THE UTILITY OF CRT-A SUB-SCALES FOR UNDERSTANDING AND PREDICTING AGGRESSIVE BEHAVIORS

A Dissertation
Presented to
The Academic Faculty

by

Patrick Dean McNiel

In Partial Fulfillment
of the Requirements for the Degree
Doctor of Philosophy in
Psychology

Georgia Institute of Technology

August, 2014

THE UTILITY OF CRT-A SUB-SCALES FOR UNDERSTANDING AND PREDICTING AGGRESSIVE BEHAVIORS

Approved by:

Dr. Lawrence R. James, Advisor
School of Psychology
Georgia Institute of Technology

Dr. Rustin D. Meyer
School of Psychology
Georgia Institute of Technology

Dr. James S. Roberts
School of Psychology
Georgia Institute of Technology

Dr. Nathan Bennett
J. Mack Robinson College of Business
Georgia State University

Dr. Jack M. Feldman
School of Psychology
Georgia Institute of Technology

Date Approved: June 27, 2014

ACKNOWLEDGMENTS

This work was enriched by the contributions of many people, and I thank them all. My advisor Larry James gave me excellent guidance (as always). Jim Roberts took the time to converse with me on a number of statistical issues. The other members of my dissertation committee Jack Feldman, Rustin Meyer, and Nate Bennett were all willing to discuss potential issues and to give advice on how to avoid or solve these issues. And finally, all of the researchers whose data was incorporated into this study went the extra mile to ensure that I had clean understandable data to work with.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF ABBREVIATIONS

AID	Active Interpersonal Deviance
AOD	Active Organizational Deviance
BA	Blatant Aggression
CMP	Composite Excessive Behavior
CoA	Count of Absences
COM	Complaining
CRT-A	Conditional Reasoning Test for Aggression
DIS	Dishonest Reporting
DR	Dishonest Reporting
EH	Expressions of Hostility
EP	Excessive Pressure on Game Controller
GRV	Grievances
ICAR	Inappropriate Corrective Action Report
NSB	Negative Socioemotive Behavior
OA	Overt Aggression
OB	Obstructionism
PA	Physical Altercations
PAU	Pausing
PID	Passive Interpersonal Deviance
POD	Passive Organizational Deviance
TF	Technical Foul
THE	Theft

UR	Unreliability
VH	Verbal Harassment

SUMMARY

The purpose of this study is to re-analyze existing findings in order to demonstrate and summarize relationships between criteria and the sub-scales/dimensions of the Conditional Reasoning Test for Aggression (CRT-A): Externalizing, Internalizing, and Powerlessness. A CRT-A sub-scale was expected to relate more strongly with criteria categorized as being more relevant to the dimension that is represented by that sub-scale. For criteria that were categorized as relevant to only a subset of the dimensions represented by CRT-A sub-scales, the regression of a criterion on all three sub-scales was expected to create a better fitting model than the regression of a criterion on the CRT-A total score alone. Scales were also expected to interact to predict criteria. This was expected to be most likely when multiple dimensions of implicit aggression were activated by environmental factors to influence specific behaviors. Support was found for all expectations.

CHAPTER 1

INTRODUCTION

The implicit personality is made up those motivational forces, desires, needs, goals, habituated thought processes, scripts, defense mechanisms, and conditioned responses that we are not consciously aware of. To define implicit personality, we borrow from James and LeBreton (2012) who describe it as "…the dynamic mental structures and processes that influence an individual's behavioral adjustments to his or her environments *that are not accessible via introspection."*

It has become increasingly clear that the implicit personality is key to understanding the psychology of individuals. Implicit cognitions and cognitive-affective systems have been shown to have powerful effects on attention, emotion, perceptions, judgments, attitudes, and behaviors (Bargh & Chartrand, 1999; Bargh, 2008; Greenwald & Banaji, 1995; Haidt, 2001; McClelland, Koestner, & Weinberger, 1989; Nisbett & Wilson, 1977). The study of individual differences in these unconscious systems provides valuable perspectives and creates a more complete picture of how personality effects behavior. Though, effective study of implicit personality requires a measurement system that is easy to administer, psychometrically robust, and that can be theoretically tied to constructs of interest.

In an effort to address these needs, James and colleagues developed a way to measure implicit personality via Conditional Reasoning (James, 1998; James & Mazerolle, 2002; James & LeBreton, 2012). This method is founded on the idea that

people who have certain dispositional tendencies or who favor certain types of behaviors are able to comfortably act and feel the way they do because they have developed supporting cognitive mechanisms. These mechanisms enhance the rational appeal of certain types of behaviors over others. They enable the release of these behaviors and make them seem more acceptable and justifiable than competing behavior choices. These supporting cognitions are referred to as Justification Mechanisms (JMs). And, they can influence the reasoning process at every step. What is attended to, what is recalled from memory, how perceptions are formed, how situations are framed, and what decisions and behaviors are chosen are all influenced by a person's JMs.

Opposing JMs bias the reasoning process in different directions. The effect of this is that assumptions, conclusions, decisions, and behaviors make more sense to people when they are congruent with their needs, motives, and dispositions. This creates an avenue for measuring implicit personality. For example, opposing personality traits such as achievement orientation and fear of failure are necessarily supported by incompatible JMs. If people must make choices based on reasoning that would flow from the JMs they possess vs. reasoning that flows from JMs that support opposing traits, they will regard the former reasoning as more tenable. Reasoning that is produced or preferred is thus conditional based on the use of different JMs.

So far, conditional reasoning (CR) has been implemented through the use of inductive reasoning problems. These problems ask people to reason from a set of premises to conclusions that are not absolutely supported by logic provided within an item. Respondents must make inferences and fill in logical gaps. These items have two correct answer choices that can be arrived at by filling in assumptions and logic that

would flow from one of two opposing JMs. Personality trait levels are indicated by the degree to which people prefer answer choices designed to evoke JMs that are supportive of particular traits.

Notably, the CR technique has been applied to measuring individual differences in the proclivity to behave aggressively. Aggression is defined here as behavior initiated with intent to harm. Specifically, the Conditional Reasoning Test for Aggression (CRT-A) measures biases that support, guide, and shape reasoning that rationalizes aggression. The CRT-A is based on six JMs identified through searches of both academic literatures and press accounts of aggressive activities (James, 1998; James & Mazerolle, 2002; James & McIntyre, 2000). However, recent research has found that items designed to assess these JMs are best grouped into three scales (Ko, Thompson, Shim, Roberts, & McIntyre, 2008; James & LeBreton, 2012).

These scales represent different dimensions of the implicit cognitive mechanisms that people use to create rationalizations for aggression (James & LeBreton, 2012). The first dimension is labeled Externalizing, as it represents biases that are externally oriented. People use biases from this dimension to justify aggression based on feelings of victimization and perceptions of threat. The second dimension is labeled Internalizing, as it represents biases that are internally oriented. People use biases from this dimension to justify aggression based on a need to feel powerful and potent as well as a preference for retribution rather than reconciliation. The third dimension is labeled Powerlessness, as it represents biases that stem from anger, frustration, and a lack of control over one's life. It is likely that people use biases from this dimension to justify aggression based on a need to discount social structures and tear down sources of negative affect.

These three CRT-A sub-scales are relatively new, and validation information at the scale level is sparse. This of course is because previous studies have not included analyses at the scale level. Additionally, the meaning behind the dimensionality of these scales suggests that scales may in some cases have utility that is independent of or greater than that of the CRT-A total score. Accordingly, the purpose of this study is to examine existing data sets using both the CRT-A and its subscales as predictors. The primary concerns will be in identifying criteria and/or conditions that are particularly relevant to one or more of the CRT-A subscales and in identifying when additive or interactive effects of subscales may have more explanatory power than the CRT-A total score.

1.1 The Conditional Reasoning Test for Aggression

The CRT-A represents a new way of measuring individual differences in aggressive tendencies. It is fully described in James and McIntyre (2000) and in James and Mazerolle (2002). This test is notable for several reasons. First, the CRT-A is an implicit test. This means that it measures aspects of the aggression construct that are not accessible through introspection. These aspects of aggression are hidden to the individuals who possess them because of the following fundamental conflict. People like to see themselves as basically good. However, engaging in anti-social and harmful behaviors threatens this image. This sets up a conflict between a desire to harm others and a desire to hold a favorable view of the self. In order to resolve this conflict, aggression must be rationalized as a socially acceptable, necessary, or justifiable action. JMs for aggression facilitate these rationalizations. At the same time, JMs mask true motives to harm with alternative motives that are congruent with a positive self-view.

Second, the CRT-A is resistant to faking. This is in part due to the implicit nature of this test, but it is also due to how the test is presented. Respondents believe that they are taking a reasoning test, but they are not aware that the test is designed to assess aspects of personality as well.

Third, the CRT-A taps into aspects of the aggression construct that have not been easily and reliably measured before. It is the first implicit test of aggression that is easy to administer, easy to score, rationally ties the object of measurement to behavior, and has good psychometric properties.

1.1.1 Justification Mechanisms

As noted previously, the CRT-A measures the use of six JMs for aggression. These are described here in turn. The hostile attribution bias is a propensity to see malevolent intent in others. It promotes the illusion that aggression is justified as defense against these intentions and resulting behaviors. The victimization by powerful others bias is a predilection to see powerful others as exploitative, unfair, or oppressive. Aggression is seen a means to right injustice, defend against oppression, and guard against exploitation. The potency bias is a fixation on dominance, strength, and power vs. submissiveness, weakness, and helplessness. For people with this bias, potency is highly tied to self-worth, and others are judged as inadequate if they show weakness. Aggression is considered justified when it proves strength or when it enables one to avoid looking weak. The retribution bias is penchant to view retaliation as a more rational course than reconciliation. Aggression is a means to restore honor, to overcome wounded pride, and to restore loss of respect.

The social discounting bias is a tendency to see societal rules, social norms, and strong expectations as confining, repressive, and restrictive. Aggression is justified as a means to free oneself of these unreasonable strictures and to express oneself as one wishes. The derogation of target bias facilitates the release of aggression onto targets by casting them as immoral, evil, or corrupt. Aggression against these targets is then partially justified because such target are more deserving of punishment.

In each CRT-A item, reasoning that would flow from the use of one or more of the JMs for aggression is set against reasoning that a pro-social or non-aggressive person would use. This reasoning tends to be socially adaptive and promoting of peace, harmony, and cooperation. It supports normative behaviors and is based on assumptions that hold being nice, polite, trusting, accommodating, responsible, and helpful as practical and rational.

1.1.2 CRT-A items

Each CRT-A item is composed of three parts. These parts are the stem, the question, and the answer choices. The stem is composed of premises, and if appropriate for the question type, a conclusion. The question will ask about appropriate conclusions or the assumptions underlying the stem, the conclusion of the stem, or the answer choices and how they relate to the stem. By convention there are four answer choices, and two will be logical (correct). One logical answer choice is designed so that it is seen as more logical to someone who is influenced by a specific implicit assumption (JM). The other is designed so that it is seen as more logical to a person who does not possess JMs for aggression and who has a pro-social disposition. In this way, two logical answers can be pitted against each other and the one that follows from an assumption held by a

respondent should seem more logical and be chosen as the correct answer. The rest of the answer choices are typical reasoning question distracters (James & Mazerolle, 2002).

1.1.3 Scoring

The CRT-A is composed of 25 inductive reasoning questions. Three of these questions are pure inductive reasoning questions placed at the beginning of the measure to enhance the face validity of the test, and 22 of these questions are inductive reasoning questions that have been designed to assess the use of JMs for aggression. The 22 scored items are each given a score of 1 if the aggressive response is chosen or a score of 0 if any other response is chosen. Each test is scored by summing the number of aggressive responses. However, if more than five illogical responses are chosen the response set will be considered invalidated. This method of scoring the CRT-A is recommended in James and McIntyre (2000).

Although the potential range of scores on the CRT-A goes from 0 to 22, very few people score higher than 12 (James & McIntyre, 2000). Even if a person has a high predisposition toward aggressive behavior, initial studies suggest that their use of JMs does not tend to cover the entire spectrum of JM possibilities. Instead, aggressive people tend to have one or two JMs associated with aggressive behavior (Minton, 2006). This makes it highly unlikely that any one individual will choose the aggressive response on most questions in the CRT-A. Based on an analysis of score level and its relation to behavior characteristics, scores on the CRT-A are interpreted as being low from 0-2, medium from 3-7, and high for scores above 8. Scores represent individual differences in cognitive readiness to aggress (James and McIntyre, 2000).

1.1.4 Factor Structure of the CRT-A

The factor structure of the CRT-A was initially assessed using a sample of 1,603 individual responses (James et al., 2005). A five factor solution was deemed most appropriate at the time. These factors each reflected one of five JMs. The derogation of target JM was not represented in the factor solution because only one item designed to assess this JM made it into the final measure. The five factor solution was a good match to the theory used to design the CRT-A. However, several of the JMs tended to co-occur, and it was thought that a larger sample might reveal a more informative or parsimonious dimensionality (James & Lebreton, 2012).

Ko, Thompson, Shim, Roberts, and McIntyre (2008) examined the dimensionality of the CRT-A again when a much larger sample size was available. They used 4,772 participant responses across 16 separate studies. The number of dominant dimensions was ascertained by comparing the eigenvalues from a principle component analysis of real question data to the average eigenvalues generated from bootstrapped random responses to these same questions. This analysis suggested the existence of three dominant dimensions. To create scales associated with these three dimensions, a factor analysis was done in which three factors were extracted. Questions were then assigned to a dimension based on their highest factor loading. This factor analysis utilized principle axis factoring and Promax rotation. Factor names and substantive meanings were derived from both the content of questions and the JMs questions were designed to assess. These are the Externalizing, Internalizing, and Powerlessness factors described in the introductory paragraphs of this manuscript. Further details on these factors are provided in the following sections.

1.1.5 Basic Psychometrics of the CRT-A and its Sub-scales

James and LeBreton (2012) report internal consistency reliabilities based upon previous studies to range from .81 to .87 (factorial reliabilities) for each of the three factors that make up the CRT-A. Intercorrelations between factors were also reported as .50 between Externalizing and Internalizing, .21 between Externalizing and Powerlessness, and .17 between Internalizing and Powerlessness. For the complete 22 item CRT-A scale, the estimate of reliability was .76 using a Kuder-Richardson (Formula 20) coefficient that used the average item-total polyserial correlation coefficients. Previous external validity estimates for the CRT-A were also reported in this volume as ranging from .11 to .64 with a mean of .41. These validity estimates were based on criteria that were judged to be the best behavioral indicators of aggression.

1.2 Sub-factors of the CRT-A and Scale/Behavior Relationships

1.2.1 Externalizing

The Externalizing Scale includes questions designed to assess the victimization by powerful others bias and the hostile attribution bias (James & Lebreton, 2012). Both JMs work to form one's perceptions of intentions. They engender a general tendency to see oneself as being threatened or a victim of unjust and oppressive powers. This creates a perceived need to protect oneself from outside forces, and provides rationales for aggression that are based on this need. While this helps to define the parameters under which aggression is justifiable and acceptable to enact, it also limits the circumstances in which rationalizations stemming from these JMs are applicable. Based on this, expectations can be formed about the types of behaviors that are most related to the

Externalizing Scale. These are aggressive behaviors intended to prevent victimization, exploitation, oppression, endangerment, malevolence, loss of resources, and loss of status. And, these are the aggressive behaviors that are most likely to be functional for these purposes. Essentially, aggressive behaviors are expected to be most related to the Externalizing Scale when they are defensive in nature and when contextual factors are most likely to evoke a defensive response.

1.2.2 Internalizing

The Internalizing Scale includes questions designed to assess the potency bias and the retribution bias (James & Lebreton, 2012). The retribution bias makes retaliation seem more reasonable than reconciliation and functions to sustain elevated pride and self-esteem. The potency bias makes aggressive behavior seem appealing as a way to show bravery, strength, or assertiveness. Both JMs reflect intrinsically generated motivations and primarily enable aggression to be used for the function of self-enhancement. They reflect a need to see oneself as powerful, daring, dominant, and unassailable. They also aid in maintaining an illusion of superiority to others. Internalizing JMs protect a sense of entitlement, and they mask the aggressive behaviors used to assert this entitlement as expressions of virtue. While this helps to define when aggressive behavior will be most useful, it also limits the circumstances in which rationalizations stemming from these JMs are applicable.

Based on the above, expectations can be formed about the types of behaviors that are most related to the Internalizing Scale. These are aggressive behaviors that function to establish dominance, win competitions, increase power, show strength/bravery/fearlessness, assuage wounded pride, and elevate self-esteem.

Essentially, aggressive behaviors are expected to be most related to the Internalizing Scale when they are self-enhancing in nature and when contextual factors are most likely to evoke self-enhancement needs.

1.2.3 Powerlessness

The Powerlessness Scale primarily includes questions designed to assess the social discounting bias, which creates a tendency to see social norms and structures as repressive, confining, and restrictive of free will. Additionally, one item in this scale assesses the derogation of target bias, which is an unconscious propensity to apply negative characteristics to people in order to make them seem more deserving of harmful outcomes. This makes it easier to justify aggression against targets. Based on these JMs and the content of the questions which load on this dimension, James and LeBreton (2012) suggest that it represents anger and frustration stemming from a lack of control over one's life, a sense of powerlessness, or an inability to gain desired outcomes. The need for Powerlessness JMs may also indicate a tendency to become impatient and frustrated with situations and social interactions that are difficult for a person to understand or resolve effectively. This dimension may be more generally relevant to aggressive behaviors than the Externalizing and Internalizing dimensions. Frustration or an immature understanding of social implications can occur to some extent in most contexts and during most tasks. However, the Powerlessness dimension is probably more relevant when situations and social relationships are confusing and/or when tasks are difficult.

Based on the above, expectations can be formed about the types of behaviors that are most related to the Powerlessness Scale. These are primarily reactionary aggressive

behaviors intended address or eliminate sources of frustration and confusion. And, these are the aggressive behaviors that undermine, inappropriately ignore, or harm important social conventions (norms, rules, laws, processes, etc.).

1.3 Hypotheses

The expected behavioral relationships with CRT-A sub-scales described above lead to several questions about the utility of these scales for predicting various criteria. CRT-A sub-scales may represent dimensions of aggression that are expressed in different ways and in different circumstances. Additionally, each scale may be examined alone, in conjunction with the other scales, or as a possible moderator of the relationships between the other scales and criteria. Therefore, scale level analysis may be able to add something to our understanding of aggressive behaviors that examinations involving only the CRT-A total score cannot.

1.3.1 Hypothesis 1

First, if the sub-scales of the CRT-A are indicative of important dimensions of aggression that can be differentiated in terms of associated cognitive mechanisms and biases, then each scale should also predict expressions of aggression that are congruent with the nature of the dimension of aggression measured by that scale. Additionally, a scale should be less useful in predicting expressions of aggression that are not congruent with the nature of the dimension of aggression measured by that scale. This leads to the first hypothesis.

Hypothesis 1: CRT-A scales will correlate more strongly with criteria/behaviors that logically follow from the perspectives, framing proclivities, assumptions, biases, and

JMs inherent in the dimensions of aggression associated with particular scales. More specific predictions are represented in the following sub-hypotheses.

Hypothesis 1a: The Externalizing scale will correlate more strongly with aggressive criteria that are self-protective or defensive in nature and that serve primarily as a response to external forces. And, this scale will correlate less strongly with criteria that do not have these qualities.

Hypothesis 1b: The Internalizing scale will correlate more strongly with aggressive criteria that are self-enhancing and enable one to assert feelings of entitlement or superiority. And, this scale will correlate less strongly with criteria that do not have these qualities.

Hypothesis 1c: The Powerlessness scale will correlate more strongly with aggressive criteria in situations that are constraining, difficult, confusing, or frustrating. Because anger and frustration are a key component of the powerlessness dimension, associated aggressive criteria are also expected to be defined by being reactive, expressive of frustration, impulsive, thoughtless, affect laden, irrational, and/or dysfunctional. And, this scale will correlate less strongly with criteria in situations that are not frustrating and that do not evoke feelings of constraint.

1.3.2 Hypothesis 2

Second, the overall relationship between the CRT-A and a criterion may sometimes be less informative than the component relationships between the CRT-A sub-scales and a criterion. A criterion may have characteristics that are only important to one or two dimensions measured by the CRT-A. If only part of the CRT-A has relevancy to a criterion, then the remaining questions in the CRT-A test are introducing irrelevant

variance which will suppress its correlation with the criterion. Additionally, specific options for expressing aggression may not be compatible. For example, if we looked at the relationships between Internalizing and Externalizing with random responding (the criterion) during a grueling test that is part of an experiment, we may find that Internalizing channels behavior into self-affirming displays of aggression (such as telling the test creator that the questions are boring and easy). At the same time, the Internalizing mechanisms may influence the test taker to defeat the test (i.e. do well) and show potency. Meanwhile, Externalizing may channel behaviors into defensive tactics which relieve the burden on a participant while damaging experimental results (i.e. random responding). Defeating the test and random responding are incompatible with one another. Behavior choice should reflect scores on the Internalizing and Externalizing scales. If random responding is the criterion of interest, then it may be positively related to Externalizing but negatively related to Internalizing.

These relationships may both be valid, but they will at least partially cancel out when scales are added to form the CRT-A total score. Total scores will then indicate a lessened or nil relationship between aggression and a criterion, even though the criterion is specifically related to some aspects of aggression. Allowing B weights in a regression equation to vary for each of the three sub-scales will make specific relationships between dimensions and criteria more visible, and it will keep opposing dimensional relationships with criteria from disappearing in the aggregated score. This leads to the second hypothesis.

Hypothesis 2: The more complex model created by using the three CRT-A sub-scales instead of the CRT-A total score when regressing criteria on these scales will

likely be more appropriate and informative in cases where criteria are only relevant to one or two of the three dimensions assessed.

1.3.3 Hypothesis 3

Third, possession of JMs associated with multiple dimensions of aggression may engender more complicated processes with regard to the release of aggression and the choice of aggressive behaviors. Situations that are evocative of multiple dimensions of aggression may result in a confluence of cognitive processes that all support a particular behavioral response, or they may result in competing cognitive processes that flatten, redirect, or eliminate a behavioral response. As a result, the relationships between behaviors and CRT-A subscales may depend on what dimensions are evoked by a situation, the relevance of behaviors to evoked dimensions, and how JMs associated with multiple dimensions interact to guide behavior.

