30 observations by the participants. These observations provided a means for inter-rater comparisons according to participant group. Lastly, an open ended question was asked to a random selection of 10 graduate students following completion of their participation in the study. The question asked "What are the salient features in the video recordings that informed your rating?" Total time for each participant within phase two was approximately 120 minutes.

Variables

Within the present study, the independent variables were the characteristics of each individual as described by observational ratings and performance on presented measures. These characteristics are representative of evidence for or against the presence of response bias. Response bias is a general term that can be defined as including many influences that bias the way an individual responds to a given task or situation. Response bias was assessed using two methods, observationally (Observational Likert-type rating scale) and through two standardized tests of effort (the TOMM and WMT). Effort, which is a component of response bias, can be defined as the amount of exertion, physical, cognitive, or emotional, an individual puts forth during a given event. Independent variable data collected was comprised of observational ratings on a ratio scale, scores from the TOMM and scores from the WMT.

The dependent variable utilized within this study included the conditions for the participants. One condition was the optimal effort group while the other condition was the suboptimal group. The condition for the optimal effort group included instructions

which stated that the individual should put forth their best effort on all tasks administered to them. The condition for the suboptimal effort group included instructions which stated that the individual should try to complete the given tasks with poor effort.

Research Hypotheses

H1 The observations by the participants were not able to discriminate between full effort and suboptimal effort participants.

H2 The observations by the participants did not add to the overall model in the differential prediction of full effort or suboptimal effort participants.

H3 No differences existed between the graduate students with training in assessment group and the graduate students without training in assessment group.

Due to the open-ended nature of the fourth research question, no specific hypothesis was tested within this study.

Data Analysis

Within the present study, three statistical procedures were used to analyze the data. A fourth procedure to address the open-ended question was also used to analyze results.

1. Means, percentages, and other descriptive statistics were calculated for each condition.

2. As mentioned previously, the second portion of the observational rating asked participants to provide an overall classification of the individual as either putting forth optimal or suboptimal effort. This dichotomous classification allowed for the use of a Fisher's exact test for each condition of the study. The Fisher's exact test provides a method for testing whether significant differences existed between classifications of effort. The calculated *p*-value from this analysis illustrated if significant differences

existed between the classifications. This analysis was applied to the two groups of graduate students (i.e. with training in assessment and without training in assessment) to examine if differences existed.

3. Discriminant analysis, a statistical process that involves the prediction of an observation into a group based on a model, was also used in this study. This statistical analysis requires that certain assumptions are fulfilled. These assumptions includes: normal distribution of each predictor variable, equality of the covariance matrices, and independence of data (McLachlan, 1992). The model that is created through discriminant analysis is based on a set of interrelated variables (i.e. independent variables) and provides a prediction value or estimate of group membership (i.e. dependent variable). Groups within discriminant analysis refer to the categorical dependent variable which can include two or more classes. The models that are created can also be evaluated within discriminant analysis using the jackknife method. This method removes one of the independent variables at a given time and creates a model based on the remaining independent variables. These new models are then averaged to form a final model or discriminant function. In the present study, individuals were classified into one of two groups (full effort or suboptimal effort) based on prior performance on measures. Using the independent variables to formulate a discriminant function, future individuals can then be predicted to have either the full effort or suboptimal group membership. These rules are based on the observational ratings as well as scores from the TOMM and WMT.

4. The open-ended question that was asked to the psychology graduate student participants provided a descriptive component to the study. The data collected via an open-ended question was analyzed for categories. A random selection of 10 participants'

responses to this question was analyzed, half of the participants were from the group with training in cognitive assessment and half were from the other group. These categories were created using information from the open-ended question that was coded and analyzed for frequency among participants (Merriam, 1998). Significant statements from the question were noted and made into a list where larger units of information were then found thus creating categories (Creswell, 2007).

CHAPTER IV

RESULTS

The purpose of this study was to examine the assessment of effort and response bias through two means, observationally and objectively. More specifically, the design of the study was to investigate both the utility of an observational rating scale and the accuracy of observers through experimental means where confederates were placed within conditions and instructed to either exert full effort or suboptimal effort during a commonly administered cognitive assessment subtest. The levels of effort that the confederates were instructed to exhibit were corroborated using two well researched and established tests of effort, including the Test of Memory Malingering (TOMM; Tombaugh, 1996) and the Word Memory Test (WMT: Green, 2005). Graduate student participants then rated the effort they perceived the confederates exhibited on both a nine-point Likert-type scale and a dichotomous rating scale of effort. Presented in this chapter are the results of the analyses completed for this study. It is divided into six sections: (a) description, (b) accuracy rates, (c) Fisher's exact test, (d) discriminant analysis, (e) open-ended question analysis, and (f) summary of the statistical findings.

Demographics

Demographic data regarding the total sample of the study of graduate student participants are shown in Table 1. The age of participants ranged from 23-41 years old. The total sample of 30 graduate student participants attended the same university and were actively enrolled in graduate programs in the areas of school psychology,

counseling psychology, counselor education and supervision, and special education. The participants were divided into two groups of 15 based on whether they had completed training in cognitive assessment.

Table 1

Demographics for the Total Sample

Variable	N	%
Gender		
Females	24	80
Males	6	20
Graduate Program		
School Psychology – Ed.S.	4	13.3
School Psychology – Ph.D.	8	26.7
Counseling Psychology	5	16.7
Clinical Counseling – Master's	1	3.3
Counselor Education and Supervision	4	13.3
Special Education	8	26.7
Cognitive Assessment Training		
Yes	15	50
No	15	50

Accuracy Rates

The overall accuracy rates of the dichotomous ratings made by the groups of participants are displayed in Table 2. Within the sample group that has received training in cognitive assessment, a total of 60 observations (100%) were rated correctly as

exhibiting full effort and 5 observations (8.3%) were rated correctly as exhibiting suboptimal effort. The control group that has not received training in cognitive assessment had a total of 60 observations (100%) were rated correctly as exhibiting full effort and 9 observations (15%) were rated correctly as exhibiting suboptimal effort. Overall, a total of 120 observations (100%) were rated correctly as exhibiting full effort and 14 observations (11.7%) were rated correctly as exhibiting suboptimal effort.

Table 2

Accuracy of Dichotomous Ratings

	Cog Training	Control	Total Identified
Full Effort	60	60	120
	100%	100%	100%
Suboptimal	5	9	14
	8.3%	15.0%	11.7%

Fisher's Exact Test

Fisher's exact test is a procedure that is utilized within experimental studies to determine whether the differences in the data between groups are due to chance. This statistical procedure allows for analysis of small sample sizes. Within the current study, the dichotomous ratings of effort by the groups of participants provided a means to compare against the known condition assignment of each confederate. Fisher's exact test was conducted for the full effort and suboptimal effort conditions using the data from correct and incorrect dichotomous ratings in each condition. Contingency tables for the

full effort and suboptimal effort conditions are listed below in Tables 3 and 4, respectively.

The participant group that had training in cognitive assessment correctly rated 100% of dichotomous ratings as exhibiting full effort. However, only 8.3% of dichotomous ratings were correctly identified as exhibiting suboptimal effort. The participant group that had not received training in cognitive assessment also had 100% of dichotomous ratings correctly identified as exhibiting full effort. Only 15% of dichotomous ratings were correctly identified as exhibiting suboptimal effort. Overall, a combined total of 100% of the dichotomous ratings were correctly identified as exhibiting full effort and 11.7% of the dichotomous ratings were correctly identified as exhibiting suboptimal effort.

Table 3

Contingency Table for Full Effort Condition

Full Effort	With Training	Without Training	% of Ratings
Correct Identification	60	60	100
Incorrect Identification	0	0	0
% of Ratings	50	50	

Table 4

Contingency Table for Suboptimal Effort Condition

Suboptimal Effort	With Training	Without Training	% of Ratings
Correct Identification	5	9	11.67
Incorrect Identification	55	51	88.33
% of Ratings	50	50	

Data from the full effort condition Fisher's exact test revealed that the calculated p-value was 1 illustrating a lack of significance. This finding answers a research question as significant differences do not exist between groups in the full effort condition beyond chance probability. The lack of significant differences between groups illustrates that both groups performed similarly in their dichotomous ratings of effort in the full effort condition.

Data from the suboptimal effort condition Fisher's exact similarly revealed that the calculated *p*-value of 0.3945 was not within the statistically significant range. This finding also answers a research question as significant differences do not exist between groups in the suboptimal effort condition beyond chance probability. An effect size estimation using Cramer's phi also illustrated a calculated value of 0.104 which is considered to be small (Cohen, 1988). As noted in the previous section, the accuracy rates of detecting suboptimal effort were low in both groups illustrating a similar difficulty in assessing effort through observational means regardless of training in cognitive assessment.

Discriminant Analysis

Discriminant analysis is a procedure that provides a means to classify subjects or observations into groups using one or more independent variables. Within the current study, the independent variable used within the analysis was the Likert-type scale ratings of effort completed by the participants. Since each confederate was randomly assigned to each condition or group, the group affiliation is known, thus allowing for a discriminant function to be created based on the rating scores of the participants. The discriminant function serves as a method to separate the observations into two groups based on the data provided by the independent variable, either through linear means or quadratic means. In other words, the discriminant function that categorizes observations may be a straight line or curved line depending on the grouping of the data. Within the current study, a jackknife holdout method of discriminant analysis was used as it provides a more realistic estimate of error rates created by the discriminant function (Stevens, 1986). This is performed by removing an observation from the data and then using it with the discriminant function that was created with the remaining data in a repeated fashion to calculate the classification error estimates in a more realistic fashion. Within the current study, classification summaries for each group of participants (with or without training in cognitive assessment) are displayed; the jackknife method of discriminant analysis was completed two times to provide a means of comparison between the sample groups. The classification summaries for the jackknife holdout method of discriminant analysis using a linear function and quadratic function are listed within Tables 5 and 6, respectively, for the sample group with training in cognitive assessment. Additionally, the error count estimates for the linear and quadratic function are listed in Table 7 for the same sample

group. The classification summaries for the linear and quadratic discriminant functions using the sample group without training in cognitive assessment are displayed within Tables 8 and 9 respectively. The error count estimates for the same group for each discriminant function are listed in Table 10.

The classification outcomes of the discriminant function analyses based on the Likert-type scale ratings from the group with cognitive assessment training illustrated a higher total of correct full effort condition classifications using the linear discriminant function in comparison to the quadratic discriminant function. Conversely, a higher total of correct suboptimal effort condition classifications was noted from the quadratic discriminant function rather than the linear discriminant function.

A total probability of misclassification can also be calculated within discriminant analysis to illustrate the overall accuracy of a given discriminant function. This total probability of misclassification allows for one to compare between linear and quadratic discriminant functions and identify the more appropriate model. Overall, the total probability of misclassification using the linear discriminant function was lower than the total probability of misclassification using the quadratic discriminant function. The linear discriminant function is thus deemed a more appropriate model, as it is less likely to misclassify an observation than the quadratic discriminant function.

