AN ANALYSIS OF THE ITEM CHARACTERISTICS OF THE CONDITIONAL REASONING TEST OF AGGRESSION

A Dissertation
Presented to
The Academic Faculty

by

Justin A. DeSimone

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

Georgia Institute of Technology
December, 2012

COPYRIGHT 2012 BY JUSTIN A DESIMONE

AN ANALYSIS OF THE ITEM CHARACTERISTICS OF THE CONDITIONAL REASONING TEST OF AGGRESSION

Approved by:

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

Dr. Jack Feldman
School of Psychology
Georgia Institute of Technology

Dr. Nathan Bennett
College of Business
Georgia State University

Dr. Rustin Meyer
School of Psychology
Georgia Institute of Technology

Dr. Susan Embretson
School of Psychology
Georgia Institute of Technology

Date Approved: October 30, 2012

ACKNOWLEDGEMENTS

I wish to thank everyone who has supported me in my academic endeavors. Specifically, I would like to acknowledge the mentorship of my advisor, Dr. Lawrence R. James, whose support and assistance enabled every academic achievement I have earned to date. I also acknowledge the unique contributions of each of my other committee members during the completion of my comprehensive exams and dissertation. I would like to recognize the assistance of Robert Cookson and Dr. Jeremy Schoen. These fellow students were invaluable as colleagues, co-authors, and friends throughout my graduate education. Finally, I would like to thank my wife and my parents for their support throughout the years.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

Page

LIST OF SYMBOLS AND ABBREVIATIONS

θ	θ indicates the level of a latent trait
CRT-A	CRT-A stands for Conditional Reasoning Test of Aggression
CTT	CTT stands for Classical Test Theory
DIF	DIF stands for Differential Item Functioning
ICC	ICC stands for Item Characteristic Curve
IRT	IRT stands for Item Response Theory

SUMMARY

This dissertation employs modern psychometric techniques to estimate the item characteristics of the Conditional Reasoning Test of Aggression (CRT-A) and three related subscales. Using a sample size of 5,511 respondents, this dissertation provides an accurate assessment of the capability of the CRT-A to measure latent aggression. Additionally, a differential item functioning analysis compares student and applied samples. Results suggest that 21 of the 22 CRT-A items work as intended for the measurement of aggression and all 22 of the CRT-A items are good indicators of their respective factor subscales. Information curves indicate that the CRT-A is best suited for use with individuals who are high on latent aggression.

CHAPTER 1

INTRODUCTION TO THE CONDITIONAL REASONING TEST OF AGGRESSION

Since its introduction in the 1960s, the use of Item Response Theory (IRT) for psychometric analysis has become standard in cognitive measurement and educational assessment. Increased availability of computing power in recent decades has improved the ability of researchers to examine mathematically complex psychometric models. By conducting such analyses, test developers can access additional (often superior) methods of examining of test items. IRT analyses are capable of estimating person-invariant item characteristics, reducing the reliance on parallel tests and representative samples. IRT provides psychometricians the opportunity to shorten tests, use multiple item formats, and develop computer adaptive testing (Embretson & Reise, 2000). IRT also allows researchers to estimate item-invariant person characteristics instead of defining ability as relative position in a norm group. By defining the probability of a response as a function of an individual's latent trait level, IRT also allows researchers to develop test items targeted to specific portions of the population.

IRT is gaining popularity in the assessment of personality measures. Polytomous IRT models are capable of providing psychometric assessment of item formats often used in the assessment of personality such as interval scales (e.g. Samejima, 1969) and Likert-type measures (see Masters, 1992). While, IRT is often used to design and develop commercial tests in education and psychology, there are many tests which have not yet benefited from IRT analysis.

Many implicit personality measurement techniques such as Conditional Reasoning (James & LeBreton, 2012) are still developed and analyzed using CTT. Analyses such as inter-

item correlations and item-total correlations have been used to examine relationships between a test item and other items or the test as a whole. IRT (contingent on the model chosen for item analysis) provides the capability of providing more information about each item; such as the difficulty of items or how well an item discriminates among respondents. IRT analyses can also determine the latent trait level at which a test provides the most information (alternatively stated, the level of a latent trait at which the standard error of measurement is minimized; see Embretson & Reise, 2000).

Although CTT concepts such as test reliability are still useful, IRT analyses allow researchers to examine items in greater depth. CTT and IRT analyses should be combined to yield a more rigorous examination of item properties. This dissertation uses IRT to supplement the psychometric analysis of the Conditional Reasoning Test of Aggression (CRT-A); a test developed using CTT.

The Conditional Reasoning Test of Aggression

The CRT-A (James & McIntyre, 2000; James & LeBreton, 2012) is an implicit personality assessment intended to measure the propensity to behave aggressively (see James et al., 2005; James, McIntyre, Glisson, Bowler, & Mitchell, 2004). Theoretically, the CRT-A assumes that an implicit preparedness to aggress is rationalized through the use of justification mechanisms (JMs). Despite social pressure to avoid acting aggressively, individuals with an implicit tendency to aggress tend to view their aggressive actions as rational (Bergman, McIntyre, & James, 2004). JMs are defined as reasoning processes that an individual uses to validate behaviors consistent with dispositional tendencies (see James, 1998). While aggression may take many forms, James et al. (2005) describe the following six JMs used to rationalize such behaviors.

The first JM is known as hostile attribution bias, which involves the tendency to view the actions of others as malevolent. The second JM is called potency bias, which involves the tendency to view social interactions as contests to establish dominance. The third JM is named retribution bias, which reflects the preference for retaliating against others instead of attempting to reconcile. The fourth JM is known as victimization by powerful others bias, indicating the proclivity to view oneself as oppressed or exploited by those in positions of authority. The fifth JM is termed derogation of target bias, which reflects a tendency to view others as deserving of aggression due to their perceived immorality or other undesirable characteristics. The sixth JM is known as social discounting bias, which is a tendency to eschew social norms and ethical ideals.

The developers of the CRT-A recognize that individual differences among aggressive individuals may yield differential expression of aggressive behaviors influenced by different sets of JMs. The CRT-A is designed to measure the propensity to repeatedly engage in reasoning processes consistent with the above-mentioned JMs.

Format of the Conditional Reasoning Test of Aggression

The CRT-A contains 22 inductive reasoning items intended to assess an individual's proclivity to reason using an aggressive JM. Although the purpose of the test is to assess implicit aggression, each item is intended to appear to the respondent as a logical reasoning item. This lack of face validity serves to circumvent measurement issues such as faking, inaccuracy of self-perception, and social desirability. Three unscored logical reasoning items are included to encourage the perception of the test as an assessment of reasoning. In total, the CRT-A contains 25 items, 22 of which give the respondent the opportunity to choose between a logical response that serves to rationalize aggressive behavior and a logical response void of such rationalization.

Each item contains four response options. The three unscored items have one logical response and three illogical responses. The 22 keyed items have two logical responses and two illogical responses. The two illogical responses are written to appear blatantly incorrect to discourage their selection by the respondent (James & McIntyre, 2000). A recent item analysis concluded that the proportion of respondents who select an illogical response ranges from 0.25% to 4.70% (DeSimone & James, 2011), with most proportions less than 2%.

The two logical response options are intended to serve different purposes. One of the logical response options requires reasoning that relies on the respondent's use of aggressive JMs while the other is bereft of this requirement. The theoretical framework behind the CRT-A assumes that individuals who are predisposed to rationalize aggressive cognitions and behaviors will be more likely to select the aggressive response option than the nonaggressive response option. Although the response to a single item may reflect any number of potential causes, individuals who select aggressive responses to many items are more likely to rationalize aggressive behaviors than those who seldom select such responses (see James & Mazerolle, 2002; James et al., 2004).

The scoring system for the CRT-A reflects its theoretical framework. A respondent is awarded one point for each aggressive response option chosen. A nonaggressive or illogical response yields zero points. As a result, individuals who choose many aggressive options receive higher scores on the CRT-A than individuals who choose few. Those with higher scores are predicted to be more likely to engage in behavioral manifestations of aggression than those with lower scores (James & McIntyre, 2000).

The indirect nature of the CRT-A is necessary due to the socially undesirable nature of the target construct. Morgeson, Campion, Dipboye, Hollenbeck, Murphy, and Schmitt (2007)

reviewed issues involved in the use of personality tests for selection. One of their conclusions addressed the well-documented complications involved with the use of self-report personality tests. These complications include the idea that individuals are not always aware of the reasons for their actions (Haidt, 2001; Nisbett & Wilson, 1977), self-perception is subject to a variety of ego-enhancing biases (Dunning, Heath, & Suls, 2004), and responses to transparent measures are easily faked (Ellingson, Sackett, & Hough, 1999; Leary & Kowalski, 1990; Rosse, Stecher, Miller, & Levin, 1998; Snell, Sydell, & Lueke, 1999; Zickar & Robie, 1999). A respondent who perceives an assessment of reasoning should not be influenced by self-perceptions pertaining to aggressive tendencies. Therefore, ego-related biases and social desirability concerns should not impact responses with respect to the target construct of aggression (Bergman et al., 2004).

Morgesson et al. (2007) mention conditional reasoning as a potential alternative to self-report tests of personality, but other indirect measurement techniques can also circumvent or attenuate the detrimental measurement effects associated with self-report tests. Unfortunately, these alternatives suffer from a variety of complications as well. Projective tests such as the Rorschach (Rorschach, 1921) and TAT (Morgan & Murray, 1935) have been criticized for lack of evidence concerning validity (Kinslinger, 1966) and reliability (Entwisle, 1972). Forced-choice tests tend to be ipsative in nature, which can attenuate reliability estimates due to violations of CTT assumptions (see Meade, 2004). Conditional reasoning tests, on the other hand, have demonstrated strong psychometric properties and are not subject to the statistical constraints placed on ipsative measures.

Psychometric Properties of the Conditional Reasoning Test of Aggression

Initial estimates of psychometric properties of the CRT-A were reported in the Test Manual (James & McIntyre, 2000). These estimates relied heavily on developmental versions of

the CRT-A and analyses performed on small samples. A recent project updated these psychometric estimates using a sample of 5,238 individuals who took the current form of the CRT-A (DeSimone & James, 2011; James & LeBreton, 2012). The latter analysis informs the estimates reported in this section.

CRT-A scores range from zero to fourteen with a mean of 3.89 and a standard deviation of 2.19. Using the cutoff scores reported in James & McIntyre (2000), 28.07% of respondents demonstrate a low level of aggression (indicated by a score of zero to two), 65.71% of respondents demonstrate a moderate level of aggression (indicated by a score of three to seven), and 6.22% of respondents demonstrate a high level of aggression (indicated by a score of eight or above). Nonaggressive logical responses are chosen more often than aggressive logical responses in each of the items, with response ratios for some items exceeding 20:1. Item-total correlations for the 22 keyed CRT-A items range from .23 to .50. Internal consistency reliability is calculated as .76[1].

Early reports of the mean criterion-related validity of the CRT-A were reported to be .43 (James & McIntyre, 2000) or .44 (James et al., 2004; 2005). Two more recent (and more comprehensive) meta-analyses were independently conducted on the CRT-A. One examined the criterion-related validity of the CRT-A for predicting what the authors classified as counterproductive work behavior (CWB), reporting an average criterion-related validity of .16 (Berry, Sackett, & Tobares, 2010). Another examined the relationship of the CRT-A to aggressive criteria and criteria related to CWB, finding criterion-related validities of .29 and .27, respectively (DeSimone, Lee, & James, 2012; James & LeBreton, 2012). Although the CRT-A was not developed to predict job performance, both meta-analyses examined this relationship as

[1] Estimated using a modification of the Kuder-Richardson formula 20 (see Gulliksen, 1950).

well, with the former reporting an average criterion-related validity of .14 and the latter providing an average criterion-related validity of -.18.

Although there is variance in criterion-related validity estimates for predicting aggressive criteria, it is not within the purview of this paper to discuss differences between meta-analytic findings (though these differences are discussed in the manuscripts cited above). For the purpose of this dissertation, it is sufficient to highlight the fact that the CRT-A serves as a valid predictor of aggressive criteria. While the magnitude of the relationship is subject to debate, there is a positive correlation between CRT-A scores and aggression as measured by a variety of criteria (see DeSimone, Lee, & James, 2012).

The factor structure of the CRT-A has been examined independently by two projects. The first (reported in James et al., 2005) obliquely rotated the results of a principal components analysis to examine which items fit with which JMs. A five-component solution was produced reflecting five of the six JMs specified in the development of the CRT-A (see James, 1998; James & McIntyre, 2000). The derogation of target bias is the only JM not represented by the results of this analysis. James et al. (2005) reported alpha coefficients for these five components ranging from .74 to .87. Component intercorrelations were generally statistically significant and ranged from .06 to .25.

More recently, a factor analysis was conducted using an updated CRT-A response database which provided access to a larger sample (4,772 as opposed to 1,603 in the original analysis; Ko, Thompson, Shim, Roberts, & McIntyre, 2008; James & LeBreton, 2012). This analysis compared the results of an iterated principal components analysis to a bootstrapped parallel analysis and concluded that a three-factor solution should be extracted. James and LeBreton (2012) believe that these three factors encompass the JMs for aggression. Specifically,

the first factor, termed externalizing controls, reflects external justifications of aggression by comprising victimization by powerful others and hostile attribution. The second, called internalizing controls, reflects internal justifications of aggression by comprising retribution and potency. The third, known as powerlessness, is primarily related to social discounting. Internal consistency estimates range from .81 to .87[2] and factor intercorrelations are statistically significant and range from .17 to .50.

While the three-factor solution is preferable to the five-factor solution for the purposes of current psychometric analyses, the five-factor solution may be useful if future developers wish to extract and expand sections of the CRT-A in order to target specific JMs. The factor intercorrelations indicate that distinct processes may be occurring in the minds of aggressive individuals, however they are not low enough to discount the idea that aggression can be considered a unidimensional construct.

It would be unwise to dissect the CRT-A into its component factor subscales for the purpose of assessment due to the small number of items in some of the resulting scales. Doing so using the first factor analysis would result in five subscales with lengths between two and eight items. Doing so using the second factor analysis would result in three subscales ranging from five to eleven items. Only one study to date has examined the differential prediction of CRT-A factor subscales. DeSimone and Cookson (2011) found that superfluous corrective action requests could be predicted using the powerlessness subscale while the subscales measuring internal and external justifications failed to yield statistically significant prediction. While future research examining the differential utility of the subscales may be informative, it would be wise to expand these subscales to full-length stand-alone tests before using them individually.

[2] Using a larger sample (N = 5,238), DeSimone and James (2011) found slightly lower reliability estimates ranging from .68 to .75).

Intent of the Current Examination of the Conditional Reasoning Test of Aggression

This dissertation aims to expand previous psychometric analyses of the CRT-A by examining the test using IRT. As previously mentioned, the CRT-A was developed and analyzed using a CTT framework, which revealed that the test is a reliable and valid predictor of aggression. However, certain properties of CRT-A items remain unknown. Using IRT provides an examination of item-level properties such as discrimination and difficulty. The analysis will allow items to be compared directly with one another. Finally, this analysis can provide researchers and developers the opportunity to understand which portion of a population is best assessed using each item. Before this endeavor can be completed, it is important to ensure that the correct IRT model is used for the analysis.

CHAPTER 2

SELECTION OF AN ITEM RESPONSE THEORY MODEL

A number of factors were considered when choosing the most appropriate IRT model for the psychometric analysis. The recommended scoring system for the CRT-A awards one point for an aggressive response and zero points for responses that are nonaggressive, illogical, or missing (see James & McIntyre, 2000; James & LeBreton, 2012). This scoring system collapses the two illogical and one nonaggressive response options in order to conceptualize each CRT-A item as a dichotomous choice between selecting an aggressive option or selecting any other response option. A second scoring system (mentioned in James et al., 2005) awards one point for each aggressive response option and deducts one point for each nonaggressive response option (illogical responses are not scored). This scoring system reflects a conceptualization of the CRT-A in which aggressive responses are offset by nonaggressive responses.