The most likely case is that dimensions will influence behavior in a synergistic fashion. In this case, JMs associated with different dimensions reinforce one another and create more elaborated and powerful reasons to engage in a particular type of aggressive behavior. This elaboration and reinforcement creates an effect that is greater than one that would be created by the sum influence of individual dimensions.

Another likely case is that dimensions will influence behavior in a competitive fashion. In this case, JMs associated with different dimensions create countervailing influences that effect the expression of particular aggressive behaviors. These behaviors will be suppressed when two sources of countervailing influence are present, but will be increasingly stimulated with reductions in the suppressing influence.

Although dimensional influences may interact to influence the expression of any particular instance of behavior, these interactions are unlikely to be seen in statistical tests unless the contextual forces evoking these influences are similar across people. Additionally, different types of behaviors are likely to be influenced in different ways by interacting dimensions. Unless criteria represent behaviors that are likely to be influenced in the same manner, statistical tests will reflect an amalgam of influences on different types of behaviors that may not all go in the same direction and are likely to be less distinct. This leads to the third hypothesis.

Hypothesis 3: CRT-A sub-scales will interact to predict criteria, and are most likely to do so when criteria are measured within highly similar situations across people and represent behaviors that are either singular or highly similar in nature.

These hypotheses are tested using multiple datasets. These data have been gathered primarily from previous studies which have used the CRT-A as a covariate and an indicator of aggression as a criterion.

CHAPTER 2

METHOD

2.1 Search for Data Sets

To find data sets and studies that include the CRT-A, a multi-pronged approach was utilized. First, PsychINFO was used to search for primary studies that cite one of the five seminal sources of information on the CRT-A. These include the first Journal article to explicate the conditional reasoning system for measuring personality (James, 1998), the Test manual for the CRT-A (James & McIntyre, 2000), a more comprehensive book on conditional reasoning (James & Mazerolle, 2002) and two Journal articles that summarize the conditional reasoning system for measuring aggression (James, McIntyre, Glisson, Bowler, & Mitchell, 2004; James, et al., 2005). Data sets referenced in these five sources were also noted. Second, PsychInfo was searched using the keywords Conditional Reasoning Test, CRT-A, and, Justification Mechanism/s. Third, authors who have been involved with conditional reasoning research were contacted and asked about current research on the CRT-A that they or those they know of are involved with. And fourth, a review of the reference sections of all articles found was used to find any additional data sets that might be relevant.

2.2 Inclusion and Exclusion of Data Sets

As data sets and studies that include the CRT-A were found, they were examined to determine their appropriateness for inclusion in this study. Eleven studies have

includable data. To be included, two things must be true. First, the CRT-A used must be the final 22 item version described in the test manual for the CRT-A. This is necessary as the Internalizing, Externalizing, and Powerlessness subscales that will be analyzed where created through a factor analysis of these 22 items. Second, the data set must include at least one criterion that measures behavior that is indicative of aggression.

Accordingly, studies that only include conduct violations or performance indicators as criteria were not included. Conduct violations are not open to close scrutiny and can occur for many reasons that are not related to the aggression construct. As the reason for a violation is not generally obtainable, it is essentially not known what if any percentage of those violations are related to aggression in some way. Similarly, Performance indicators do not reflect a specific occurrence or mode of aggression. Rather, they are determined by multiple factors including; cognitive ability, social skill, energy level, neuroticism, domain of responsibility, job complexity, visibility, etc. Performance is also problematic because aggressive behaviors are likely useful for improving scores on typical performance indicators or for covering up a lack of performance.

It is important to note that inclusion of a study or dataset into this study requires a relevant criterion, scores on the three subscales of the CRT-A, and some explanation of how the data were created. As subscale scores are not usually reported, the source of reporting was generally insufficient for an appropriate analysis and necessitated the acquisition of the source data in all cases. One study had to be dropped because source data could not be obtained.

2.3 Coding for Moderators

Two sets of codes were applied to each criterion. The first set categorized criteria according to perceived associations with Externalizing, Internalizing, and Powerlessness within the context of measurement. Categorizations were made independently by three raters who read descriptions of these dimensions that included general examples of when these dimensions are most likely to be relevant. Raters also read a description of each criterion and the study conditions in which it was measured. The raters then noted which of the three dimensions criteria are most likely to relate to. Raters could indicate that a criterion is likely to relate to one, two, or all three of the dimensions depending upon how relevant they thought dimensions where to the criterion in comparison to the other dimensions. Raters could also indicate that none of the dimensions are relevant to a criterion. In cases where raters disagreed, raters discussed reasons for disagreement until a consensus was reached. In cases where a consensus could not be reached, criteria were not categorized as relevant to the dimension under dispute. Each rater had a master's degree or PhD in Psychology.

A second set of codes was assigned by the author. These codes categorize criteria in two ways. First, a criterion may consist of a specific behavior/behavior type, or it may consist of non-specific behaviors. Specific behaviors are categorized as such if they meet one of two conditions. They can either be precisely defined behaviors such as the amount of pressure placed on a button, absenteeism, or Theft of a particular item, or they can be a specified range of behaviors that are categorically defined such as verbal hostility or physical altercations. Criteria are categorized as non-specific when measurement does not reflect specific behaviors or behavior ranges. An example of this would be ratings of a

person on their tendency to engage in interpersonal aggression. Interpersonal aggression can be achieve in an innumerable number of qualitatively different ways, and ratings of interpersonal aggression are likely to reflect the gamut of behaviors. So, this criterion would be categorized as non-specific.

Second, a criterion may be measured within a context that is regular and highly similar across instances of measurement or it may be measured within contexts that change in substantial ways across instances of measurement. Regular and similar contexts will be categorized as specific. Contexts that change in substantial ways across instances of measurement will be categorized as non-specific. Examples of specific contexts are experiments that fix all stimuli for each participant, different instances of a game with fixed rules and an established culture of play (like basketball in a league), or a single social unit that is the target of measurement (such as a single company). Examples of when contexts are non-specific include measurement of behaviors across different companies, or amalgams of measures from people who have substantively different perspectives on behaviors of interest (such as 360 scores).

2.4 Analysis

Multiple data sources allow for the examination of particular relationships between CRT-A scales and many different types of criteria that are situated in a diverse array of contextual factors. Multiple data sets also allow for the examination of trends across studies. In order to examine both specific relationships and trends, two approaches to analyzing the data were taken.

2.4.1 Study level Analyses

The first approach focused on individual studies. In this approach, the relationships between all relevant criteria and CRT-A scales were analyzed for each study. Hypothesis 1 (H1) was tested by looking at three tests for differences between dependent correlations using the T_2 test proposed by Williams (1959) as cited and recommended by Steiger (1980). The correlation between each subscale and the criterion was compared with the average correlation of the other two subscales and the criterion. The expectation is that significant differences will align with how a criterion was categorized in terms of its relevance to Externalizing, Internalizing, and Powerlessness.

There are three different expectations depending on how a criterion was categorized. The first expectation is developed when only one dimension is relevant to a criterion. In this case, the CRT-A subscale associated with that dimension is expected to have a significantly greater correlation with the criterion compared to the average of the other two subscale/criterion correlations. The second expectation is developed when only two dimensions are relevant to a criterion. In this case, the subscale associated with the unassigned dimension is expected to have a significantly smaller correlation with the criterion compared to the average of the other two subscale/criterion correlations. The third expectation is developed when all three dimensions are relevant to a criterion. In this case, the null hypothesis of no differences between dependent correlations is expected to be supported. Criteria that could not be categorized as being relevant to any of the three dimensions were not considered relevant to the study and were excluded from analysis.

Hypothesis 2 (H2) was tested by comparing the model fit of two models. Model 1 is derived from the regression of a criterion on the CRT-A total score. Model 2 is derived from the regression of a criterion on all three subscales of the CRT-A. Model comparisons were made using AICc, which is used exactly as AIC but includes an adjustment for smaller sample sizes. The AIC method is recommended by Cohen, Cohen, West, and Aiken when regression models that are not nested are compared (2003, p. 509). For this study, a nested design will lead to matrix inversion problems because scores on the subscales of the CRT-A and the CRT-A total score are created using the same information. AIC is designed to correct for the upward bias in the number of parameters targeted by maximized log-likelihood when it is used for model selection. It is also an approximately unbiased estimator of the expected difference between a model and the true mechanisms generating the observed data (Burnham & Anderson, 2002). It is appropriate to use in selecting regression models for prediction purposes because it is based on expectations that take into account model complexity. This gives it a cross-validation property (Burnham & Anderson, 2002).

Because AIC is based on a different paradigm than most hypothesis testing techniques, there is not a significance test associated with its use. Instead, models are compared according to the likelihood of a particular model given the data. The model with the lowest AIC value is considered to be more likely and is taken to be a more optimal fit to the data than models with higher AIC values. Therefore, model 2 will be considered a better fit when the AICc for model 1 is greater than the AICc for model 2. Delta AICc will be set up throughout this paper to reflect a preference for model 1 when it is negative and to reflect a preference for model 2 when it is positive.

The expectations are that AICc will show a better fit for model 1 when criteria are categorized as being relevant for all three dimensions, and that AICc will show a better fit for model 2 when criteria are categorized as being relevant to only one or two of the three dimensions.

Hypothesis 3 (H3) was tested by looking at ΔR^2 from linear regression when going from all three scale scores (Model 2) to all three scale scores plus the three two way interaction terms derived from the three scales (Model 3). The expectation is that when criteria are specific in terms of behavior and when they are embedded within a specific context, the F test for the ΔR^2 will be significant. It should be noted that the power to detect interactions is always an issue. And, in many cases a specific behavior is not further influenced by dimensional interactions. Likewise, a specific context does not always elicit additional activations and dimensional interactions. Consequently, this hypothesis is tentative in terms of consistency. However, significant interactions are expected to occur across multiple criteria and studies, and they are not expected to occur when both behavior and context are non-specific.

2.4.2 Aggregate level Analyses

A second approach focused on the aggregation of study results through moderated meta-analyses and summary statistics. Hypothesis 1 was tested by performing three meta-analyses of effect sizes using criteria categorizations as moderators. This resulted in a meta-analysis for each CRT-A subscale. For each meta-analysis, the relevancy or irrelevancy of criteria to the dimension represented by a subscale acted as the moderator. The expectation is that the average correlation between a CRT-A subscale and those criteria that have been categorized as relevant to the dimension it represents will be

higher than the average correlation between a CRT-A subscale and those criteria that have not been categorized as being relevant to the dimension it represents. Important details about how the meta-analyses were conducted follow.

When more than one criterion from a single sample can be assigned to a dimension, the correlations for these criteria were averaged together. Likewise, if multiple criteria from a single sample could not be assigned to a dimension, the correlations for these criteria were also averaged together. In cases where a study has criteria that were categorized into the dimension of interest as well as criteria that were not, only the criteria that have been associated with the dimension of interest were utilized in the initial analysis. Because sample sizes varied substantially across studies, both sample weighted average correlations and average correlations are reported.

Hypothesis 2 was tested by looking at counts of when fit was better for model 1 vs. model 2 (using AICc) in cases where criteria are assigned to all three dimensions vs. only one or two dimensions. The expectation is that the probability of better model fit for model 2 will be significantly greater when only one or two dimensions are assigned to a criterion. Fisher's exact test was used to determine if the cell proportions created by counts in the resulting 2x2 table support this expectation. This test was used instead of the chi-squared test because the sample size is small (Hays, 1994).

Hypothesis 3 was tested by looking at counts of when interactions were significant vs. non-significant in cases where criteria are specific in terms of behavior and context vs. non-specific in terms of behavior or context. The expectation is that the probability of a significant interaction will be significantly greater when criteria are

specific in terms of behavior and context. As with Hypothesis 2, Hypothesis 3 was tested using a Fisher's exact test.

CHAPTER 3

RESULTS

3.1 Basic Statistics

For the ten studies found that meet the requirements for inclusion into this study, data for thirteen independent samples were obtained. Across these thirteen samples, twenty seven criteria were related to the aggression construct. A brief description of each study and criterion is provided in Table 1 along with basic statistics for the CRT-A total score, the three subscales of the CRT-A, and the criteria from each study. Results will be presented at the level of the individual study first, and then they will be described at a summary level. A more complete description of the studies, associated criteria, and rational for criteria categorizations may be found in the Appendix along with statistics that are auxiliary to the study level hypotheses and more in depth descriptions of findings for each study.

Table 1: Basic Descriptions and Statistics for Included Samples

Sample Description			Variable Statistics									
				CRT-A[a]		EXT[b]		INT[c]		POW[d]		Criterion[e]	
Source	Sample	Criterion Description	*N*[f]	*M*	*SD*	*M*	*SD*	*M*	*SD*	*M*	*SD*	*M*	*SD*
Baysinger (2013)	Undergraduates	Negative Socioemotive Behavior	473	4.24	2.08	1.99	1.38	0.81	0.85	1.44	1.04	1.85	0.85
Bing (2007)	Hospital Workers	Passive Organizational Deviance	173	3.14	1.95	1.61	1.31	0.66	0.87	0.88	0.80	-0.02	0.98
Bing (2007)		Active Organizational Deviance										-0.08	0.49
Bing (2007)		Passive Interpersonal Deviance										0.01	1.01
Bing (2007)		Active Interpersonal Deviance										-0.03	0.96
Bing (2007)		Filing Complaints										0.10	0.90
DeSimone (2010)	Nuclear Power plant Operators	Inappropriate Corrective Action Report	112	4.40	2.47	1.79	1.52	1.07	1.13	1.52	0.91	0.05	0.22
Frost (2002)	Basketball Players	Technical Fouls	185	3.97	1.98	1.89	1.41	0.81	0.89	1.27	1.03	0.10	0.35

Table 1 (Continued)

Sample Description			Variable Statistics										
				CRT-A[a]		EXT[b]		INT[c]		POW[d]		Criterion[e]	
Source	Sample	Criterion Description	*N*[f]	*M*	*SD*	*M*	*SD*	*M*	*SD*	*M*	*SD*	*M*	*SD*
Frost (2002)		Verbal Harassment										0.20	0.50
Frost (2002)		Composite of Excessive Behaviors										0.38	0.81
Frost (2002)		Physical Altercations										0.10	0.35
Frost (2005)	Basketball Players	Obstruction	177	3.98	2.23	1.84	1.33	0.79	0.92	1.36	1.04	0.14	0.38
Frost (2005)		Overt Aggression										0.24	0.57
Frost (2005)		Expressions of Hostility										0.29	0.61
James and McIntyre (2000) - Study 7	Package Handlers	Absences	105	3.88	2.37	1.98	1.42	0.92	0.97	0.94	0.93	4.91	5.24
James and McIntyre (2000) - Study 8	Temporary Employees	Blatant Aggression	111	2.78	1.77	1.45	1.28	0.61	0.77	0.71	0.78	0.12	0.32
James and McIntyre	Temporary Employees	Unreliability (w Blatant										0.78	0.64

Table 1 (Continued)

Sample Description			Variable Statistics										
				CRT-A[a]		EXT[b]		INT[c]		POW[d]		Criterion[e]	
Source	Sample	Criterion Description	*N*[f]	*M*	*SD*	*M*	*SD*	*M*	*SD*	*M*	*SD*	*M*	*SD*
(2000) - Study 8		Aggression)											
Mc Niel (2009)	Undergraduates	Pausing	164	4.60	2.18	2.05	1.43	0.94	0.89	1.62	1.12	0.17	0.38
Mc Niel (2009)		Complaining										0.10	0.31
Mc Niel (2009)		Excessive Pressure 5 min										-0.01	1.01
Mc Niel (2009)		Dishonest Reporting										0.18	0.39
Mc Niel (New Sample) - Sample 1 Pilot	Undergraduates	Lack of effort	97	4.74	2.50	1.76	1.39	1.06	0.96	1.63	1.20	5.53	5.26
Mc Niel (New Sample) - Sample 2	Undergraduates	Lack of effort	66	4.89	2.27	2.49	1.57	0.88	0.85	1.55	1.16	8.71	9.33
Mc Niel (New Sample) - Sample 3	Undergraduates	Lack of effort	120	4.62	2.31	2.06	1.44	0.91	1.02	1.68	1.12	6.29	8.93

Table 1 (Continued)

Sample Description			Variable Statistics										
				CRT-A[a]		EXT[b]		INT[c]		POW[d]		Criterion[e]	
Source	Sample	Criterion Description	*N*[f]	*M*	*SD*	*M*	*SD*	*M*	*SD*	*M*	*SD*	*M*	*SD*
Mc Niel (New Sample) - Sample 4	Undergraduates	Lack of effort	118	4.43	2.23	2.06	1.39	0.80	0.83	1.63	1.08	6.42	7.91
Sablynski (2001)	Undergraduates	Theft	91	3.92	2.32	2.11	1.62	0.63	0.73	1.19	1.01	0.05	0.23
Sablynski (2001)		Disobedience										1.59	1.18

[a]CRT-A refers to the total score across all 22 scored items of the CRT-A.
[b]EXT refers to the score on the externalizing scale of the CRT-A.
[c]INT refers to the score on the internalizing scale of the CRT-A.
[d]POW refers to the powerlessness scale of the CRT-A.
[e]The CRT-A, EXT, INT, and POW scale descriptive statistics only change within a study when multiple samples are part of that study.

3.2 Categorization of Criteria

Each criterion was independently categorized by three raters as being relatively more or less relevant to each dimension of aggression. These ratings were based both on the nature and substance of the measured behaviors that made up the individual criteria, and the context in which these measurements were made. A criterion might have been relevant to one, two, or all three of the aggression dimensions. For each criterion, this created three dichotomous decision points for each rater. These decision points correspond to the three dimensions of justification mechanisms. The decisions were either "yes" this criterion is more relevant to a particular dimension than one or both of the other dimensions, or "no" this criterion is not likely to be as relevant to this dimension as compared to the other two dimensions. There were twenty four unique criteria. So, for each criterion a decision had to be made about its relevance to each of the three dimensions. In total, each rater had to make 72 dichotomous decisions.

Fleiss' Kappa was used to assess agreement among the three raters. This statistic takes into account agreement by chance among a number of raters, and is an appropriate indicator of agreement on nominal categorizations (Fleiss, 1971). Fleiss' Kappa for categorizations into dimensions by the three raters was .44. This is widely considered to be "Moderate agreement", which is less than "Substantial agreement" but more than "Fair agreement". These anchors are suggested by Landis and Koch (1977). They are not universally accepted, however "Moderate agreement" does fall in line with the average of 74% raw agreement found among the three possible pairs of raters. These indicators of agreement are for categorizations that were made independently by each rater. In cases where raters did not agree, the categorization of criteria were carefully discussed by the

three raters in an attempt to come to consensus. After discussion, all three raters were in perfect agreement.

An additional set of codes was assigned by the author to each criterion according to either its specificity in terms of behavior, or its specificity in terms of context. This was done according to a set of rules that required little judgment. Criteria that reflect composites of different behavior types were not considered specific. While singular behaviors and behaviors that belonged to a clear and well defined category of behaviors were considered to be specific. Likewise, ratings where perspectives change substantively between raters are not considered specific in terms of context, nor are behaviors that are measured across organizations that are clearly different in terms of activity, culture, goals, etc. Measurements of behaviors that target a specific organization, are generated within a specific experiment, or are generated within the confines of a specific type of activity are considered specific in context.

3.3 Study Level Results

Table 2 presents the statistics associated with the tests of hypotheses 1 – 3 at the study level. Statistics that are bold conform to expectations that derive from the categorizations of criteria that are displayed on the left side of the table.

General relationships between all variables were assessed using product moment, biserial, polyserial, or polychoric correlations as is appropriate given the scales and the underlying nature of the scales being related (Jöreskog & Sörbom, 1999). PRELIS 2.8 was used to estimate all correlations, which were calculated pairwise (Jöreskog & Sörbom, 2008).

Interactions terms and sub-scale validities for dichotomous criteria were estimated using both ordinary least squares (OLS) and logistic regression. Differences in results from these two analysis procedures did not change interpretations of outcomes. Therefore, results for hypotheses one and three based on OLS are reported for dichotomous criteria to maintain consistency across studies. Additionally, AICc was calculated for models 1 and 2 using both OLS and logistic regression techniques for dichotomous criteria. In five of the comparisons between models 1 and 2 the delta AICc switched signs when going from OLS to logistic regression. As delta AICc is not an indicator of effect size and because results stemming from logistic regression techniques appear to be more conservative for hypothesis 2, delta AICc values for dichotomous criteria are reported from the logistic regression results.