Table 5

Cognitive Training Group Classification Summary using Linear Discriminant Function

Condition	Full Effort	Suboptimal Effort	Total
Full Effort	47	13	60
	78.33%	21.67%	100.00%
Suboptimal Effort	29	31	60
	48.33%	51.67%	100.00%
Total	76	44	120
	63.33%	36.67%	100.00%
Priors	0.5	0.5	

Table 6

Cognitive Training Group Classification Summary using Quadratic Discriminant Function

Condition	Full Effort	Suboptimal Effort	Total
Full Effort	54	6	60
	90.00%	10.00%	100.00%
Suboptimal Effort	43	17	60
	71.67%	28.33%	100.00%
Total	97	23	120
	80.83%	19.17%	100.00%
Priors	0.5	0.5	

Table 7

Cognitive Training Group Error Count Estimates using Linear Discriminant Function

	Linear Discriminant Function		
	Full Effort	Suboptimal Effort	Total
Rate	0.2167	0.4833	0.3500
Priors	0.5000	0.5000	
	Quadratic Discriminant Function		
	Full Effort	Suboptimal Effort	Total
Rate	0.1000	0.7167	0.4083
Priors	0.5000	0.5000	

The classification outcomes of the discriminant function analyses based on the Likert-type scale ratings from the control group (participants without cognitive training) illustrated a higher total of correct full effort condition classifications using the quadratic discriminant function in comparison to the linear discriminant function. Conversely, a higher total of correct suboptimal effort condition classifications was noted from the linear discriminant function rather than the quadratic discriminant function. Overall, the total probability of misclassification using the linear discriminant function was lower than the total probability of misclassification using the quadratic discriminant function. This result illustrates that the linear discriminant function is more appropriate as it is less likely to misclassify observations.

Table 8

Without Training Group Classification Summary using Linear Discriminant Function

Condition	Full Effort	Suboptimal Effort	Total
Full Effort	49	11	60
	81.67%	18.33%	100.00%
Suboptimal Effort	26	34	60
	43.33%	56.67%	100.00%
Total	75	45	120
	62.50%	37.50%	100.00%
Priors	0.5	0.5	

Table 9

Without Training Group Classification Summary using Quadratic Discriminant Function

Condition	Full Effort	Suboptimal Effort	Total
Full Effort	55	5	60
	91.67%	8.33%	100.00%
Suboptimal Effort	40	20	60
	66.67%	33.33%	100.00%
Total	95	25	120
	79.17%	20.83%	100.00%
Priors	0.5	0.5	

Table 10

Without Training Group Error Count Estimates using Linear and Quadratic Discriminant Functions

	Linear Discriminant Function		
	Full Effort	Suboptimal Effort	Total
Rate	0.1833	0.4333	0.3083
Priors	0.5000	0.5000	
	Quadratic Discriminant Function		
	Full Effort	Suboptimal Effort	Total
Rate	0.0833	0.6667	0.3750
Priors	0.5000	0.5000	

Open-Ended Question Analysis

Responses to the single open-ended question asked to each participant after completing their ratings of effort for each video was analyzed for significant statements. As mentioned previously, the question asked "What behaviors did you observe that led you to your rating?" as a means of gaining insight into what salient features that informed each individual's ratings. A random selection of 10 participants' responses to this question was analyzed, half of the participants were from the group with training in cognitive assessment and half were from the other group. The categories were formed using the same process of first noting the frequency of statements among participant responses (Merriam, 1998). The most frequent statements were deemed significant and made into a list, from which they could be grouped into the prominent categories (Creswell, 2007). Overall categories elicited from the responses of the 10 participants

were coded into two main groups, adequate effort rating categories and inadequate effort rating categories.

From the response data within the adequate effort ratings, a total of five major categories could be identified. The first categories referred to the confederate as being persistent or not giving up prematurely. This category also included responses that referred to trying multiple approaches to solving the item that was presented. The second category seen within the data referred to the speed at which an individual responded to a given item. Interestingly, speed was noted to be an attribute for adequate effort, that is, the person responded quickly or slowly. The third category noted within the data referred to the accuracy or number of correct responses an individual was observed to display during the video. The fourth category within the data referred to an individual's checking of completed work and self-correcting errors. The fifth and final category noted within the adequate effort ratings referred to individuals who leaned forward during tasks.

Responses to the open-ended question on ratings of inadequate effort revealed a much different picture as very few participants perceived any of the confederates to put forth inadequate effort. The lack of data did not allow for a robust analysis. A single category was elicited from this data which referred to behaviors associated with distractability or disengagement with the task. These behaviors included checking a cell phone or "fiddling" with one's hair, ring, or shoes.

Summary

The results presented in this chapter are summarized in terms of the research questions previously stated in Chapter I.

Q1 Are the observations by the participants able to discriminate between full effort and suboptimal effort group membership?

The results from this study suggest that the participants demonstrated minimal ability to discriminate accurately between full and suboptimal group membership. Accuracy rates illustrated that both groups of participants could identify full effort membership more often in comparison to their ability to accurately identify suboptimal effort membership.

Q2 Does the addition of the observational rating scale improve the differential prediction of full effort or suboptimal effort group membership?

The results from this study suggest that the addition of the observational rating scale did improve the differential prediction of full effort or suboptimal effort group membership. Discriminant analysis revealed that the Likert-type scale ratings of effort allowed for an increased probability of accurate classification beyond that of the rates of the dichotomous ratings or chance probability.

Q3 Does training in cognitive assessment aid in the ability to discriminate between groups using observational methods?

With reference to the overall dichotomous ratings of effort, participants within the group who had received training in cognitive assessment had a slightly lower accuracy rate in comparison to the control group's ratings. Additionally, in reference to the Likert-type scale ratings of effort, the discriminant function created had a slightly higher overall probability of misclassification in comparison to the control group's discriminant function.

Q4 What salient features (i.e. behaviors) informed the observational rating of effort?

The results from this study suggested that six primary categories exist in reference to the salient features that inform one's observational ratings of effort. These categories

included speed of task completion, persistence, accuracy or number of correct items, self-correcting or checking behaviors, and attentive behaviors (e.g., leaning in). Conversely, perceived suboptimal effort was identified through behaviors associated with distractability or disengagement.

CHAPTER V

DISCUSSION

The purpose of this study was to investigate methods of assessing effort within experimental conditions where confederates were instructed to either try their best or purposely exhibit suboptimal effort. The specific methods of assessing effort that were examined were observations with the use of a Likert-type scale rating of effort and a dichotomous rating of effort. Data were collected from two groups of graduate students who either had training in cognitive assessment or did not. This chapter summarizes the findings of this study and discusses these results in relation to other current literature on this topic. Additionally, recommendations for application and future research directions are provided.

Literature examining the effect that effort can have on performance test scores (Green et al., 2001; Hunt et al., 2007; Kirkwood et al., 2012) is an important area of study in the field of psychology. Even highly trained and seasoned professionals have difficulty in accurately detecting suboptimal or malingering behaviors (Faust et al., 1988b; Faust et al., 1988c). Several measures of effort, such as the Test of Memory Malingering (Tombaugh, 1996) and the Word Memory Test (Green, 2005) have been developed to allow for more objective and accurate means of evaluating effort during psychological evaluations. Efforts to identify response bias have also been embedded into indices within already established and commonly used tests have (Lange et al., 2006; Kim et al.,

2010; Young, Caron, Baughman, & Sawyer, 2012). This study provided an experimental approach to determine the effectiveness of observational methods to assess effort.

Findings

As mentioned previously, research has demonstrated that psychologists have difficulty in assessing effort or malingering across populations (Faust et al., 1988b; Faust et al., 1988c). Although there is a paucity of school psychology research in this area, one could assume that school psychologists would have the same difficulty.

The findings from this study corroborated previous research; it revealed that raters using observational methods showed the same difficulty in accurately assessing the effort that examinees put forth. Overall, accuracy rates of effort ratings through a dichotomous approach were found to be low in the correct identification of individuals putting forth suboptimal effort. Also, accuracy rates of adequate effort ratings were vastly overestimated across the sample, as there was evidence of a high false positive rate. In other words, graduate participants tended to perceive all confederates as putting forth adequate effort and had limited ability to determine inadequate effort. Surprisingly, examination of the accuracy rates of dichotomous ratings by the sample group revealed that the group without training in cognitive assessment performed slightly better (15% accuracy compared to 8.3% accuracy) than the experimental group, which had received graduate level training in cognitive assessment. However, the difference was not significant. Taken together these findings suggest that observational efforts on their own are not effective, regardless of whether an individual is familiar with cognitive assessment.

Researchers have also noted in the literature that effort, which is a part of the greater construct of response bias, can be viewed as part of a continuum or spectrum that ranges from full effort to malingering, with suboptimal effort referring to the span between the two (Dunn, 2006). For this reason, a Likert-type scale was developed by the researcher and included in the study, to allow participants to rate the perceived level of effort exerted. A discriminant analysis indicated that creation of a prediction model based on the Likert-type scale ratings marginally aided in the differential prediction of full effort or suboptimal effort group membership. Unfortunately, the predictive power of the model using ratings from either the control or experimental group was only slightly higher than chance probability in the specific classification of suboptimal effort. As such, these findings also found evidence consistent with the previous research that individuals, independent of tests of effort, have significant difficulty in accurately detecting suboptimal effort and malingering behaviors (Faust et al., 1988b; Faust et al., 1988c; Harrison et al., 2007; Harrison et al., 2008). Specifically, previous research has determined that not only are children, adolescents, and adults adept at feigning deficits on neuropsychological tests, but also professionals could not detect malingering by reviewing cases and scores (Faust et al., 1988b; Faust et al., 1988c). This study is consistent with previous findings from these studies as participants had significantly low rates of accurate detection of suboptimal effort. Previous studies have also discovered that normal individuals can perform on measures of attention-deficit/hyperactivity disorder and dyslexia in such a manner that is indistinguishable to individuals properly diagnosed with such conditions (Harrison et al., 2007; Harrison et al., 2008). Similarly,

this study also found that professionals-in-training also experience this same difficulty in noting response bias on an ability-based test using observational methods.

Although the ratings from both groups were found to be generally inaccurate in the detection of suboptimal effort, it appeared that the control group (without training in cognitive assessment) was slightly more accurate in comparison to the experimental group (with training in cognitive assessment). Specifically, estimations of classifications based on the Likert-type scale derived discriminant functions revealed a lower total error count estimate for the control group at 0.30 in comparison to the experimental group total error count estimate of 0.35.

Through inclusion of a single open-ended question, a descriptive component allowed for greater insight from the participants in the study. Unfortunately, due to the small amounts of data from accurate ratings of suboptimal effort, only one category related to perceived suboptimal effort behaviors could be elicited. This category involved perceiving the individual as distracted, as evidenced by checking one's cell phone and "playing" with materials. However, there were more categories noted in the responses for perceived full effort behaviors. These categories included speed of completion, accuracy of responses, trying multiple attempts to solve item, leaning in, and correcting mistakes. Within the research that has been completed and was available for search in common databases in the English language, no other study has attempted to include a component which would allow for descriptive analysis of the judgment process one makes when making decisions on levels of effort one perceives.

In the practice of psychological assessment, one goal of the evaluation is to produce both reliable and valid scores which are indicative of a person's true abilities.