The distinction between the two scoring systems lies in the decision of whether to treat illogical responses as conceptually similar or distinct from nonaggressive logical responses. The two scoring systems rank order respondents similarly due to the dearth of respondents who select illogical options to each item. From a CTT perspective, the two scoring systems produce similar criterion-related validity coefficients. However, the decision of whether or not to collapse the illogical and nonaggressive response options has implications for the selection of an IRT model.

Dichotomous IRT models (such as Rasch, 1PL, 2PL, and 3PL) do not differentiate between distractors, as they are intended for tests in which one correct answer is compared to a set of equally undesirable distractors. Polytomous IRT models are able to differentiate between distractors. CRT-A items feature two conceptually different types of responses that can be

classified as distractors (nonaggressive logical and illogical). However, the recommended scoring model for the CRT-A treats all nonaggressive response options identically. As a result, this dissertation examines the CRT-A using both dichotomous and polytomous models. The dichotomous models will serve to determine whether (and how well) each CRT-A item is able to differentiate aggressive individuals from nonaggressive individuals. The polytomous models will serve to provide analyses of each response option in order to determine whether they work as intended.

The primary differences between the major dichotomous IRT models involve the number of estimated parameters (Lord & Novick, 1968). Models containing a single parameter (item difficulty) assume that each test item is equally able to discriminate among respondents and that guessing does not impact item responses. Models containing two parameters also assume that guessing has a negligible effect on responses, but differ from single parameter models in that they account for variation in item discrimination. By estimating variation in item discrimination two-parameter models are able to estimate which items best differentiate between individuals with various levels of a latent trait. Models containing three parameters account for item difficulty, item discrimination, and the effects of guessing behavior.

Polytomous IRT models are somewhat more complex due to the fact that they can account for items that have more than two response categories. Thissen and Steinberg (1986) describe a taxonomy in which they propose three categories for polytomous IRT models. Their "difference" model category includes the Graded Response Model (GRM; Samejima, 1969), which requires the response options to be ordered (graded) in terms of the desirability of response. The GRM calculates the probability that an individual will respond in response category k or above, conditional on trait level.

The "divide-by-total" model category includes models such as the Nominal Response Model (NRM; Bock, 1972), Rating Scale Model (RSM; Andrich, 1978), and Partial Credit Model (PCM; Masters, 1982). These models calculate the probability that an individual will respond in each response category, conditional on trait level. Thissen and Steinberg (1986) describe the RSM and PCM as constrained versions of the NRM. Specifically, the RSM is a version of the NRM that assumes equal discrimination parameters across items. The PCM is a version of the NRM that requires slope parameters to increase in steps of unity. While the RSM and PCM can be used on ordinal responses, the NRM is intended to be used with tests containing nominal response options.

Thissen and Steinberg's third polytomous IRT model category, "left side added divide-by-total," includes models that are similar to those in their "divide-by-total" category, but also account for guessing. These models include Samejima's (1979) Guessing Model (SGM) and Thissen and Steinberg's (1984) Multiple Choice Model (MCM). Like the RSM and PCM, both the SGM and MCM can be considered modified versions of the NRM. The SGM adds a response category for individuals who guess due to their inability to determine the correct response. In the SGM, guessed responses are equally distributed among all response options. The MCM allows for variability in the distribution of guessed responses.

Guessing in Conditional Reasoning Test of Aggression Responses

Evidence suggests that there is little guessing in CRT-A responses, and it is unlikely that the benefit of modeling guessing behavior outweighs the costs associated with increasing the number of estimated parameters. Previous item analyses indicate that the vast majority of respondents choose the aggressive or nonaggressive logical responses. Illogical responses account for only 0.25% to 4.70% of the responses chosen (DeSimone & James, 2011). This

pattern of responses is likely indicative of the obvious incongruity of the illogical responses with the stem. If guessing does occur, it is more likely to take place between the two logical response options than between all four response options.

Another concern with modeling guessing on the CRT-A is the nature of guessing. While many IRT models were developed to assess items on transparent multiple choice tests, the CRT-A masks its true purpose. If participants are making "educated guesses" (as opposed to guessing at random), they are doing so in an attempt to portray themselves as knowledgeable in a different construct (logic or inferential reasoning) than the one assessed by the test (aggression). As a result, it is likely that guessing operates differently on the CRT-A than on the multiple choice tests for which the 3PL model, SGM, and MCM were developed.

The aforementioned issues yield two conclusions. First, guessing behavior on the CRT-A is unlikely to be completely at random or consistent across items. As a result, the SGM would likely be an inappropriate model for CRT-A responses. Second, the differences between the CRT-A and traditional multiple choice tests render the benefits of modeling guessing behavior tenuous. The potential for guessing behavior to be governed by a dimension not assessed by the test indicates that neither the SGM nor the 3PL model may adequately model guessing behavior on the CRT-A. It is possible that modeling guessing may do more harm than good. Consequently, IRT models which model guessing behavior are unsuitable for item analysis of the CRT-A.

Item Difficulty and Discrimination on the Conditional Reasoning Test of Aggression

The proportions of individuals who endorse the logical aggressive option range from .04 to .37. The proportions of individuals who endorse the logical nonaggressive option range from

.60 to .95 (DeSimone & James, 2011). These analyses indicate that CRT-A items are likely to vary in difficulty.

Previous item analyses have yielded item-total correlation estimates ranging from .23 to .50 for the CRT-A, .42 to .57 on the externalizing controls subscale, .60 to .74 on the internalizing controls subscale, and .59 to .71 on the powerlessness subscale (DeSimone & James, 2011). Factor loadings range from .24 to .59 on the externalizing controls subscale, .21 to .63 on the internalizing controls subscale, and .23 to .48 on the powerlessness subscale (Ko et al., 2008). These analyses suggest that CRT-A items are likely to differ in their ability to discriminate between individuals with various levels of latent aggression as well as individuals with different levels of the latent traits assessed by the three CRT-A factors.

CHAPTER 3

MOST APPROPRIATE MODELS FOR THE CURRENT ANALYSES

Previous analyses suggest that CRT-A items vary in item difficulty and item discrimination while guessing behavior is likely to have a negligible effect on responses. In choosing a dichotomous IRT model, single parameter models would be insufficient due to expected variance in item discrimination. Three parameter models would be ill-advised due to the anticipated negligible effect of guessing as well as the nature of guessing behavior on the CRT-A. As a result, the most appropriate dichotomous model for the CRT-A is the 2PL model.

In choosing a polytomous model for the CRT-A, the Rating Scale Model is considered inappropriate for the same reasons as single parameter dichotomous models: the assumption of equal discrimination parameters across items. The "left side added divide-by-total models" are unsuitable for the same reasons as three-parameter dichotomous models: the nature of guessing on CRT-A responses. The Graded Response Model, Nominal Response Model, and Partial Credit Model are three polytomous IRT models which are potentially theoretically appropriate for use with the CRT-A.

The Two-Parameter Logistic Model

The 2PL model (Lord & Novick, 1968) is intended to calculate the probability that a respondent will respond in a particular manner conditional on the respondent's latent trait level and the difficulty and discrimination of the item. Mathematically, this probability is calculated as $p(X_i) = \frac{e^{a_i(\theta - b_i)}}{1+e^{a_i(\theta - b_i)}}$, where θ is the latent trait level of the respondent, a_i is the discrimination of item i, and b_i is the difficulty of item i. Since the 2PL model is a dichotomous model, the responses are constrained to one keyed response and all other responses. Traditionally,

educational and psychological testing have designated these response categories as correct and incorrect. For use with the CRT-A, the keyed response will indicate the logical aggressive response option while the other three options will be combined into the category of non-keyed responses.

The Graded Response Model

The GRM (Samejima, 1969) calculates the probability that an individual will respond to an item using a response option at or above a certain level. Kieftenbeld and Natesan (2012) claim that the GRM is appropriate for use with ordered categorical data. The GRM assumes that K response options are ordered and calculates the probability that an individual will respond in response category k or above, conditional on trait level. Embretson and Reise (2000) describe the GRM as a series of 2PL models calculated on k – 1 dichotomies. Mathematically, the response probabilities are calculated using the formula $p(\theta, a_i, \lambda_{it}) = \frac{e^{a_i(\theta-\lambda_{it})}}{1+e^{a_i(\theta-\lambda_{it})}}$, where a_i indicates the difficulty of item i and λ_{it} indicates the boundary between two response categories for item i (similar to item difficulty in the 2PL model). The probability that an individual will choose response category k is obtained by subtracting the probability for one boundary from the probability for the boundary below it: $p(X = k) = p\left(\theta, a_i, \lambda_{i(t-1)}\right) - p(\theta, a_i, \lambda_{it})$.

CRT-A items are not naturally ordered into distinct categories. The two illogical items must be combined due to the fact that they are not designed to be distinguished from one another. Therefore, three response categories can be examined using the GRM: nonaggressive logical responses, illogical responses, and aggressive logical responses. These response categories do not naturally conform to an ordered scale due to the irrelevance of the illogical responses to the measurement of aggression. For the purpose of analysis, the response categories are ordered

theoretically to reflect the level of latent aggression indicated by each response. This scoring method is identical to the trichotomous scoring system for the CRT-A (see James et al., 2005).

The Nominal Response Model

The NRM was derived with the intention of acquiring information from both "correct" responses and distractors (see Bock, 1972). It is intended to be used in tests where the response options are nominal. While it may be used for other types of scales (e.g. ordinal, interval), there may be more appropriate models for those scales (e.g. GRM, PCM). Similar to other IRT models, the NRM assumes independence of items.

In a given test with n polytomous items, each item has m alternatives such that a vector of responses can be expressed as $\boldsymbol{k} = [k_1 \quad k_2 \quad k_3 \quad \ldots \quad k_n]$**.** This corresponds to the test containing $\prod_{j=1}^{n} m_j$ possible response patterns. The probability that an individual will respond to item j using response option k_i is a function of θ as well as two item parameters representing the slope (a) and intercept (c). This probability is given by the function $P_{jk_i}(\theta) = \frac{e^{z_{jk_i}(\theta)}}{\sum_{h=1}^{m_j} e^{z_{jh}(\theta)}}$, where $z_{jh}(\theta) = c_{jh} + a_{jh}(\theta)$ and h = 1, 2, 3, …, k_j, …, m_j. The NRM is classified by Thissen and Steinberg (1986) as a "divide-by-total" model because it divides the probability of a given response (conditional on a_j, c_j, and θ) by the sum of these probabilities for each of the m_j response options (also conditional on a_j, c_j, and θ). $\sum_{h=1}^{m_j} z_{jh}(\theta)$ is usually constrained to equal zero which imposes the same restriction on the slope and intercept parameters such that $\sum_{h=1}^{m_j} a_{jh} = 0$ and $\sum_{h=1}^{m_j} c_{jh} = 0$.

Since the NRM allows both correct responses and distractors to provide information about each item, early research on the NRM attempted to assess its utility in computer adaptive testing (CAT). De Ayala (1989) compared CAT procedures using the NRM and the 3PL model.

He found that the NRM provided more information for individuals with θ levels below -0.25, the models provided similar information for individuals with θ levels between -0.25 and 1.5, and the 3PL model provided more information for individuals with θ levels between 1.5 and 2.0. Both models required a similar number of items to converge on person estimates, though the NRM converged more often than the 3PL model. The two models provided similar results in cases where they both converged. A replication study (De Ayala, 1992) found similar results, concluding that the models are mostly comparable, but the NRM is preferable to the 3PL model for θ levels of 1.0 and below (also see De Ayala, 1993).

The NRM is useful for analyzing data when response options can be considered a nominal scale. As a result, the NRM can be used to analyze CRT-A item data without the need to combine response options into categories. Using the NRM on dichotomous data is equivalent to using the 2PL model (see Mellenbergh, 1995), so doing so would be redundant in this analysis. Using the NRM on data scored using the trichotomous system is also possible. Doing so will make it possible to directly compare the NRM to the GRM using model fit indices.

The Partial Credit Model and Generalized Partial Credit Model

The Partial Credit Model (PCM; Masters, 1992) was designed to analyze data from items in which the response options reflect ordered performance levels. The quintessential example of such an item is a multi-step mathematics problem in which success on the item depends on the respondent's ability to successfully complete a sequence of calculations. Masters also believed that the PCM could be used to analyze items with a Likert-type scale due to the fact that the available response alternatives are ordered in a hierarchical progression.

Mathematically, the PCM calculates the probability of an individual selecting a specific response option: $p(x_i|\theta) = \frac{e^{\sum_{k=0}^{x}(\theta-\delta_{ik})}}{\sum_{h=0}^{m} e^{\sum_{k=0}^{h}(\theta-\delta_{ik})}}$, where $\sum_{k=0}^{0}(\theta-\delta_{ik}) \equiv 0$. According to Masters

(1992), the item step difficulties (δ_{ik}) are theoretically similar to the boundaries in the GRM, but can be estimated separately from person parameters.

Since the PCM lacks a discrimination parameter, it is comparable to a single parameter dichotomous model, and therefore not suitable for the present analysis. Adding a discrimination parameter (a_i) to the PCM yields the Generalized Partial Credit Model (GPCM; Muraki, 1992). Mathematically, it is identical to the PCM except for the addition of this discrimination parameter: $p(x_i|\theta) = \frac{e^{\sum_{k=0}^{x} a_i(\theta-\delta_{ik})}}{\sum_{h=0}^{m} e^{\sum_{k=0}^{h} a_i(\theta-\delta_{ik})}}$, where $\sum_{k=0}^{0} a_i(\theta - \delta_{ik}) \equiv 0$. By adding a_i, the GPCM allows items to vary in discrimination, making the GPCM a more appropriate model for the current analyses.

Similar to the GRM, the GPCM requires response options to be theoretically ordered. As a result, the trichotomous scoring method must be employed due to the nature of the illogical responses.

CHAPTER 4

RESEARCH PLAN

Optimizing the Utility of the Polytomous Item Response Theory Models

A key decision involved in use of the polytomous IRT models is the choice of estimation techniques. Wollack, Bolt, Cohen, and Lee (2002) used simulated data to compare two popular estimation methods, Marginal Maximum Likelihood (MML) and Markov Chain Monte Carlo (MCMC), across multiple sample sizes and test lengths. Though results produced by the two methods were both acceptable, they demonstrated that MML consistently produced intercept and slope estimates with lower root mean square error (RMSE) than MCMC. Kieftenbeld and Natesan (2012) found that MML and MCMC produced similar results when sample sizes exceeded 300, noting that larger samples were associated with smaller RMSE.

Kang, Cohen, and Sung (2009) compared four model selection indices in addition to replicating the findings of lower RMSE and similarity of MML and MCMC parameter estimates when sample size is high. The authors concluded that the Akaiki Information Criterion (AIC; see Akiake, 1974) and Bayesian Information Criterion (BIC; see Schwarz, 1978) outperformed other model selection indices. While the AIC and BIC were both found to be quite accurate, Kang et al. (2009) note that the AIC tends to select a less parsimonious model and the BIC tends to select a more parsimonious model when large samples are used.

De Ayala and Sava-Bolesta (1999) examined the relationship of sample size to parameter estimates under various distributions of θ. They concluded that a higher ratio of observations to parameter estimates yielded better results. The authors recommended a ratio of at least 5:1 for a uniformly-distributed θ, at least 10:1 for a normally-distributed θ, and between 10:1 and 20:1 for

a skewed distribution of θ. DeMars (2003) replicated and extended this finding, but concluded that the number of parameters per item had the highest impact on the RMSE of parameter values, followed by sample size. Item discrimination and θ distribution had a smaller impact on RMSE. DeMars concluded that the conditions for optimal parameter recovery include high sample size, low number of parameters per item, low discrimination, and non-skewed (e.g. uniform, normal) θ distribution. Reise and Yu (1990) found similar results. Parameter recovery is optimized when sample size is high, particularly for items with a higher number of response options. Skewed distributions yield higher RMSE than uniform or normal distributions.