Table 2: Categorizations by Dimension and Specificity for Each Criterion and study level statistics associated with the tests of hypotheses 1 – 3

Criterion and Categorizations							Statistics for Hypotheses 1 - 3				
		Specificity		Dimension			H1			H2	H3
Source	Criterion Description	*Behavior*	*Context*	*Externalizing*	*Internalizing*	*Powerlessness*	T_2 EXT[a]	T_2 INT[b]	T_2 POW[c]	Model 1 – Model 2 ΔAICc	Model 2 to Model 3 ΔR^2
Baysinger (2013)	Negative Socioemotive Behavior	No	Yes			X	-0.74	-0.71	1.46	-0.74	.007
Bing (2007)	Passive Organizational Deviance	No	No			X	-1.51	1.53	-0.01	**0.04**	.02
Bing (2007)	Active Organizational Deviance	No	No	X		X	-0.28	-0.81	1.06	-2.61	.01
Bing (2007)	Passive Interpersonal Deviance	No	No	X	X	X	**-0.43**	**0.51**	**-0.07**	-3.99	.03
Bing (2007)	Active Interpersonal Deviance	No	No	X	X	X	**-0.42**	**0.23**	**0.19**	-4.33	.003
Bing (2007)	Filing Complaints[d]	Yes	Yes	X	X		-1.62	**2.39**[*]	-0.77	-3.23	**.137**[**]

Table 2 (Continued)

Criterion and Categorizations							Statistics for Hypotheses 1 - 3				
		Specificity		Dimension			H1			H2	H3
Source	Criterion Description	*Behavior*	*Context*	*Externalizing*	*Internalizing*	*Powerlessness*	T_2 EXT[a]	T_2 INT[b]	T_2 POW[c]	Model 1 – Model 2 ΔAICc	Model 2 to Model 3 ΔR^2
DeSimone (2010)	Inappropriate Corrective Action Report[d]	Yes	Yes			X	-1.91*	-1.18	**3.04***	-1.08	**.206****
Frost (2002)	Technical Fouls	Yes	Yes			X	-0.57	-3.22*	**3.99****	**17.08**	**.072***
Frost (2002)	Verbal Harassment	Yes	Yes			X	-1.16	-0.86	**2.07***	**2.76**	**.082***
Frost (2002)	Composite of Excessive Behaviors	No	Yes	X	X	X	**0.45**	**-1.52**	**1.06**	-3.41	.054
Frost (2002)	Physical Altercations	Yes	Yes	X	X	X	2.10*	-1.93	-0.17	1.21	.044
Frost (2005)	Obstruction	Yes	Yes	X		X	1.26	**-2.35***	1.24	**14.92**	**.143****
Frost (2005)	Overt Aggression	Yes	Yes	X	X	X	**1.46**	**-1.53**	**0.20**	-3.02	**.072****
Frost (2005)	Expressions of Hostility	Yes	Yes			X	1.00	-1.52	0.58	-0.08	.014

Table 2 (Continued)

Criterion and Categorizations							Statistics for Hypotheses 1 - 3				
		Specificity		Dimension			H1			H2	H3
Source	Criterion Description	*Behavior*	*Context*	*Externalizing*	*Internalizing*	*Powerlessness*	T_2 EXT[a]	T_2 INT[b]	T_2 POW[c]	Model 1 – Model 2 ΔAICc	Model 2 to Model 3 ΔR^2
James and McIntyre (2000) - Study 7	Absences	Yes	Yes	X		X	0.52	-1.15	0.70	-2.55	.044
James and McIntyre (2000) - Study 8	Blatant Aggression[d]	No	No	X	X		2.39*	-0.50	**-1.64***	-0.33	.04
James and McIntyre (2000) - Study 8	Unreliability (w Blatant Aggression)	No	No	X	X	X	**0.67**	**0.24**	**-0.87**	2.43	.018
Mc Niel (2009)	Pausing[d]	Yes	Yes		X		-0.54	**1.68***	-1.17	-3.23	.039
Mc Niel (2009)	Complaining[d]	Yes	Yes	X	X		2.12**	-1.12	-1.06	-2.62	.0398
Mc Niel (2009)	Excessive Pressure 5 min	Yes	Yes			X	-0.36	-1.25	**1.67***	-3.55	.0039
Mc Niel (2009)	Dishonest Reporting[d]	Yes	Yes	X	X		-2.54**	-0.78	3.54**	**2.44**	.001
Mc Niel (New	Lack of effort	Yes	Yes	X			**2.97***	-1.58	-1.24	**7.74**	.011

Table 2 (Continued)

Criterion and Categorizations							Statistics for Hypotheses 1 - 3				
		Specificity		Dimension			H1			H2	H3
Source	Criterion Description	*Behavior*	*Context*	*Externalizing*	*Internalizing*	*Powerlessness*	T_2 EXT[a]	T_2 INT[b]	T_2 POW[c]	Model 1 – Model 2 ΔAICc	Model 2 to Model 3 ΔR^2
Sample) - Sample 1 Pilot											
Mc Niel (New Sample) - Sample 2	Lack of effort	Yes	Yes	X			**2.15***	-1.19	-1.02	**0.27**	.0381
Mc Niel (New Sample) - Sample 3	Lack of effort	Yes	Yes	X			**2.18***	-2.91*	0.64	**6.26**	.0408
Mc Niel (New Sample) - Sample 4	Lack of effort	Yes	Yes	X			0.41	-0.76	0.35	-2.27	.018
Sablynski (2001)	Theft[d]	Yes	Yes	X	X	X	**-1.31**	**-0.15**	**1.48**	-2.41	**.458****
Sablynski (2001)	Disobedience	Yes	Yes	X	X	X	**0.55**	**-1.41**	**0.91**	-3.31	**.126***

Note: Bold statistics conform to the expectations derived from Hypotheses 1 – 3.
[a]T_2 EXT refers the test for differences between dependent correlations between the EXT/Criterion correlation and the Average of the INT/Criterion and POW/Criterion correlations.
[b]T_2 INT refers the test for differences between dependent correlations between the INT/Criterion correlation and the Average of the EXT/Criterion and POW/Criterion correlations.

[c]T_2 POW refers the test for differences between dependent correlations between the POW/Criterion correlation and the Average of the EXT/Criterion and INT/Criterion correlations.
[d]Dichotomous variable - ΔAICc is derived from logistic regression techniques

3.4 Summary Results

The first hypothesis of this study is concerned with the predictive utility of CRT-A sub-scales when these scales are being used to predict criteria that either are or are not relevant to a dimension that a particular scale is associated with. At the study level, 27 separate predictions were made about the relative strength of correlations between criteria and the Externalizing, Internalizing, and Powerlessness scales. The T_2 tests associated with 18 of these predictions were either significant in the expected direction or supported the expected null hypothesis of no difference between dependent correlations. The T_2 tests for 4 additional predictions trended in the expected direction, but were not significant. And, the T_2 tests for 5 of these predictions did not trend in expected directions or were significant in unexpected directions.

Because T_2 results across individual studies are effected by differences in sample size, CRT-A scale intercorrelations, and effect size magnitudes, a more stable summary statistic across studies is also appropriate to look at. A meta-analytic approach was taken to summarize each of the three CRT-A sub-scale associations with criteria when moderated by alignment or misalignment between a scale and criteria on the dimension the sub-scale represents. Hypotheses 1a – 1c separate the expectations for hypothesis 1 into those for the Externalizing, Internalizing, and Powerlessness scales respectively, and were tested using this meta-analytic approach. The expectation for H1a is that the Externalizing scale will have a larger average correlation with criteria that have been categorized as being more relevant to the Externalizing dimension (aligned correlations) compared to criteria that have not been categorized as being as relevant to the Externalizing dimension (misaligned correlations). The expectation for H1b is that the

Internalizing scale will have a larger average correlation with criteria that have been categorized as being more relevant to the Internalizing dimension (aligned correlations) compared to criteria that have not been categorized as being as relevant to the Internalizing dimension (misaligned correlations). And, the expectation for H1c is that the Powerlessness scale will have a larger average correlation with criteria that have been categorized as being more relevant to the Powerlessness dimension (aligned correlations) compared to criteria that have not been categorized as being as relevant to the Powerlessness dimension (misaligned correlations).

A separate moderated meta-analysis of the relationships between each of the three scales and criteria was performed. Both average correlations and sample weighted average correlations for each grouping of correlations (i.e. from aligned or misaligned correlations) were calculated. If a sample had more than one correlation that was appropriate for a grouping, those correlations were averaged together. Initially, correlations from misaligned relationship were only included in an analysis if correlations from aligned relationships were absent in the same study. However, for each of the three sub-scales, there were a limited number of studies that only had misaligned correlations. Therefore, a separate set of analyses were performed that included misaligned correlations from studies that also contributed aligned correlations. Table 3 shows the categorizations by dimension for each criterion and the correlations of each criterion with the three CRT-A sub-scales.

Table 4 shows the unweighted and sample weighted mean correlations between the three CRT-A sub-scales and criteria according to alignment or misalignment with dimensions. All sample weighted correlations and associated statistics were calculated

using the random effects model recommended by Hedges and colleagues (Hedges & Olkin, 1985; Hedges & Vevea, 1998; DerSimonian & Laird, 1986). For misaligned correlations, means are presented that come from two analyses. In the first analysis, only correlations from studies without an included aligned correlation are averaged together. In the second analysis correlations from all studies with a misaligned correlation are included. Results were highly similar for both analyses, so only results from the latter method will be discussed. The results from simple averages and the sample weighted analyses were also highly similar, so only results from the sample weighted analyses will be discussed.

For the relationships between the Externalizing scale and criteria, the mean correlation when scales align with criteria is .24, and the mean correlation when scales do not align with criteria is .03. The 95% confidence intervals for these two means do not overlap. This indicates that alignment on dimensionality moderates the relationships between the Externalizing scale and criteria. For the relationships between the Powerlessness scale and criteria, the mean correlation when scales align with criteria is .24, and the mean correlation when scales do not align with criteria is .02. The 95% confidence intervals for these two means do not overlap. This indicates that alignment on dimensionality moderates the relationships between the Powerlessness scale and criteria. For the relationships between the Internalizing scale and criteria, the mean correlation when scales align with criteria is .14, and the mean correlation when scales do not align with criteria is .07. The difference between these two means is not significant. This indicates that alignment on dimensionality may not moderate the relationships between the Internalizing scale and criteria. However, these results trend in the expected direction.

Table 3: Categorizations by dimension for Each Criterion and the Correlations of Each Criterion with the Three CRT-A Sub-scales

Sample Description			Dimension and Scale Correlations					
			Dimensions			CRT-A Sub-Scale		
Source	Sample	Criterion Description	*Externalizing*	*Internalizing*	*Powerlessness*	EXT[b]	INT[c]	POW[d]
Baysinger (2013)	Undergraduates	Negative Socioemotive Behavior			X	0.08	0.08	0.17
Bing (2007)	Hospital Workers	Passive Organizational Deviance			X	-0.04	-0.07	0.05
Bing (2007)		Active Organizational Deviance	X		X	-0.04	0.02	-0.02
Bing (2007)		Passive Interpersonal Deviance	X	X	X	-0.03	0.02	0.01
Bing (2007)		Active Interpersonal Deviance	X	X	X	0.14	0.05	0.08
Bing (2007)		Filing Complaints	X	X		0.21	0.22	-0.13
DeSimone (2010)	Nuclear Power plant Operators	Inappropriate Corrective Action Report			X	-0.01	0.04	0.38

Table 3 (Continued)

Sample Description			Dimension and Scale Correlations					
			Dimensions			CRT-A Sub-Scale		
Source	Sample	Criterion Description	*Externalizing*	*Internalizing*	*Powerlessness*	EXT[b]	INT[c]	POW[d]
Frost (2002)	Basketball Players	Technical Fouls			X	0.18	0.01	0.46
Frost (2002)		Verbal Harassment			X	0.04	0.07	0.26
Frost (2002)		Composite of Excessive Behaviors	X	X	X	0.18	0.05	0.22
Frost (2002)		Physical Altercations	X	X	X	0.32	0.05	0.17
Frost (2005)	Basketball Players	Obstructionism	X		X	0.50	0.30	0.50
Frost (2005)		Overt Aggression	X	X	X	0.45	0.28	0.38
Frost (2005)		Expressions of Hostility			X	0.01	-0.15	-0.01
James and McIntyre (2000) - Study 7	Package Handlers	Absences	X		X	0.28	0.15	0.30
James and McIntyre (2000) - Study 8	Temporary Employees	Blatant Aggression	X	X		0.50	0.28	0.18
James and McIntyre (2000) - Study 8	Temporary Employees	Unreliability (w Blatant	X	X	X	0.33	0.30	0.20

Table 3 (Continued)

Sample Description			Dimension and Scale Correlations					
			Dimensions			CRT-A Sub-Scale		
Source	Sample	Criterion Description	*Externalizing*	*Internalizing*	*Powerlessness*	EXT[b]	INT[c]	POW[d]
		Aggression)						
Mc Niel (2009)	Undergraduates	Pausing		X		-0.24	-0.09	-0.28
Mc Niel (2009)		Complaining	X	X		0.21	-0.02	-0.01
Mc Niel (2009)		Excessive Pressure 5 min			X	0.08	0.02	0.22
Mc Niel (2009)		Dishonest Reporting	X	X		-0.22	-0.10	0.18
Mc Niel (New Sample) - Sample 1 Pilot	Undergraduates	Lack of effort	X			0.33	-0.05	-0.03
Mc Niel (New Sample) - Sample 2	Undergraduates	Lack of effort	X			0.29	-0.08	-0.07
Mc Niel (New Sample) - Sample 3	Undergraduates	Lack of effort	X			0.20	-0.21	0.07

Table 3 (Continued)

Sample Description			Dimension and Scale Correlations					
			Dimensions			CRT-A Sub-Scale		
Source	Sample	Criterion Description	*Externalizing*	*Internalizing*	*Powerlessness*	EXT[b]	INT[c]	POW[d]
Mc Niel (New Sample) - Sample 4	Undergraduates	Lack of effort	X			0.17	0.08	0.17
Sablynski (2001)	Undergraduates	Theft	X	X	X	0.21	0.30	0.43
Sablynski (2001)		Disobedience	X	X	X	0.28	0.11	0.31

[b]EXT refers to the score on the externalizing scale of the CRT-A.
[c]INT refers to the score on the internalizing scale of the CRT-A.
[e]POW refers to the powerlessness scale of the CRT-A.

Table 4: Unweighted and Sample Weighted Mean Correlations between the Three Scales and Criteria According to Alignment or Misalignment on Dimension Association

Grouping	Total *N*	*k*	r_{mean}	SE_r	95% CI
Externalizing					
Aligned	1419	11	.24 (.24)	(.048)	(.14, .34)
Misaligned[a]	585	2	-.01 (.06)	(.041)	(-.02, .14)
Misaligned all studies[b]	1293	6	.03 (.03)	(.039)	(-.07, .08)
Internalizing					
Aligned	987	6	.14 (.14)	(.056)	(.03, .25)
Misaligned[a]	1094	7	.00 (.06)	(.03)	(.0, .12)
Misaligned all studies[b]	1897	11	.01 (.07)	(.024)	(.02, .12)
Powerlessness					
Aligned	1677	9	.24 (.24)	(.04)	(.16, .31)
Misaligned[a]	404	4	.04 (.05)	(.051)	(-.05, .15)
Misaligned all studies[b]	938	7	.02 (.02)	(.042)	(-.07, .11)

Note: Statistics in parentheses are from sample weighted analyses.
[a]Misaligned correlations are only included if aligned correlations are not drawn from the same study
[b]Misaligned correlations are included across all studies

Hypothesis 2 was tested using ΔAIC_c. The supposition was that irrelevant variance and countervailing relationships will reduce the relationship between the CRT-A total score and criteria that are only relevant to one or two dimensions. At the same time, countervailing influences will be picked up by looking at each CRT-A sub-scale, and irrelevant variance from scales not aligned with a criterion will primarily effect the

regression weights of those scales. In cases where countervailing influences and irrelevant variance are extreme enough, ΔAIC_c will be smaller for a model that includes all three subscales instead of the single CRT-A total score.

However, factors other than the pattern of relationships between the CRT-A, its sub-scales, and criteria effect ΔAIC_c (model 1 – model 2). If all relationships are weak, AICc will prefer the simpler model, even if the effects mentioned above are occurring. Additionally, if scales have differing relationship strengths with a criterion, but account for variance in a highly additive way, then AIC_c will also prefer the simpler model. Because of these considerations, an element of randomness is expected in result that pertain to H2 predictions. However, ΔAIC_c will still only be smaller for model 2 when subscales are indicating countervailing influences on a criterion or random associations with a criterion.

To take this randomness into consideration, counts of all ΔAICc results were gathered into a 2x2 table. Positive and negative ΔAIC_c values are contrasted with criteria that are associated with either all three or only one or two dimensions. The expectation is that a 1 tailed Fisher's exact test will show a significant relationship between these two categorizations, where positive ΔAICc values are more likely for criteria that do not associate with all three functions. Table 5 shows the 2x2 cell counts and the associated Fisher's exact test statistics. Although the results trend in the expected direction, the 1 tailed test is not significant with a *p* value of .349.

Table 5: Cell Counts and Fisher's Exact Test for the 2x2 contingency table categorizing ΔAICc values and Criteria Associations with Dimensions.

	ΔAICc < 0	ΔAICc > 0	Marginal Counts
Criterion associated with 1 or 2 dimensions	11	8	19
Criteria associated with all 3 dimensions	6	2	8
Marginal Counts	17	10	27
		P value 1 tailed	.349
		P value 2 tailed	.665

The expectation for hypothesis 3 is that significant interaction terms are more likely when criteria are categorized as being specific in terms of behavior and context. Like the summary test for hypothesis 2, this expectation was assessed using a 2x2 contingency table and the Fisher's exact test. The expectation is that a 1 tailed Fisher's exact test will show a significant relationship between the categories specific vs. not specific and interaction present vs. no interaction present, where an interaction is more likely for criteria that are specific in terms of behavior and context. Table 6 shows the 2x2 cell counts and the associated Fisher's exact test statistics. The 1 tailed test is significant with a *p* value of .034.

Table 6: Cell Counts and Fisher's Exact Test for the 2x2 contingency table categorizing criteria in terms of Specificity and the Presence of Significant Interactions.

	No Interaction	Interaction Present	Marginal Counts
Specific	11	8	19
Not Specific	8	0	8
Marginal Counts	19	8	27
		P value 1 tailed	.034
		P value 2 tailed	.061

CHAPTER 4

DISCUSSION

This study had two primary objectives. The first was to see if criteria can be expected to have different relationships with CRT-A subscales based on the characteristics of criteria and the contexts in which criteria are measured. The second was to examine the utility of the CRT-A subscales for predicting and understanding behavior.

4.1 Hypotheses

The results for Hypothesis 1 provided consistent support for idea that the relationships between the three CRT-A sub-scales and behaviors are relatively larger when there is alignment between sub-scales and the perceived relevance of criteria to Externalizing, Internalizing, and Powerlessness. However, support was much clearer when alignment was on the Externalizing and Powerlessness dimensions. It may be that the alignment of behaviors with Internalizing is simply more difficult to judge. This makes sense as behaviors stemming from Internalizing mechanisms are by definition less provoked by external cues and more influenced by internal motivations and objectives. In studies, frustrating circumstances and offensive external factors are easy to see, but what a person actually wants can be well hidden. Despite the less compelling evidence associated with the Internalizing dimension, it appears that if dimensions can be accurately associated with behaviors within specific contexts, then realistic predictions can be made about the relative associations between dimensions of aggression and outcomes.

Results from hypotheses 2 and 3 provide strong evidence that the three sub-scales of the CRT-A can contribute to our understanding of aggression when considered separately from the CRT-A total score. About 40% of the criteria were described better by a model that included the three CRT-A sub-scores rather than the total score. While this did not always happen when expected, results did trend in the expected direction. Violations of expectations for the overall test of hypothesis 2 are likely due to difficulties in categorizing criteria and in difficulties judging effect sizes and the comparative effects of dimensions on criteria.

Interactive effects of sub-scales on criteria were also fairly common and occurred for approximately 30% of the criteria. Additionally, it was possible to some degree to predict when interactions would occur based on the specificity of context and behavior. Also, it should be noted that an interaction did not occur in any of the criteria that were not specific in either behavior or context.

Combined, 55% of the criteria examined are either described better by model 2 or had a significant amount of variance associated with the interaction terms in model 3. These results generally support the idea that scales should be looked at individually as well as in aggregation in the form of the CRT-A total score. By looking at scales individually, it is possible to see more effects that are relevant to our understanding of why and under what circumstances aggression occurs. This is further demonstrated by the post hoc analyses of significant interaction terms presented in the Appendix.

4.2 Dimensional Interactions

Post hoc analyzes of interactions revealed some interesting and consistent findings involving the Powerlessness scale. The discussion on interactions will therefore

start with Powerlessness. Because this scale is associated with frustration and uncomfortable feelings of constraint, an expectation can be made that Powerlessness will be related to an activation of systems that make all types of aggressive actions more preferable. This expectation is in line with research stemming from the frustration aggression hypothesis. The current version of this theory is Berkowitz's Cognitive Neoassociation Theory (CNT) (1989, 1990, 1993). CNT links aversive stimuli, such as heat, loud noises, unpleasant odors, provocations, and frustrations to negative affect. Negative affect induces automated activation of systems associated with fight or flight such as expressive motor reactions, physiological responses, memories and thoughts. If the fight system is more activated, then vague feelings of anger may cause a search for the source of this anger. The result is that feelings of frustration and discomfort may make unfair circumstances seem more onerous, pressures from others seem more malevolent, and barriers keeping more vicious tactics in check weaker. In this way, Powerlessness can be expected to synergize with Externalizing and Internalizing when influencing behavior. For six criteria, this exact relationship was found. Inappropriate corrective action reports, technical fouls, verbal hostility, Obstructive behaviors, disobedience, and theft were all influenced by a synergistic interaction between powerlessness and either Externalizing, Internalizing, or both.

While a synergistic relationship between the Powerlessness scale and the Internalizing and Externalizing scales is likely the most common. The analysis of grievances suggests that other interactive relationships are likely as well. In Bing (2007) the Powerlessness by Externalizing interaction term was significance for the grievances criterion. In this sample, powerlessness suppressed the relationship between externalizing

and grievances. It can be speculated that trait powerlessness creates distrust in official channels, reducing the likelihood that a person will use them to address perceived problems. This creates a series of interesting questions. Can this relationship be found in other settings? Does it generalize to uses of other formal systems? What are the preferred behaviors when formal systems are not used?

Finally, both synergistic and suppressive relationships are found when Externalizing and Internalizing have interactive effects on criteria. In some cases, it is difficult to tell what is going on because the effects are too weak. This is true for the inappropriate corrective action report and verbal hostility criteria. However, for technical fouls and verbal hostility during a basketball game, Externalizing suppresses the effect of Internalizing. This may be due to an increased awareness of authorities that would call a foul. For obstructionism during a basketball game, Internalizing and Externalizing have a synergistic effect. In this case, a general need to get revenge coupled with feelings of harassment by officials may simply create a greater impetus to obstruct. For overt aggression during a basketball game, Internalizing and Externalizing also have a synergistic effect. It is highly likely in this case that two reasons to aggress that come from two motivational systems provides a much stronger impetus to act. And finally for theft, Externalizing suppresses the effect of Internalizing. As with technical fouls and verbal hostility, it may be that higher Externalizing makes a person more perceptive of powers that might exert punishment. If this is the case, then two motives that both influence one to aggress may cause a degree of dissonance when choosing aggressive actions that could lead to punishment.

The reasons given for these interactions are speculative, as examining the interactive effects of the implicit subsystems of a broader construct (in this case aggression) is new territory. Educated guesses can be made based on situational familiarity, psychological insight, and, where possible, ties to existing lines of literature. However, further study of these systems that narrow the scope of examination to specific types of interactions between specific sets of sub-scales has the potential to increase our knowledge of aggression as a construct and its use in the world.

4.3 Limitations and Future Directions

The major limitation of this study was the number of available samples that were appropriate to include. This issue was compounded by the need to gather source data in order to obtain the statistics of interest. 14 samples were found that had criteria that were indicative of aggressive behavior, but only 13 data sets could be obtained. This sample limitation issue was ameliorated by the analyses conducted at the study level. To the extent that sample level findings conformed to expectations, uncertainty about the stability of overall findings due to the limitation of available samples should be reduced.