Psychological evaluations can consist of a variety of different assessments that measure any number of characteristics of a given construct, but the scores are always used to inform decisions. School psychologists often administer cognitive tests to diverse pediatric populations as part of the special education identification process. However, the validity of the results gained from such tests often hinges on blanket statements referring to behavioral observations associated with the perceived level of effort of the individual and might be based on some of the same behaviors (e.g., leaning forward, making multiple attempts) as described by the participants in this study. However, the accuracy and validity of clinical judgment has been questioned (Dawes et al., 1989). The results of this current study corroborate, as well as add to the argument that judgment is highly susceptible to bias and error even when there are attempts to create more actuarial methods of classification and prediction using scores derived from clinical judgment.

Implications

Accurate detection of suboptimal effort cannot rely on observational means alone. Subsequently, when completing psychological evaluations, psychologists should include multiple measures of response bias to ensure validity of results. Formal training in best practices of evaluating effort is essential within graduate psychology programs and should not be limited to neuropsychology and forensic psychology. This training should include different methods of evaluating effort, such as embedded measures and objective tests that are well researched and appropriate for different populations such as children. Additionally, there is a dearth in the literature regarding the number of school psychologists who utilize tests of effort. Given this lack of information on the practice of formal effort assessment in school-based evaluations, one can speculate that tests of

effort are not commonly utilized despite their ability to inform the validity of other test results.

Another implication of this study was that familiarity with the task being observed provided no added utility in the accurate detection of suboptimal effort. One could speculate that this may be related to some specific concepts learned in cognitive assessment training. For instance, in cognitive assessment training, graduate students are taught that cognitive diagnostic assessments are used to identify ipsative strengths and weaknesses within a cognitive profile (Sattler, 2001). Understanding that an individual may be naturally weak or perform more poorly on a given subtest could make one think that the poor performance on Block Design may be a natural part of their presentation rather than suboptimal effort. Examination of the accuracy of dichotomous ratings also showed that there was generally a positive bias for full effort as it was drastically over-identified. There were no instances of false identification of suboptimal effort. One could also speculate that this positive bias may be correlated with the understanding that individual cognitive profiles consist of both ipsative strengths and weaknesses.

Measures of response bias that have been developed and researched include both embedded indexes within commonly administered tests, as well as independent tests designed specifically to measure response bias. This study examined the utility of observational means in the assessment of effort and found only marginally better than chance accuracy of classification. Another implication of this study is that efforts to improve the assessment of response bias should focus on alternative methods other than observational ratings.

Limitations

The current study included the following limitations, which may have affected the generalizability of the results. The first limitation is that the graduate student participants were from a single university. Although graduate programs may have some similarities in training due to requirements for specific national certifications or endorsements, programs can have more specialized areas of study and training depending on the faculty. Expertise that individual faculty members hold produces differences in the skill sets of the students within their respective programs. Similarly, the small sample size of graduate student participants is also a limitation of the study which limits generalizability of the results.

The Likert-type scale measures which allowed participants to rate behaviors of each confederate are another limitation of the study. These measures were developed by the researcher and were not tested through pilot studies previous to the current study. This may have limited the validity and reliability of the results gained.

Another limitation to note within the current study is that only a single subtest was viewed for each confederate, allowing for a relatively short period of time to form judgments on effort. Psychological evaluations very often consist of a battery of tests, depending on the referral concern. Thus, having individuals make decisions on effort based on only a fragment of what would possibly exist in a more typical evaluation is not equivalent to actual practice. Research has also illustrated that analysis of performance across a battery is commonly used as a method of detection of response bias (Iverson & Binder, 2000). The lack of additional subtests to view as well as access to scaled scores to

visually analyze prevented participants from being able to engage in this method of evaluating effort.

Similarly, another limitation to consider is the reliance on a single subtest which relies on nonverbal output rather than verbal output. It is possible to consider a lack of verbal responses a limitation as some individuals may place significant importance on different characteristics of speech within the observational method of evaluating effort. This reliance on nonverbal behaviors may affect the generalizability of the results to encompass all of the components of evaluating effort through observational means.

The use of video recordings rather than live observations is another limitation of this study. Due to the angle of recording, participants may not have been able to observe subtle behavioral indicators of effort which may have affected their ratings. Additionally, graduate student participants may have become fatigued watching video recordings as they were not interacting with the confederate.

The order in which the videos were viewed and rated is another limitation of the study. Because all but one participant viewed the videos in the same order, one can hypothesize that an ordering effect may have influenced the ratings. This effect would affect any subsequent analyses.

Another significant limitation of the current study to note was the use of linear and quadratic discriminant analysis. As mentioned previously, an assumption that is required within discriminant analysis is normal distribution of the data. However, the data within the current study was not normally distributed which then violates the required assumption. Steps were also not made to artificially normalize the data through cubing

the data to allow for appropriate use of quadratic discriminant analysis. Use of logistic regression would have allowed for a more appropriate analysis of the data.

Future Directions

Research in the area of evaluating effort has become an area of focus within the fields of neuropsychology and forensic psychology but still has not garnered significant attention within the field of school psychology. Current research has continued to focus on both new measures for evaluating effort which include embedded indexes within commonly administered tests (Young et al., 2012) as well as using well-established tests such as the Test of Memory Malingering (Tombaugh, 1996) with pediatric populations (Brooks et al., 2012). Creating and validating multiple measures, as well as methods of evaluating effort, are still warranted within research; studies have shown that a diverse range of populations can effectively put forth inadequate effort or even malinger. Thus, future researchers should continue to investigate other pediatric populations practicing school psychologists encounter and frequently evaluate within the school setting.

Within the current study, observational analysis of an individual's behavior during a nonverbal subtest yielded little to no utility as an effective means of accurately evaluating effort. However, no published study available in common databases in the English language has examined the potential for observational analysis of an individual's behavior during verbal tasks. An individual's speech characteristics while actively putting forth suboptimal effort may provide more information to aid in the detection of suboptimal effort. Thus, the substitution of a verbal task within a similar experimental design warrants investigation.

Another future direction that researchers should investigate in relation to the current study is the inclusion of additional subtests, a full standardized test of cognitive ability, or a full battery of tests within the video stimuli for participants to view and rate. The addition of varied subtests may produce different observable factors which may be informative in the detection of suboptimal effort or malingering. Selection of a standardized test with embedded indicators of suboptimal effort within a similar experimental design may also be beneficial, as it would remove the assumption that confederates exhibited the appropriate level of effort during all portions of the evaluation. Furthermore, increasing the similarity to actual psychological evaluations would improve the generalizability of the results.

Although some studies have utilized coaching of confederates in regards to faking a specific disability or medical condition (Dunn, Shear, Howe & Ris, 2003; Faust et al., 1988b; Faust et al., 1988c; Harrison et al., 2008; Sollman et al., 2010), none have specified directions for confederates to inform behavioral functioning for suboptimal effort. Inclusion of specific directions which outline both overt and subtle behavior examples may provide video stimuli with increased face validity. Subsequently, the increased face validity may have a positive effect on the accuracy of the observational ratings.

Finally, research on current practices within school psychology with regard to the evaluation of effort is an area that needs to be further investigated. Examination of the current implementation of effort testing within evaluations by school psychologists should be considered, as well as possible inclusion of effort testing practices within the framework of the Response to Intervention paradigm. Along these same lines, research

on the frequency of suboptimal effort within special education identification evaluations should occur to further bolster the rationale for specialized training in the evaluation of response bias for school psychologists.

Conclusions

The current study illustrated the importance of assessing effort and response bias in psychological evaluations in ways that are not reliant on observational input. The implications of the study are notably relevant to the field of school psychology as the evaluation of response bias has been an area largely overlooked within school psychology literature as well as graduate training programs. In the school setting, school psychologists frequently complete cognitive evaluations as part of the special education identification process without any other data to support statements of effort exerted by the student in the evaluation. Future research is needed to investigate rates of suboptimal effort within school-based evaluations to provide additional support for inclusion of specialized training on measures of response bias in school psychology programs. Additionally, future research is needed in the potential role of response bias and effort within assessment methods. Overall, this study emphasized the importance of understanding response bias, the subtle presentation of suboptimal effort, and the critical importance of sound methods of detection.

REFERENCES

Abdelfattah, F. (2010). The relationship between motivation and achievement in low-stakes examinations. *Social Behavior and Personality, 38*(2), 159-168. doi: 10.2224/sbp.2010.38.2.159

Aiken, L (1994). *Psychological testing and assessment*. Needham Heights, MA: Allyn and Bacon. American Psychiatric Association. (2000). *Diagnostic and statistical manual of mental disorders* (4th ed.). Washington, DC: Author.

American Psychological Association. (2010). *Ethical Principles of Psychologists and Code of Conduct*. Retrieved from http://www.apa.org/ethics/code/index.aspx

Baker, R., Donders, J., & Thompson, E. (2000). Assessment of incomplete effort with the California Verbal Learning Test. *Applied Neuropsychology, 7*(2), 111-114. doi: 0.1207/S15324826AN0702_8

Barrash, J., Suhr, J., & Manzel, K. (2004). Detecting poor effort and malingering with an expanded version of the Auditory Verbal Learning Test (AVLTX): Validation with clinical samples. *Journal of Clinical and Experimental Neuropsychology, 26*(1), 125-140. doi: 10.1076/jcen.26.1.125.23928

Bauer, L., O'Bryant, S. E., Lynch, J. K., McCaffrey, R. J., & Fisher, J. M. (2007). Examining the Test of Memory Malingering trial 1 and Word Memory Test immediate recognition as screening tools for insufficient effort. *Assessment, 14*(3), 215-222. doi: 10.1177/1073191106297617

Binder, L. M., & Rohling, M. L. (1996). Money matters: A meta-analytic review of the effects of financial incentives on recovery after closed-head injury. *The American Journal of Psychiatry, 153*, 7-10. Retrieved from http://0-proquest.umi.com.source.unco.edu

Blaskewitz, N., Merten, T., & Brockhaus, R. (2009). Detection of suboptimal effort with the Rey Complex Figure Test and recognition trial. *Applied Neuropsychology, 16*, 54-61. doi: 10.1080/09084280802644227

Blaskewitz, N., Merten, T., & Kathmann, N. (2008). Performance of children on symptom validity tests: TOMM, MSVT, and FIT. *Archives of Clinical Neuropsychology, 23*, 379-391. doi: 10.1016/j.acn.2008.01.008

Brooks, B. L., Sherman, E. M. S., Krol, A. L. (2012). Utility of TOMM trial 1 as an indicator of effort in children and adolescents. *Archives of Clinical Neuropsychology, 27*, 23-29. doi: 10.1093/ Discriminant analysisarclin/acr086

Bush, S. S., Ruff, R. M., Tröster, A. I., Barth, J. T., Koffler, S. P., Pliskin, N. H., Reynolds, C. R., & Silver, C. H. (2005). Symptom validity assessment: Practice issues and medical necessity. *Archives of Clinical Neuropsychology, 20*, 419-426. doi: 10.1016/j.acn.2005.02.002

Butcher, J. N., Dahlstrom, W. G., Graham, J. R., Tellegen, A, Kaemmer, B., & McGargee, E. I. (2001). *Minnesota Multiphasic Personality Inventory – Second Edition*. San Antonio, TX: Pearson Education, Inc.