Limitations of the Polytomous IRT Models

All IRT models have a number of limitations. As stated above, the GRM and GPCM assume that the response options follow a logical ordered progression. The NRM is primarily intended for use with polytomous items at the nominal level of measurement. When used with dichotomous items, the NRM is equivalent to the more parsimonious 2PL model. When used with ordinal responses, the NRM is inappropriate because it treats response options as unordered and does not preserve the ordinal nature of the responses (Baker, 1993; Mellenbergh, 1995).

Use of a polytomous model requires more parameters to be estimated (De Ayala & Sava-Bolesta, 1999; DeMars, 2003; Suh & Bolt, 2010). Another limitation of the GRM, NRM, and GPCM is their lack of suitability for multidimensional tests (see Thissen, 1993; Thissen, Cai, & Bock, 2010). Finally, due to the fact that the NRM imposes the constraints $\sum_{h=1}^{m} a = 0$ and $\sum_{h=1}^{m} c = 0$, item characteristic curves tend to contain one response option that has a monotonic increasing relationship to θ (usually the keyed response), another response option that has a monotonic decreasing relationship to θ, and other response options with negligible relationships to θ (Thissen & Steinberg, 1984).

In conclusion, multiple IRT models may be appropriate for the proposed analyses. Since the CRT-A is scored dichotomously, the 2PL model will provide an assessment of whether each item is able to differentiate aggressive individuals from nonaggressive individuals. Use of polytomous models is also important in order to determine whether the nonaggressive responses are performing as theorized. The set of responses to a given CRT-A item is best characterized as nominal due to the unordered nature of the distractors on the trait of interest. The NRM is a theoretically appropriate polytomous model due to the lack of a requirement to combine the illogical response options and order the response categories. As a result, the NRM is expected to outperform the GRM and GPCM.

The CRT-A was designed to measure aggression. While previous factor analyses have attempted to partition aggression into separate factors, these factors are correlated with one another (see James et al., 2005; Ko, Thompson, Shim, Roberts, & McIntyre, 2008). Previous psychometric analyses indicate that the distribution of scores is positively skewed, so a large sample is required in order to obtain reliable item parameter estimates. The CRT-A response database contains more than 5,000 respondents, making it an ideal sample for this analysis. Finally, unlike most tests, the CRT-A is designed to contain one response option with a monotonic increase, one response with a monotonic decrease, and two responses with no relationship to θ. As a result, the NRM characteristic that Thissen & Steinberg (1984) saw as a liability is perfectly suited to the format of the CRT-A.

Analyses, Expectations, and Hypotheses

This dissertation contains multiple analyses, each of which is expected to provide evidence that the CRT-A works as theoretically intended. Prior to the item-specific analyses, however, it is important to determine which models are most appropriate for use with the CRT-

A. When considering dichotomous models, the interest in discrimination parameters necessitates the use of the 2PL model. When examining polytomous models, however, the choice of model is less clear. Since the GRM, NRM, and GPCM are all polytomous extensions of the 2PL model, each could potentially provide adequate item parameters.

Comparing the GRM, NRM, and GPCM requires the use of model fit indices. Since none of these models is nested within another, direct comparison of chi-square statistics cannot provide an adequate evaluation of model superiority. Both AIC and BIC are examined to determine which model best fit the data. As mentioned above, when sample size is large, AIC tends to favor complex models (such as the NRM) due to the increased explanatory power associated with adding additional components to a model. BIC tends to favor more parsimonious models due to the penalty for additional components.

The GRM and GPCM impose the restriction that CRT-A response options must be ordered logically while the NRM allows these response options to be treated as nominal. It is expected that the three response categories available on a given CRT-A item do not conform to this logical ordered progression, so the GRM and GPCM will both be too restrictive. Therefore, it is expected that the NRM will display superior fit on both the AIC and BIC. Since the NRM is a more general model, the AIC should favor the NRM even if all three models were equally applicable to the CRT-A. The increase in model fit due to theoretical appropriateness is expected to outweigh the complexity penalty of the BIC.

Unfortunately, the 2PL model cannot be directly compared to the GRM, NRM, or GPCM due to the different scoring systems required for use with dichotomous and polytomous models. Even model fit indices such as the AIC and BIC would be inappropriate due to the differing number of parameters estimated in the models. Modifying the scoring system for the purpose of

this comparison would be impractical since the NRM is equivalent to the 2PL model when used with dichotomous data. As a result, the results are presented using the 2PL model and the most appropriate polytomous model.

The confirmatory portion of this dissertation involves examining the current CRT-A scoring system by demonstrating that the probability of choosing the aggressive response option increases as latent aggression increases. It is also expected that the probability of selecting the nonaggressive logical response option decreases as estimates of latent aggression increase. There is no reason to suspect that the selection of illogical options is influenced by latent aggression. If the CRT-A works as theorized, then more aggressive individuals will be more likely to endorse aggressive logical responses and less likely to endorse nonaggressive logical responses. Demonstration of these relationships using IRT would be consistent with positive item-total correlations found in previous analyses. Illogical responses are not expected to relate to estimates of latent aggression.

As stated above, the CRT-A contains 22 items, each of which contains one aggressive logical and one nonaggressive logical option. Therefore, 22 individual hypotheses (one per item) which will be tested using IRT. In the interest of space, these hypotheses are stated in a general form below:

Hypothesis: *For every keyed CRT-A item, the logical aggressive response option will display a positive relationship with estimates of latent aggression as evidenced by the slope coefficient for the item.*

The relationship of nonaggressive logical response options to latent aggression will also be examined with the expectation of a negative association. The analysis proposed above can also be extended to each of the subscales of the CRT-A. While the same pattern of relationships

is expected in each of the subscales, confirmation of these relationships is not as important due to the recommendation that further development is necessary before these subscales serve as stand-alone tests examining specific mechanisms of aggression.

Although the primary hypothesis for this dissertation reflects the theoretical underpinnings of the CRT-A, the large sample available for this analysis allows for a more comprehensive examination of item characteristics. In addition to computing the direction of the item slope estimates, this analysis should also be able to provide somewhat accurate point estimates of the slope parameters. Since a steeper slope indicates a stronger relationship to latent aggression estimates, this dissertation should reveal the items which best discriminate among respondents. Unlike the parallel CTT analysis of item-total correlations, the IRT analysis provides information regarding the levels of latent aggression at which each item best discriminates.

A final analysis aims to determine whether CRT-A respondents from student samples respond differently to items than respondents from non-student samples. A differential item functioning (DIF) analysis using the 2PL model yields a comparison of the two groups on each item to determine the items on which differences exist and the nature of those differences.

CHAPTER 5

METHOD

Sample

The CRT-A response database contains 5,511 individuals who provided a valid set of responses to items on the CRT-A (as defined by James & McIntyre, 2000). The database contains responses from participants in 22 studies from 2000 to 2011, each of which used the current version of the CRT-A. The psychometric analysis produces between 44 and 132 item parameter estimates for the 22 items on the test, depending on the model used. Using De Ayala and Sava-Bolesta's (1999) conservative estimate of 20 subjects per parameter estimate, a sample of 2,640 is required to obtain reliable parameter estimates when the distribution of θ is skewed. Using the CRT-A response database will provide a ratio greater than 41:1.

Differential Item Functioning Analysis

The analysis of DIF between student and non-student (applied) samples uses 21 of the 22 samples in the CRT-A response database for which respondent information is available. This analysis compares 3,816 student respondents to 1,288 applied respondents using the larger group as the reference group. Wald test results indicate the items for which DIF occurs. Since the models being compared contain two parameters, the chi-square statistic associated with the Wald test is also partitioned into slope and intercept components, with differences in slope coefficients being the primary analysis of interest.

Conditional Reasoning Test of Aggression Scoring

For analyses involving the 2PL model, the CRT-A was scored using the dichotomous scoring system described above. For analyses involving the GRM, NRM, and GPCM, the CRT-A was scored using the trichotomous scoring system described above.

Modeling Software

All analyses are conducted using IRTPRO 2.1, a program written by Li Cai, David Thissen, and Stephen du Toit, which is available on SSICentral.com. Parameter estimates are obtained using Maximum Likelihood estimation based on the findings of Wollack et al. (2002). Unless otherwise stated, iterative analyses are allowed 10,000 cycles with a convergence criterion of .00001, 5,000 M-Step iterations with a convergence criterion of .000001, and 600 rectangular quadrature points.

Four different IRT models are analyzed using the IRTPRO software: the 2PL model, GRM, NRM, and GPCM. Results from the 2PL model include slope and intercept parameter estimates along with standard error. As stated in the hypothesis, slope parameters for all items are expected to be statistically greater than zero. Graphs for each item characteristic curve (ICC) and information function are presented for visual representation.

The three polytomous models are compared using the AIC and BIC model fit indices, based on the findings of Kang et al. (2009). Linear slope estimates based on the ICCs are provided to examine whether the response options are working as intended. Results from the model with the lowest AIC and BIC estimate are presented along with ICC and information graphs for each item.

The DIF analysis compares student and applied respondents using the 2PL model. Wald test results are presented to determine whether each item displays evidence of DIF. Wald test

chi-square values are partitioned into slope and intercept components. A graph of the student and applied ICCs is provided for comparison purposes.

Each item is also analyzed in the context of its factor subscale. 2PL model slope and intercept estimates are presented along with graphs of the ICC similar to the 2PL CRT-A analysis mentioned above.

CHAPTER 6

RESULTS

Model Fit Analyses

The first set of analyses is intended to provide model comparisons using the AIC and BIC model fit indices. The IRTPRO software also includes the M_2 statistic along with an associated chi-square test and root estimate of the root mean square error of approximation (RMSEA). Model fit results for each model are reported in Tables 1 and 2.

Table 1

Model Fit Statistics for Models using the Conditional Reasoning Test of Aggression.

Model	M_2	df	p-value	RMSEA
2PL	633.91	209	0.0001	0.02
GRM	3569.33	902	0.0001	0.02
NRM	2202.66	880	0.0001	0.02
GPCM	3921.49	902	0.0001	0.02

Note. df = degrees of freedom; RMSEA = root mean square error of approximation; 2PL = two-parameter logistic model; GRM = graded response model; NRM = nominal response model; GPCM = generalized partial credit model.

Table 2

Model Fit Indices for Models using the Conditional Reasoning Test of Aggression.

Model	-2lnL	AIC	BIC
2PL	103635.47	103723.47	104014.51
GRM	122978.7	123110.7[a]	123547.26[b]
NRM	122683.3	122859.3[a]	123441.37[b]
GPCM	123064.17	123196.17[a]	123632.73[b]

Note. -2lnL = -2 log likelihood, AIC = Akiake information criterion; BIC = Bayesian information criterion; 2PL = two-parameter logistic model; GRM = graded response model; NRM = nominal response model; GPCM = generalized partial credit model.
[a]Models can be directly compared using AIC. [b]Models can be directly compared using BIC.

RMSEA estimates indicate good fit for each of the four models. Consistent with expectations, both the AIC and BIC are slightly lower for the NRM than the GRM or the GPCM, although results for the three models are comparable (AIC and BIC indices for each model are within one percent of one another). -2lnL and RMSEA estimates for each of the models are also similar. These results indicate that the three polytomous IRT models are comparable, but that the NRM fits slightly better than the GRM or GPCM.

Item Characteristics of the Conditional Reasoning Test of Aggression

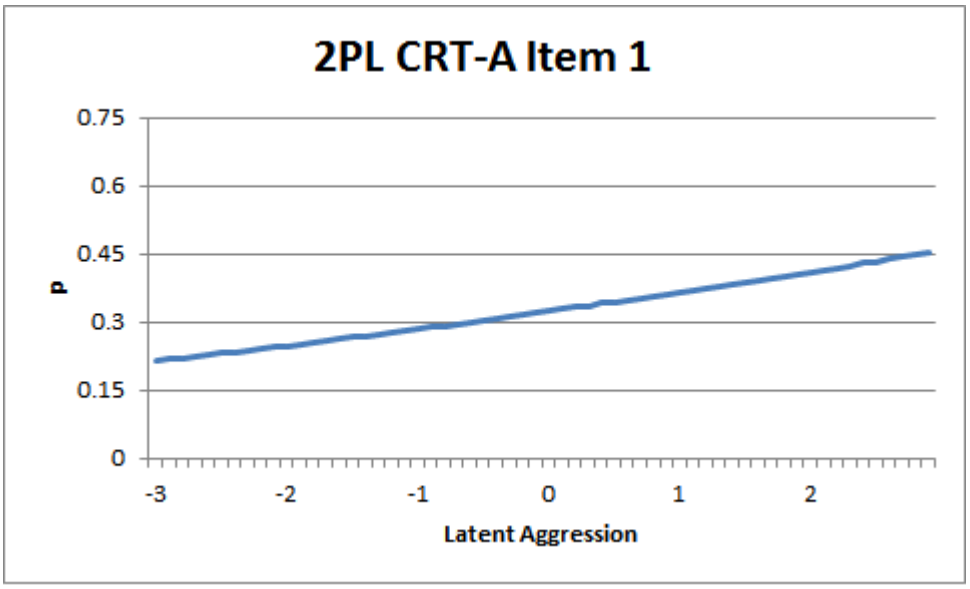
Analyses involving individual item characteristics are presented by item to allow the reader to examine each CRT-A item in depth. The text of the item is first presented alongside its CTT psychometric characteristics (updated to reflect the current dataset). The 2PL model and NRM analyses are then presented to determine whether the item performance is consistent with theoretical expectations. The DIF analysis examines whether the item performs differently in student and applied samples, and whether any differences are due to slope, intercept, or both. Finally, the item is presented within the framework of their respective factors.

Item 1

The first CRT-A item stem is "Joe is usually on time for work and for meetings with his boss and clients. He is also on time for appointments with his doctor, dentist, and priest. However, Joe is always five or more minutes late for meetings with Bill. Which of the following is the most logical explanation for Joe being late for meetings with Bill?" The aggressive logical response option is "Joe is usually on time for people he respects, so he must not respect Bill." The nonaggressive logical response option is "Joe and Bill are friends, so they don't care about being on time for each other." The two illogical options are "Bill gets up later than Joe" and "Joe and Bill are both self-employed." Of the 5,511 responses, 32.46% chose the aggressive logical option, 65.41% chose the nonaggressive logical option,

1.91% chose one of the illogical options, and 0.22% of responses were missing. The item-total correlation for item one and the dichotomously scored CRT-A is 0.06.

The 2PL slope estimate is 0.19 with a standard error of 0.05. The data are consistent with the hypothesis in that the slope coefficient is statistically greater than zero and monotonically increases with estimates of latent aggression. The 2PL intercept estimate is 3.93 with a standard error of 0.95. Information monotonically increases up to $\theta = 2.9$, indicating that this item is best suited for individuals high on latent aggression. The 2PL ICC and information graph are presented in Figure 1.