Further, the criteria from all studies analyzed were initially conceived of as being behavioral indicators of overall aggression. In the source studies, relationships with specific CRT-A subscales were not considered as the studies were being designed. The only exception to this was for the passive aggressive lack of effort criteria that was first reported on in this document. As described in the Appendix in the McNiel (New Samples) section, this criteria was designed to be irrelevant to powerlessness, and to activate Externalizing and Internalizing in such a way as to result in opposite behaviors. While much information can be gleaned by looking at general aggression, results were

likely muted by the fact that previous criteria could be relevant to all three dimensions of aggression to some extent, if in differing magnitudes. To get a clearer picture of what is going on in terms of dimensional relationships with behaviors, more criteria must be conceived that theoretically align with specific dimensions while either being irrelevant to other dimensions or being non-compatible with other dimensions. This should result in stronger measured relationships that are much clearer in terms of meaning. And, to the extent that these specific relationships can be replicated across similar behaviors in similar contexts, we can start to build up a lexicon that describes the conditions under which certain types of aggressive behaviors are most likely to occur (or not) and the reasons for such behavior.

The building of a lexicon will be greatly aided by studies that attempt to vary the associations of a criterion with dimensions of aggression or that attempt to interchange criteria and impose specific associations through contextual factors. This may be done through careful experimental manipulation, or it may be done by careful scouting of situations that have naturally occurring variance in behavioral choices and/or environmental pressures that are relevant to aggression. A simple example of this may be seen in a hypothetical extension of the Sablynski (2001) study. In this study two groups of students competed for prizes and the ability to leave early while getting full extra credit. One group had an obvious disadvantage on the task, and was made to feel like they were being treated unfairly. In this situation, it was expected that all three dimensions of aggression would be related to theft of a prize by members of the loosing group. Theft is relevant to Externalizing because loosing group members are keeping themselves from being victimized and treated unfairly. Theft is relevant to Internalizing

because the prizes were desirable, and theft acts as a way to get revenge and possess proof of potency. And, theft is related to Powerlessness because the task was highly frustrating (especially since the other group had an easier task). If we created a second condition where both groups were treated the same (fairly), and the task was easy, then the theft of the prize by members of the losing group would no longer be relevant to Externalizing or Powerlessness. Instead, it would only be an expression of a need to gain something because it is desired, ego enhancing, and proof of potency (i.e. a win). This would lead to the expectation that only the Internalizing scale would be related to theft.

Currently, the subscales of the CRT-A are the best way to measure individual differences in Externalizing, Internalizing, and Powerlessness. However, further research would benefit if these sub-scales were expanded in scope to become independent measures of these aspects of aggression. An increased number of related items for each scale would improve not only the psychometric characteristics of the scales, but also the potential for individuals who are higher on these dimensions of aggression to see scenarios that are particularly evocative for them. If the sub-scales can more reliably and accurately identify trait level, then studies that use more elaborated methodologies will be more likely to produce interpretable outcomes.

As research on aggression using conditional reasoning techniques continues, there are many potential directions investigations might take. Three of these are described below that either flow naturally from the existing literature, have immediate implications for research, or have potential for providing answers to existing questions.

The integrative model for aggression extended the work of Winter, John, Stewart, Klohnen, and Duncan (1998). Winter et al.'s original model hypothesized that the explicit

trait extroversion/introversion channels the expression of the implicit motives to affiliate and seek power. In a similar vein, the integrative model for aggression suggests that self-reported levels of aggression channel how implicit aggression is expressed (Bing et al., 2000; James & Mazerolle, 2002; Frost, Ko, & James, 2007). Those who are high on both implicit and explicit aggression see themselves as aggressive and want to be aggressive. They should therefore engage in all sorts of aggressive behavior without experiencing a great deal of dissonance. Conversely, those who are low on both implicit and explicit aggression do not see themselves as aggressive and don't have a drive to aggress. They should therefore be unlikely to engage in aggressive behaviors. Those who are high in explicit and low in implicit aggression will engage in socially acceptable pseudo aggressive behaviors, but are unlikely to actually attempt harm on others. And, those who are high in implicit and low in explicit aggression will likely engage in aggressive behaviors that are highly defensible or might be misinterpreted as being non-aggressive.

For Externalizing, Internalizing, and Powerlessness, one can make similar hypotheses about the channeling of motives from implicit aspects of aggression through explicit levels of aggression (or congruent explicit dimensions of aggression). For example, people who are high in powerlessness and high in self-reported aggression might be more likely to aggress in uncomfortable and frustrating situations. They may even blame the situation for their behavior and expect others to work around them when they are frustrated and uncomfortable. However, people who are high in powerlessness and low in explicit aggression may feel shame at impulsive behaviors and seek to curtail or redirect these impulses. Further, they may be more likely to make amends after aggressive acts.

There are a significant number of combinations to be explored here. A separate integrative model can be created for each dimension of implicit aggression measured by the CRT-A, and for each combination of the sub-facets of implicit aggression. Additionally, it may be that channeling is most relevant when explicit aggression is measured in such a way that it aligns with implicit aspects of aggression in terms of content (i.e. does a person see themselves as harming others for defensive purposes, to feel superior, or because they are easily frustrated)

The results centered on passive aggressive lack of effort reported in this document suggest an application in experimental design. In many studies it is very difficult to see if a person is withholding effort. Controls and checks may ensure to some extent that results have integrity even if some people are not trying as hard, but in repeated measures designs where relatively few people act as their own control, a few people withholding effort might seriously compromise a study. The predictions made in this study were based on classifying passive aggressive lack of effort as being relevant to the Externalizing dimension. This classification appears to be accurate within the context of our data collection, but is also likely accurate with respect to many other experimental designs. A study that looks at a variety of experimental tasks and compares the strength and interpretability of results for people who are low in externalizing vs. high in externalizing might be informative. The same might be done using the Powerlessness scale for experimental tasks that are highly frustrating.

Finally, full measures of Externalizing, Internalizing, and Powerlessness may be used with various measures of explicit aggression to cultivate a more complete picture of how trait aggression develops and the circumstances that encourage this development.

Dodge and Coie (1987) and Crick and Dodge (1994) have already created an elaborated theory of how the hostile attribution bias contributes to a cyclical pattern in which aggressive behavior in children occurs from attributing hostile intent in other children, who subsequently alienate and harass the aggressor, who is now more likely to see hostile intent and aggress (see also, Crick & Dodge, 1996; Dodge, Price, Bachorowski, & Newman, 1990). Eventually, this leads to habitual or trait aggression. This bias is part of what is measured in the Externalizing sub-scale of the CRT-A. Having more refined measures of specific aspects of aggression (that would need to be modified for different ages) allows us to ask different questions about the development of the construct. Some of these questions are: What circumstances that compel the use of harm to others are related to specific dimensions of aggression? How do these circumstances affect and modify the cognitive mechanisms that precipitate further aggression? What circumstances discourage the use of harm that is precipitated by different dimensions of aggression? How do these discouraging circumstances affect or modify the cognitive mechanisms that precipitate further aggression?

4.4 Conclusions

Associations between dimensions of implicit aggression and categorized criteria make it clear that the use of aggression is at least partially dependent upon an alignment between specific motivational tendencies and the usefulness of behaviors for expressing these motivations. Consistent with this, CRT-A sub-scale associations with criteria were not always in the same direction, and some scales had much stronger relationships with specific criteria than others. As a result, an analysis model that takes into account differential scale/criteria relationships often described aggressive behaviors more

effectively. Additionally, scales commonly interacted with one another to account for significantly more variance in criteria. It may be that some situations activate multiple dimensions of implicit aggression which create effects that are greater than those that would be expected from a purely additive model. Although CRT-A subscales do not always add useful information beyond that provided by the CRT-A total score, for certain types of criteria, the utility of these subscales for explaining and predicting behavior seems to be quite high.

APPENDIX

Analysis of Individual Studies

In the following sections, the ten studies included in this review will be examined in turn. A more detailed description of the studies and the criteria are provided along with rational for the categorization of each criterion. Statistical tests for hypotheses 1, 2, and 3 for each of the ten studies are presented, and Intercorrelations between the CRT-A, CRT-A scales, and criteria are provided.

The interactive effects of two CRT-A subscales on a criterion are depicted using a technique recommended by Cohen, Cohen, West, and Aiken (2003). Two slopes show the relationships between a criterion and one CRT-A subscale for people at one standard deviation above the mean and one standard deviation below the mean on the other subscale that goes into the depicted interaction.

Baysinger (2013)

This study consisted of a group exercise in which three to five people must individually rank order 15 items in order of importance for survival and then come to consensus as a group. Participants were told that their performance would be compared to other groups and that the goal was to reach the best decision possible. The criterion was self-rated negative socio-emotive behaviors (NSB) on a five item scale. This was assessed after the group exercise. Each person rated the degree to which they expressed frustration, criticized others, and felt that their opinions and suggestions were rejected. The behaviors and emotions rated were inherently counterproductive for this task and were unlikely to increase a sense of self-worth or function to defend against hostility or

victimization. Additionally, expressed frustration and feelings of rejection are in line with what would be expected for a person who is high on trait powerlessness. For these reasons, the NSB criterion was categorized as being more relevant to the Powerlessness dimension. As the ratings were indicative of a wide variety of possible feelings and behaviors, the NSB criterion was coded as non-specific. And, as this criterion was gathered within a fixed experimental design, the context was coded as specific. Table 7 shows the coding decisions, variable intercorrelations, and results by hypothesis for this study.

Because of the categorizations of NSB, this criterion was expected to have a significantly stronger correlation with the Powerlessness scale as compared to the Externalizing and Internalizing scales (H1); the regression of this criterion on Model 2 was expected to show better fit than the regression on model 1 (H2); and the ΔR^2 due to the addition of interaction terms in model 3 was expected to be significant. These expectations were not supported by the results shown in table 7. However, the H1 analysis results trended in the expected direction, and AICc, used for the H2 analysis, always gives preference to the simpler model. Therefore, when all relationships between predictors and criteria are small, as they are for this study, AICc will default to model 1.

Table 7: Coding Decisions, Variable Intercorrelations, and Results by Hypothesis for Baysinger (2013)

Coding Decisions

Criterion (C)	Externalizing	Internalizing	Powerlessness	Behavioral Specificity	Context Specificity
NSB	-	-	X	No	Yes

Variable Intercorrelations (N = 473)

C	CRxE	CRxI	CRxP	ExI	ExP	IxP	CRxC	ExC	IxC	PxC
NSB	0.75**	0.60**	0.64**	0.10*	0.06	0.15**	0.17**	0.08	0.08	0.17**

Hypothesis 1 – T_2 Tests for Differences Between Dependent Correlations

C	E vs. Avg. of I & P	I vs. Avg. of E & P	P vs. Avg. of E & I
NSB	-0.74	-0.71	1.46

Hypothesis 2 – Comparison of Model fit Between Model 1[a] (M1) and Model 2[b] (M2)

C	M1 R^2	M2 R^2	ΔR^2	M1 AIC_c	M2 AIC_c	M1 – M2 AIC_c
NSB	.029**	.036**	.007	-159.51	-158.77	-0.74

Hypothesis 3 – Significance of the F Test for ΔR^2 From M2 to Model 3[c] (M3)

C	M2 R^2	M3 R^2	ΔR^2	B[d] ExI	B ExP	B IxP
NSB	.036**	.043*	.007	.047	-.025	-.01

Note. CR = Total score on the CRT-A; E = score on the Externalizing scale from the CRT-A; I = score on the Internalizing scale from the CRT-A; P = score on the Powerlessness scale from the CRT-A. T_2 Tests are 1 tailed.
[a]Model 1 is the regression of the criterion on the CRT-A total score.
[b]Model 2 is the regression of the criterion on the Externalizing, Internalizing, and Powerlessness scales.
[c]Model 3 adds the Externalizing by Internalizing, Externalizing by Powerlessness, and Internalizing by Powerlessness interaction terms to the regression of the criterion on model 2.
[d]Beta weight for the interaction term from the indicated adjacent scales.
* $p < .05$; ** $p < .01$.

Bing (2007)

In this study, hospital workers were rated in a 360 style assessment by five peers with whom they had worked most frequently. The first four resulting criteria are the mean ratings from the five co-workers on four Likert scales that allow the co-workers to assess the target on negative work behaviors. These work behaviors were on: passive organizational deviance (POD) ("Intentionally worked slower than they could have worked"); active organizational deviance (AOD) ("Took property from work without permission"); passive interpersonal deviance (PID) ("Repeated a rumor or gossip about a coworker or manager at work"); and active interpersonal deviance (AID) ("Acted rudely toward someone at work"). In addition to the 360 style ratings, whether or not a hospital worker filed a complaint was used as a proxy for an indirect attack against the organization (Grievances – GRV). The assumption was that some complaints will be unjustifiable and groundless, but be taken seriously by the organization, thus potentially damaging it.

POD is unlikely to gain positive results for a person and is likely to result in negative social consequences as others in the organization will have to deal with or fix problems that occur due to lack of effort, withholding of information, and other negative passive behaviors. POD behaviors also betray a lack of maturity and understanding of the importance of formal and informal work expectations. For these reasons, the POD criterion was categorized as being more relevant to the Powerlessness dimension. AOD behaviors also disregard organizational rules and expectations, however they are more directed against an organization that has power over the actor. For these reasons, the AOD criterion was categorized as being more relevant to the Externalizing and

Powerlessness dimensions. Both PID and AID behaviors represent a wide array of behaviors that are likely to be used to deal with peers or emotional needs for a variety of reasons. Because of this, they were both categorized as being equally relevant to all three Dimensions. Finally, GRV has utility both as a means to change oppressive and victimizing organizational practices, and as a means to shape the organization according to an actors will. It represents a formal process done without any direct social engagement or immediate feedback that a person who is high on Powerlessness is less likely to view as effective. For these reasons, the GRV criterion was categorized as being more relevant to the Externalizing and Internalizing dimensions. Because the four 360 style ratings represent a multitude of behaviors, perspectives, and targets, they were all coded as non-specific in terms of both behavior and context. Whereas filing grievances is a specific behavioral act that is directed at the organization. GRV was therefore coded as specific in terms of both behavior and context. Table 8 shows the coding decisions, variable intercorrelations, and results by hypothesis for this study.

Because of the categorizations of POD, this criterion was expected to have a significantly stronger correlation with the powerlessness scale as compared to the externalizing and internalizing scales (H1), however the one tailed T_2 test for this comparison was not significant. Additionally, the regression of this criterion on Model 2 was expected to show better fit than the regression on model 1 (H2), however ΔAICc indicates that model 1 should be preferred in this case. And finally, the ΔR^2 due to the addition of interaction terms in model 3 was not expected to be significant, and results support this expectation.

Because of the categorizations of AOD, this criterion was expected to have a significantly weaker correlation with the internalizing scale as compared to the externalizing and powerlessness scales (H1), and while the one tailed T_2 test for this comparison trended in the expected direction, results were not significant. Additionally, the regression of this criterion on Model 2 was expected to show better fit than the regression on model 1 (H2), however ΔAICc indicates that model 1 should be preferred in this case. And finally, the ΔR^2 due to the addition of interaction terms in model 3 was not expected to be significant, and results support this expectation.

Because of the categorizations of both PID and AID, these criteria were not expected to have significantly different correlations with the three CRT-A subscales (H1); the regressions of these criteria on Model 2 were not expected to show better fit than the regressions on model 1 (H2), and ΔR^2 due to the addition of interaction terms in model 3 was not expected to be significant. Results support these expectations.

Because of the categorizations of GRV, this criterion was expected to have a significantly weaker correlation with the Powerlessness scale as compared to the Externalizing and Internalizing scales (H1), and the one tailed T_2 test for this comparison was significant ($T_2 = -3.47$, $p < .01$). Additionally, the regression of this criterion on Model 2 was expected to show better fit than the regression on model 1 (H2), however the ΔAIC_c of -0.33 does not support this expectation. And finally, the ΔR^2 due to the addition of interaction terms in model 3 was expected to be significant, and results support this expectation ($\Delta R^2 = .137$, $p < .01$). Only the Externalizing by Powerlessness interaction term was significant. The B weight for Externalizing was .178 and the B

weight for Powerlessness was -.169. The interaction term was -.358. As seen in figure 1, this suggest that Powerlessness buffers the relationship between Externalizing and GRV.

Table 8: Coding Decisions, Variable Intercorrelations, and Results by Hypothesis for Bing (2007)

Coding Decisions

Criterion (C)	Externalizing	Internalizing	Powerlessness	Behavioral Specificity	Context Specificity
POD	-	-	X	No	No
AOD	X	-	X	No	No
PID	X	X	X	No	No
AID	X	X	X	No	No
GRV	X	X	-	Yes	Yes

Variable Intercorrelations (N = 173)

C	CRxE	CRxI	CRxP	ExI	ExP	IxP	CRxC	ExC	IxC	PxC
POD	0.81**	0.68**	0.59**	0.21*	0.11	0.14*	-0.07	-0.14	0.06	-0.04
AOD	"	"	"	"	"	"	-0.03	-0.04	-0.07	0.05
PID	"	"	"	"	"	"	-0.04	-0.04	0.02	-0.02
AID	"	"	"	"	"	"	-0.05	-0.03	0.02	0.01
GRV	"	"	"	"	"	"	0.21	0.21	0.22	-0.13

Hypothesis 1 – T_2 Tests for Differences Between Dependent Correlations

C	E vs. Avg. of I & P	I vs. Avg. of E & P	P vs. Avg. of E & I
POD	-1.51	1.53	-0.01
AOD	-0.28	-0.81	1.06
PID	-0.43	0.51	-0.07
AID	-0.42	0.23	0.19
GRV	1.70*	1.81*	-3.47**

Table 8 (Continued)

Hypothesis 2 – Comparison of Model fit Between Model 1[a] (M1) and Model 2[b] (M2)

C	M1 R^2	M2 R^2	ΔR^2	M1 AIC_c	M2 AIC_c	M1 – M2 AIC_c
POD	.005	.029	.024	-3.92	-3.95	0.04
AOD	.000	.01	.01	-242.26	-239.63	-2.62
PID	.002	.003	.001	7.40	11.40	-3.99
AID	.002	.001	.001	-8.85	-4.53	-4.33
GRV	.043	.105*	.062*	116.75	117.0852	-0.33

Hypothesis 3 – Significance of the F Test for ΔR^2 From M2 to Model 3[c] (M3)

C	M2 R^2	M3 R^2	ΔR^2	B[d] ExI	B ExP	B IxP
POD	.029	.046	.02	-.028	-.041	-.143
AOD	.01	.022	.005	.018	.014	.06
PID	.003	.031	.03	-.022	-.118	.186
AID	.001	.028	.004	.037	.004	-.027
GRV	.105*	.242**	.137**	-.031	-.358**	.014

Note. CR = Total score on the CRT-A; E = score on the Externalizing scale from the CRT-A; I = score on the Internalizing scale from the CRT-A; P = score on the Powerlessness scale from the CRT-A. T_2 Tests are 1 tailed.
[a]Model 1 is the regression of the criterion on the CRT-A total score.
[b]Model 2 is the regression of the criterion on the Externalizing, Internalizing, and Powerlessness scales.
[c]Model 3 adds the Externalizing by Internalizing, Externalizing by Powerlessness, and Internalizing by Powerlessness interaction terms to the regression of the criterion on model 2.
[d]Beta weight for the interaction term from the indicated adjacent scales.
* $p < .05$; ** $p < .01$.

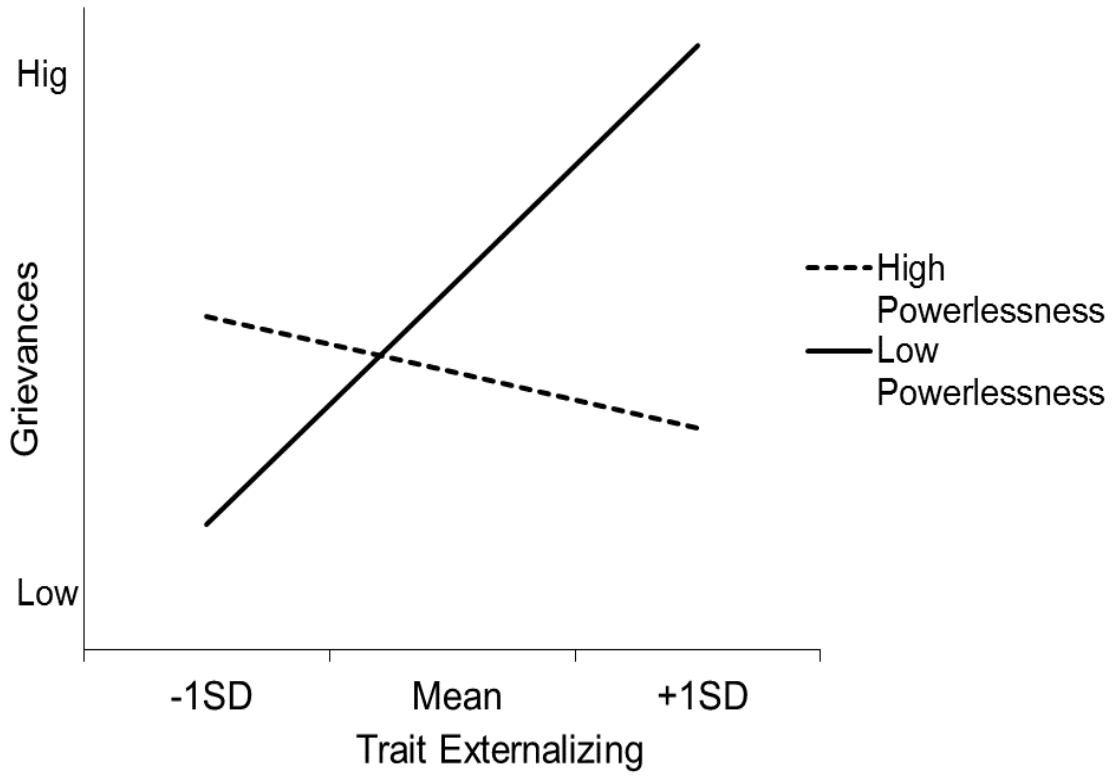

Figure 1. Interactive relationship between Externalizing/Powerlessness and GRV

Desimone (2010)

The criterion for this study was obtained from records of corrective action reports submitted by operators in a nuclear power plant. These reports were classified on a scale that ranged from Urgent "significant issues adverse to quality, issues of significant regulatory concern or public interest, or issues with significant economic impact." to Stupid "not adverse to quality that can be corrected with minimal, if any, evaluation through routine work activities, or that can be closed to actions taken or to trending". The latter category included corrective action suggestions that were irresponsible, ill considered, inconsiderate of existing processes and rules, and that should not have been

submitted. The criterion in this study was whether or not a person submitted a report that falls into this category. This will be referred to as Inappropriate corrective action reports (ICAR). This criterion represents an action that is likely to cause dysfunction and that is indicative of a lack of respect and consideration for both rules and consequences to others. For these reasons, the ICAR criterion was categorized as being more relevant to the Powerlessness dimension. Because ICARs are a narrow type of behavior that takes place in the same cultural environment for all people who write corrective action reports, this criterion was coded as specific in terms of both behavior and context. Table 9 shows the coding decisions, variable intercorrelations, and results by hypothesis for this study.