Cohen, J. (1988). *Statistical Power Analysis for the Behavioral Sciences – Second Edition.* Hillsdale, NJ: Lawrence Erlbaum Associates

Constantinou, M. & McCaffrey, R. J. (2003). Using the TOMM for evaluating children's effort to perform optimally on neuropsychological measures. *Child Neuropsychology, 9*(2), 81-90. doi: 10.1076/chin.9.2.81.14505

Cragar, D. E., Berry, D. T. R., Fakhoury, T. A., Cibula, J. E., & Schmitt, F. A. (2006). Performance of patients with epilepsy or psychogenic non-epileptic seizures on four measures of effort. *The Clinical Neuropsychologist, 20,* 552-566. doi: 10.1080/13854040590947380

Creswell, J. (2007). *Qualitative Inquiry & Research Design: Choosing Among Five Approaches*. Thousand Oaks, CA: Sage Publications, Inc.

Dawes, R. M., Faust, D., & Meehl, P. E. (1989). Clinical versus actuarial judgment. *Science, 243*(4899), 1668-1674. Retrieved from http://xt9lp6eh4r.search.serialsso lutions.com

Demakis, G. J., Sweet, J. J., Sawyer, T. P., Moulthrop, M., Nies, K., & Clingerman, S. (2001). Discrepancy between predicted and obtained WAIS-R IQ scores discriminates between traumatic brain injury and insufficient effort. *Psychological Assessment, 13*(2), 240-248. doi: 10.1037//1040-3590.13.2.240

Donders, J. (2005). Performance on the Test of Memory Malingering in a mixed pediatric sample. *Child Neuropsychology, 11,* 221-227. doi: 10.1080/09297040490917298

Dunn, T. (2006). Current issues in response bias during neuropsychological assessment: Incomplete effort to malingering. In J. Dupri (Ed.), *Focus on neuropsychology research* (pp. 199-217). New York: Nova Sciences Publishers, Inc.

Dunn, T., Shear, P., Howe, S., & Ris, M. (2003). Detecting neuropsychological malingering: Effects of coaching and information. *Archives of Clinical Neuropsychology, 18*(2), 121-134. doi: 10.1093/arclin/18.2.121

Duhon, G. J., Noell, G. H., Witt, J. C., Freeland, J. T., Dufrene, B. A., & Gilbertson, D. N. (2004). Identifying academic skill and performance deficits: The experimental analysis of brief assessments of academic skills. *School Psychology Review, 33*(3), 429-443. Retrieved from http://0-search.proquest.com.source.unco.edu

Farkas, M., Rosenfeld, B., Robbins, R., & van Gorp, W. (2006). Do tests of malingering concur? Concordance among malingering measures. *Behavioral Sciences & the Law, 24,* 659-671. doi: 10.1002/bsl.730

Faust, D., Guilmette, T. J., Hart, K., Arkes, H. R., Fishburne, F. J., & Davey, L. (1988a). Neuropsychologists' training, experience, and judgment accuracy. *Archives of Clinical Neuropsychology, 3*, 145-163. doi: 10.1016/0887-6177(88)90060-1

Faust, D., Hart, K., & Guilmette, T. J. (1988b). Pediatric malingering: The capacity of children to fake believable deficits on neuropsychological testing. *Journal of Consulting and Clinical Psychology, 56*(4), 578-582. doi: 10.1037/0022-006X.56.4.578

Faust, D., Hart, K., Guilmette, T. J., & Arkes, H. R. (1988c). Neuropsychologists' capacity to detect adolescent malingerers. *Professional Psychology: Research and Practice, 19*(5), 508-515. doi: 10.1037/0735-7028.19.5.508

Flaro, L., Green, P., & Robertson, E. (2007). Word Memory Test failure 23 times higher in mild brain injury than in parents seeking custody: The power of external incentives. *Brain Injury, 21*(4), 373-383. doi: 10.1080/02699050701311133

Frederick, R. I. (1997). *Validity Indicator Profile*. San Antonio, TX: Pearson Education, Inc.

Green, P. (2005). *Green's Word Memory Test for Microsoft Windows® User's Manual.* Edmonton, Alberta, CA: Green's Publishing

Green, P., Allen, L. M., & Astner, K. (1996). *The Word Memory Test: A users guide to the oral and computer-administered forms, US Version 1.0.* Edmonton, Alberta, CA: Green's Publishing.

Green, P. & Flaro, L. (2003). Word Memory Test performance in children. *Child Neuropsychology, 9*(3), 189-207. doi: 10.1076/chin.9.3.189.16460

Green, P., Flaro, L., & Courtney, J. (2009). Examining false positives on the Word Memory Test in adults with mild traumatic brain injury. *Brain Injury, 23*(9), 741-750. doi: 10.1080/02699050903133962

Green, P., Iverson, G. L., & Allen, L. (1999). Detecting malingering in head injury litigation with the Word Memory Test. *Brain Injury, 13*(10), 813-819. doi: 10.1080/026990599121205

Green, P., Rohling, M. L., Lees-Haley, P. R., & Allen III, L. M. (2001). Effort has a greater effect on test scores than severe brain injury in compensation claimants. *Brain Injury,15*(12), 1045-1060. doi: 10.1080/02699040110088254

Greiffenstein, M. F., Greve, K. W., Bianchini, K. J., & Baker, W. J. (2008). Test of Memory Malingering and Word Memory Test: A new comparison of failure concordance rates. *Archives of Clinical Neuropsychology, 23,* 801-807. doi:10/10/16/j.acn.2008.07.005

Grove, W. M., & Meehl, P. E. (1996). Comparative efficiency of informal (Subjective, impressionistic) and formal (Mechanical, algorithmic) prediction procedures: The clinical-statistical controversy. *Psychology, Public Policy, and Law*, *2*(2), 293-323. doi:10.1037/1076-8971.2.2.293

Gunn, D., Batchelor, J., & Jones, M. (2010). Detection of simulated memory impairment in 6- to 11-year-old children. *Child Neuropsychology, 16,* 105-118. doi: 10.1080/09297040903352564

Harrison, A. G., Edwards, M. J., & Parker, K. C. H. (2007). Identifying students faking ADHD: Preliminary findings and strategies for detection. *Archives of Clinical Neuropsychology, 22*, 577-588. doi: 10.1016/j.acn.2007.03.008

Harrison, A. G., Edwards, M. J., & Parker, K. C. H. (2008). Identifying students feigning dyslexia: preliminary findings and strategies for detection. *Dyslexia, 14*, doi: 10.1002/dys.366

Haynes, S. & O'Brien, W. (2000). *Principles and practice of behavioral assessment.* New York, NY: Kluwer Academic/Plenum.

Heilbronner, R. L., Sweet, J. J., Morgan, J. E., Larrabee, G. J., Millis, S. R., & Conference Participants (2009). American academy of clinical neuropsychology consensus conference statement on the neuropsychological assessment of effort, response bias, and malingering. *The Clinical Neuropsychologist*, *23*, 1093-1129, doi: 10.1080/13854040903155063

Heubrock, D. & Petermann, F. (1998). Neuropsychological assessment of suspected malingering: Research results, evaluation techniques, and further directions of research and application. *European Journal of Psychological Assessment, 14*(3), 211-225, doi: 10.1027/1015-5759.14.3.211

Hosp, J. L. & Reschly, D. J. (2002). Regional differences in school psychology practice. *School Psychology Review, 31*(1), 11-29. Retrieved from http://0web.ebscohost.com.source.unco.edu

Hunt, T. N., Ferrara, M. S., Miller, L. S., & Macciocchi, S. (2007). The effect of effort on baseline neuropsychological test scores in high school football athletes. *Archives of Clinical Neuropsychology, 22,* 615-621. doi: 10.1016/j.acn.2007.04.005

Iverson, G. L., & Binder, L. M. (2000). Detecting exaggeration and malingering in neuropsychological assessment. *The Journal of Head Trauma Rehabilitation, 15,* 829-858. doi: 10.1097/00001199-200004000-00006

Jacob, S., & Hartshorne, T. S. (2007). *Ethics and law for school psychologists - Fifth edition*. Hoboken, NJ: John Wiley & Sons, Inc..

Kamphaus, R. W., Reynolds, C. R., & Imperato-McCammon, C. (1999). Roles of diagnosis and classification in school psychology. In C. R. Reynolds & T. B. Gutkin (Eds.), *The handbook of school psychology (3rd ed., pp. 292-306).* New York: John Wiley & Sons, Inc.

Kaplan, R., & Saccuzzo, D. (2001). *Psychological testing: Principles, applications, and issues*. Belmont, CA: Wadsworth/Thomson Learning.

Kim, N., Boone, K. B., Victor, T., Lu, P., Keatinge, C., & Mitchell, C. (2010). Sensitivity and specificity of a digit symbol recognition trial in the identification of response bias. *Archives of Clinical Neuropsychology, 25*, 420-428. doi: 10.1093/arclin/acq040

Kirkwood, M. W. & Kirk, J. W. (2010). The base rate of suboptimal effort in a pediatric mild TBI sample: Performance on the Medical Symptom Validity Test. *The Clinical Neuropsychologist, 24*, 860-872. doi: 10.1080/13854040903527287

Kirkwood, M. W., Hargrave, D. D., & Kirk, J. W. (2011). The value of the WISC-IV digit span subtest in detecting noncredible performance during pediatric neuropsychological examinations. *Archives of Clinical Neuropsychology, 26*, 377-384. doi: 10.1093/arclin/acr040

Kirkwood, M., Yeates, K., Randolph, C., & Kirk, J. (2012). The implications of symptom validity test failure for ability-based test performance in a pediatric sample. *Psychological Assessment, 24*(1), 36-45. doi: 10.1037/a0024628

Lange, R. T., Iverson, G. L., Sullivan, K., & Anderson, D. (2006). Suppressed working memory on the WMS-III as a marker for poor effort. *Journal of Clinical and Experiemental Neuropsychology, 28*, 294-305. doi: 10.1080/13803390490918156

Larrabee, G. L. (2007). *Assessment of malingered neuropsychological deficits.* New York, NY. Oxford University Press, Inc.

Laurent, J. (1997). Characteristics of the standard and supplemental batteries of the Woodcock-Johnson Tests of Cognitive Ability-Revised with a college sample. *Journal of School Psychology, 35*(4), 403-416. doi: 10.1016/S0022-4405(97)00015-0

Lezak, M. D., Howieson, D. B., & Loring, D. W. (2004). *Neuropsychological assessment* (4th ed.). New York, NY: Oxford University Press, Inc.

Lu, P. H. & Boone, K. B. (2002). Suspect cognitive symptoms in a 9-year-old child: Malingering by proxy? *The Clinical Neuropsychologist, 16,* 90-96. doi: 10.1076/clin.16.1.90.8328

MacAllister, W. S., Nakhutina, L., Bender, H. A., Karantzoulis, S., & Carlson, C. (2009). Assessing effort during neuropsychology evaluation with the TOMM in children and adolescents with epilepsy. *Clinical Neuropsychology, 15*(6), 521-531 doi: 10.1080/09297040902748226

Machek, G. R. & Nelson, J. M. (2010). School psychologists' perceptions regarding the practice of identifying reading disabilities: Cognitive assessment and response to intervention considerations. *Psychology in the Schools, 47*(3), 230-245. doi: 10.1002/pits.20467

McConaughy, S. H. (2005). Direct observational assessment during test sessions and child clinical interviews. *School Psychology Review, 34*(4), 490-506. Retrieved from http://0-vnweb.hwwilsonweb.com.source.unco.edu

McGrath, R. E., Mitchell, M., Kim, B. H., & Hough, L. (2010). Evidence for response bias as a source of error variance in applied assessment. *Psychological Bulletin, 136*(3), 450-470. doi: 10.1037/a0019216

McLachlan, G. J. (1992). *Discriminant Analysis and Statistical Pattern Recognition.* New York, NY: Wiley-Interscience

Merriam, S. B. (1998). *Qualitative Research and Case Study Applications in Education*. San Francisco, CA: Jossey-Bass Publishers.