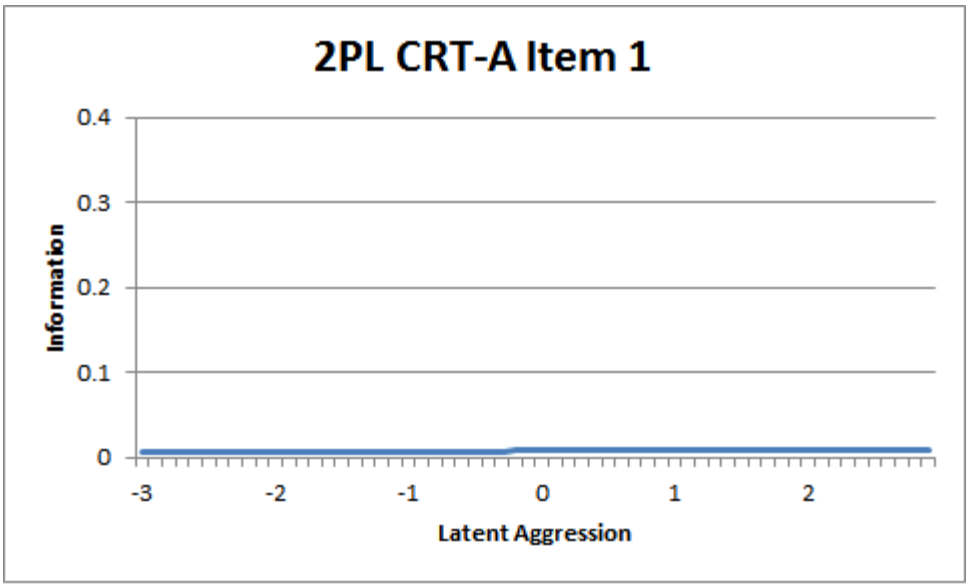
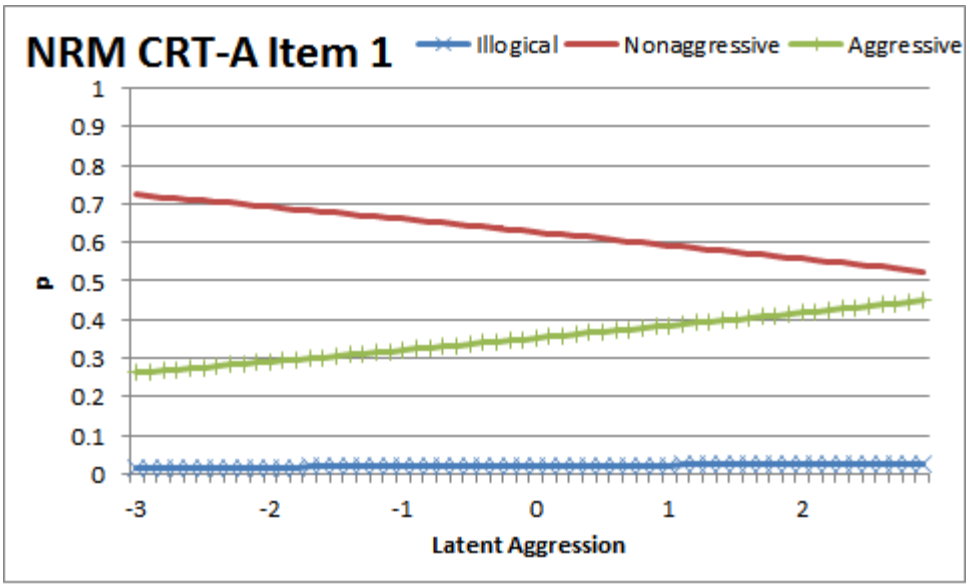


Figure 1. Two-Parameter Logistic Model item characteristic curve and information function for item 1 of the Conditional Reasoning Test of Aggression.

The linear NRM slope estimates are .032 for the aggressive logical response option, -.034 for the nonaggressive logical response option, and .002 for the combined illogical options. This is consistent with the expectation that the aggressive option will have a positive relationship with latent aggression while the nonaggressive option has a negative relationship with latent aggression. Information monotonically increases up to $\theta = 2.9$, indicating that this item is best suited for individuals high on latent aggression. The NRM ICC and information graph are presented in Figure 2.



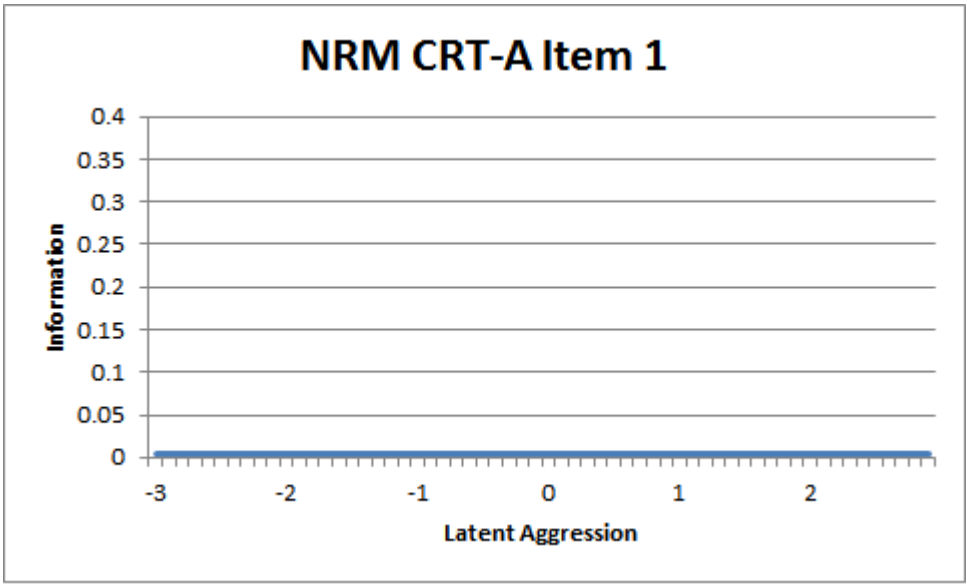


Figure 2. Nominal Response Model item characteristic curve and information function for item 1 of the Conditional Reasoning Test of Aggression.

Item one displays evidence of DIF (X^2, $2df = 7.7$, $p < .05$). This chi-square value can be partitioned into slope (X^2, $1df = 0.4$, $p > .05$) and intercept (X^2, $1df = 7.2$, $p < .05$) components, indicating that student and applied samples statistically differ in terms of intercept, but not slope. A graph displaying the ICCs for each group can be found in Figure 3.

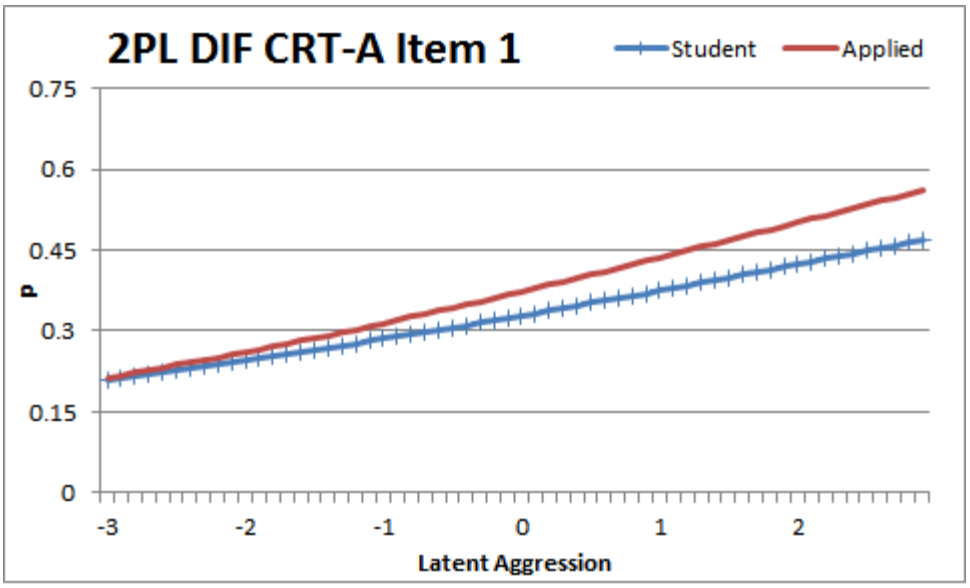


Figure 3. Two-Parameter Logistic Model differential item functioning item characteristic curve for item 1 of the Conditional Reasoning Test of Aggression.

The factor loadings for item one are 0.004 for externalizing controls, 0.051 for internalizing controls, and 0.233 for powerlessness, indicating that item one is best suited for the powerlessness factor. The factor-specific 2PL slope estimate for item one is 0.37 with a standard error of 0.07. The 2PL intercept estimate is 2.02 with a standard error of 0.36. This indicates that item one works well as an indicator of latent powerlessness as well as latent aggression. The factor-specific 2PL ICC for item one can be found in Figure 4.

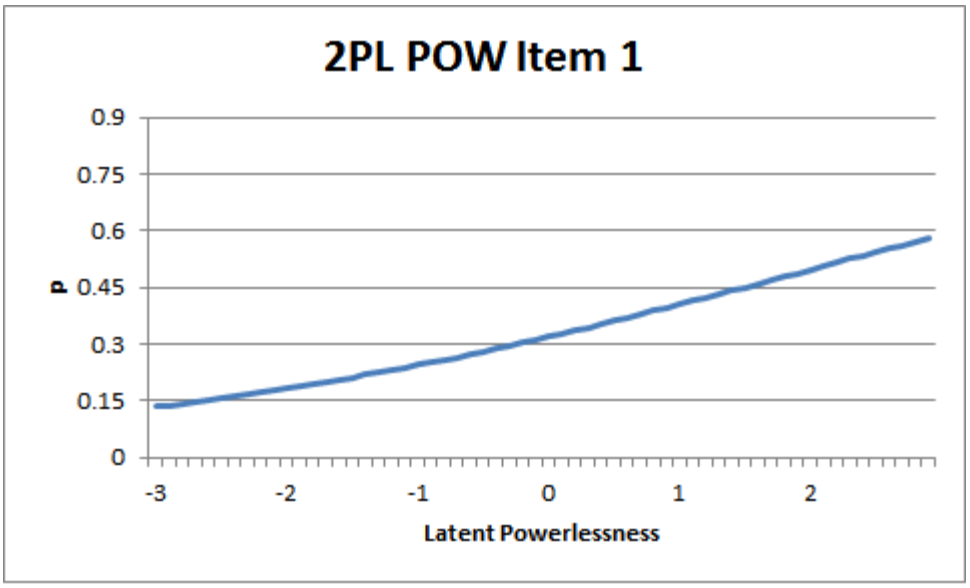


Figure 4. Two-Parameter Logistic Model item characteristic curve and information function for item 1 of the Powerlessness subscale of the Conditional Reasoning Test of Aggression.

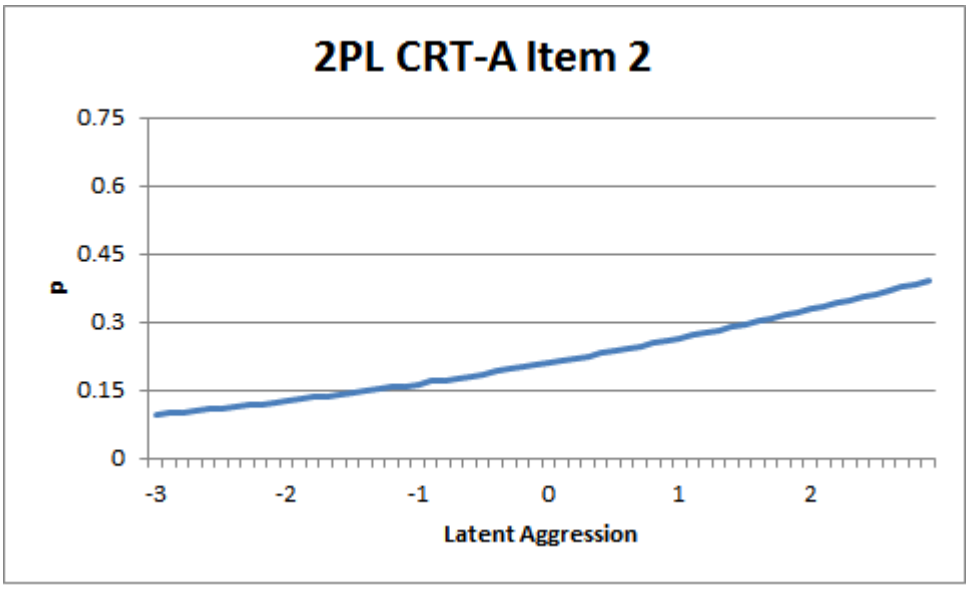
Overall, item one has strong item characteristics. Responses to item one are well suited to indicate both latent aggression and latent powerlessness. There is some evidence of DIF, but it is constrained to difference in intercept, meaning that item one discriminates among student and applied respondents similarly.

Item 2

The second CRT-A item stem is "People who are pushy about getting what they want are often disliked by others. However, aggressively going after customers is often needed to be successful in sales. People who are successful in sales are usually respected by others. Which of the following is the most logical conclusion based on the above?" The aggressive logical response option is "Salespeople who are not pushy will not be successful or respected." The nonaggressive logical response options is "Pushy salespeople may be successful but will often be disliked." The two illogical options are "Doctors are not respected by most people" and "Sales is the only job that requires pushiness." Of the 5,511 responses, 21.45% chose the aggressive logical option, 76.39% chose the nonaggressive logical

option, 2.05% chose one of the illogical options, and 0.11% of responses were missing. The item-total correlation for item two and the dichotomously scored CRT-A is 0.08.

The 2PL slope estimate is 0.30 with a standard error of 0.05. The data are consistent with the hypothesis in that the slope coefficient is statistically greater than zero and monotonically increases with estimates of latent aggression. The 2PL intercept estimate is 4.37 with a standard error of 0.72. Information monotonically increases up to $\theta = 2.9$, indicating that this item is best suited for individuals high on latent aggression. The 2PL ICC and information graph are presented in Figure 5.



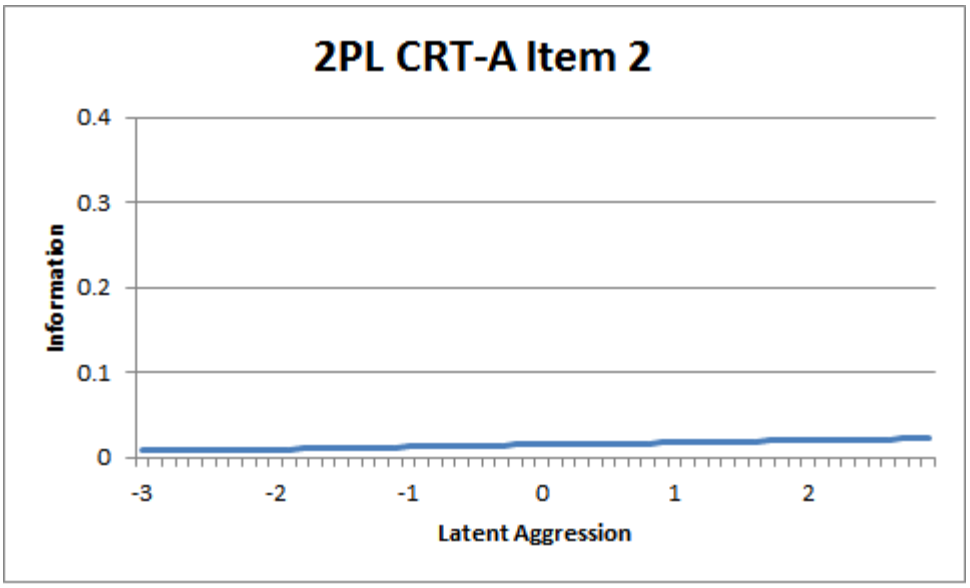
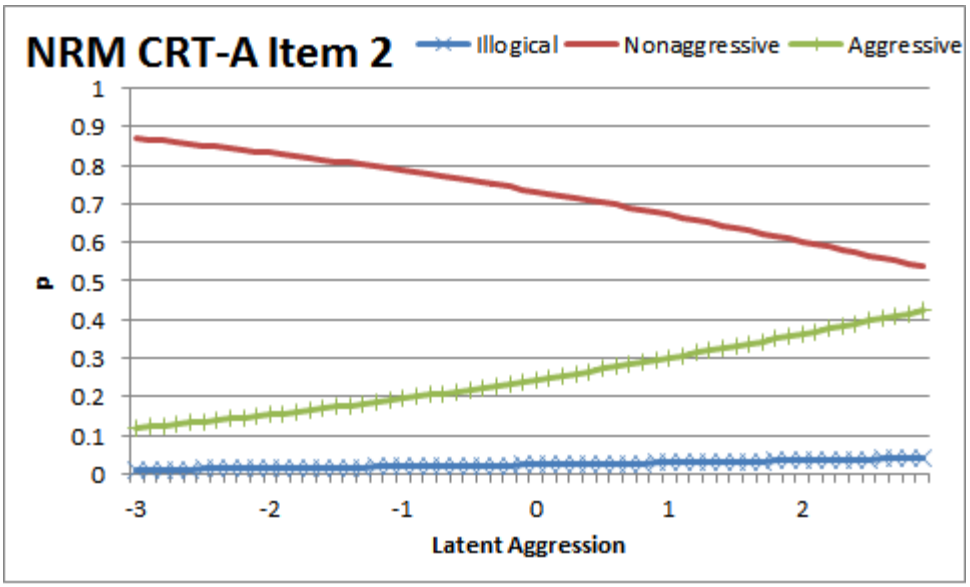


Figure 5. Two-Parameter Logistic Model item characteristic curve and information function for item 2 of the Conditional Reasoning Test of Aggression.