Because of the categorizations of ICAR, this criterion was expected to have a significantly stronger correlation with the Powerlessness scale as compared to the Externalizing and Internalizing scales (H1), and the one tailed T_2 test for this comparison was significant ($T_2 = 3.04$, $p < .01$). Additionally, the regression of this criterion on Model 2 was expected to show better fit than the regression on model 1 (H2), However the ΔAIC_c of -1.08 does not support this expectation. And finally, the ΔR^2 due to the addition of interaction terms in model 3 was expected to be significant, and results support this expectation ($\Delta R^2 = .206$, $p < .01$). Both the Externalizing by Internalizing and the Internalizing by Powerlessness interaction terms were significant. The B weight for Externalizing was -.034, the B weight for Internalizing was .066, and the B weight for Powerlessness was .417. The Externalizing by Internalizing interaction term was -.16, and the Internalizing by Powerlessness interaction term was .462. As shown in Figure 2, this suggests that Internalizing and Externalizing interfere with one another with regard to

their relationship with ICAR. And as shown in figure 3, this also suggests that Internalizing and Powerlessness have a synergistic relationship with ICAR.

Table 9: Coding Decisions, Variable Intercorrelations, and Results by Hypothesis for Desimone (2010)

Coding Decisions

Criterion (C)	Externalizing	Internalizing	Powerlessness	Behavioral Specificity	Context Specificity
ICAR	-	-	X	Yes	Yes

Variable Intercorrelations (N = 112)

C	CRxE	CRxI	CRxP	ExI	ExP	IxP	CRxC	ExC	IxC	PxC
ICAR	0.84**	0.73**	0.55**	0.32**	0.14	0.08	0.19*	-0.01	0.04	0.38**

Hypothesis 1 – T_2 Tests for Differences Between Dependent Correlations

C	E vs. Avg. of I & P	I vs. Avg. of E & P	P vs. Avg. of E & I
ICAR	-1.91*	-1.18	3.04*

Hypothesis 2 – Comparison of Model fit Between Model 1[a] (M1) and Model 2[b] (M2)

C	M1 R^2	M2 R^2	ΔR^2	M1 AIC_c	M2 AIC_c	M1 – M2 AIC_c
ICAR	.035*	.148**	.098*	50.44	51.51574	-1.08

Hypothesis 3 – Significance of the F Test for ΔR^2 From M2 to Model 3[c] (M3)

C	M2 R^2	M3 R^2	ΔR^2	B[d] ExI	B ExP	B IxP
ICAR	.148**	.354**	.206**	-.016	-.065	.462**

Note. CR = Total score on the CRT-A; E = score on the Externalizing scale from the CRT-A; I = score on the Internalizing scale from the CRT-A; P = score on the Powerlessness scale from the CRT-A. T_2 Tests are 1 tailed.
[a]Model 1 is the regression of the criterion on the CRT-A total score.
[b]Model 2 is the regression of the criterion on the Externalizing, Internalizing, and Powerlessness scales.

[c]Model 3 adds the Externalizing by Internalizing, Externalizing by Powerlessness, and Internalizing by Powerlessness interaction terms to the regression of the criterion on model 2.
[d]Beta weight for the interaction term from the indicated adjacent scales.
* $p < .05$; ** $p < .01$.

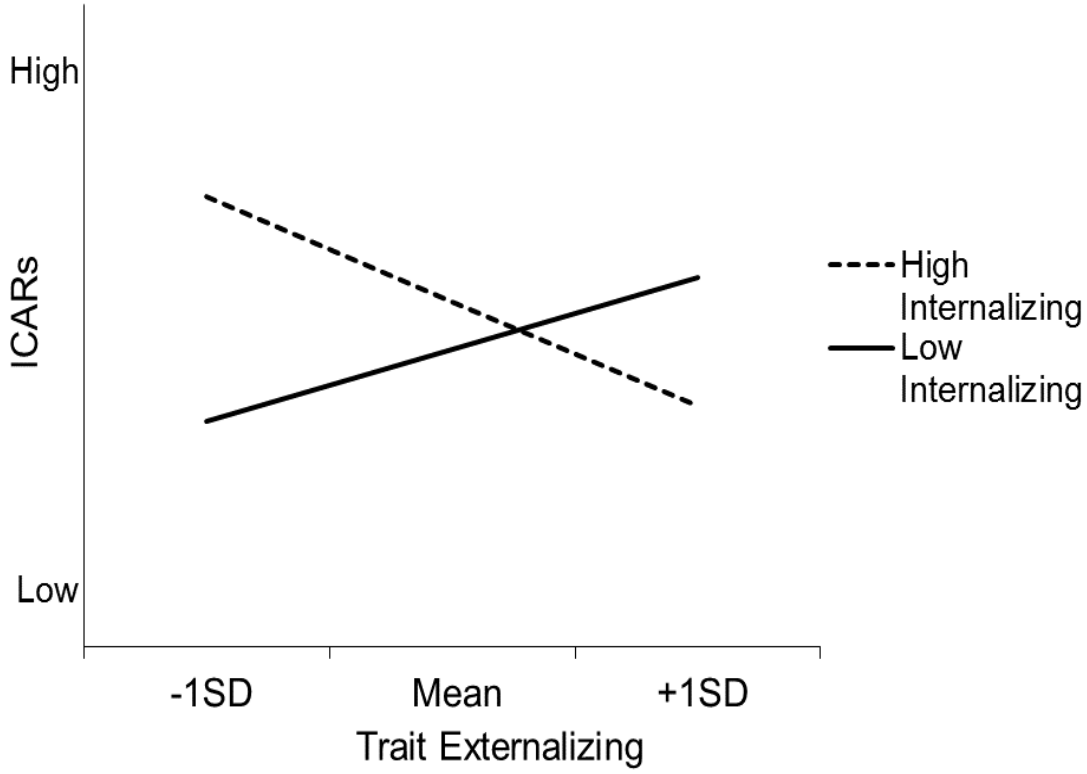

Figure 2. Interactive relationship between Internalizing/Externalizing and ICARs

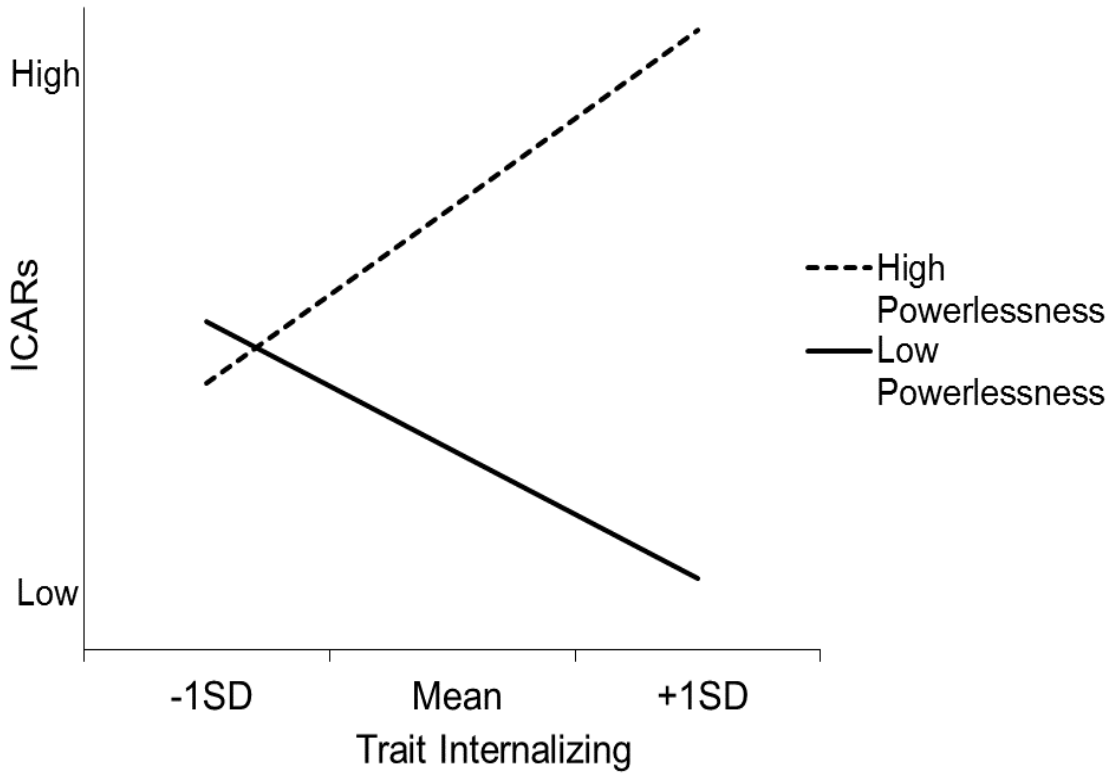

Figure 3. Interactive relationship between Internalizing/Powerlessness and ICARs

Frost (2002)

In this study, all criteria were collected from Intramural basketball player behaviors as they were playing basketball in seasonal games and in tournaments. Trained observers tracked the players and noted technical fouls (TF), incidents of verbal harassment (VH), and incidents of fighting or intense physical altercations (PA). Additionally, a composite criterion that included the most extreme cases of the previous three categories was created (CMP). Variables were created from these incidents based on the number of recorded incidences of each type observed through a playing season for a particular individual. A 0 was assigned when no behavior was displayed, a 1 was

assigned when a behavior was displayed once, and a 2 was assigned when a behavior was displayed multiple times. This study provided detailed examples of the types of behaviors that were classified as TF, VH, and PA.

Almost all of the TF examples given are likely result in a negative penalty during gameplay for the person enacting them, and are unlikely to give the player a significant edge. Additionally, most of these behaviors appear to be emotional or impulsive in nature. For these reasons, the TF criterion was categorized as being more relevant to the Powerlessness dimension. Recorded VH behaviors also tended to be intense, affect laden, and dysfunctional. An argument was made among the raters that these behaviors could be used to establish dominance over other players and to intimidate them. While this is true, most of these behaviors were directed at a referee. For these reasons, it was agreed that the VH criterion was categorized as being more relevant to the Powerlessness dimension. PAs were not coded in this study for who initiated the behaviors. Severe pushing, shoving, and fighting behaviors might have been initiated in a reactionary manner, they might have been defensive, or they might have been an attempt to dominate and intimidate other players. However, PAs were always between players. Because of this, PA was categorized as being equally relevant to all three dimensions. Finally, CMP was deliberately created from a variety of different behavior types. Because of this, it was categorized as being equally relevant to all three dimensions. Because TF, VH, and PA were all defined behavior classes that occurred as part of a well-defined activity, they were all coded as specific in terms of both behavior and context. Whereas CMP is a compilation of explicitly different behavior types that occurred as part of a well-defined activity. CMP was therefore coded as non-specific in terms of behavior but specific in

terms of context. Table 10 shows the coding decisions, variable intercorrelations, and results by hypothesis for this study.

Because of the categorizations of TF, this criterion was expected to have a significantly stronger correlation with the Powerlessness scale as compared to the Externalizing and Internalizing scales (H1), and the one tailed T_2 test for this comparison was significant ($T_2 = 3.99$, $p < .01$). Additionally, the regression of this criterion on Model 2 was expected to show better fit than the regression on model 1 (H2), and a ΔAICc of 17.08 supports this expectation. And finally, the ΔR^2 due to the addition of interaction terms in model 3 was expected to be significant, and the F test for ΔR^2 from model 2 to model 3 supports this expectation ($F = 3.1$, $p < .05$). Both the Externalizing by Internalizing and the Externalizing by Powerlessness interaction terms were significant. The B weight for Externalizing was .222, the B weight for Internalizing was -.042, and the B weight for Powerlessness was .459. The Externalizing by Internalizing interaction term was -.117, and the Externalizing by Powerlessness interaction term was .121. As shown in Figure 4, this suggests that Internalizing buffers the relationship of Externalizing with TF. And as shown in figure 5, this also suggests that Externalizing and Powerlessness have a synergistic relationship with TF.

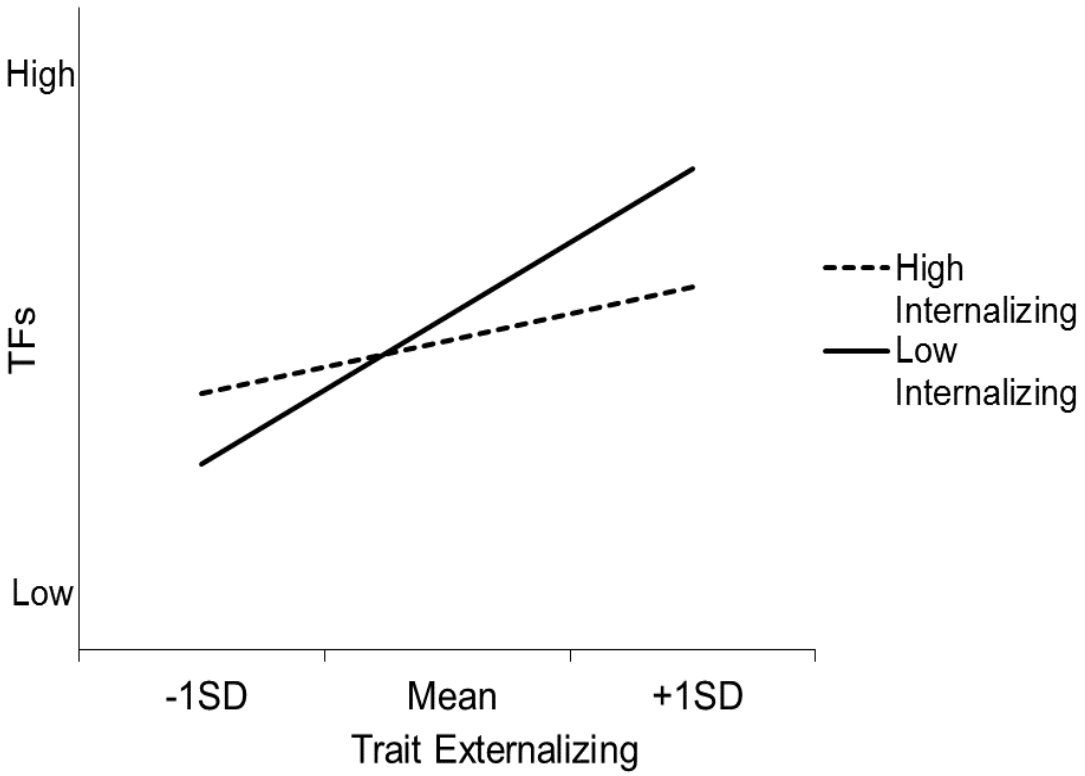

Figure 4. Interactive relationship between Externalizing/Internalizing and TFs

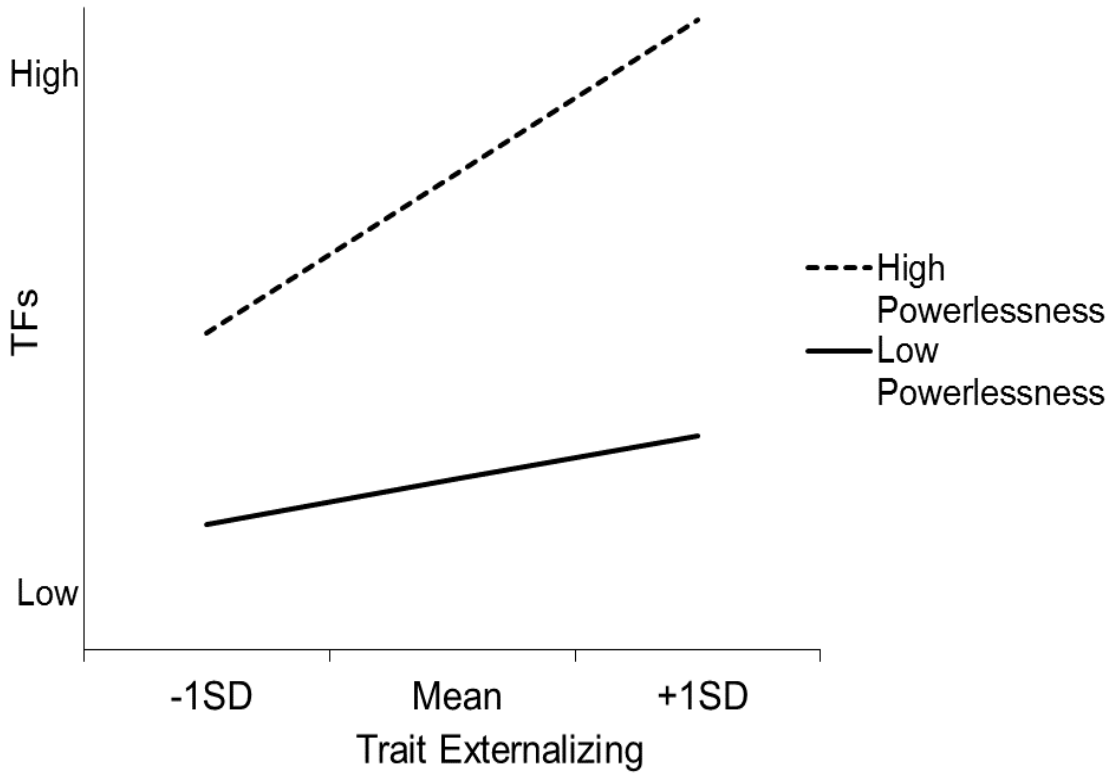

Figure 5. Interactive relationship between Externalizing/Powerlessness and TFs

Because of the categorizations of VH, this criterion was expected to have a significantly stronger correlation with the Powerlessness scale as compared to the Externalizing and Internalizing scales (H1), and the one tailed T_2 test for this comparison was significant ($T_2 = 2.07$, $p < .05$). Additionally, the regression of this criterion on Model 2 was expected to show better fit than the regression on model 1 (H2), and a ΔAICc of 2.76 supports this expectation. And finally, the ΔR^2 due to the addition of interaction terms in model 3 was expected to be significant, and the F test for ΔR^2 from model 2 to model 3 supports this expectation ($F = 2.83$, $p < .05$). Both the Externalizing by Internalizing and the Externalizing by Powerlessness interaction terms were significant. The B weight for Externalizing was .083, the B weight for Internalizing was

.032, and the B weight for Powerlessness was .249. The Externalizing by Internalizing interaction term was -.128, and the Externalizing by Powerlessness interaction term was .143. As shown in Figure 6, this suggests that Internalizing buffers the relationship of Externalizing with VH. And as shown in figure 7, this also suggests that Externalizing and Powerlessness have a synergistic relationship with VH.

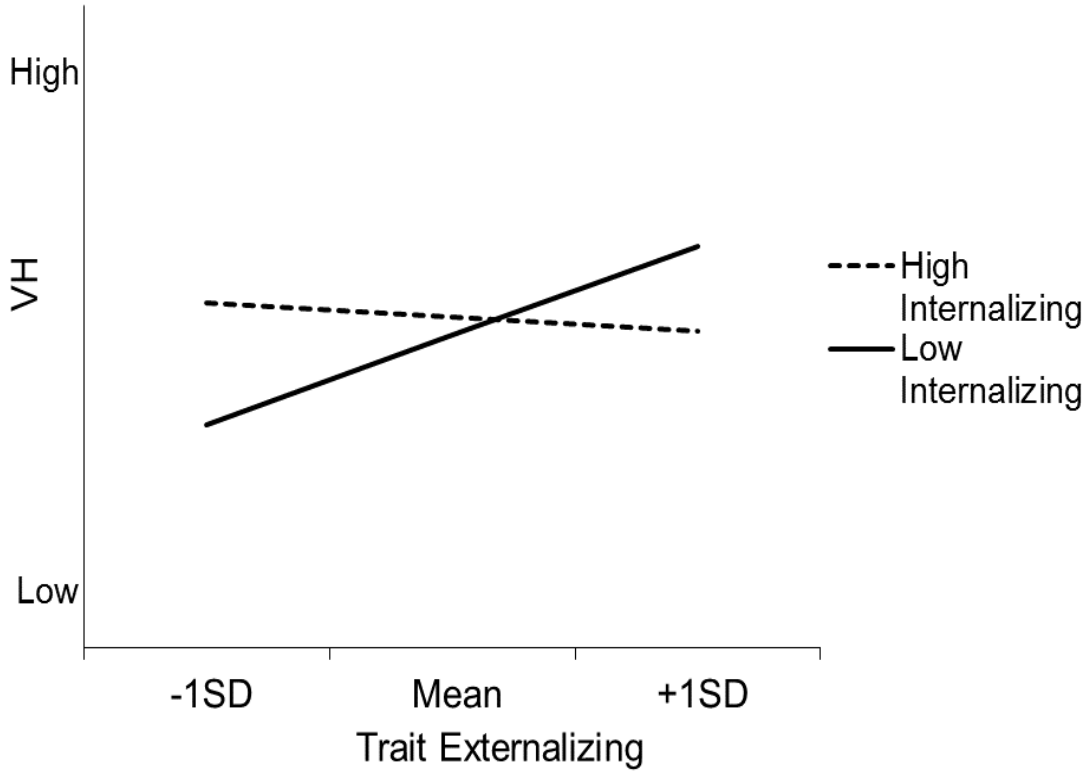

Figure 6. Interactive relationship between Externalizing/Internalizing and VH

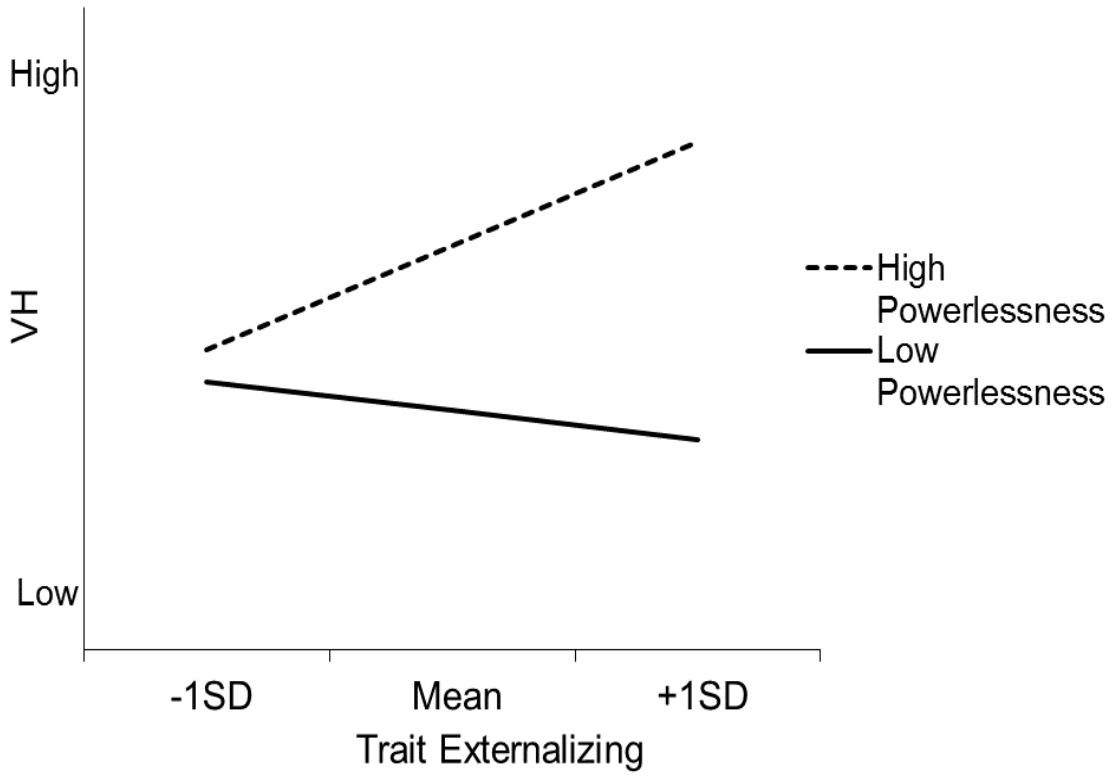

Figure 7. Interactive relationship between Externalizing/Powerlessness and VH

Because of the categorizations of PA, this criterion was not expected to have significantly different correlations with the three sub-scales of the CRT-A (H1), however the one tailed T_2 test for the comparison between Externalizing and the average of Internalizing and Powerlessness was significant ($T_2 = 2.10$, $p < .05$), as was the one tailed T_2 test for the comparison between Internalizing and the average of Externalizing and Powerlessness ($T_2 = -1.93$, $p < .05$). Additionally, the regression of this criterion on Model 2 was not expected to show better fit than the regression on model 1 (H2), and results support this expectation. And finally, the ΔR^2 due to the addition of interaction terms in model 3 was expected to be significant, however the F test for ΔR^2 from model 2 to model 3 does not supports this expectation.