Millon, T., Millon, C., Davis, R., & Grossman, S. (2009). *Millon Clinical Multiaxial Inventory –Third Edition*. San Antonio, TX: Pearson Education, Inc.

Mittenberg, W., Patton, C., Canyock, E. M., & Condit, D. C. (2002). Base rates of malingering and symptom exaggeration. *Journal of Clinical and Experimental Neuropsyholgy, 24*(8), 1094-1102. doi: 10.1076/jcen.24.8.1094.8379

Morey, L. C. (2007). *The Personality Assessment Inventory professional manual*. Lutz, FL: Psychological Assessment Resources.

Morgan, J. E. , & Sweet, J. J. (Ed.). (2009). *Neuropsychology of malingering casebook.* New York, NY: Psychology Press.

National Association of School Psychologists. (2010). *Principles for Professional Ethics*. Retrieved from http://www.nasponline.org/standards/2010standards/1_%20Ethical%20Principles.pdf

Ord, J. S., Greve, K. W., Bianchini, K. J. & Aguerrevere, L. E. (2009). Executive dysfunction in traumatic brain injury: The effects of injury severity and effort on the Wisconsin Card Sorting Test. *Journal of Clinical and Experimental Neuropsychology,* doi: 10.1080/13803390902858874

Rees, L., Tombaugh, T., Gansler, D., & Moczynski, N. (1998). Five validation experiments of the Test of Memory Malingering (TOMM). *Psychological Assessment, 10,* 10-20. doi: 10.1037/1040-3590.10.1.10

Rogers, R. (1984). Towards an empirical model of malingering and deception. *Behavioral Sciences and the Law, 2*(1), 93-111. doi: 10.1002/bsl.2370020109

Sattler, J. M. (2001). *Assessment of children: cognitive applications - fourth edition*. San Diego, CA: Jerome M. Sattler, Publisher, Inc.

Sattler, J. M. (2008). *Assessment of children: cognitive foundations – fifth edition.* San Diego, CA: Jerome M. Sattler, Publisher, Inc.

Sharland, M. J., & Gfeller, J. D. (2007). A survey of neuropsychologists' beliefs and practices with respect to the assessment of effort. *Archives of Clinical Neuropsyhology, 22*, 213-223. doi: 10.1016/j.acn.2006.12.004

Sollman, M. J., Ranseen, J. D., & Berry, D. T. R. (2010). Detection of feigned ADHD in college students. *Psychological Assessment, 22*(2), 325-335. doi: 10.1037/a0018857

Stevens, J. (1984). *Applied multivariate statistics for the social sciences.* Hillsdale, NJ: Lawrence Erlbaum Associates, Publishers

Sullivan, B. K., May, K., & Galbally, L. (2007). Symptom exaggeration by college adults in attention-deficit hyperactivity disorder and learning disorder assessments. *Applied Neuropsychology, 14*(3), 189-207. doi: 10.1080/09084280701509083

Teglasi, H., & Freeman, R. W. (1983). Rapport pitfalls of beginning testers. *Journal of School Psychology, 21*, 229-240. doi: 10.1016/0022-4405(83)90018-3

Tombaugh, T. (1996). *Test of Memory Malingering*. North Tonawanda, NY: Multi-Health Services.

VanDerHeyden, A. M. & Witt, J. C. (2008). Best practices in can't do/won't do assessment. In Thomas, A. & Grimes, J. (Eds.), *Best Practices in School Psychology V* (pp. 131-140). Bethesda, MD: NASP Publications

VanDerHeyden, A. M., Witt, J. C., & Naquin, G. (2003). Development and validation of a process for screening referrals to special education. *School Psychology Review, 32*(2), 204-227. Retrieved from http://xt9lp6eh4r.search.serialssolutions.com

Wechsler, D. (1999). *Wechsler Abbreviated Scale of Intelligence*. San Antonio, TX: Pearson Education, Inc.

Wechsler, D. (2008). *Wechsler Adult Intelligence Scale – Fourth Edition*. San Antonio, TX: Pearson Education, Inc.

Weinborn, M., Orr, T., Woods, S., Conover, E., & Feix, J. (2003). A validation of the Test of Memory Malingering in a forensic psychiatric setting. *Journal of Clinical and Experimental Neuropsychology, 25*, 979-990. doi: 10.1076/jcen.25.7.979.16481

Whitney, K. A., Hook, J. N., Steiner, A. R., Shepard, P. H., & Callaway, S. (2008). Is the Rey 15-Item Test II (Rey II) a valid symptom validity test?: Comparison with the TOMM. *Applied Neuropsychology, 15*, 287-292. doi: 10.1080/09084280802325215

Witmer, J. M., Bornstein, A. V., & Dunham, R. M. (1971). The effects of verbal approval and disapproval upon the performance of third and fourth grade children on four subtests of the Wechsler Intelligence Scale for Children. *Journal of School Psychology, 9*(3), 347-356. doi: 10.1016/0022-4405(71)90093-8

Wolf, L. F., & Smith, J. K. (1995). The consequence of consequence: Motivation, anxiety, and test performance. *Applied Measurement in Education, 8*(3), 227-252. doi: 10.1207/s15324818ame0803_3

Young, J. C., Caron, J. E., Baughman, B. C., & Sawyer, R. J. (2012). Detection of suboptimal effort with Symbol Span: Development of a new embedded index. *Archives of Clinical Neuropsychology, 27*, 159-164. doi: 10.1093/arclin/acr109

Zeldow, P. (2009).In defense of clinical judgment, credentialed clinicians, and reflective practice. *Psychotherapy Theory, Research, Practice, Training, 46*, 1-10. doi: 10.1037/a0015132

APPENDIX A

INSTITUTIONAL REVIEW BOARD APPROVAL

UNIVERSITY of NORTHERN COLORADO
Institutional Review Board (IRB)

February 24, 2012

TO: Gary Heise
School of Sport and Exercise Science

FROM: The Office of Sponsored Programs

RE: Exempt Review of *An Experimental Study Examining Observational and Objective Methods of Assessing Effort in Undergraduate Populations,* submitted by Aaron Schrader (Research Advisor: Thomas Dunn)

The above proposal is being submitted to you for exemption review. When approved, return the proposal to Sherry May in the Office of Sponsored Programs.

I recommend approval.

Gary D. Heise 12 Mar 2012
Signature of Co-Chair Date

The above referenced prospectus has been reviewed for compliance with HHS guidelines for ethical principles in human subjects research. The decision of the Institutional Review Board is that the project is exempt from further review.

IT IS THE ADVISOR'S RESPONSIBILITY TO NOTIFY THE STUDENT OF THIS STATUS.

Comments:

25 Kepner Hall ~ Campus Box #143
Greeley, Colorado 80639
Ph: 970.351.1907 ~ Fax: 970.351.1934

APPENDIX B

INFORMED CONSENT FOR CONFEDERATES

UNIVERSITY of
NORTHERN COLORADO

CONSENT FORM FOR HUMAN PARTICIPANTS IN RESEARCH
UNIVERSITY OF NORTHERN COLORADO

Project Title: An experimental study examining observational and objective methods of assessing effort in undergraduate populations

Lead Researcher: Aaron Schrader B.A., School of Applied Psychology and Counselor Education
Phone: 720-771-2396 E-mail: schr2990@bears.unco.edu
Research Advisor: Dr. Thomas Dunn
Phone: 970-351-1501 E-mail: Thomas.dunn@unco.edu

Purpose and Description: The primary purpose of this study is to examine the assessment of effort and response bias through two means, observationally and objectively. During this single session, I will begin by collecting basic demographic information. I will then provide you specific instructions that will tell you how I would like you to approach taking portions of commonly administered cognitive measures. Results from these measures are only analyzed to ensure that adherence to instructions and will not be used in any other way. You will then be video recorded taking portions of cognitive measures.

Although complete anonyminity is impossible given the video recordings, your name will not be stated during the session. We will assign a subject number to you and your name will not be linked in any way to this number. As mentioned previously, the data collected will not be used in any other fashion than to ensure adherence to the instructions provided. The video recordings and data for this study will be kept at the residence of the lead investigator.

Potential risks in this project are minimal. As when completing an exam like instrument, one may experience some stress. However, as your participation is voluntary, you may decide to stop and withdraw at any time.

School psychologists as well as graduate programs in school psychology will be the populations who most benefit from the results of this study. This study is intended to begin examining the phenomenon of response bias and effort within the field of school psychology.

As mentioned previously, your participation is voluntary. You may decide not to participate in this study and if you begin participation you may still decide to stop and withdraw at any time. Your decision will be respected. Having read the above and having had an opportunity to ask any questions, please sign below if you would like to participate in this research. A copy of this form will be given to you to retain for future reference. If you have any concerns about your selection or treatment as a research participant, please contact the Office of Sponsored Programs, Kepner Hall, University of Northern Colorado Greeley, CO 80639; 970-351-2161.

_____

Subject's Signature Date

_____

Researcher's Signature Date

APPENDIX C

INFORMED CONSENT FOR PARTICIPANTS

UNIVERSITY of
NORTHERN COLORADO

CONSENT FORM FOR HUMAN PARTICIPANTS IN RESEARCH
UNIVERSITY OF NORTHERN COLORADO

Project Title: An experimental study examining observational and objective methods of assessing effort in undergraduate populations
Researcher: Aaron Schrader, B.A., School of Applied Psychology and Counselor Education
Phone: 720-771-2396 E-mail: schr2990@bears.unco.edu
Advisor: Thomas Dunn, Ph.D., Department of Psychology
Advisor Phone: 970-351-1501 E-mail: thomas.dunn@unco.edu

Purpose and Description: The primary purpose of this study is to examine the assessment of effort and response bias through two means, observationally and objectively. During this single session, I will begin by collecting basic demographic information and information regarding your graduate training program. I will then ask you to watch a series of video recordings of individuals completing portions of commonly administered cognitive measures. Upon completion of these recordings, I will ask you to then rate the effort you perceived the individual to put forth during the given task(s) as well as describe what behaviors you observed to lead you to your conclusion. Overall, your participation in the study should take approximately 120 minutes.

We will take every precaution in order to protect your anonymity. We will assign a subject number to you and your name will not be linked in any way to this number. Data collected and analyzed for this study will be kept at the residence of the lead investigator with the exception of this consent form which will be kept with the research advisor.

Potential risks in this project are minimal due to the observational nature of the data collection. As when observing any individual or completing an exam-like instrument, one may experience some stress. However, as your participation is voluntary, you may decide to stop and withdraw at any time.

School psychologists as well as graduate programs in school psychology will be the populations who most benefit from the results of this study. This study is intended to begin examining the phenomenon of response bias and effort within the field of school psychology.