The linear NRM slope estimates are .052 for the aggressive logical response option, -.057 for the nonaggressive logical response option, and .005 for the combined illogical options. This is consistent with the expectation that the aggressive option will have a positive relationship with latent aggression while the nonaggressive option has a negative relationship with latent aggression. Information monotonically increases up to $\theta = 2.9$, indicating that this item is best suited for individuals high on latent aggression. The NRM ICC and information graph are presented in Figure 6.



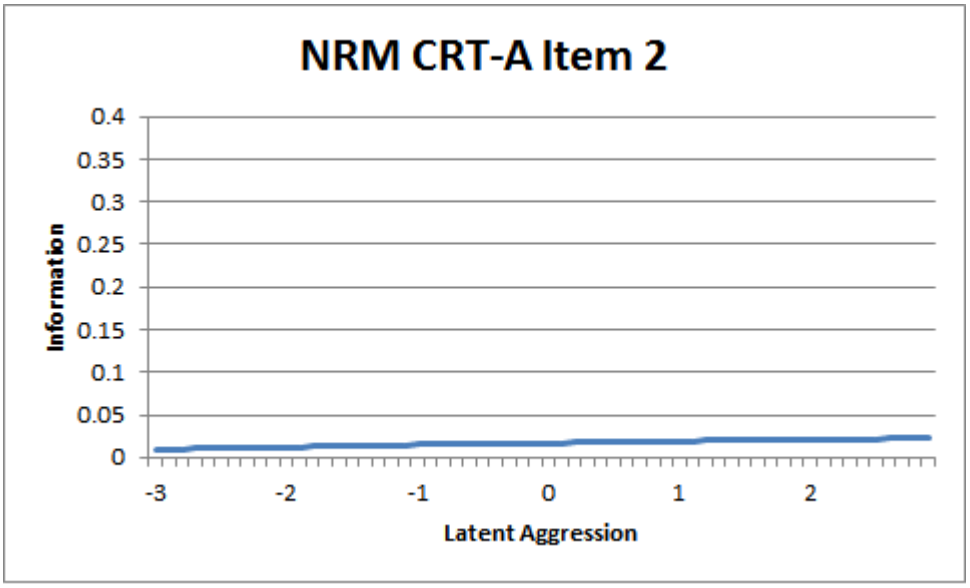


Figure 6. Nominal Response Model item characteristic curve and information function for item 2 of the Conditional Reasoning Test of Aggression.

Item two displays evidence of DIF (X^2, $2df = 10.9$, $p < .05$). This chi-square value can be partitioned into slope (X^2, $1df = 0.7$, $p > .05$) and intercept (X^2, $1df = 10.2$, $p < .05$) components, indicating that student and applied samples statistically differ in terms of intercept, but not slope. A graph displaying the ICCs for each group can be found in Figure 7.

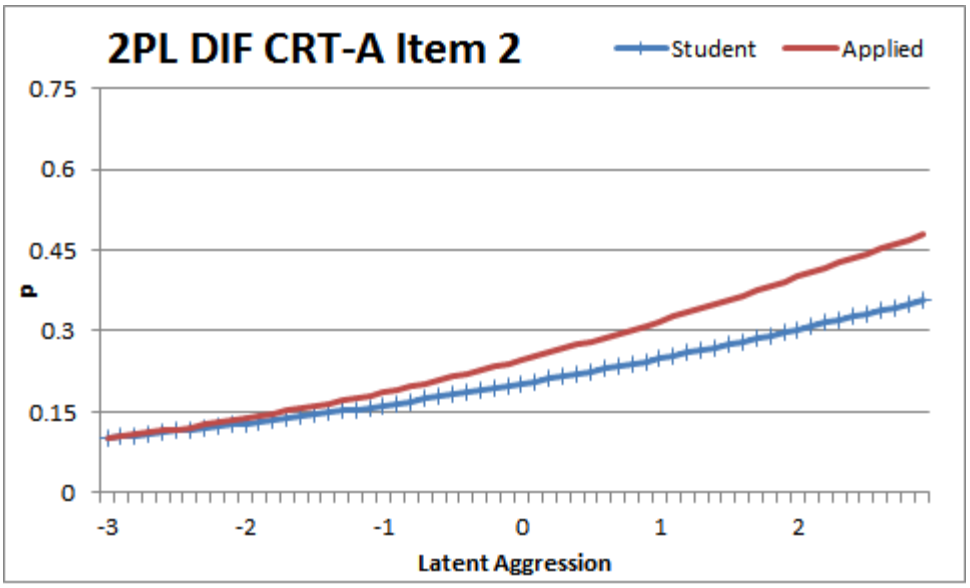


Figure 7. Two-Parameter Logistic Model differential item functioning item characteristic curve for item 2 of the Conditional Reasoning Test of Aggression.

The factor loadings for item two are 0.121 for externalizing controls, 0.211 for internalizing controls, and 0.087 for powerlessness, indicating that item two is primarily related to internalizing controls, but also related to externalizing controls. The factor-specific 2PL slope estimate for item two is 0.30 with a standard error of 0.07. The 2PL intercept estimate is 4.39 with a standard error of 1.00. This indicates that item two works well as an indicator of latent internalizing controls as well as latent aggression. The factor-specific 2PL ICC for item two can be found in Figure 8.

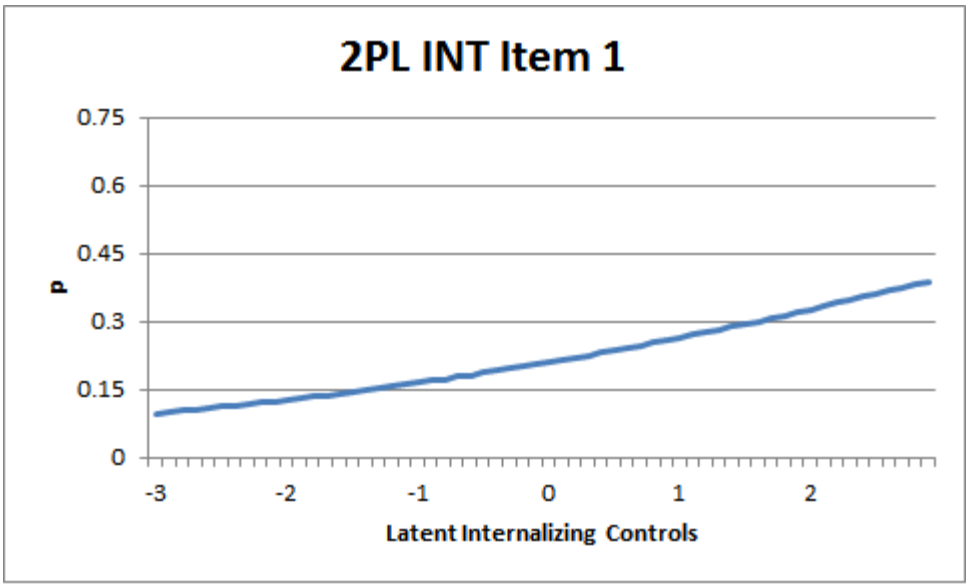


Figure 8. Two-Parameter Logistic Model item characteristic curve and information function for item 1 of the Internalizing Controls subscale of the Conditional Reasoning Test of Aggression.

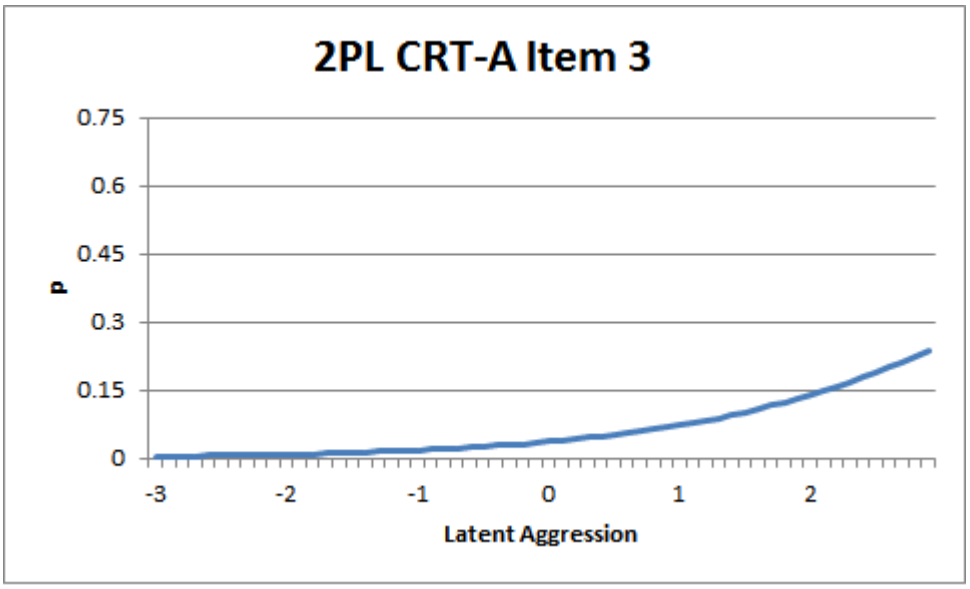
Overall, item two has strong item characteristics. Responses to item two are well suited to indicate both latent aggression and latent internalizing controls. There is some evidence of DIF, but it is constrained to difference in intercept, meaning that item two discriminates among student and applied respondents similarly.

Item 3

The third CRT-A item stem is "History shows that many generals who were good leaders in war were not as good during peacetime. Also, many generals who were promoted during peacetime were not good at leading soldiers in war. Which of the following is the most logical conclusion based on the above?" The aggressive logical response option is "Weak people with friends in high places are often chosen to be generals during peacetime." The nonaggressive logical response option is "It is hard to know how officers will do in battle until they are actually in a war." The two illogical options are "Generals and privates usually sit together at meals" and "Modern wars are more often fought at sea than in the air." Of the 5,511 responses, 4.61% chose the aggressive logical option, 94.92% chose the nonaggressive logical option, 0.33% chose one of the illogical options, and 0.15% of responses

were missing. The item-total correlation for item three and the dichotomously scored CRT-A is 0.11.

The 2PL slope estimate is 0.72 with a standard error of 0.10. The data are consistent with the hypothesis in that the slope coefficient is statistically greater than zero and monotonically increases with estimates of latent aggression. The 2PL intercept estimate is 4.52 with a standard error of 0.55. Information monotonically increases up to $\theta = 2.9$, indicating that this item is best suited for individuals high on latent aggression. The 2PL ICC and information graph are presented in Figure 9.



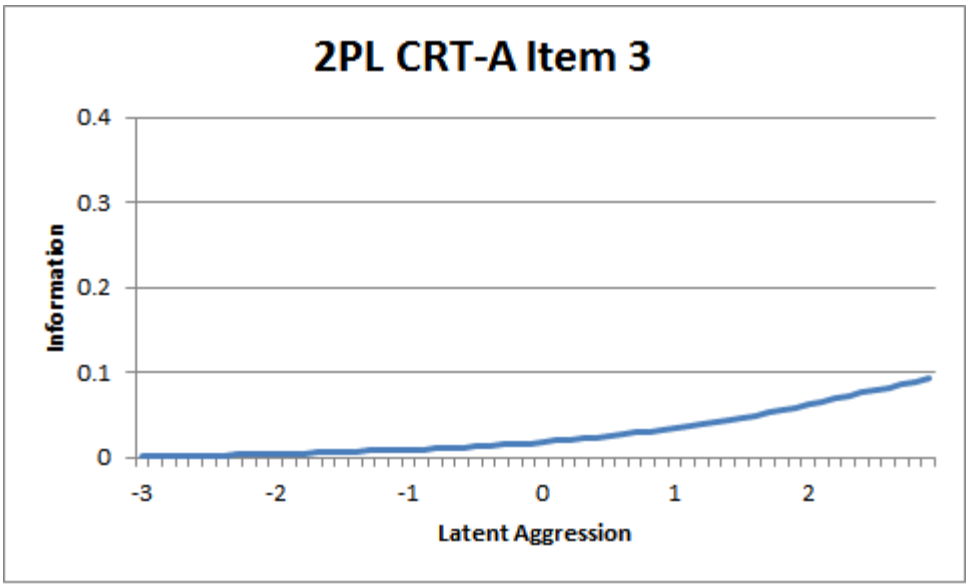
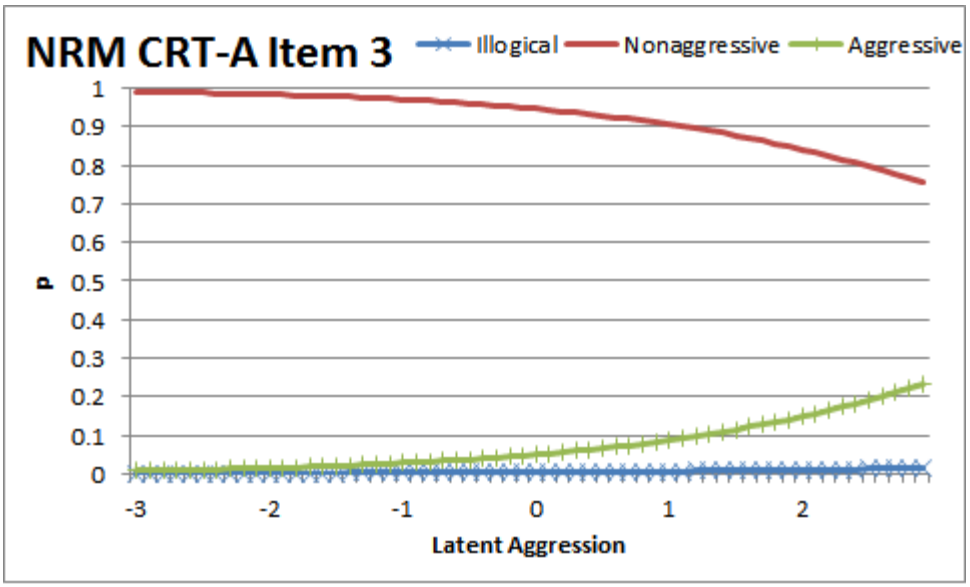


Figure 9. Two-Parameter Logistic Model item characteristic curve and information function for item 3 of the Conditional Reasoning Test of Aggression.