Because of the categorizations of CMP, this criterion was not expected to have significantly different correlations with the three sub-scales of the CRT-A (H1), and results support this expectation. Additionally, the regression of this criterion on Model 2 was not expected to show better fit than the regression on model 1 (H2), and results support this expectation. And finally, the ΔR^2 due to the addition of interaction terms in model 3 was expected to be significant, however the F test for ΔR^2 from model 2 to model 3 does not supports this expectation.

Table 10: Coding Decisions, Variable Intercorrelations, and Results by Hypothesis for Frost (2002)

Coding Decisions

Criterion (C)	Externalizing	Internalizing	Powerlessness	Behavioral Specificity	Context Specificity
TF	-	-	X	Yes	Yes
VH	-	-	X	Yes	Yes
PA	X	X	X	Yes	Yes
CMP	X	X	X	No	Yes

Variable Intercorrelations (N = 185)

C	CRxE	CRxI	CRxP	ExI	ExP	IxP	CRxC	ExC	IxC	PxC
TF	0.74^{**}	0.55^{**}	0.58^{**}	0.02	-0.02	0.07	0.4^{**}	0.18^{*}	0.01	0.46^{**}
VH	"	"	"	"	"	"	0.2^{**}	0.04	0.07	0.26^{**}
PA	"	"	"	"	"	"	0.33^{**}	0.32^{**}	0.05	0.17^{*}
CMP	"	"	"	"	"	"	0.29^{**}	0.18^{*}	0.05	0.22^{**}

Hypothesis 1 – T_2 Tests for Differences Between Dependent Correlations

C	E vs. Avg. of I & P	I vs. Avg. of E & P	P vs. Avg. of E & I
TF	-0.57	-3.22	3.99
VH	-1.16	-0.86	2.07
PA	2.10	-1.93	-0.17
CMP	0.45	-1.52	1.06

Hypothesis 2 – Comparison of Model fit Between Model 1[a] (M1) and Model 2[b] (M2)

C	M1 R^2	M2 R^2	ΔR^2	M1 AIC_c	M2 AIC_c	M1 – M2 AIC_c
TF	$.16^{**}$	$.251^{**}$	$.091^{**}$	-27.07	-44.15	17.08

Table 10 (Continued)

VH	.038**	.074**	.036	-2.02	-4.78	2.76
PA	.107**	.132**	.025	-15.72	-16.92	1.21
CMP	.082**	.086**	.004*	-10.64	-7.23	-3.41

Hypothesis 3 – Significance of the F Test for ΔR^2 From M2 to Model 3[c] (M3)

C	M2 R^2	M3 R^2	ΔR^2	B[d] ExI	B ExP	B IxP
TF	.251**	.323**	.072*	-.117**	.121**	.113
VH	.074**	.156**	.082*	-.128**	.143**	.0862
PA	.132**	.176**	.044	.038	-.018	.243**
CMP	.086**	.052**	.054	-.057	-.148**	.056

Note. CR = Total score on the CRT-A; E = score on the Externalizing scale from the CRT-A; I = score on the Internalizing scale from the CRT-A; P = score on the Powerlessness scale from the CRT-A. T_2 Tests are 1 tailed.
[a]Model 1 is the regression of the criterion on the CRT-A total score.
[b]Model 2 is the regression of the criterion on the Externalizing, Internalizing, and Powerlessness scales.
[c]Model 3 adds the Externalizing by Internalizing, Externalizing by Powerlessness, and Internalizing by Powerlessness interaction terms to the regression of the criterion on model 2.
[d]Beta weight for the interaction term from the indicated adjacent scales.
* $p < .05$; ** $p < .01$.

Frost (2007)

In this study, all criteria were collected from Intramural basketball player behaviors as they were playing basketball in seasonal games and in tournaments. Trained observers tracked the players and noted obstructive behaviors (OB), overt aggression (OA), and expressions of hostility (EH). Variables were created from these incidents based on the number of recorded incidences of each type observed through a playing season for a particular individual. A 0 was assigned when no behavior was displayed, a 1

was assigned when a behavior was displayed once, and a 2 was assigned when a behavior was displayed multiple times. This study provided detailed examples of the types of behaviors that were classified as OB, OA, and EH.

OB behaviors were passive aggressive in nature and tended to be antisocial and/or reactionary. It was not clear if these behaviors were more relevant to Externalizing or Powerlessness. However, given that the behaviors in this category would likely not have happened without an external impetus, it is less likely that they were relevant to the Internalizing dimension. For these reasons, the OB criterion was categorized as being less relevant to the Internalizing dimension. OA behaviors were not coded in this study for who initiated the behaviors. Severe pushing, shoving, yelling, and fighting behaviors might have been initiated in a reactionary manner, they might have been defensive, or they might have been an attempt to dominate and intimidate other players. Because of this, OA was categorized as being equally relevant to all three dimensions. Recorded EH behaviors tended to be intense, affect laden, and dysfunctional. An argument was made among the raters that these behaviors could be used to establish dominance over other players and to intimidate them. While this is true, the recorded behaviors went beyond the bounds of what was socially acceptable as part of play, and more often than not were directed at a referee. For these reasons, it was agreed that the EH criterion was categorized as being more relevant to the Powerlessness dimension. Because OB, OA, and EH were all defined behavior classes that occurred as part of a well-defined activity, they were all coded as specific in terms of both behavior and context. Table 11 shows the coding decisions, variable intercorrelations, and results by hypothesis for this study.

Because of the categorizations of OB, this criterion was expected to have a significantly weaker correlation with the Internalizing scale as compared to the Externalizing and Powerlessness scales (H1), and the one tailed T_2 test for this comparison was significant ($T_2 = -2.35$, $p < .05$). Additionally, the regression of this criterion on Model 2 was expected to show better fit than the regression on model 1 (H2), and a ΔAICc of 14.92 supports this expectation. And finally, the ΔR^2 due to the addition of interaction terms in model 3 was expected to be significant, and the F test for ΔR^2 from model 2 to model 3 supports this expectation ($F = 9.19$, $p < .01$). The Externalizing by Internalizing, the Externalizing by Powerlessness, and the Internalizing by Powerlessness interaction terms were significant. The B weight for Externalizing was .333, the B weight for Internalizing was .215, and the B weight for Powerlessness was .377. The Externalizing by Internalizing interaction term was .182, the Externalizing by Powerlessness term was .175, and the Internalizing by Powerlessness interaction term was .193. As shown in figures B8, B9, and B10, each pairing of sub-scores have a synergistic relationship with OB.

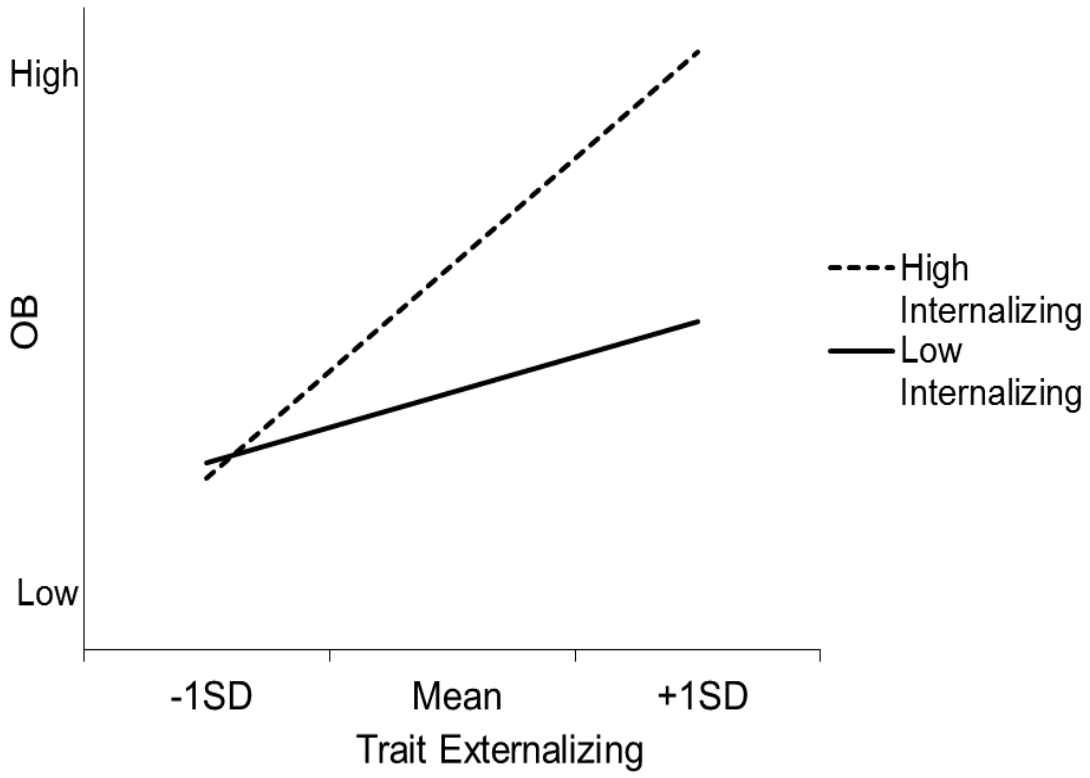

Figure 8. Interactive relationship between Externalizing/Internalizing and OB

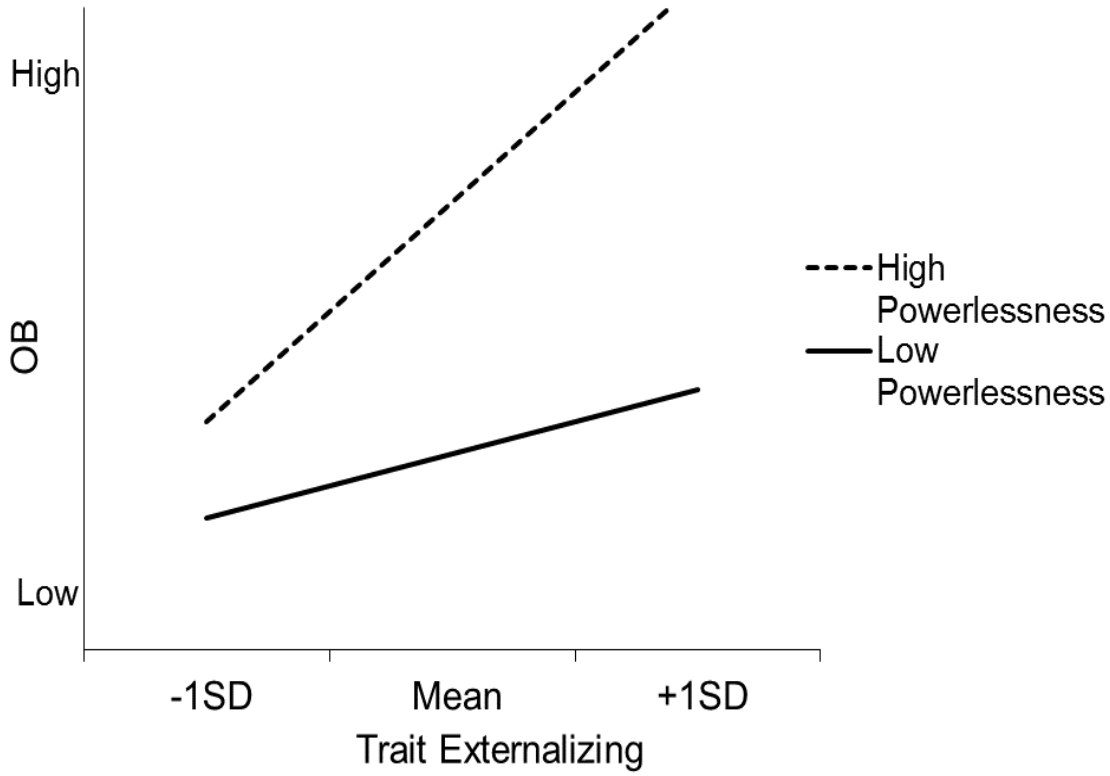

Figure 9. Interactive relationship between Externalizing/Powerlessness and OB

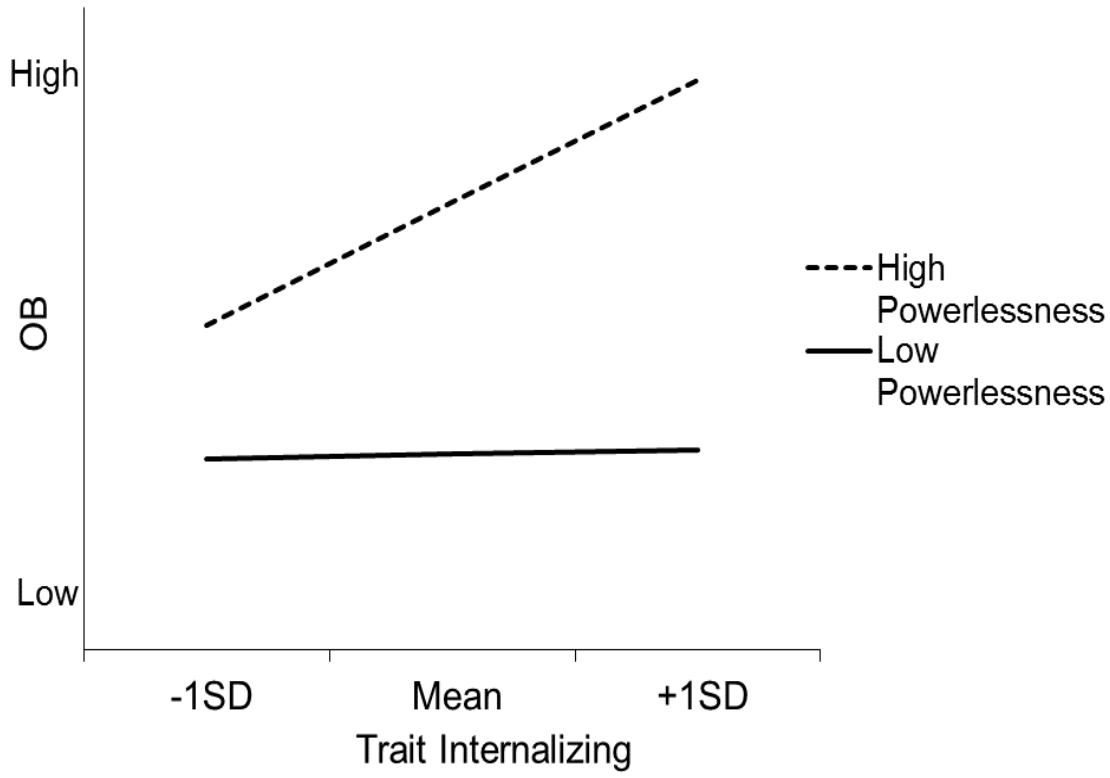

Figure 10. Interactive relationship between Internalizing/Powerlessness and OB

Because of the categorizations of OA, this criterion was not expected to have significantly different correlations with the three sub-scales of the CRT-A (H1), and results support this expectation. Additionally, the regression of this criterion on Model 2 was not expected to show better fit than the regression on model 1 (H2), and results support this expectation. And finally, the ΔR^2 due to the addition of interaction terms in model 3 was expected to be significant, and the F test for ΔR^2 from model 2 to model 3 supports this expectation ($F = 3.212$, $p < .05$). The Externalizing by Internalizing interaction term was significant. The B weight for Externalizing was .316, the B weight for Internalizing was .198, and the B weight for Powerlessness was .275. The

Externalizing by Internalizing interaction term was .228. As shown in Figure B11, this suggests that Externalizing and Internalizing have a synergistic relationship with OA.

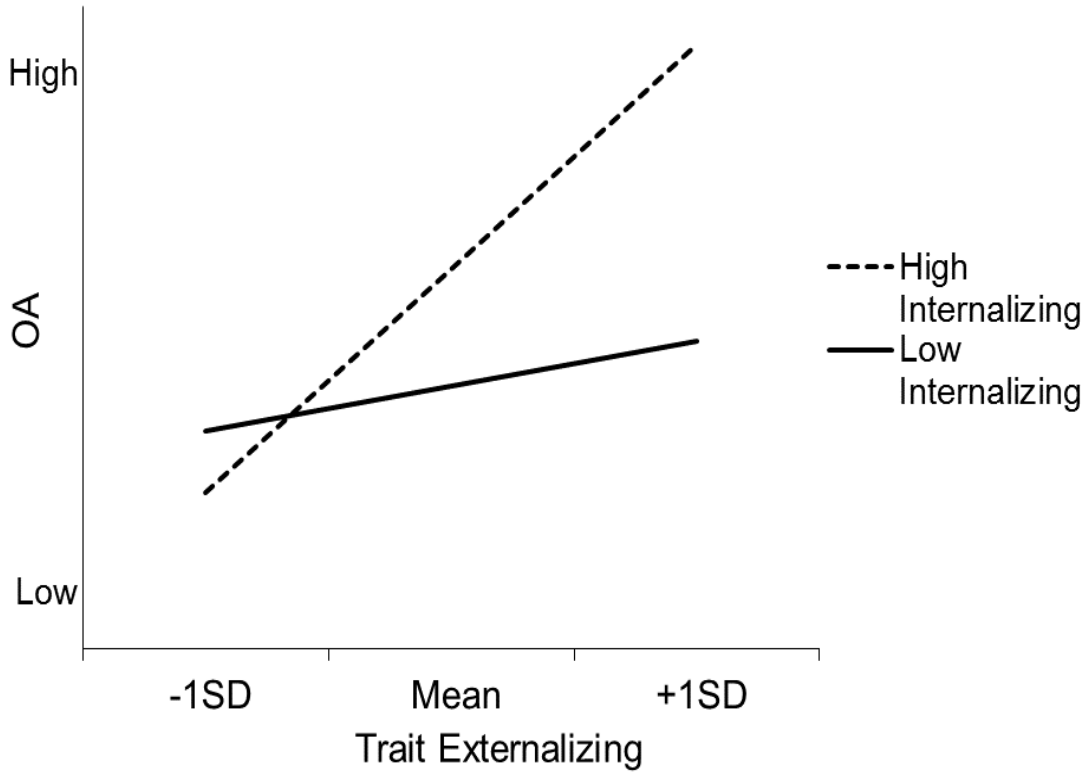

Figure 11. Interactive relationship between Externalizing/Internalizing and OA

Because of the categorizations of EH, this criterion was expected to have a significantly stronger correlation with the Powerlessness scale as compared to the Externalizing and Internalizing scales (H1), however the one tailed T_2 test for this comparison was not significant. Additionally, the regression of this criterion on Model 2 was expected to show better fit than the regression on model 1 (H2), however the ΔAICc does not supports this expectation. And finally, the ΔR^2 due to the addition of interaction

terms in model 3 was expected to be significant, however the F test for ΔR^2 from model 2 to model 3 does not support this expectation.