Participation is voluntary. You may decide not to participate in this study and if you begin participation you may still decide to stop and withdraw at any time. Your decision will be respected and will not result in loss of benefits to which you are otherwise entitled. Having read the above and having had an opportunity to ask any questions, please sign below if you would like to participate in this research. A copy of this form will be given to you to retain for future reference. If you have any concerns about your selection or treatment as a research participant, please contact the Office of Sponsored Programs, Kepner Hall, University of Northern Colorado Greeley, CO 80639; 970-351-2161.

_____
Subject's Signature Date

_____
Researcher's Signature Date

APPENDIX D

PARTICIPANT DEMOGRAPHIC DATA FORM

Participant Information

Please provide the following demographic information.

Participant #: _____

Age: _____

Sex: (Circle)

Male Female

Highest degree currently held (ex. MA in School Counseling):

_____

Graduate program:

_____

University:

_____

Year in the program:

_____

Have you received graduate level training specifically in the area of:

Cognitive (IQ) Assessment: Y N

Effort / Malingering: Y N

APPENDIX E

PARTICIPANT RATING FORM

Participant Rating Form

Video # _____

Have you completed specialized training in cognitive assessment? (Circle one)

YES NO

Part A:

1. **How would you rate this individual's effort they put forth during the video on the following scale?**

Please circle ONE number:

1 2 3 4 5 6 7 8 9

1 = Least amount of effort an individual could put forth
9 = Most amount of effort an individual could put forth

2. **How would you rate this individual's overall level of effort they put forth during the video? (Circle One)**

Adequate Effort Inadequate effort

Part B:

What behaviors did you observe that lead you to your rating? Please describe below.

_____

_____

_____

_____

_____

APPENDIX F

PHASE ONE STATISTICS

Phase One Statistics

Table 11

Demographics for Phase One Confederates

Variable	N	%
Gender		
Male	4	50.0
Female	4	50.0
Age		
18	2	25.0
19	5	62.5
20	1	12.5

Table 12

Descriptive Statistics for Phase One Performance

	WASI	TOMM			WMT		Self-Report
Confederate	BD T-Score	T1%C	T2%C	IR%C	DR%C	CNS%C	Likert Rating
				Full Effort			
1	58	100	100	100	97.5	97.5	9.0
2	55	100	100	100	100	100	9.0
3	64	98.0	100	100	100	100	8.0
4	59	90.0	100	97.5	100	97.5	8.0
Mean	59	97.0	100	99.4	99.4	98.8	8.5
SD	3.7	4.8	0	1.3	1.3	1.4	0.6
				Suboptimal Effort			
1	31	50.0	58.0	57.5	60.0	57.5	8.0
2	39	92.0	98.0	52.5	62.5	55.0	5.0
3	60	80.0	82.0	95.0	77.5	77.5	6.0
4	51	78.0	100	55.0	97.5	52.5	6.0
Mean	45.3	75.0	84.5	65.0	74.4	60.6	6.3
SD	12.8	17.8	19.4	20.1	17.2	11.2	1.3

Note. BD T-Score = Block Design T-Score; T1%C = Trial One Percent Correct; T2%C = Trial Two Percent Correct; IR%C = Immediate Recall Percent Correct; DR%C = Delayed Recall Percent Correct; CNS%C = Consistency Percent Correct; SD = Standard Deviation

APPENDIX G

SUMMARY OF STUDY IN ARTICLE FORMAT

An Experimental Study Examining Observational and Objective Methods of Assessing Effort in an Undergraduate Sample

Aaron G. Schrader

University of Northern Colorado

Abstract

This study examined the assessment of effort and response bias through two means, observationally and objectively, with confederates who were instructed to put forth their best effort or suboptimal effort. While being video recorded, eight confederates were administered Block Design subtest from the Wechsler Abbreviated Scales of Intelligence (WASI) along with two well established tests of effort, the Word Memory Test (WMT) and Test of Memory Malingering (TOMM). The videos were then viewed by a total of 30 graduate students and the perceived effort by the undergraduate students was then rated using a Likert-type scale, a dichotomous rating scale, and a single open-ended question. Similar to previous research, participants had difficulty in accurately detecting suboptimal effort. More specifically, results illustrated that the ratings had a high level of misclassification of suboptimal effort but accurately identified all confederates putting forth adequate effort. Discriminant analysis of the results showed marginal improvement in the detection of suboptimal effort in comparison to chance probability. The results provide a rationale for the addition of research based methods of effort testing within evaluations as well as implications for inclusion of training in such methods in school psychology training programs.

Keywords: response bias, suboptimal effort, detection, discriminant analysis

Introduction

Understanding that assessment remains a dominant practice in school psychology, one must consider ethical obligations of appropriate evaluation techniques. According to the American Psychological Association (APA) ethical standards 9.01 and 9.02 (APA, 2010) and NASP ethical standard II.3.2 (NASP, 2010), psychologists are expected to use appropriate and responsible assessment procedures (Jacob & Hartshorne, 2007). Therefore, response bias should be evaluated in an appropriate and responsible manner, just as a clinician would gather data on other aspects of cognitive or emotional functioning. Response bias is general or umbrella term that is multifarious and can include many factors related to an individual which will affect or influence the way that individual responds to a given task or situation. This term encapsulates many different constructs depending on the task at hand which can include but is not limited to negative impression management, positive impression management, suboptimal effort, feigning, and malingering (Dunn, 2006; McGrath, Mitchell, Kim, & Hough, 2010).

Although some might believe that the different forms of response bias are a more significant issue in adult populations, research has also illustrated that children and adolescents are adept at feigning commonly encountered disorders (Faust, Hart, & Guilmette, 1988a; Faust, Hart, Guilmette, & Arkes, 1988b). More current research continues to examine the phenomena of pediatric malingering, as well as other forms of response bias, within psychological evaluations. A study by Lu and Boone (2002) illustrated that children as young as nine were capable of feigning cognitive deficits. Another study which examined 193 pediatric patients who had sustained mild traumatic brain injuries, it was discovered that suboptimal effort had a base rate of 17% (Kirkwood

& Kirk, 2010). Because of this potential to exhibit response bias, many efforts have also been made to provide evidence for valid use of already established tests of effort as well as embedded measures with pediatric populations (Donders, 2005; Blaskewitz, Merten, & Kathmann, 2008); Gunn, Batchelor, & Jones, 2010; Kirkwood, Hargrave, & Kirk, 2011; Brooks, Sherman, & Krol, 2012). It is important to keep in mind, however, that children who do not put forth maximum effort may be exhibiting response bias due to less sinister issues such as fatigue or disinterest. Nonetheless, establishing an efficient and accurate means of evaluating effort within pediatric psychological evaluations continues to be an important area of focus within research.

Effort that is exerted by the individual during a psychological evaluation can be assessed through many different methods. Specifically, within the fields of clinical, forensic, and neuropsychology, multiple methods of evaluating effort have been created and are used in practice regularly. Mittenberg, Patton, Canyock, and Condit (2002), as well as Sharland and Gfeller (2007), reviewed the various methods used within psychological practice; they identified nine to ten different approaches to support diagnostic impression of response bias depending on the study. These methods included severity of impairment which is inconsistent with condition, inconsistent performance on cognitive tests, performance scores below empirical cutoffs on forced-choice tests, discrepancies among reported and observed behavior, unexplainable reported symptoms in an interview, performance scores below empirical cutoffs on other malingering tests, unexplainable changes in performance across multiple examinations, and performance scores above validity scale cutoffs on personality tests (Mittenberg et al., 2002; Sharland & Gfeller, 2007).

Research from the field of neuropsychology has also examined clinicians' ability to detect malingering or "faking bad" through informal means. Participants in these studies where provided a brief background summary of the individual, cognitive test results, and neuropsychological test results. They were then asked to review the case to which no clinician was able to detect malingering beyond chance probability (Faust et al., 1988a; Faust et al., 1988b). These research studies provide a clear impetus for additional research and a call for attention in the assessment of response bias whenever a professional is providing a comprehensive evaluation of an individual.

On a daily basis, school psychologists within secondary education settings must make decisions about test accommodations, qualification for special education, and other types of support programs such as vocational rehabilitation. Because of this, it is important to consider the ways that response bias might affect student performance in the school setting and ultimately the overall validity of results gained from school based evaluations. Unfortunately, in the field of school psychology there is a lack of research pertaining to the assessment of effort and response bias during school based evaluations for special education services.

The purpose of this study was to examine the ability of observers (i.e., graduate students in an applied field of psychology) to assess and evaluate effort through informal observational means and compare their ratings to standardized measures of effort. Specifically, psychology graduate students with or without training in assessment evaluated the perceived effort put forth by undergraduates on a common cognitive task that were also instructed to either put forth their best effort or exert suboptimal effort on two different items. The first item was a nine point Likert-type scale which required

participants to rate the perceived level of effort from one to nine with one being the least amount of effort an individual could put forth and nine being the maximum level of effort. The other item was a dichotomous rating that required participants to rate the perceived effort as either adequate or inadequate.

The Likert-type scale observational ratings served as the independent variable within the study and the condition assignment served as the dependent variable. Using these variables a statistical procedure of discriminant analysis was completed to allow for a model that predicts group membership to be created and evaluated. The jackknife method within discriminant analysis allows for a systematic evaluation of classification models to occur. This method withholds an independent variable from the model and attempts to classify individuals based on the model created by the remaining independent variables. This process is repeated with every independent variable one at a time. The final result is an average of all classification models created to form a final model. The following hypotheses will be tested using descriptive statistics as well as discriminant analysis with the jackknife method:

H1 The observations by the participants are not able to discriminate between full effort and suboptimal effort participants.

H2 The observations by the participants does not add to the overall model in the differential prediction of full effort or suboptimal effort participants.

Fisher's exact test allows researchers to examine if there is a significant difference that exists between two classification groups. Within this statistical analysis, the groups of graduate student participants with training in assessment or without training in assessment were examined for differences based on the observational dichotomous rating scale. These differences were examined as simulated response bias research has not

incorporated participants that have no training with any form of assessment. The following hypothesis will be tested using the Fisher's exact test:

H3 No differences exist between the graduate students with training in assessment group and the graduate students without training in assessment group.

Lastly, using thematic analysis, researchers were able to identify common categories or trends among responses. Within the current study, participants were asked to identify the salient features which informed their observational ratings in an open format. These categories were created through the use of information from the open ended question that that was coded and analyzed for frequency among participants (Merriam, 1998). No specific hypothesis was tested through the open ended question.

Methods

Prior to beginning the study, permission was obtained from the university Institutional Review Board (IRB). This study utilized two phases where the first involved the generation of video stimuli for the participants to rate in the second phase. The specific procedures within each phase are delineated below.

Phase One Confederates

A total of nine undergraduate students were utilized within this study as confederates. These college students were deemed to be an appropriate population for school psychology students to observe and rate as they may have been served by a school psychologist in a secondary education setting a year or two prior. They were obtained through a volunteer sampling of an undergraduate psychology participant pool and were provided course credit for their participation.

Phase One Conditions

The undergraduate participants were randomly assigned into one of two conditions which include full effort and suboptimal effort groups. Subjects within the full effort condition were presented with the following instructions:

> "Today I would like you to try your best on all of the measures you are given. If you are unsure of an answer, please feel free to take guesses. It is important that you try your best on all the items."