The linear NRM slope estimates are .034 for the aggressive logical response option, -.036 for the nonaggressive logical response option, and .002 for the combined illogical options. This is consistent with the expectation that the aggressive option will have a positive relationship with latent aggression while the nonaggressive option has a negative relationship with latent aggression. Information monotonically increases up to $\theta = 2.9$, indicating that this item is best suited for individuals high on latent aggression. The NRM ICC and information graph are presented in Figure 10.



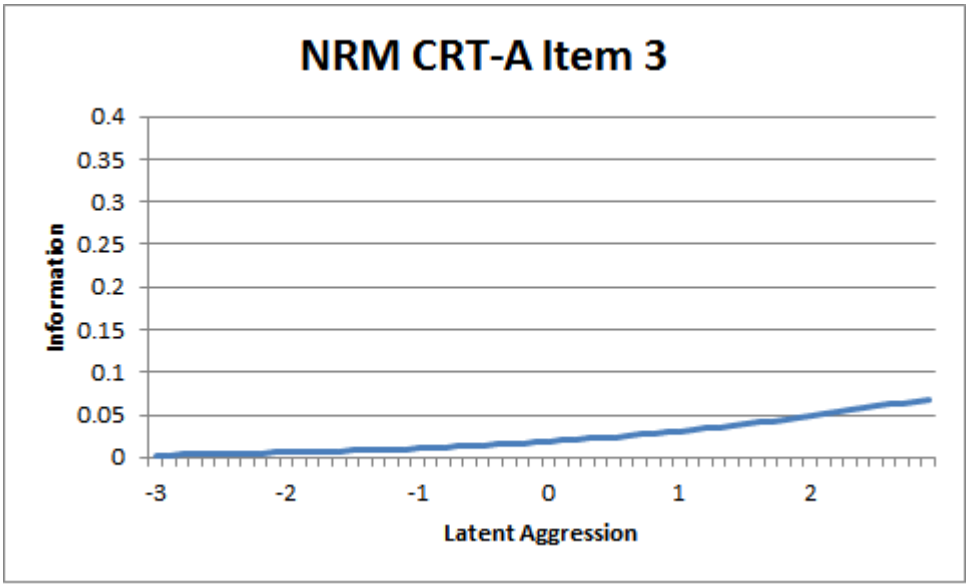


Figure 10. Nominal Response Model item characteristic curve and information function for item 3 of the Conditional Reasoning Test of Aggression.

Item three displays evidence of DIF (X^2, $2df = 10.9$, $p < .05$). This chi-square value can be partitioned into slope (X^2, $1df < 0.1$, $p > .05$) and intercept (X^2, $1df = 10.9$, $p < .05$) components, indicating that student and applied samples statistically differ in terms of intercept, but not slope. A graph displaying the ICCs for each group can be found in Figure 11.

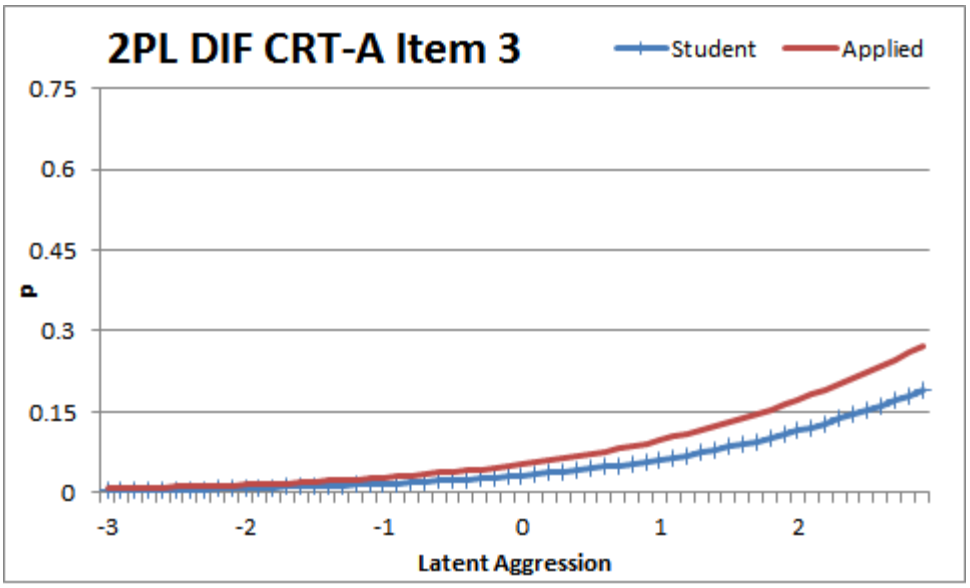


Figure 11. Two-Parameter Logistic Model differential item functioning item characteristic curve for item 3 of the Conditional Reasoning Test of Aggression.

The factor loadings for item three are 0.375 for externalizing controls, 0.476 for internalizing controls, and 0.078 for powerlessness, indicating that item three is primarily related to internalizing controls, but also related to externalizing controls. The factor-specific 2PL slope estimate for item three is 0.72 with a standard error of 0.14. The 2PL intercept estimate is 4.5 with a standard error of 0.76. This indicates that item three works well as an indicator of latent internalizing controls as well as latent aggression. The factor-specific 2PL ICC for item three can be found in Figure 12.

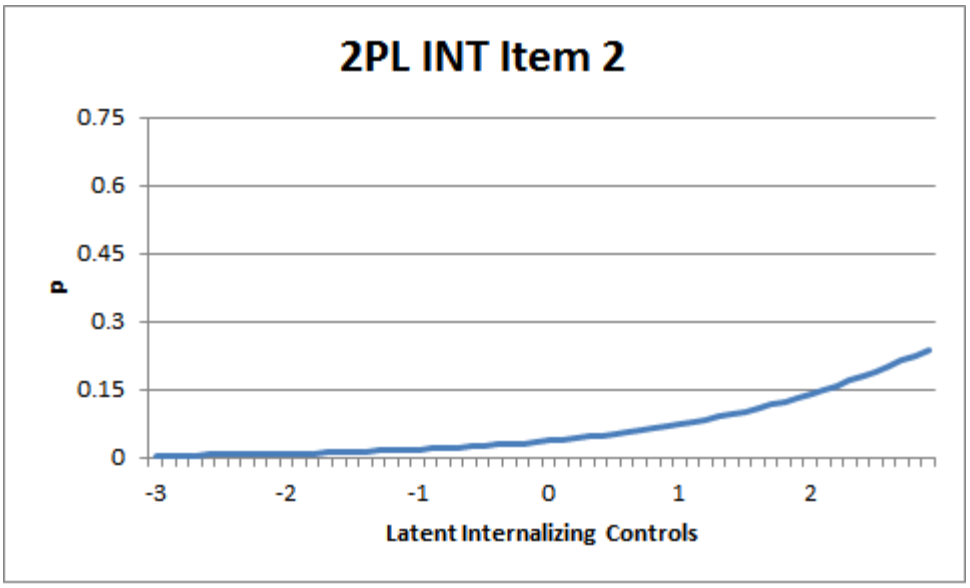


Figure 12. Two-Parameter Logistic Model item characteristic curve and information function for item 2 of the Internalizing Controls subscale of the Conditional Reasoning Test of Aggression.

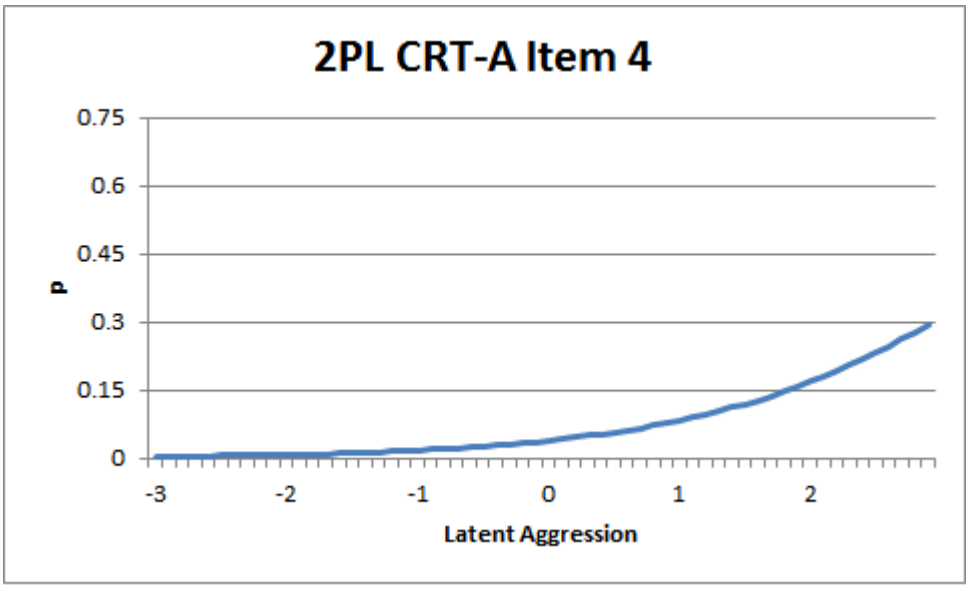
Overall, item three has strong item characteristics. Responses to item three are well suited to indicate both latent aggression and latent internalizing controls. There is some evidence of DIF, but it is constrained to difference in intercept, meaning that item three discriminates among student and applied respondents similarly.

Item 4

The fourth CRT-A item stem is "The old saying, 'an eye for an eye,' means that if someone hurts you, then you should hurt that person back. If you are hit, then you should hit back. If someone burns your house, then you should burn that person's house. Which of the following is the biggest problem with the 'eye for an eye' plan?" The aggressive logical response option is "People have to wait until they are attacked before they can strike." The nonaggressive logical response option is "It offers no way to settle a conflict in a friendly manner." The two illogical options are "It tells people to 'turn the other cheek'" and "It can only be used at certain times of the year." Of the 5,511 responses, 5.14% chose the aggressive logical option, 93.52% chose the nonaggressive logical option, 1.29% chose one of

the illogical options, and 0.05% of responses were missing. The item-total correlation for item four and the dichotomously scored CRT-A is 0.12.

The 2PL slope estimate is 0.80 with a standard error of 0.10. The data are consistent with the hypothesis in that the slope coefficient is statistically greater than zero and monotonically increases with estimates of latent aggression. The 2PL intercept estimate is 3.98 with a standard error of 0.41. Information monotonically increases up to $\theta = 2.9$, indicating that this item is best suited for individuals high on latent aggression. The 2PL ICC and information graph are presented in Figure 13.



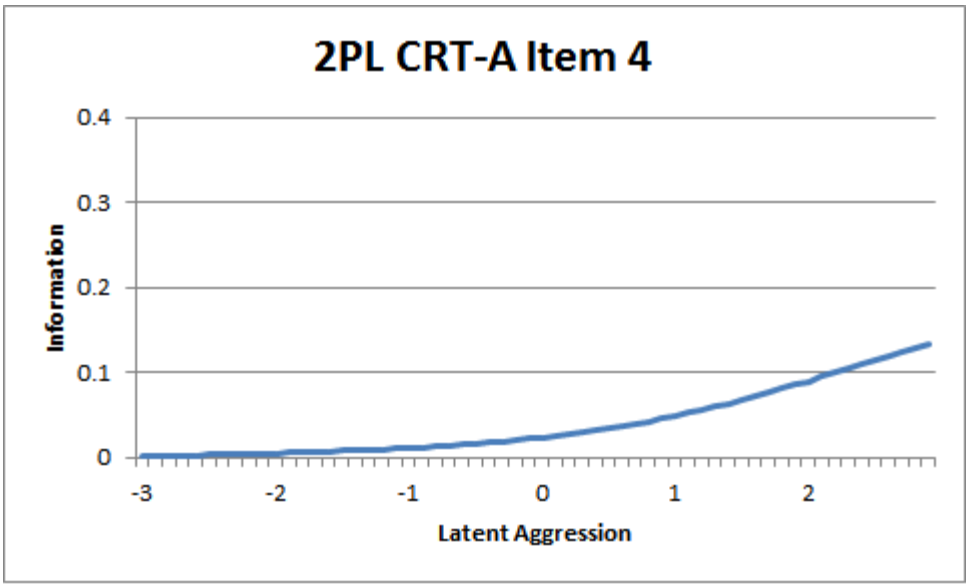
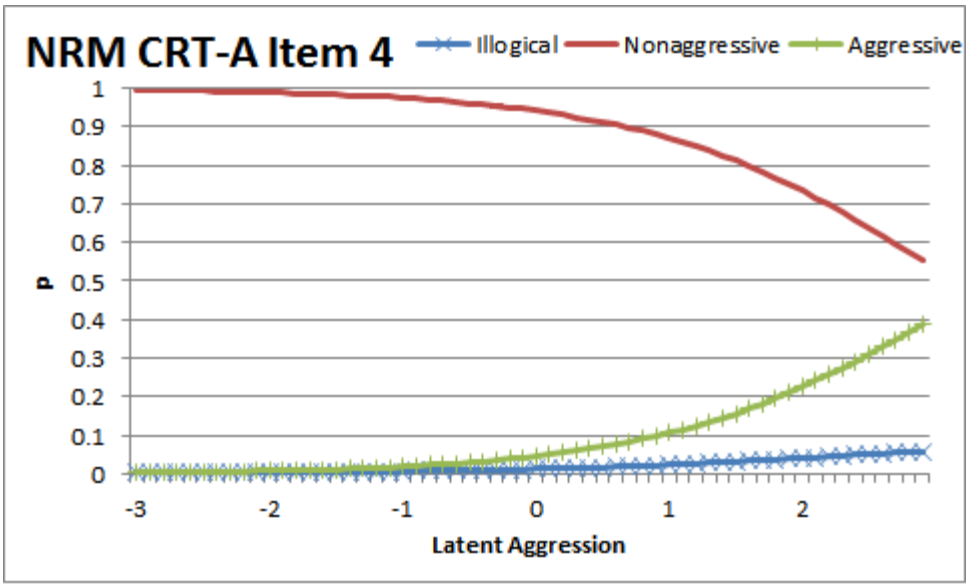


Figure 13. Two-Parameter Logistic Model item characteristic curve and information function for item 4 of the Conditional Reasoning Test of Aggression.

The linear NRM slope estimates are .056 for the aggressive logical response option, -.065 for the nonaggressive logical response option, and .009 for the combined illogical options. This is consistent with the expectation that the aggressive option will have a positive relationship with latent aggression while the nonaggressive option has a negative relationship with latent aggression. Information monotonically increases up to $\theta = 2.9$, indicating that this item is best suited for individuals high on latent aggression. The NRM ICC and information graph are presented in Figure 14.



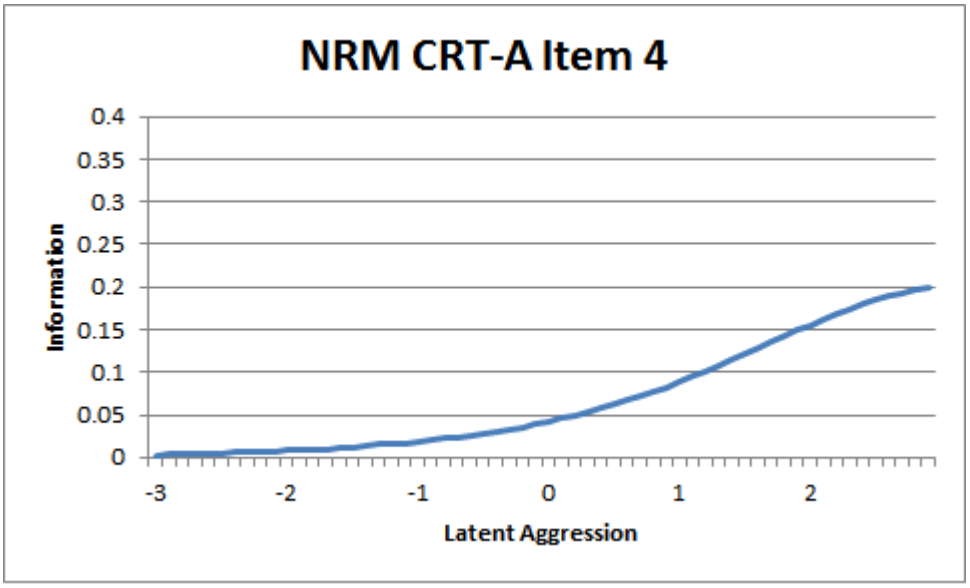


Figure 14. Nominal Response Model item characteristic curve and information function for item 4 of the Conditional Reasoning Test of Aggression.