Table 11: Coding Decisions, Variable Intercorrelations, and Results by Hypothesis for Frost (2005)

Coding Decisions

Criterion (C)	Externalizing	Internalizing	Powerlessness	Behavioral Specificity	Context Specificity
OB	X	-	X	Yes	Yes
OA	X	X	X	Yes	Yes
EH	-	-	X	Yes	Yes

Variable Intercorrelations (N = 177)

C	CRxE	CRxI	CRxP	ExI	ExP	IxP	CRxC	ExC	IxC	PxC
OB	0.84**	0.65**	0.67**	0.26**	0.28**	0.06	0.60**	0.50**	0.30**	0.50**
OA	"	"	"	"	"	"	0.55**	0.45**	0.28**	0.38**
EH	"	"	"	"	"	"	-.04	0.01	-0.15*	-0.01

Hypothesis 1 – T_2 Tests for Differences Between Dependent Correlations

C	E vs. Avg. of I & P	I vs. Avg. of E & P	P vs. Avg. of E & I
OB	1.26	-2.35*	1.24
OA	1.46	-1.53	0.20
EH	1.00	-1.52	0.58

Hypothesis 2 – Comparison of Model fit Between Model 1[a] (M1) and Model 2[b] (M2)

C	M1 R^2	M2 R^2	ΔR^2	M1 AIC_c	M2 AIC_c	M1 – M2 AIC_c
OB	.359**	.424**	.065**	-73.47	-88.39	14.92
OA	.299**	.304**	.005	-57.73	-54.71	-3.02
EH	.002	.025	.023	4.80	4.88	-0.08

Hypothesis 3 – Significance of the F Test for ΔR^2 From M2 to Model 3[c] (M3)

Table 11 (Continued)

C	M2 R^2	M3 R^2	ΔR^2	B^d ExI	B ExP	B IxP
OB	.424**	.567**	.143**	.182**	.175*	.193*
OA	.304**	.376**	.072**	.228**	-.006	.088
EH	.025	.039	.014	-.069	-.014	.128

Note. CR = Total score on the CRT-A; E = score on the Externalizing scale from the CRT-A; I = score on the Internalizing scale from the CRT-A; P = score on the Powerlessness scale from the CRT-A. T_2 Tests are 1 tailed.
[a]Model 1 is the regression of the criterion on the CRT-A total score.
[b]Model 2 is the regression of the criterion on the Externalizing, Internalizing, and Powerlessness scales.
[c]Model 3 adds the Externalizing by Internalizing, Externalizing by Powerlessness, and Internalizing by Powerlessness interaction terms to the regression of the criterion on model 2.
[d]Beta weight for the interaction term from the indicated adjacent scales.
* $p < .05$; ** $p < .01$.

James and McIntyre – Study 7 (2000)

The context for this study was the daily work routine of package handlers in the warehouses of a major package delivery service. This type of work includes a considerable amount of heavy lifting in a repetitive fashion for long periods of time. The criterion was habitual absences during the first 90 days of work. This was a count of absences (CoA). Absences in this environment were construed as a form of counterproductive work behavior. From an aggression standpoint, absences are a way of avoiding any unfair, negative, or constricting aspects of the work environment. This is done at the expense of co-workers, who are left to cover for the absent person, and the employer, who must pay for sick leave and spend resources to get the absence covered or lose productivity. For these reasons, the CoA criterion was categorized as being more relevant to both the Externalizing and Powerlessness dimensions. Because CoA in this

case is a narrow type of behavior that takes place in the exact same work environment across the people sampled, this criterion was coded as specific in terms of both behavior and context. Table 12 shows the coding decisions, variable intercorrelations, and results by hypothesis for this study.

Because of the categorizations of CoA, this criterion was expected to have a significantly weaker correlation with the Internalizing scale as compared to the Externalizing and Powerlessness scales (H1), and while results trend in this direction, the one tailed T_2 test for this comparison was not significant. Additionally, the regression of this criterion on Model 2 was expected to show better fit than the regression on model 1 (H2), However the ΔAIC_c does not supports this expectation. And finally, the ΔR^2 due to the addition of interaction terms in model 3 was expected to be significant, However results do not support this expectation.

Table 12: Coding Decisions, Variable Intercorrelations, and Results by Hypothesis for James and McIntyre – Study 7 (2000)

Coding Decisions

Criterion (C)	Externalizing	Internalizing	Powerlessness	Behavioral Specificity	Context Specificity
CoA	X	-	X	Yes	Yes

Variable Intercorrelations (N = 105)

C	CRxE	CRxI	CRxP	ExI	ExP	IxP	CRxC	ExC	IxC	PxC
CoA	0.90**	0.66**	0.63**	0.40**	0.35**	-0.05	0.34**	0.28**	0.15*	0.30**

Hypothesis 1 – T_2 Tests for Differences Between Dependent Correlations

C	E vs. Avg. of I & P	I vs. Avg. of E & P	P vs. Avg. of E & I
CoA	0.52	-1.15	0.70

Hypothesis 2 – Comparison of Model fit Between Model 1[a] (M1) and Model 2[b] (M2)

C	M1 R^2	M2 R^2	ΔR^2	M1 AIC_c	M2 AIC_c	M1 – M2 AIC_c
CoA	.118**	.134**	.016	341.44	343.99	-2.55

Hypothesis 3 – Significance of the F Test for ΔR^2 From M2 to Model 3[c] (M3)

C	M2 R^2	M3 R^2	ΔR^2	B[d] ExI	B ExP	B IxP
CoA	.134**	.178*	.044	-.717	.149	1.049

Note. CR = Total score on the CRT-A; E = score on the Externalizing scale from the CRT-A; I = score on the Internalizing scale from the CRT-A; P = score on the Powerlessness scale from the CRT-A. T_2 Tests are 1 tailed.

[a]Model 1 is the regression of the criterion on the CRT-A total score.

[b]Model 2 is the regression of the criterion on the Externalizing, Internalizing, and Powerlessness scales.

[c]Model 3 adds the Externalizing by Internalizing, Externalizing by Powerlessness, and Internalizing by Powerlessness interaction terms to the regression of the criterion on model 2.

[d]Beta weight for the interaction term from the indicated adjacent scales.

* $p < .05$; ** $p < .01$.

James and McIntyre – Study 8 (2000).

In this study, all criteria were gathered from people who had been selected to be part of a temporary work pool for entry level jobs at local businesses. Essentially, the study was interested in the behaviors of temp workers. An overall unreliability criterion (UR) was created as a scale from 0 to 2 based on the following: 0 = Performed reliably, 1 = Failed to accept offer for temporary employment or accepted offer but failed to show up for assignment, 2 = Engaged in blatantly unreliable/aggressive behavior like threatening supervisor, false injury claims, and lying to staff of organization. Additionally, a blatant aggression (BA) criterion was created as a scale from 0 to 1 based on a subset of the previous categorizations where 0 = Performed reliably, and 1 = Engaged in blatantly unreliable/aggressive behavior like threatening supervisor, false injury claims, and lying to staff of organization.

The UR criterion consisted of behaviors that were highly diverse in terms of quality and intensity. It was therefore categorized as being equally relevant to all three dimensions of aggression. The BA criterion has diverse behaviors as well, however the examples given seem to have a more deliberate tone and are more goal oriented. Because of this, BA was categorized as being less relevant to the Powerlessness dimension. Because UR and BA were made up of diverse behaviors that were not part of defined behavior classes and that occurred across a variety of work environments, both criteria were coded as non-specific in terms of both behavior and context. Table 13 shows the coding decisions, variable intercorrelations, and results by hypothesis for this study.

Because of the categorizations of BA, this criterion was expected to have a significantly weaker correlation with the Powerlessness scale as compared to the average

of the Internalizing and Externalizing scales (H1), and the one tailed T_2 test for this comparison barely supports this expectation ($T_2 = -1.64$, $p < .05$). Additionally, the regression of this criterion on Model 2 was expected to show better fit than the regression on model 1 (H2), however a ΔAICc of -3.90 does not support this expectation. And finally, the ΔR^2 due to the addition of interaction terms in model 3 was not expected to be significant, and the F test for ΔR^2 from model 2 to model 3 supports this expectation.

Because of the categorizations of UR, this criterion was not expected to have significantly different correlations with the three sub-scales of the CRT-A (H1), and the one tailed T_2 tests for these comparisons support this expectation. Additionally, the regression of this criterion on Model 2 was not expected to show better fit than the regression on model 1 (H2), however a ΔAICc of 2.43 does not supports this expectation. And finally, the ΔR^2 due to the addition of interaction terms in model 3 was not expected to be significant, and the F test for ΔR^2 from model 2 to model 3 supports this expectation.

Table 13: Coding Decisions, Variable Intercorrelations, and Results by Hypothesis for James and McIntyre – Study 8 (2000)

Coding Decisions

Criterion (C)	Externalizing	Internalizing	Powerlessness	Behavioral Specificity	Context Specificity
BA	X	X	-	No	No
UR	X	X	X	No	No

Variable Intercorrelations (N = 111)

C	CRxE	CRxI	CRxP	ExI	ExP	IxP	CRxC	ExC	IxC	PxC
BA	0.84**	0.49**	0.59**	0.06	0.12	-0.13	0.53**	0.50**	0.28**	0.18
UR	"	"	"	"	"	"	0.42**	0.33**	0.30**	0.20*

Hypothesis 1 – T_2 Tests for Differences Between Dependent Correlations

C	E vs. Avg. of I & P	I vs. Avg. of E & P	P vs. Avg. of E & I
BA	2.39*	-0.50	-1.64*
UR	0.67	0.24	-0.87

Hypothesis 2 – Comparison of Model fit Between Model 1[a] (M1) and Model 2[b] (M2)

C	M1 R^2	M2 R^2	ΔR^2	M1 AIC_c	M2 AIC_c	M1 – M2 AIC_c
BA	.289**	.342**	.053*	69.21	73.108	-3.90
UR	.176**	.225**	.049	-16.32	-18.75	2.43

Hypothesis 3 – Significance of the F Test for ΔR^2 From M2 to Model 3[c] (M3)

C	M2 R^2	M3 R^2	ΔR^2	B[d] ExI	B ExP	B IxP
BA	.342**	.382**	.04	.058	.170*	.099
UR	.225**	.243**	.018	.103	.082	.033

Note. CR = Total score on the CRT-A; E = score on the Externalizing scale from the CRT-A; I = score on the Internalizing scale from the CRT-A; P = score on the Powerlessness scale from the CRT-A. T_2 Tests are 1 tailed.

[a]Model 1 is the regression of the criterion on the CRT-A total score.
[b]Model 2 is the regression of the criterion on the Externalizing, Internalizing, and Powerlessness scales.
[c]Model 3 adds the Externalizing by Internalizing, Externalizing by Powerlessness, and Internalizing by Powerlessness interaction terms to the regression of the criterion on model 2.
[d]Beta weight for the interaction term from the indicated adjacent scales.
* $p < .05$; ** $p < .01$.

McNiel (2009)

This was a lab study in which participants played Tetris for 5 minutes and then 20 minutes. During the 5 minute period play was normal and the goal was to practice getting 4 line eliminations. During the 20 minute period play was made more difficult by random control remapping. So, the button for move a piece left might become move a piece down…etc. The controls changed randomly and at random time intervals, however they did not change enough to make the goal of getting a four line elimination impossible. The idea was to make the task very frustrating but achievable. Participants were given a token to win 500 dollars for each four line elimination they managed to get. Participant were also asked not to pause the game unless it was an absolute emergency, as this would likely invalidate their data. At the end of the 20 minute time period, participants were given the opportunity to comment and were asked to report how many 4 line eliminations they managed to get so that they could be given the right number of tokens to win a drawing for the $500. The four resulting criteria are pausing (PAU), complaints (COM), dishonest reporting (DR), and excessive pressure during the 5 minute time period (EP). PAU was simply whether or not a participant paused the game after being told this might invalidate the results. Pausing makes it easier to strategize in Tetris and accomplish 4 line eliminations. COM was created by assessing participant comments. Participants were

asked to give constructive feedback. They were given a 0 if feedback was missing or constructive and a 1 if feedback was simply complaining or meant to be insulting. DR was coded as follows: underreporting or accurate reporting of the number of 4 line eliminations was scored as a 0 and over reporting of 4 line eliminations was scored as a 1. EP represents the amount of time spent pressing very hard on buttons. Pressure sensors were used to collect data on this and various algorithms extracted a single indicator of excessive pressure.

Pausing (PAU) within the context of this experiment aids in the accomplishment of gaining a four line elimination. It is an approach behavior that can aid a person who has a wish to defeat the task set before them. It was speculated in this study that those who are high in either Powerlessness or Externalizing would aggress by withdrawing from rather than engaging in the task. For these reasons, the PAU criterion was categorized as being more relevant to the Internalizing dimension. Complaints (COM) in this study gave participants an opportunity to describe the task as unfair, stupid, poorly thought out, etc. Most of the complaints were a direct reaction to study conditions or an attempt to restore ego. Additionally, the opportunity for complaints was removed from the study by about 5 minutes, making pure frustration less of an impetus for complaining. For these reasons, the COM criterion was categorized as being more relevant to the Externalizing and Internalizing dimensions. EP was designed to be an indicator of frustration. Because of this, it was categorized as being more relevant to the Powerlessness dimension. Finally, DR represents a clear goal oriented choice that can either be a reaction to the perceived unfairness of the task or an expression of the desire to win money. For these reasons, the DR criterion was categorized as being more relevant

to the Externalizing and Internalizing dimensions. Because these four criteria were all well-defined and singular in nature and because all participants experienced the exact same environment when generating these behaviors, they were all coded as specific in terms of both behavior and context. Table 14 shows the coding decisions, variable intercorrelations, and results by hypothesis for this study.

Because of the categorizations of PAU, this criterion was expected to have a significantly more positive correlation with the internalizing scale as compared to the Externalizing and Powerlessness scales (H1), and the one tailed T_2 test for this comparison was significant ($T_2 = 1.68$, $p < .05$). It should be noted however that the Internalizing relationship with this criterion was simply less negative than the Externalizing/Powerlessness relationship. Additionally, the regression of this criterion on Model 2 was expected to show better fit than the regression on model 1 (H2), however, ΔAICc does not support this expectation. And finally, the ΔR^2 due to the addition of interaction terms in model 3 was expected to be significant, however results do not support this expectation.

Because of the categorizations of COM, this criterion was expected to have a significantly weaker correlation with the Powerlessness scale as compared to the Externalizing and Internalizing scales (H1), and while the one tailed T_2 test for this comparison trended in the expected direction, results were not significant. Additionally, the regression this criterion on Model 2 was expected to show better fit than the regression on model 1 (H2), however ΔAICc does not support this expectation. And finally, the ΔR^2 due to the addition of interaction terms in model 3 was expected to be significant, however results do not support this expectation.

Because of the categorizations of EP, this criterion was expected to have a significantly stronger correlation with the Powerlessness scale as compared to the Externalizing and Internalizing scales (H1), and the one tailed T_2 test for this comparison was significant ($T_2 = 1.67$, $p < .05$). Additionally, the regression of this criterion on Model 2 was expected to show better fit than the regression on model 1 (H2), however ΔAICc does not supports this expectation. And finally, the ΔR^2 due to the addition of interaction terms in model 3 was expected to be significant, however the F test for ΔR^2 from model 2 to model 3 does not support this expectation.

Because of the categorizations of DR, this criterion was expected to have a significantly weaker correlation with the Powerlessness scale as compared to the Externalizing and Internalizing scales (H1), however results do not support this expectation. Additionally, the regression this criterion on Model 2 was expected to show better fit than the regression on model 1 (H2), and a ΔAICc of 2.44 supports this expectation. And finally, the ΔR^2 due to the addition of interaction terms in model 3 was expected to be significant, however results do not support this expectation.

Table 14: Coding Decisions, Variable Intercorrelations, and Results by Hypothesis for McNiel (2009)

Coding Decisions

Criterion (C)	Externalizing	Internalizing	Powerlessness	Behavioral Specificity	Context Specificity
PAU	-	X	-	Yes	Yes
COM	X	X	-	Yes	Yes
EP	-	-	X	Yes	Yes
DR	X	X	-	Yes	Yes

Variable Intercorrelations (N = 164)

C	CRxE	CRxI	CRxP	ExI	ExP	IxP	CRxC	ExC	IxC	PxC
PAU	0.71**	0.58**	0.70**	0.03	0.10	0.22**	-0.31**	-0.24**	-0.09	-0.28**
COM	"	"	"	"	"	"	0.16*	0.21**	-0.02	-0.01
EP	"	"	"	"	"	"	0.22**	0.08	0.02	0.22**
DR	"	"	"	"	"	"	-0.10	-0.22**	-0.10	0.18*

Hypothesis 1 – T_2 Tests for Differences Between Dependent Correlations

C	E vs. Avg. of I & P	I vs. Avg. of E & P	P vs. Avg. of E & I
PAU	-0.54	1.68*	-1.17
COM	2.12**	-1.12	-1.06
EP	-0.36	-1.25	1.67*
DR	-2.54**	-0.78	3.54**

Hypothesis 2 – Comparison of Model fit Between Model 1[a] (M1) and Model 2[b] (M2)

C	M1 R^2	M2 R^2	ΔR^2	M1 AIC_c	M2 AIC_c	M1 – M2 AIC_c
PAU	.098**	.121**	.024	140.74	143.9654	-3.23

Table 14 (Continued)

COM	.024*	.043	.019	107.88	110.5015	-2.62
EP	.047**	.051*	.004	-2.09	1.46	-3.55
DR	.010	.111**	.101**	162.52	160.0797	2.44

Hypothesis 3 – Significance of the F Test for ΔR^2 From M2 to Model 3[c] (M3)

C	M2 R^2	M3 R^2	ΔR^2	B[d] ExI	B ExP	B IxP
PAU	.121**	.160**	.039	.074	-.019	-.172**
COM	.043	.083*	.0398	-.160**	-.035	-.021
EP	.051*	.055	.0039	.025	.020	.043
DR	.111**	.112*	.001	-.003	-.022	.008

Note. CR = Total score on the CRT-A; E = score on the Externalizing scale from the CRT-A; I = score on the Internalizing scale from the CRT-A; P = score on the Powerlessness scale from the CRT-A. T_2 Tests are 1 tailed.
[a]Model 1 is the regression of the criterion on the CRT-A total score.
[b]Model 2 is the regression of the criterion on the Externalizing, Internalizing, and Powerlessness scales.
[c]Model 3 adds the Externalizing by Internalizing, Externalizing by Powerlessness, and Internalizing by Powerlessness interaction terms to the regression of the criterion on model 2.
[d]Beta weight for the interaction term from the indicated adjacent scales.
* $p < .05$; ** $p < .01$.

McNiel (New Samples)

In this study, Students were given the opportunity to gain extra credit in psychology classes by completing an online battery of tests. They got credit for 3.5 hours of work regardless of how much time they spent on the tests. The majority of items were prototype conditional reasoning questions that required reading about ½ a page double spaced and moderate concentration to complete correctly. All items were designed to be very easy and had two correct answer choices out of four available responses. A total of

115 of these reasoning questions were given and were interspersed with personality questions and demographics. The last 60 items were all reasoning questions. This whole task was designed to be easy but grinding. Four separate samples were collected. Each sample had a different set of reasoning questions making up the last 60 items. The Criterion for each sample was passive aggressive lack of effort (PALE). This was operationalized by counting the number of incorrect answers across the last 60 items.

PALE was expected to occur to a greater degree when a person was higher in Externalizing. Because extra credit was given based on completion rather than performance, the most economical way to get extra credit while minimizing subjective effort was to randomly respond or to pay minimal attention to the task. This is aggressive to the extent that participants understands that withdrawing effort in such a way is damaging to the study they have agreed to help with. As mentioned in an example placed in the hypothesis section of this paper, those who are high in Internalizing but not Externalizing are not as likely to respond to external cues in an aggressive manner. They should be less attentive to the fact that they are being put through a grinding task, and they should be less concerned about issues of fairness or equal reciprocity. Instead those who are high in Internalizing will likely be concerned with establishing potency and defeating obstacles. This lead to a choice between incompatible behaviors. A participant could either withdraw effort, or attempt to defeat the test, but they could not do both. Additionally, there was an incentive to withdraw effort, but there was no incentive to try and do very well on the tasks. Finally, the task was explicitly meant to be easy and non-frustrating. For these reasons, the PALE criterion was categorized as being more relevant to the Externalizing dimension. If it is related to the Internalizing dimension, the

relationship is expected to be negative. Because PALE is a well-defined behavior and all participants experienced the same task, it was coded as specific in terms of both behavior and context. Table 15 shows the coding decisions, variable intercorrelations, and results by hypothesis for this study.

Because of the categorizations of PALE, this criterion was expected to have a significantly stronger positive correlation with the Externalizing scale as compared to the Internalizing and Powerlessness scales (H1), and the one tailed T_2 test for this comparison was significant across the first three samples (S1 - $T_2 = 2.97$, $p < .05$; S2 - $T_2 = 2.15$, $p < .05$; S3 - $T_2 = 2.18$, $p < .05$). However, the on tailed T_2 test for this comparison was not significant for S4. Additionally, the regression of this criterion on Model 2 was expected to show better fit than the regression on model 1 (H2), and the ΔAICc of 7.74 for S1 and of 6.26 for S3 supports this expectation. However the ΔAICcs for S2 and S4 were not high enough to prefer model 2 over model 1 in these samples. And finally, the ΔR^2 due to the addition of interaction terms in model 3 was expected to be significant, however results do not support this expectation for any of the four samples.

Table 15: Coding Decisions, Variable Intercorrelations, and Results by Hypothesis for McNiel (New Samples)

Coding Decisions

Criterion (C)	Externalizing	Internalizing	Powerlessness	Behavioral Specificity	Context Specificity
S1 (N = 97)	X	-	-	Yes	Yes
S2 (N = 66)	X	-	-	Yes	Yes
S3 (N = 120)	X	-	-	Yes	Yes
S4 (N = 118)	X	-	-	Yes	Yes

Variable Intercorrelations

C	CRxE	CRxI	CRxP	ExI	ExP	IxP	CRxC	ExC	IxC	PxC
S1	0.74**	0.64**	0.64**	0.28**	0.06	0.12	0.13	0.33**	-0.05	-0.03
S2	0.73**	0.58**	0.63**	0.04	-0.03	0.23	0.16	0.29*	-0.08	-0.07
S3	0.71**	0.57**	0.73**	-0.04	0.14	0.26**	0.09	0.20*	-0.21*	0.07
S4	0.79**	0.62**	0.73**	0.17	0.22*	0.24*	0.18	0.17	0.08	0.17

Hypothesis 1 – T_2 Tests for Differences Between Dependent Correlations

C	E vs. Avg. of I & P	I vs. Avg. of E & P	P vs. Avg. of E & I
S1	2.97*	-1.58	-1.24
S2	2.15*	-1.19	-1.02
S3	2.18*	-2.91*	0.64
S4	0.41	-0.76	0.35

Hypothesis 2 – Comparison of Model fit Between Model 1[a] (M1) and Model 2[b] (M2)

C	M1 R^2	M2 R^2	ΔR^2	M1 AIC_c	M2 AIC_c	M1 – M2 AIC_c
S1	.017	.129*	.112*	333.50	325.76	7.74

Table 15 (Continued)

S2	.026	.095	.069	303.68	303.41	0.27
S3	.009	.092*	.084*	530.82	524.56	6.26
S4	.031	.048	.017	480.94	483.21	-2.27

Hypothesis 3 – Significance of the F Test for ΔR^2 From M2 to Model 3[c] (M3)

C	M2 R^2	M3 R^2	ΔR^2	B[d] ExI	B ExP	B IxP
S1	.129*	.140*	.011	-.242	.131	.302
S2	.095	.133	.0381	-.714	-.435	-1.381
S3	.092*	.133*	.0408	-1.129	.730	-.685
S4	.048	.066	.018	.373	.664	-.090

Note. CR = Total score on the CRT-A; E = score on the Externalizing scale from the CRT-A; I = score on the Internalizing scale from the CRT-A; P = score on the Powerlessness scale from the CRT-A. T_2 Tests are 1 tailed.
[a]Model 1 is the regression of the criterion on the CRT-A total score.
[b]Model 2 is the regression of the criterion on the Externalizing, Internalizing, and Powerlessness scales.
[c]Model 3 adds the Externalizing by Internalizing, Externalizing by Powerlessness, and Internalizing by Powerlessness interaction terms to the regression of the criterion on model 2.
[d]Beta weight for the interaction term from the indicated adjacent scales.
* $p < .05$; ** $p < .01$.