The subjects within the suboptimal effort condition presented with the following instructions:

> "Today I would like you to pretend that you are trying very hard when completing all of the measures, but, in reality, you are not performing to the best of your ability. It is important that you still complete the tasks you are given. Some tasks may be easy but some may be more difficult. It is important that the examiner thinks that you are trying your best, but once again please do not perform well on the measures."

Phase One Procedures

The confederate undergraduates were administered the following instruments in the consistent order of the Word Memory Test (WMT: Green, 2005), Block Design subtest from the Wechsler Abbreviated Scale of Intelligence (WASI: Wechsler, 1999), and the Test of Memory Malingering (TOMM: Tombaugh, 1996). Only the administration of the Block Design subtest was videotaped to be used within phase two. Basic demographic information was also collected at this time and included gender, age, and history of head injury. Following completion of the measures, one videotaped administration was excluded as the confederate performed at a level deemed inappropriate for the condition according to the scores on both the WMT and TOMM.

Phase Two Participants

The participants selected for this study were collected through a convenience sample at a public university. The participants in this study were graduate level students. A total of 30 graduate level students that attend nationally accredited school psychology, counseling psychology, counselor education and supervision, educational psychology, and special education programs were used within the study. Two groups of graduate students were formed based on the criteria of having completed coursework in cognitive assessment or having not yet been exposed to such coursework. The group of graduate students that have not yet been exposed to coursework in cognitive assessment served as the control group. Participants were provided monetary compensation for their participation in the study.

Phase Two Procedures

The two groups of graduate student participants were first given consent forms and asked to provide demographic information about themselves which included gender, age, graduate degrees currently held, and graduate program currently enrolled in. The participants were also asked to indicate whether they had received graduate level training in cognitive assessment or the assessment of effort/malingering. Participants then viewed each of the eight video recordings in a consistent order of the Block Design subtest administration and were asked to complete the two parts of the observational rating of effort immediately following the completion of the video. The observational ratings included a Likert-type scale rating of perceived effort, a dichotomous rating of effort, and an open ended question.

Results

Presented in this section are the results of the analyses completed for this study. It is divided into eight sections: (a) phase one confederate demographics, (b) phase one description, (c) phase two sample description, (d) accuracy rates, (e) Fisher's exact test, (f) discriminant analysis, and (g) open-ended question analysis.

Confederate Demographics

Demographic data regarding the confederates used within the first phase of this study are shown in Table 1. The age range for the included confederates within the study was from 18-20 years of age. The total amount of confederates consisted of eight undergraduate students including four females (50%) and four males (50%) with the mean age of 18.875 and a standard deviation of 0.64 years.

Table 1

Demographics for Phase One Confederates

Variable	N	%
Gender		
Male	4	50.0
Female	4	50.0
Age		
18	2	25.0
19	5	62.5
20	1	12.5

Confederate Performance Descriptive Statistics

Table 2 represents the performance of the confederates on the Block Design subtest, TOMM, and WMT within their respective conditions of either full or suboptimal effort as well as their adherence to directions self-report rating. As mentioned previously, scores on the second trial of the TOMM below 90% correct are deemed to be within a range of suboptimal effort or the indication of possible malingering behaviors. Scores on any of the WMT subscales that are below 82.5% correct are considered failing and indicative of suboptimal effort or possible malingering behaviors.

After examination of these scores, trends in the data suggested that the full effort confederate group scores were higher in every measure than the suboptimal effort confederate group. Each confederate within the full effort group scored within acceptable ranges across the measures of effort. Conversely, each confederate within the suboptimal group scored within the fail range on at least one area of the WMT. The adherence to directions self-report ratings, which provided an indication of the confederates were also higher within the full effort group in comparison to the suboptimal group.

Table 2

Descriptive Statistics for Phase One Performance

	WASI	TOMM		WMT			Self-Report
Confederate	BD T-Score	T1%C	T2%C	IR%C	DR%C	CNS%C	Likert Rating
				Full Effort			
1	58	100	100	100	97.5	97.5	9.0
2	55	100	100	100	100	100	9.0
3	64	98.0	100	100	100	100	8.0
4	59	90.0	100	97.5	100	97.5	8.0
Mean	59	97.0	100	99.4	99.4	98.8	8.5
SD	3.7	4.8	0	1.3	1.3	1.4	0.6
				Suboptimal Effort			
1	31	50.0	58.0	57.5	60.0	57.5	8.0
2	39	92.0	98.0	52.5	62.5	55.0	5.0
3	60	80.0	82.0	95.0	77.5	77.5	6.0
4	51	78.0	100	55.0	97.5	52.5	6.0
Mean	45.3	75.0	84.5	65.0	74.4	60.6	6.3
SD	12.8	17.8	19.4	20.1	17.2	11.2	1.3

Note. BD T-Score = Block Design T-Score; T1%C = Trial One Percent Correct; T2%C = Trial Two Percent Correct; IR%C = Immediate Recall Percent Correct; DR%C = Delayed Recall Percent Correct; CNS%C = Consistency Percent Correct; SD = Standard Deviation

Participant Demographics

Demographic data regarding the total sample of the study of graduate student participants are shown in Table 3. The age range included participants from 23-41 years of age. The total sample size consisted of 30 graduate students including 24 females (80%) and 6 males (20%) with the mean age of 29 years and a standard deviation of 4.97 years. The total sample of participants all attended the same university and were actively enrolled in graduate programs in the areas of school psychology, counseling psychology, counselor education and supervision, and special education. The two participant groups used within the study consisted of 15 participants having completed training in cognitive assessment and 15 participants who have not received training in cognitive assessment.

Table 3

Demographics for the Total Sample

Variable	N	%
Gender		
Females	24	80
Males	6	20
Graduate Program		
School Psychology	12	40.0
Counseling Psychology	6	20.0
Counselor Education and Supervision	4	13.3
Special Education	8	26.7
Cognitive Assessment Training		
Yes	15	50
No	15	50

Observation Accuracy Rates

The overall accuracy rates of the dichotomous ratings made by the groups of participants are displayed in Table 4. Within the sample group that has received training in cognitive assessment, a total of 60 observations (100%) were rated correctly as exhibiting full effort and 5 observations (8.3%) were rated correctly as exhibiting suboptimal effort. The group that has not received training in cognitive assessment had a total of 60 observations (100%) were rated correctly as exhibiting full effort and 9 observations (15%) were rated correctly as exhibiting suboptimal effort. Overall, a total

of 120 observations (100%) were rated correctly as exhibiting full effort and 14 observations (11.7%) were rated correctly as exhibiting suboptimal effort.

Table 4

Accuracy Rates of Dichotomous Ratings

	With Training	Without Training	Total Identified
Full Effort	60	60	120
	100%	100%	100%
Suboptimal	5	9	14
	8.3%	15.0%	11.7%

Fisher's Exact Test

Fisher's exact test is a procedure that is utilized within experimental studies to see if the differences in the data between groups are due to chance probability or another factor influencing the results. In other words, if there is a significant difference between classification rates within the data, one can reject the null hypothesis of no differences exist between classification rates. This statistical procedure allows for analysis of small sample sizes. Within the current study, the dichotomous ratings of effort by the groups of participants provided a means to compare against the known condition assignment of each confederate. Fisher's exact test was conducted for the full effort and suboptimal effort conditions using the data from correct and incorrect dichotomous ratings in each condition. Contingency tables for the full effort and suboptimal effort conditions are listed below in Tables 3 and 4, respectively. Data from the full effort condition Fisher's exact test revealed that the calculated p-value is 1 illustrating a lack of significance. This

finding illustrated that the null hypothesis cannot be rejected as significant differences do not exist between groups in the full effort condition beyond chance probability. The lack of significant differences between groups illustrates that both groups performed similarly in their dichotomous ratings of effort in the full effort condition.

Data from the suboptimal effort condition Fisher's exact similarly revealed that the calculated *p*-value of 0.3945 is not within the statistically significant range. This finding also illustrated that the null hypothesis cannot be rejected as significant differences do not exist between groups in the suboptimal effort condition beyond chance probability. An effect size estimation using Cramer's phi also illustrated a calculated value of 0.104 which is considered to be small. As noted in the previous section, the accuracy rates of detecting suboptimal effort were low in both groups illustrating a similar difficulty in assessing effort through observational means regardless of training in cognitive assessment. To sum, participants with training in cognitive assessment were not significantly different in the accuracy ratings than those without cognitive assessment.

Table 5

Contingency Table for Full Effort Condition

Full Effort	With Training	Without Training
Correct Identification	60	60
Incorrect Identification	0	0

Table 6

Contingency Table for Suboptimal Effort Condition

Suboptimal Effort	With Training	Without Training
Correct Identification	5	9
Incorrect Identification	55	51

Discriminant Analysis

Classification summaries for each group of participants (with or without training in cognitive assessment) are displayed as the jackknife method of discriminant analysis was completed two times to provide a means of comparison between the sample groups. The classification summaries for the jackknife holdout method of discriminant analysis using a linear function are listed within Tables 7 for the sample group with training in cognitive assessment. Additionally, the error count estimates for the linear function are listed in Tables 8 for the same sample group. The classification summaries for the linear discriminant function using the sample group without training in cognitive assessment are displayed within Tables 9. The error count estimates for the same group for the linear discriminant function are listed in Tables 10 as well.

Using the linear discriminant function from the sample group with training in cognitive assessment, a total of 47 observations (78.33%) were correctly classified as being within the full effort condition while a total of 31 observations (51.67%) were correctly classified as being within the suboptimal effort condition. Overall, a total probability of misclassification using the linear discriminant function was 0.35.

Table 7

Cognitive Training Group Classification Summary using Linear Discriminant Function

Condition	Full	Suboptimal	Total
Full	47	13	60
	78.33%	21.67%	100.00%
Suboptimal	29	31	60
	48.33%	51.67%	100.00%
Total	76	44	120
	63.33%	36.67%	100.00%
Priors	0.5	0.5	

Table 8

Cognitive Training Group Error Count Estimates using Linear Discriminant Function

	Full	Suboptimal	Total
Rate	0.2167	0.4833	0.3500
Priors	0.5000	0.5000	

The linear discriminant function which was derived from the without training group data (without cognitive training) resulted in a total of 49 observations (81.67%) that were correctly classified as being within the full effort condition while a total of 34 observations (56.67%) were correctly classified as being within the suboptimal effort condition. Overall, a total probability of misclassification using the linear discriminant function was 0.3083.

Table 9

Without Training Group Classification Summary using Linear Discriminant Function

Condition	Full	Suboptimal	Total
Full	49	11	60
	81.67%	18.33%	100.00%
Suboptimal	26	34	60
	43.33%	56.67%	100.00%
Total	75	45	120
	62.50%	37.50%	100.00%
Priors	0.5	0.5	

Table 10

Without Training Group Error Count Estimates using Linear Discriminant Function

	Full	Suboptimal	Total
Rate	0.1833	0.4333	0.3083
Priors	0.5000	0.5000	

Descriptive Analysis

From the response data within the adequate effort ratings, a total of five major categories could be identified. The first category noted referred to the individual being observed to be persistent or not giving up prematurely. This category also included responses that referred to trying multiple approaches to solving the item that was presented. The second category seen within the data referred to the speed an individual

would respond to on a given item. Interestingly, speed was noted to be an attribute for adequate effort whether the person responded quickly or slowly. The third category noted within the data referred to the accuracy or number of correct responses an individual was observed to display during the video. The fourth category within the data referred to an individual's checking of completed work and self-correcting errors. The fifth and final category noted within the adequate effort ratings data referred to individuals who leaned forward during tasks.