Item four displays no evidence of DIF (X^2, $2df = 2.4$, $p > .05$). This chi-square value can be partitioned into slope (X^2, $1df = 2.1$, $p > .05$) and intercept (X^2, $1df = 0.4$, $p > .05$) components, indicating that student and applied samples do not statistically differ in either slope or intercept. A graph displaying the ICCs for each group can be found in Figure 15.

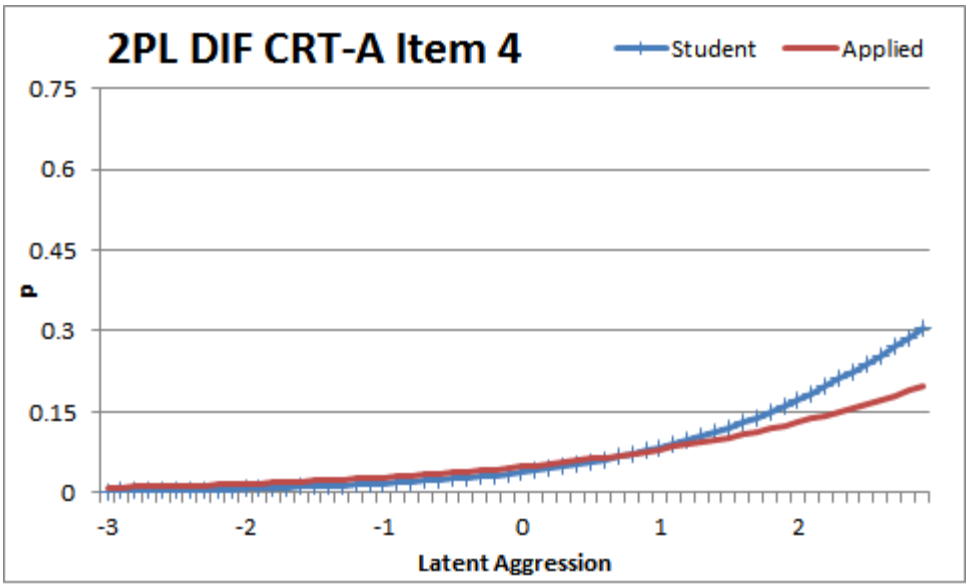


Figure 15. Two-Parameter Logistic Model differential item functioning item characteristic curve for item 4 of the Conditional Reasoning Test of Aggression.

The factor loadings for item four are 0.424 for externalizing controls, 0.633 for internalizing controls, and -0.047 for powerlessness, indicating that item four is primarily related to internalizing controls, but also related to externalizing controls. The factor-specific 2PL slope estimate for item four is 0.89 with a standard error of 0.14. The 2PL intercept estimate is 3.65 with a standard error of 0.46. This indicates that item four works well as an indicator of latent internalizing controls as well as latent aggression. The factor-specific 2PL ICC for item four can be found in Figure 16.

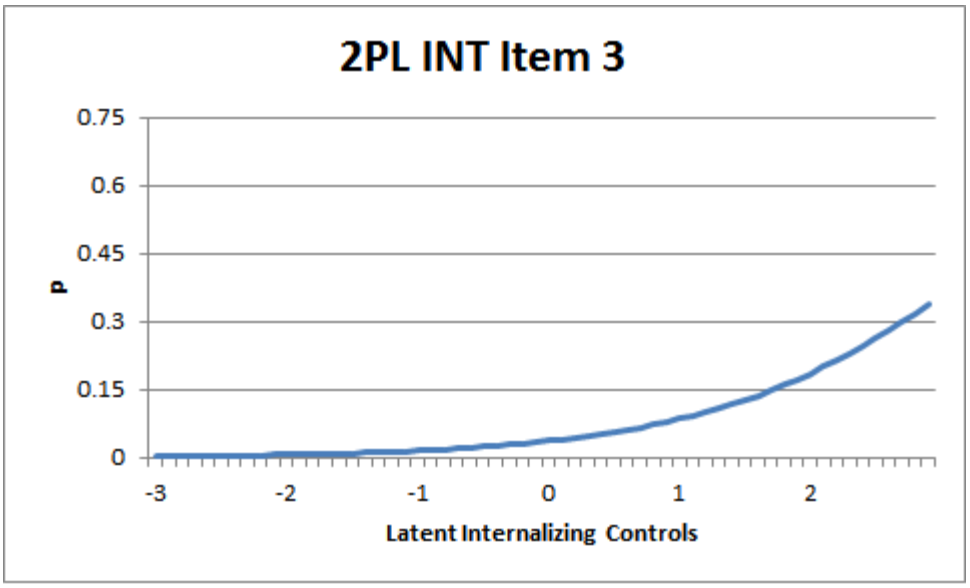


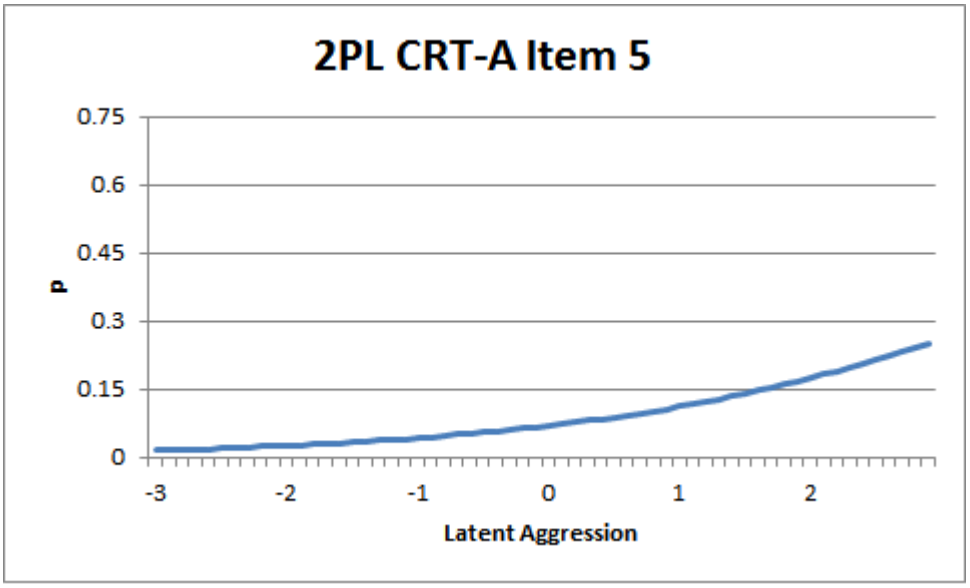
Figure 16. Two-Parameter Logistic Model item characteristic curve and information function for item 3 of the Internalizing Controls subscale of the Conditional Reasoning Test of Aggression.

Overall, item four has strong item characteristics. Responses to item four are well suited to indicate both latent aggression and latent internalizing controls. There is no evidence of DIF, meaning that item four discriminates among student and applied respondents similarly.

Item 5

The fifth CRT-A item stem is "Most bosses do not like to criticize employees. It makes both the boss and the employee uneasy. Which of the following is the most logical explanation for the above?" The aggressive logical response option is "Bosses are afraid to criticize problem workers." The nonaggressive logical response option is "Bosses and employees like a friendly place to work." The two illogical options are "Annual performance reviews happen only once a year" and "Many companies now have no-smoking policies." Of the 5,511 responses, 7.77% chose the aggressive logical option, 90.66% chose the nonaggressive logical option, 1.15% chose one of the illogical options, and 0.07% of responses were missing. The item-total correlation for item five and the dichotomously scored CRT-A is 0.10.

The 2PL slope estimate is 0.52 with a standard error of 0.08. The data are consistent with the hypothesis in that the slope coefficient is statistically greater than zero and monotonically increases with estimates of latent aggression. The 2PL intercept estimate is 5.00 with a standard error of 0.70. Information monotonically increases up to $\theta = 2.9$, indicating that this item is best suited for individuals high on latent aggression. The 2PL ICC and information graph are presented in Figure 17.



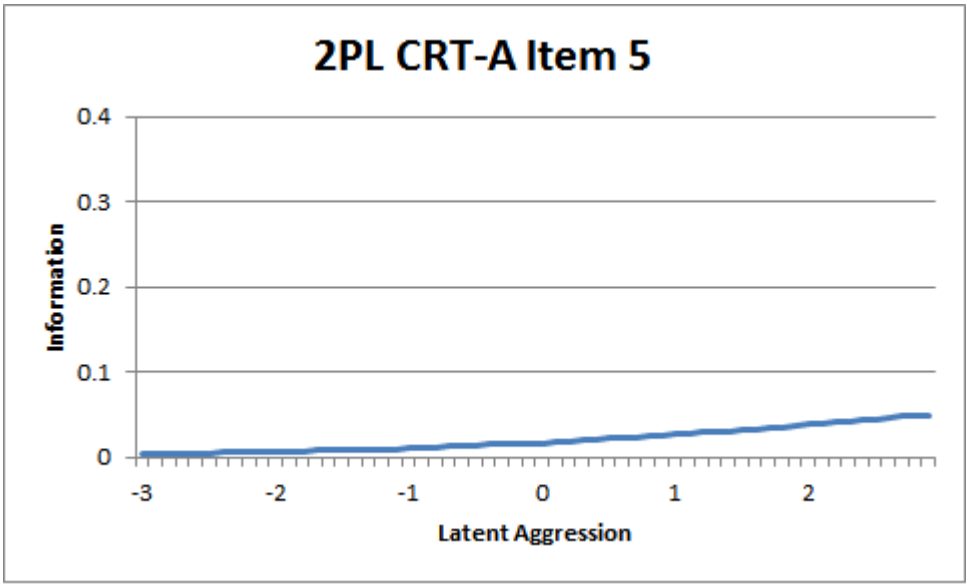
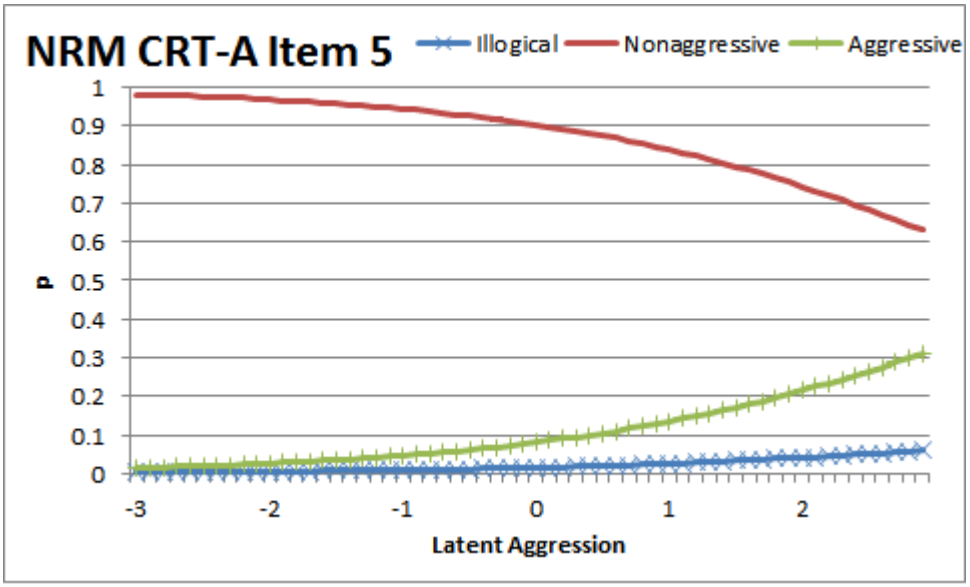


Figure 17. Two-Parameter Logistic Model item characteristic curve and information function for item 5 of the Conditional Reasoning Test of Aggression.

The linear NRM slope estimates are .047 for the aggressive logical response option, -.056 for the nonaggressive logical response option, and .009 for the combined illogical options. This is consistent with the expectation that the aggressive option will have a positive relationship with latent aggression while the nonaggressive option has a negative relationship with latent aggression. Information monotonically increases up to $\theta = 2.9$, indicating that this item is best suited for individuals high on latent aggression. The NRM ICC and information graph are presented in Figure 18.



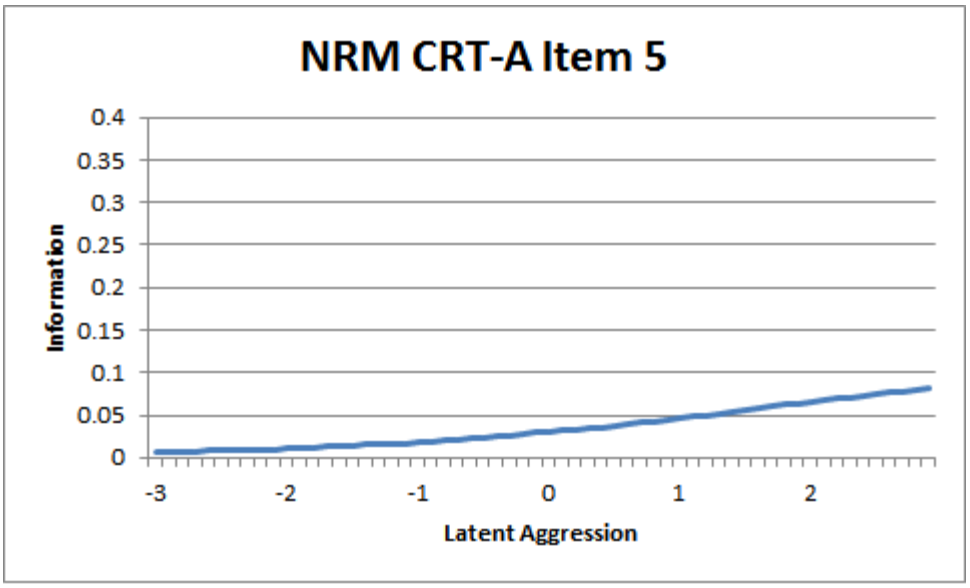


Figure 18. Nominal Response Model item characteristic curve and information function for item 5 of the Conditional Reasoning Test of Aggression.

Item five displays evidence of DIF (X^2, $2df = 10.4$, $p < .05$). This chi-square value can be partitioned into slope (X^2, $1df = 0.4$, $p > .05$) and intercept (X^2, $1df = 10.0$, $p < .05$) components, indicating that student and applied samples statistically differ in terms of intercept, but not slope. A graph displaying the ICCs for each group can be found in Figure 19.