Sablynski (2001)

This was an experimental study involving two groups in a lab setting. The two groups were seated at the same time around tables that were next to each other. The groups were asked to compete for prizes and the possibility of leaving early while getting full extra credit for participation. It was obvious to one group that the other was given a comparatively easier and more quickly completed task. The group with the more difficult task was the group that criteria were collected from. The experimenter made a point to

say that both tasks should take the same amount of time. Additionally, rules such as no talking, no getting up to leave, and no obvious digressions from the task were put forth as being essential for everyone to follow for the study to work properly. The winning group got to leave early and take a prize (this was always the group with the easier task). At the end of the task for the losing group, the experimenter left the room for a bit as the losing group was in the process of leaving. This gave the loosing group an opportunity to steal a prize, even though they were told that the experimenter paid for these prizes with her own money and needed them for other trials. Prizes were tested as being desirable to students and ranged from 5 to 20 dollars in value. The first criteria was the level of disobedience to the rules as rated by confederates on a 5 point scale (DIS). The second criteria was theft of a prize committed by a member of the losing group (THE). This was a dichotomous variable.

The context of this study was quite strong and the behaviors used to generate the criteria could have occurred for any number of reasons. For example, a person might have acted out and disobeyed rules (DIS) because he or she was frustrated, felt unfairly treated, or just wanted to show others that they could. For these reasons, the DIS criterion was categorized as being relevant to all three dimensions. Similarly, participants might have stolen a prize (THE) as a defense against unfair treatment, as a way to get back at the experimenter, or because they were frustrated and thought the competition was stupid and poorly thought out. For these reasons, both DIS and THE were considered relevant to all three dimensions. Because DIS and THE were both well-defined behaviors that occurred as part of a well-defined activity, they were both coded

as specific in terms of both behavior and context. Table 16 shows the coding decisions, variable intercorrelations, and results by hypothesis for this study.

Because of the categorizations of DIS, this criterion was not expected to have significantly different correlations with the three sub-scales of the CRT-A (H1), and results support this expectation. Additionally, the regression of this criterion on Model 2 was not expected to show better fit than the regression on model 1 (H2), and results support this expectation. And finally, the ΔR^2 due to the addition of interaction terms in model 3 was expected to be significant, and the F test for ΔR^2 from model 2 to model 3 supports this expectation ($F = 2.31$, $p < .05$). The Internalizing by Powerlessness interaction term was significant. The B weight for Externalizing was .235, the B weight for Internalizing was .037, and the B weight for Powerlessness was .248. The Internalizing by Powerlessness interaction term was .329. As shown in Figure 12, this suggests Internalizing and Powerlessness have a synergistic relationship with DIS.

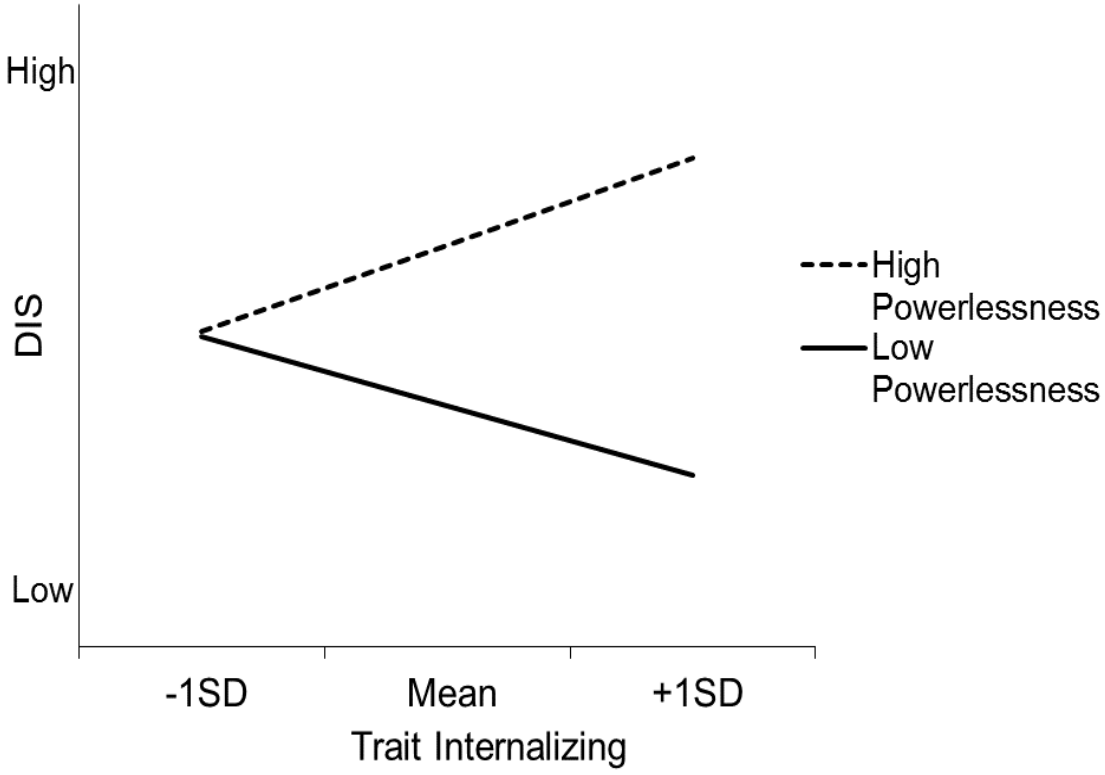

Figure 12. Interactive relationship between Internalizing/Powerlessness and DIS

Because of the categorizations of THE, this criterion was not expected to have significantly different correlations with the three sub-scales of the CRT-A (H1), and results support this expectation. Additionally, the regression of this criterion on Model 2 was not expected to show better fit than the regression on model 1 (H2), and results support this expectation. And finally, the ΔR^2 due to the addition of interaction terms in model 3 was expected to be significant, and the F test for ΔR^2 from model 2 to model 3 supports this expectation ($F = 21.321$, $p < .01$). The both the Externalizing by Internalizing and the Internalizing by Powerlessness interaction terms were significant. The B weight for Externalizing was .085, the B weight for Internalizing was .250, and the B weight for Powerlessness was .397. The Externalizing by Internalizing interaction term

was -.269, and the Internalizing by Powerlessness interaction term was .618. As shown in Figure 13, this suggests that Internalizing buffers the relationship between Externalizing and THE. As shown in Figure 14, this suggests that Internalizing and Powerlessness have a synergistic relationship with DIS.

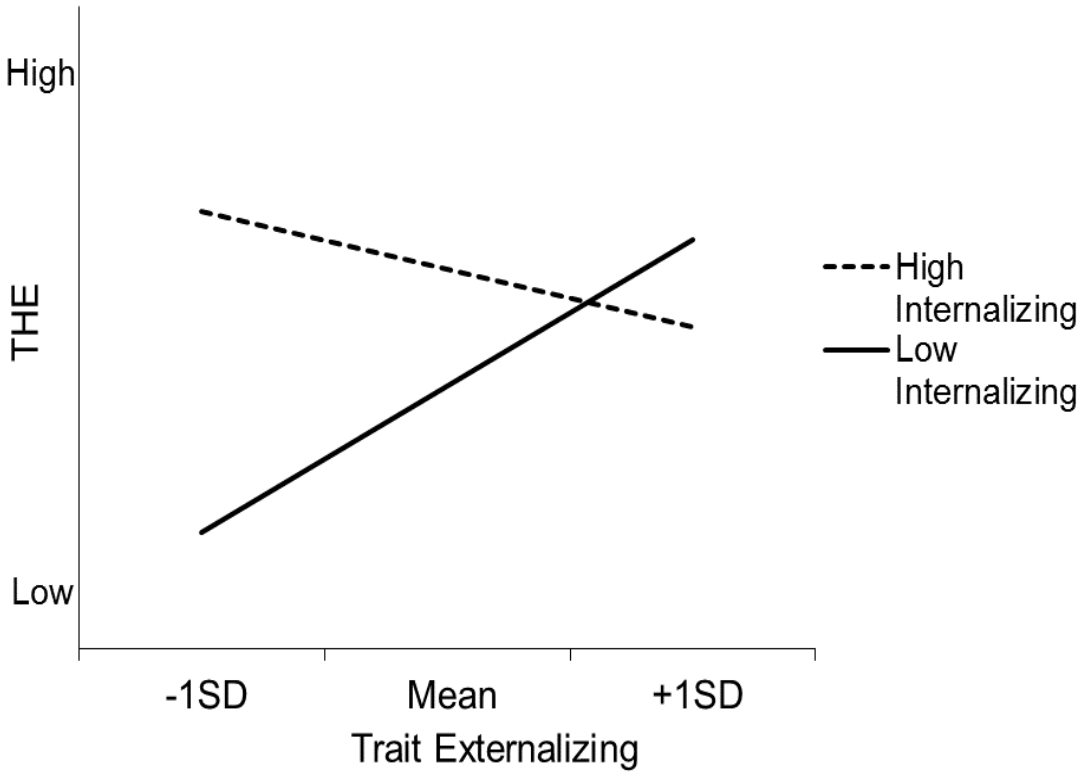

Figure 13. Interactive relationship between externalizing/internalizing and THE

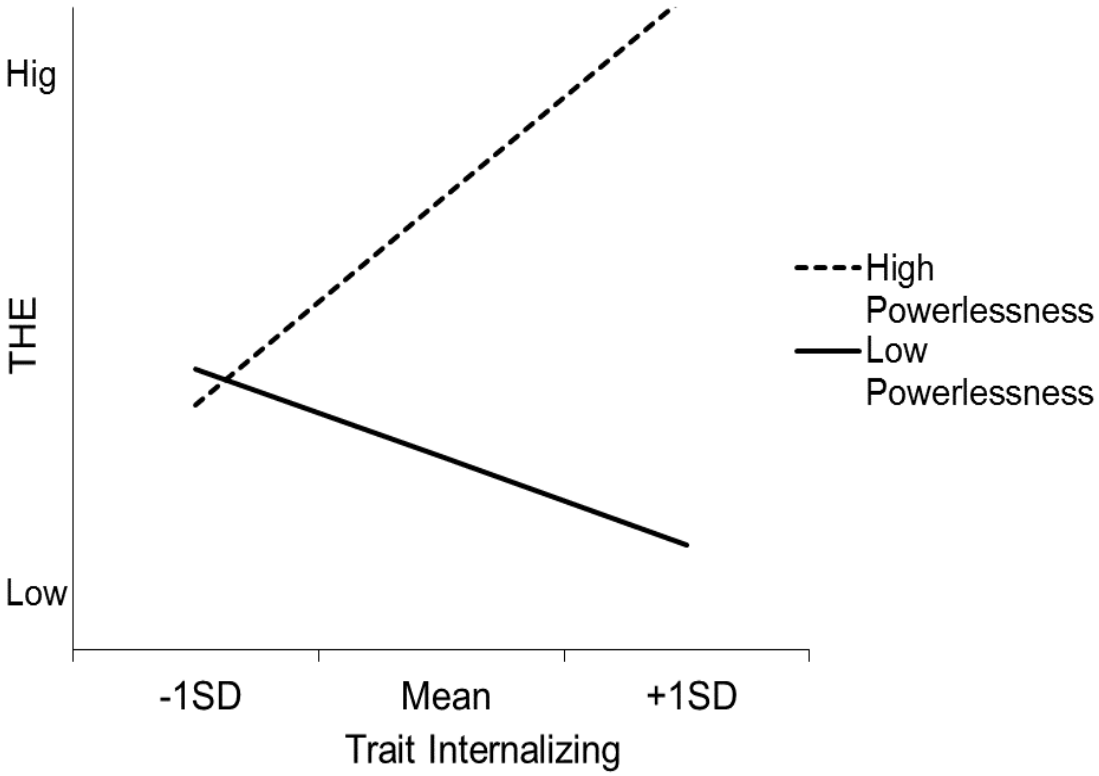

Figure 14. Interactive relationship between internalizing/powerlessness and THE

Table 16: Coding Decisions, Variable Intercorrelations, and Results by Hypothesis for Sablynski (2001)

Coding Decisions

Criterion (C)	Externalizing	Internalizing	Powerlessness	Behavioral Specificity	Context Specificity
THE	X	X	X	Yes	Yes
DIS	X	X	X	Yes	Yes

Variable Intercorrelations (N = 91)

C	CRxE	CRxI	CRxP	ExI	ExP	IxP	CRxC	ExC	IxC	PxC
THE	0.86**	0.55**	0.67**	0.20	0.25*	0.13	0.46**	0.21	0.30**	0.43**
DIS	"	"	"	"	"	"	0.36**	0.28**	0.11	0.31**

Hypothesis 1 – T_2 Tests for Differences Between Dependent Correlations

C	E vs. Avg. of I & P	I vs. Avg. of E & P	P vs. Avg. of E & I
THE	-1.31	-0.15	1.48
DIS	0.55	-1.41	0.91

Hypothesis 2 – Comparison of Model fit Between Model 1[a] (M1) and Model 2[b] (M2)

C	M1 R^2	M2 R^2	ΔR^2	M1 AIC_c	M2 AIC_c	M1 – M2 AIC_c
THE	.213**	.252**	.039	39.10	41.51	-2.41
DIS	.13**	.140**	.01	-7.37	-4.06	-3.31

Hypothesis 3 – Significance of the F Test for ΔR^2 From M2 to Model 3[c] (M3)

C	M2 R^2	M3 R^2	ΔR^2	B[d] ExI	B ExP	B IxP
THE	.252**	.71**	.458**	-.269**	.029**	.618**
DIS	.140**	.266**	.126*	-.088	.06**	.329**

Note. CR = Total score on the CRT-A; E = score on the Externalizing scale from the CRT-A; I = score on the Internalizing scale from the CRT-A; P = score on the Powerlessness scale from the CRT-A. T_2 Tests are 1 tailed.

[a]Model 1 is the regression of the criterion on the CRT-A total score.
[b]Model 2 is the regression of the criterion on the Externalizing, Internalizing, and Powerlessness scales.
[c]Model 3 adds the Externalizing by Internalizing, Externalizing by Powerlessness, and Internalizing by Powerlessness interaction terms to the regression of the criterion on model 2.
[d]Beta weight for the interaction term from the indicated adjacent scales.
* $p < .05$; ** $p < .01$.

REFERENCES

References marked with an asterisk indicate studies from which samples were drawn for analysis

Bargh J. A., & Chartrand, T. L. (1999). The unbearable automaticity of being. *American Psychologist, 54*(7), 462-479.

Bargh, J. A. & Morsella, E. (2008). The Unconscious Mind. *Perspectives on Psychological Science, 3*(1), 73-79.

*Baysinger, M. A., LeBreton, J. M. & Scherer, K. T. (2014). Exploring the Disruptive Effects of Psychopathy and Aggression on Group Processes and Group Effectiveness. *Journal of Applied Psychology, 99*(1), 48-65.

Berkowitz L. 1993. *Aggression: Its Causes, Consequences, and Control*. New York: McGraw-Hill.

Berkowitz, L. (1989). Frustration-aggression hypothesis: Examination and reformulation. *Psychological Bulletin, 106*, 59–73.

Berkowitz, L. (1990). On the formation and regulation of anger and aggression: A cognitive-neoassociationistic analysis. *American Psychologist, 45*(4), 494-503.

*Bing M. N., Stewart S. M., Davison H. K., Green P. D., McIntyre M. D., & James L. R. (2007). An integrative typology of personality assessment for aggression: Implications for predicting counterproductive workplace behavior. *Journal of Applied Psychology, 92*, 722–744.

Burnham, K. P., & Anderson, D. R. (2002). *Model Selection and Multimodal Inference A Practical Information-Theoretic Approach: Second Edition.* New York: Springer.

Cohen, J., Cohen, P., West, S. G., & Aiken, L. S. (2003). *Applied Multiple Regression/Correlation Analysis for the Behavioral Sciences: Third Edition.* New Jersey: Lawrence Erlbaum Associates.

Crick, N. R., & Dodge, K. A. (1994). A review and reformulation of social information processing mechanisms in children's social adjustment. *Psychological Bulletin, 115*, 74-101.

Crick, N. R., & Dodge, K. A. (1996). Social information-processing mechanisms in reactive and proactive aggression. *Child Development, 67*, 993–1002.

DerSimonian, R., Laird, N. (1986). Meta-analysis in clinical trials. *Controlled Clinical Trials 7*, 177-188.

*DeSimone, J. A. (2010). *Using the Conditional Reasoning Test for Aggression to Predict Corrective Action Requests in a Sample of Nuclear Power Plant Employees.* Unpublished master's thesis, Georgia Institute of Technology, Atlanta

Dodge, K. A., & Coie, J. D. (1987). Social-information-processing factors in reactive and proactive aggression in children's peer groups. *Journal of Personality and Social Psychology, 53*, 1146–1158.

Dodge, K. A., Price, J. M., Bachorowski, J., & Newman, J. P. (1990). Hostile attributional biases in severely aggressive adolescents. *Journal of Abnormal Psychology, 99*(4), 385-392.

Fleiss, J. L. (1971). Measuring nominal scale agreement among many raters. *Psychological Bulletin, 76*(5), 378–382.

*Frost B. C. (2002). *Implicit versus explicit aggression: Who's ready to rumble?* Unpublished master's thesis, University of Tennessee, Knoxville.

*Frost, B. C., Ko, C-H. E., James, L. R. (2007). Implicit and explicit personality: A test of a channeling hypothesis of aggressive behavior, *Journal of Applied Psychology, 92*(5), 1299-1319.

Greenwald, A. G., & Banaji, M. R. (1995). Implicit social cognition: Attitudes, self-esteem, and stereotypes. *Psychological Review, 102*, 4-27.

Haidt, J. (2001) The emotional dog and its rational tail: A social intuitionist approach to moral judgment. *Psychological Review, 108*(4), 814-834.

Hays, W. L. (1994). *Statistics: Fifth Edition.* Belmont CA: Wadsworth.

Hedges, L. V., & Olkin, I. (1985). *Statistical methods for meta-analysis*. Orlando, FL: Academic Press.

Hedges, L. V., & Vevea, J. L. (1998). Fixed- and random-effects models in meta-analysis. *Psychological Methods, 3*, 486–504.

James, L. R., & LeBreton, J. M. (2012). *Assessing the implicit personality through conditional reasoning*. Washington, DC US: American Psychological Association.

James, L. R. (1998). Measurement of personality via conditional reasoning. *Organizational Research Methods, 1*, 131-163.

James, L. R., & Mazerolle, M. D. (2002). *Personality in work organizations*. Thousand Oaks, CA: Sage.

*James, L. R., & McIntyre, M. D. (2000). *Conditional Reasoning Test of Aggression test manual*. San Antonio, TX: Psychological Corporation.

James, L. R., & McIntyre, M. D., Glisson, C. A., Bowler, J. L., & Mitchell, T. R. (2004). The conditional reasoning measurement system for aggression: An overview. *Human Performance, 17,* 271-295.

James, L. R., & McIntyre, M. D., Glisson, C. A., Green, P. D., Patton, T. W., LeBreton, J. M., et al. (2005). Conditional reasoning: An efficient, indirect method for assessing implicit cognitive readiness to aggress. *Organizational Research Methods, 8,* 69-99.

Jöreskog, K. G., & Sörbom, D. (1999). *PRELIS 2: User's reference guide*. Chicago: IL: Scientific Software International.

Jöreskog, K. G., & Sörbom, D. (2008). *LISREL 8.8 for Windows*. Chicago: IL: Scientific Software International.

Ko, C. E., Thompson, V. M., Shim, H. S., Roberts, J. S., & McIntyre, H. (2008). *Alternative scoring strategies for the Conditional Reasoning Test of Aggression.* Presentation at the 23rd Annual Conference of the Society for Industrial and Organizational Psychology, San Francisco, CA.

Landis, J. R. & Koch, G. G. (1977). The measurement of observer agreement for categorical data. *Biometrics, 33,* 159–174.

McClelland, D. C., Koestner, R., Weinberger, J. (1989). How Do Self-Attributed and Implicit Motives Differ? *Psychological Review, 96,* 690-702.

*McNiel, P. D. (2009). *Implicit and explicit measures: A test of a dissociative model of aggression.* Unpublished master's thesis, Georgia Institute of Technology, Atlanta.

Minton, K. M. (2006). *Toward a Typology of the Aggressive Personality*. Un-published master's thesis, Georgia Institute of Technology, Atlanta.

Nisbett, R. E., & Wilson, T. D. (1977). Telling more than we can know: Verbal reports on mental processes. *Psychological Review, 84,* 231-259.

*Sablynski C. J., & Mitchell T. R. (2006, May). Post hoc analyses of interactions between the aggression questionnaire and conditional reasoning test of aggression. In Lebreton J. M. (Chair), *Bringing the implicit personality into I-O psychology.* Symposium presented at the 21stAnnual Conference of the Society for Industrial and Organizational Psychology, Dallas, TX.

Steiger, J. H. (1980). Tests for Comparing Elements of a Correlation Matrix. *Psychological Bulletin, 87*(2), 245-251.

Winter, D. G., John, O. P., Stewart, A. J., Klohnen, E. C., & Duncan, L. E. (1998). Traits and motives: Toward an integration of two traditions in personality research. *Psychological Review, 105,* 230–250.