Responses to the open-ended question on ratings of inadequate effort revealed a much different picture as very few participants perceived any of the confederates to put forth inadequate effort. The lack of data did not allow for a very robust descriptive analysis. A single category was elicited from this data which referred to behaviors associated with distractibility or disengagement with the task. These behaviors included checking the cell phone or "fiddling" with one's hair, ring, or shoes.

Discussion

Research has demonstrated that psychologists have difficulty in assessing effort or malingering across populations. The findings from this current study corroborated this same difficulty using only observational methods in the form of dichotomous ratings and Likert-type scale ratings. Overall, accuracy rates of effort ratings through dichotomous ratings were found to be low in the correct identification of individuals putting forth suboptimal effort. Accuracy rates of adequate effort ratings were also vastly overestimated across the sample as there was evidence of a high false positive rate. Examination of the accuracy rates of dichotomous ratings by sample group revealed that control group performed slightly better (15% accuracy compared to 8.3% accuracy) than

the experimental group which had received graduate level training in cognitive assessment. Although the accuracy was low in the detection of suboptimal effort, the participants were able to discriminate between conditions. These findings from dichotomous ratings rejected the hypothesis within the present study which was the observations by the participants were not able to discriminate between full effort and suboptimal effort.

Researchers have also noted in the literature that effort which is a part of the greater construct of response bias can be viewed as being part of a continuum or spectrum that ranges from full effort to malingering with suboptimal effort referring to the span between the two (Dunn, 2006). For this reason, a Likert-type scale was included in the study to allow participants to rate the perceived level of effort exerted. A discriminant analysis indicated that creation of a prediction model based on the Likert-type scale ratings marginally aided in the differential prediction of full effort or suboptimal effort group membership. It should be noted that the predictive power of the model using ratings from either the control or experimental group was only slightly higher than chance probability in the specific classification of suboptimal effort. As such, these findings also found evidence consistent with the previously stated research that individuals independent of tests of effort have significant difficulty in accurately detecting suboptimal and malingering behaviors. These findings, although marginal, also rejected the second hypothesis that inclusion of the observational ratings did not add to the overall model in the differential prediction of full effort or suboptimal effort participants.

Although the ratings from both groups were found to be generally inaccurate in the detection of suboptimal effort, it appeared that the control group that had not received training in cognitive assessment was slightly more accurate in comparison to the experimental group who has received training in cognitive assessment. Specifically, estimations of classifications based on the liker scale derived discriminant functions revealed a lower total error count estimate for the control group at 0.30 in comparison to the experimental group total error count estimate of 0.35. These findings provide support for the rejection of the third hypothesis that no differences exist between the graduate students with training in assessment group and the graduate students without training in assessment group.

Through inclusion of a single open ended question, a descriptive component allowed for greater insight from the participants in the study. Unfortunately, due to the small amounts of data from accurate ratings of effort, only one category related to perceived suboptimal effort behaviors could be elicited. This category involved the perceiving the individual to be easily distracted as evidenced by checking a cell phone and "playing" with one's own hair or ring. However, there were more categories noted in the responses for perceived full effort behaviors. These categories included speed of completion, accuracy of responses, trying multiple attempts to solve item, leaning in, and correcting mistakes. Within the research that has been completed, no other study has attempted to include a component which would allow for descriptive analysis of the judgment process one makes when making decisions on levels of effort one perceives.

In the practice of psychological assessment, one goal of the evaluation is to produce both reliable and valid scores which are indicative of a person's true abilities.

Although psychological evaluations can consist of a variety of different assessments that measure any number of characteristics of a given construct, the scores are used to inform decisions. School psychologists often administer cognitive tests to diverse pediatric populations as part of the special education identification process. However, the validity of the results gained from such tests often hinges on blanket statements referring to behavioral observations associated with the perceived level of effort the person taking the test exerted. Literature in the field of psychology examining clinical judgment has questioned the accuracy and validity of judgment that is based on perception rather than scores (Dawes, Faust, & Meehl, 1989). The results of this current study further corroborates the argument that judgment is highly susceptible to error even when there are attempts to create more actuarial methods of classification and prediction using scores derived from clinical judgment.

One should also consider the limitations of the current study which may have affected the generalizability of the results. The first limitation is that the graduate student participants were from a single university. Although graduate programs may have some similarities in training due to requirements for specific national certifications or endorsements, programs can have more specialized areas of study and training depending on the faculty. Another limitation to note within the current study is that only a single subtest was viewed for each confederate allowing for a relatively short period of time to form judgments on effort. Psychological evaluations very often consist of a battery of tests, depending on the referral concern. Thus, having individuals make decisions on effort based on only a fragment of what would possibly exist in a more common evaluation is not equivalent to actual practice. Similarly, another limitation to consider is

the reliance on a single subtest which relies on nonverbal output rather than verbal output. It is possible to consider a lack of verbal responses a limitation as some individuals may place significant importance on different characteristics of speech within the observational method of evaluating effort.

Formal training in best practices of evaluating effort is essential within graduate psychology programs and should not be limited to neuropsychology and forensic psychology. This training should include different methods of evaluating effort such as embedded measures and objective tests that are well researched and appropriate for different populations such as children. Given the dearth of information on the practice of formal effort assessment in school based evaluations, one can speculate that tests of effort are not commonly utilized despite their ability to inform the validity of other test results.

Another implication of this study was that familiarity with the task being observed provided no added utility in the accurate detection of suboptimal effort. One could speculate that this may be related to some specific concepts learned in cognitive assessment training. In cognitive assessment training, graduate students are taught that cognitive diagnostic assessments are used to identify ipsative strengths and weaknesses within a cognitive profile (Sattler, 2001). Understanding that an individual may be naturally weak or perform more poorly on a given subtest makes one think that the poor performance on Block Design may be a natural part of their presentation rather than suboptimal effort. Examination of the accuracy of dichotomous ratings also showed that there was generally a positive bias for full effort as it was drastically over identified. There were no instances of false identification of full effort. One could also speculate that

this positive bias may be correlated with the understanding that individual cognitive profiles consist of both ipsative strengths and weaknesses.

Measures of response bias that have been developed and researched include both embedded indexes within commonly administered tests, as well as independent tests designed specifically to measure response bias. To date, no measures utilizing only observational means in the assessment of effort have been designed. This study examined the utility of observational means in the assessment of effort and found only marginally better than chance accuracy of classification. Efforts to improve the assessment of response bias should focus on alternative methods other than observational ratings.

References

American Psychological Association. (2010). *Ethical Principles of Psychologists and Code of Conduct*. Retrieved from http://www.apa.org/ethics/code/index.aspx

Blaskewitz, N., Merten, T., & Kathmann, N. (2008). Performance of children on symptom validity tests: TOMM, MSVT, and FIT. *Archives of Clinical Neuropsychology, 23*, 379-391. doi: 10.1016/j.acn.2008.01.008

Brooks, B. L., Sherman, E. M. S., Krol, A. L. (2012). Utility of TOMM trial 1 as an indicator of effort in children and adolescents. *Archives of Clinical Neuropsychology, 27*, 23-29. doi: 10.1093/arclin/acr086

Dawes, R. M., Faust, D., & Meehl, P. E. (1989). Clinical versus actuarial judgment. *Science, 243*(4899), 1668-1674. Retrieved from http://xt9lp6eh4r.search.serialssolutions.com

Donders, J. (2005). Performance on the Test of Memory Malingering in a mixed pediatric sample. *Child Neuropsychology, 11*, 221-227. doi: 10.1080/09297040490917298

Dunn, T. (2006). Current issues in response bias during neuropsychological assessment: Incomplete effort to malingering. In J. Dupri (Ed.), *Focus on neuropsychology research* (pp. 199-217). New York: Nova Sciences Publishers, Inc.

Faust, D., Hart, K., & Guilmette, T. J. (1988a). Pediatric malingering: The capacity of children to fake believable deficits on neuropsychological testing. *Journal of Consulting and Clinical Psychology, 56*(4), 578-582. doi: 10.1037/0022-006X.56.4.578

Faust, D., Hart, K., Guilmette, T. J., & Arkes, H. R. (1988b). Neuropsychologists' capacity to detect adolescent malingerers. *Professional Psychology: Research and Practice, 19*(5), 508-515. doi: 10.1037/0735-7028.19.5.508

Green, P. (2005). *Green's Word Memory Test for Microsoft Windows® User's Manual.* Edmonton, Alberta, CA: Green's Publishing

Gunn, D., Batchelor, J., & Jones, M. (2010). Detection of simulated memory impairment in 6- to 11-year-old children. *Child Neuropsychology, 16,* 105-118. doi: 10.1080/09297040903352564

Jacob, S., & Hartshorne, T. S. (2007). *Ethics and law for school psycologists - Fifth edition*. Hoboken, NJ: John Wiley & Sons, Inc.

Kirkwood, M. W. & Kirk, J. W. (2010). The base rate of suboptimal effort in a pediatric mild TBI sample: Performance on the Medical Symptom Validity Test. *The Clinical Neuropsychologist, 24,* 860-872. doi: 10.1080/13854040903527287

Kirkwood, M. W., Hargrave, D. D., & Kirk, J. W. (2011). The value of the WISC-IV digit span subtest in detecting noncredible performance during pediatric neuropsychological examinations. *Archives of Clinical Neuropsychology, 26,* 377-384. doi: 10.1093/arclin/acr040

Lu, P. H. & Boone, K. B. (2002). Suspect cognitive symptoms in a 9-year-old child: Malingering by proxy? *The Clinical Neuropsychologist, 16,* 90-96. doi: 10.1076/clin.16.1.90.8328

McGrath, R. E., Mitchell, M., Kim, B. H., & Hough, L. (2010). Evidence for response bias as a source of error variance in applied assessment. *Psychological Bulletin, 136*(3), 450-470. doi: 10.1037/a0019216

Merriam, S. B. (1998). *Qualitative Research and Case Study Applications in Education*. San Francisco, CA: Jossey-Bass Publishers.

Mittenberg, W., Patton, C., Canyock, E. M., & Condit, D. C. (2002). Base rates of malingering and symptom exaggeration. *Journal of Clinical and Experimental Neuropsyholgy, 24*(8), 1094-1102. doi: 10.1076/jcen.24.8.1094.8379

National Association of School Psychologists. (2010). *Principles for Professional Ethics*. Retrieved from http://www.nasponline.org/standards/2010standards/1_%20Ethical%20Principles.pdf

Sattler, J. M. (2001). *Assessment of children: cognitive applications - fourth edition*. San Diego, CA: Jerome M. Sattler, Publisher, Inc.

Sharland, M. J., & Gfeller, J. D. (2007). A survey of neuropsychologists' beliefs and practices with respect to the assessment of effort. *Archives of Clinical Neuropsyhology, 22*, 213-223. doi: 10.1016/j.acn.2006.12.004

Tombaugh, T. (1996). *Test of Memory Malingering*. North Tonawanda, NY: Multi-Health Services.

Wechsler, D. (1999). *Wechsler Abbreviated Scale of Intelligence*. San Antonio, TX: Pearson Education, Inc.