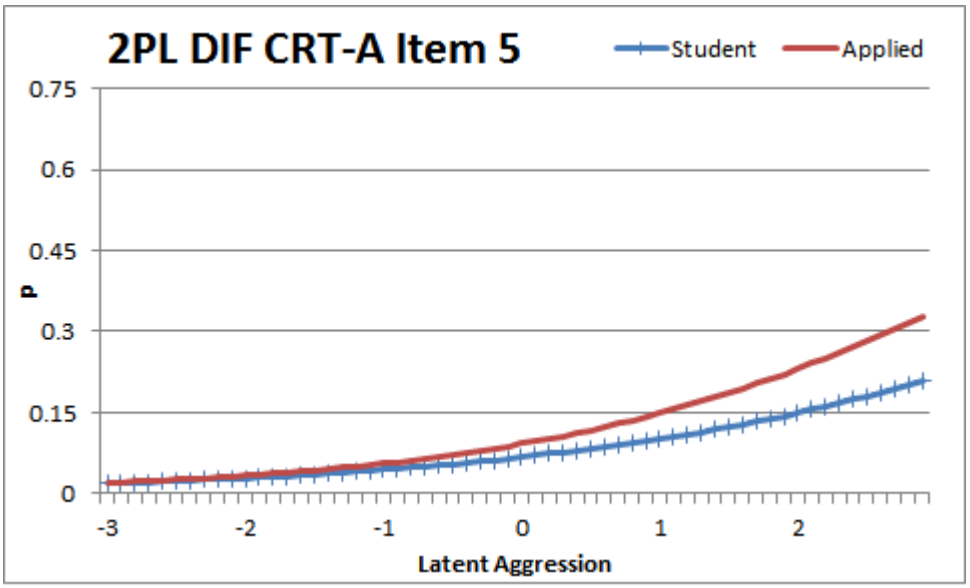


Figure 19. Two-Parameter Logistic Model differential item functioning item characteristic curve for item 5 of the Conditional Reasoning Test of Aggression.

The factor loadings for item five are 0.298 for externalizing controls, 0.366 for internalizing controls, and 0.162 for powerlessness, indicating that item five is primarily related to internalizing controls, but also related to externalizing controls and powerlessness. The factor-specific 2PL slope estimate for item five is 0.63 with a standard error of 0.11. The 2PL intercept estimate is 4.17 with a standard error of 0.64. This indicates that item five works well as an indicator of latent internalizing controls as well as latent aggression. The factor-specific 2PL ICC for item five can be found in Figure 20.

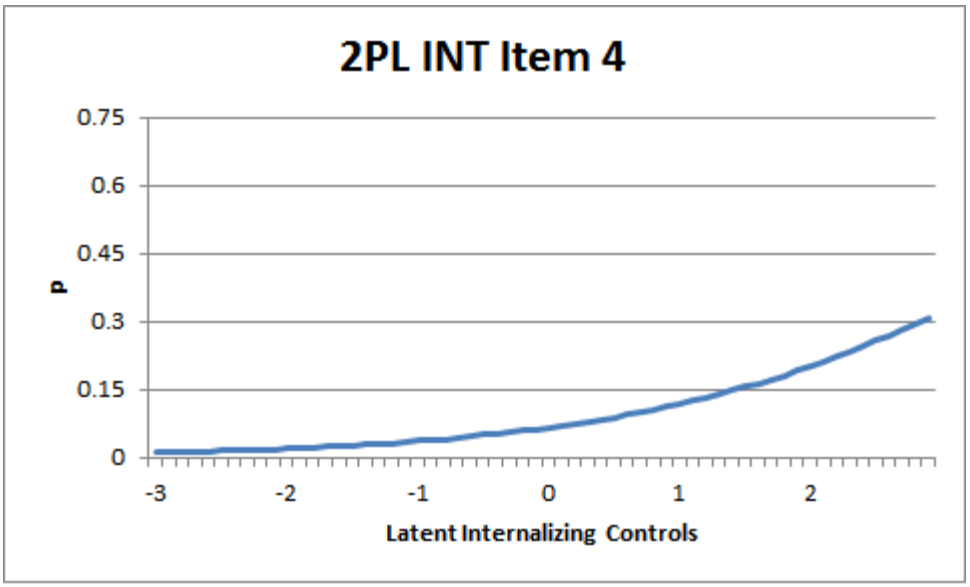


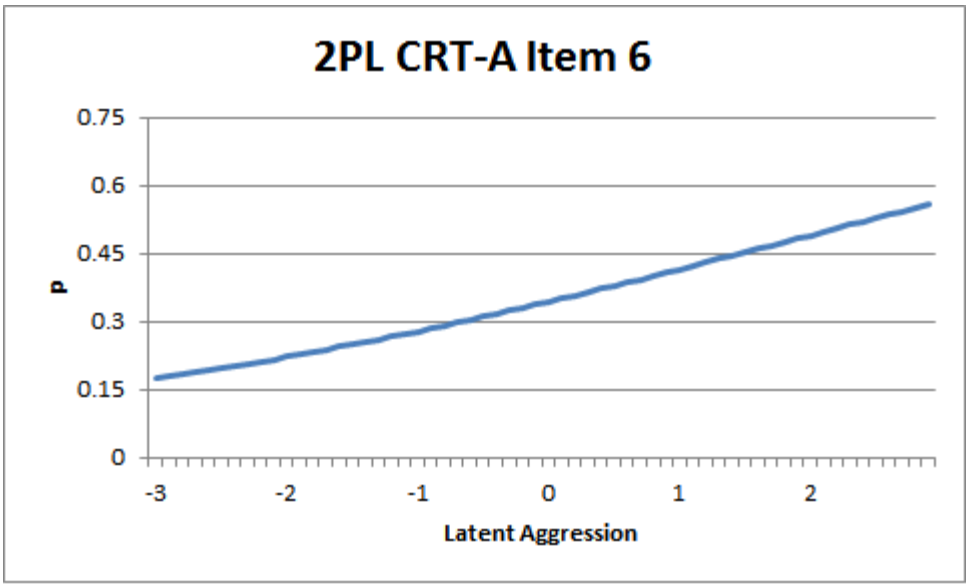
Figure 20. Two-Parameter Logistic Model item characteristic curve and information function for item 4 of the Internalizing Controls subscale of the Conditional Reasoning Test of Aggression.

Overall, item five has strong item characteristics. Responses to item five are well suited to indicate both latent aggression and latent internalizing controls. There is some evidence of DIF, but it is constrained to difference in intercept, meaning that item five discriminates among student and applied respondents similarly.

Item 6

The sixth CRT-A item stem is "New technology has changed the American workplace. A job that is here today could be gone tomorrow. People can no longer expect to work on the same job for very long. On the other hand, many new jobs are being created. Which of the following is the most logical conclusion based on the above?" The aggressive logical response option is "Trying to be steady and dependable will not be as important in future jobs." The nonaggressive logical response option is "People will spend more time in school learning new skills." The two illogical options are "More people will buy their homes rather than rent" and "The American workplace never changes." Of the 5,511 responses, 34.68% chose the aggressive logical option, 63.51% chose the nonaggressive logical option, 1.72% chose one of the illogical options, and 0.09% of responses were missing. The item-total correlation for item six and the dichotomously scored CRT-A is 0.09.

The 2PL slope estimate is 0.30 with a standard error of 0.05. The data are consistent with the hypothesis in that the slope coefficient is statistically greater than zero and monotonically increases with estimates of latent aggression. The 2PL intercept estimate is 2.13 with a standard error of 0.32. Information monotonically increases up to $\theta = 2.9$, indicating that this item is best suited for individuals high on latent aggression. The 2PL ICC and information graph are presented in Figure 21.



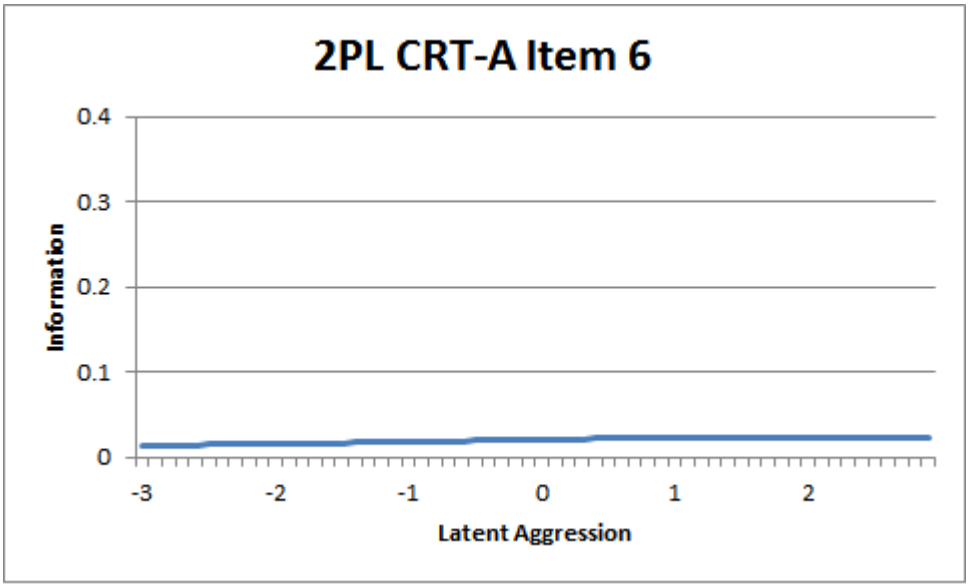


Figure 21. Two-Parameter Logistic Model item characteristic curve and information function for item 6 of the Conditional Reasoning Test of Aggression.

The linear NRM slope estimates are .080 for the aggressive logical response option, -.084 for the nonaggressive logical response option, and .004 for the combined illogical options. This is consistent with the expectation that the aggressive option will have a positive relationship with latent aggression while the nonaggressive option has a negative relationship with latent aggression. Information increases up to $\theta = 1.7$, after which it decreases through $\theta = 2.9$. This

indicates that this item is best suited for individuals moderately high to high on latent aggression. The NRM ICC and information graph are presented in Figure 22.

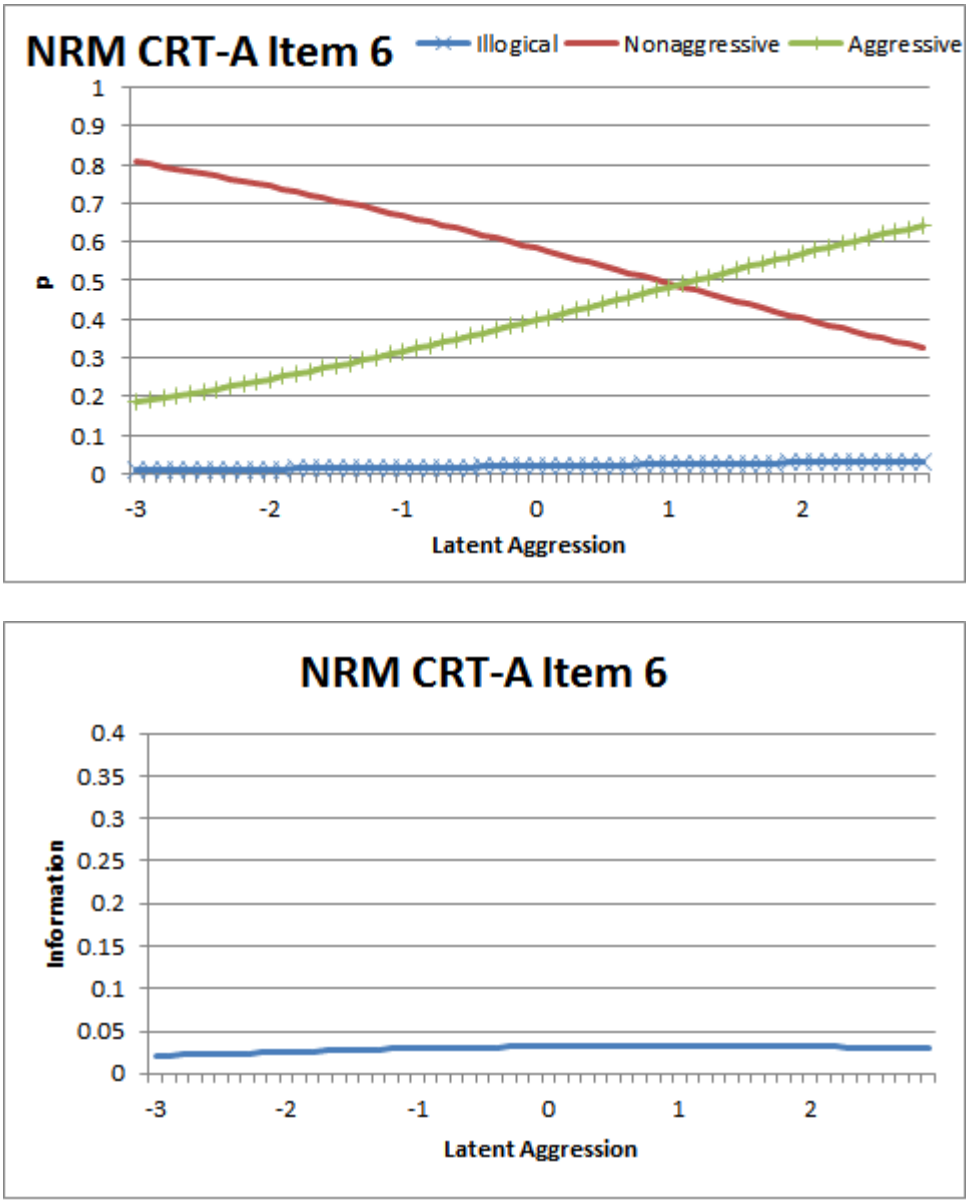


Figure 22. Nominal Response Model item characteristic curve and information function for item 6 of the Conditional Reasoning Test of Aggression.

Item six displays evidence of DIF (X^2, $2df = 36.2$, $p < .05$). This chi-square value can be partitioned into slope (X^2, $1df = 1.5$, $p > .05$) and intercept (X^2, $1df = 34.6$, $p < .05$) components,

indicating that student and applied samples statistically differ in terms of intercept, but not slope. A graph displaying the ICCs for each group can be found in Figure 23.

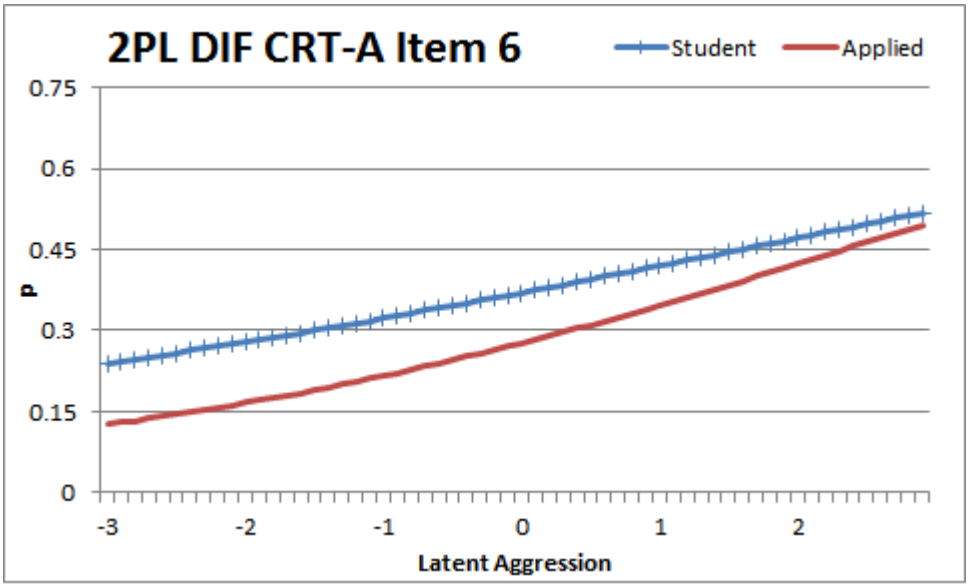


Figure 23. Two-Parameter Logistic Model differential item functioning item characteristic curve for item 6 of the Conditional Reasoning Test of Aggression.

The factor loadings for item six are 0.38 for externalizing controls, 1.89 for internalizing controls, and 2.81 for powerlessness, indicating that item six is primarily related to powerlessness, but also related to internalizing controls. The factor-specific 2PL slope estimate for item six is 0.48 with a standard error of 0.07. The 2PL intercept estimate is 1.39 with a standard error of 0.20. This indicates that item six works well as an indicator of latent powerlessness as well as latent aggression. The factor-specific 2PL ICC for item six can be found in Figure 24.

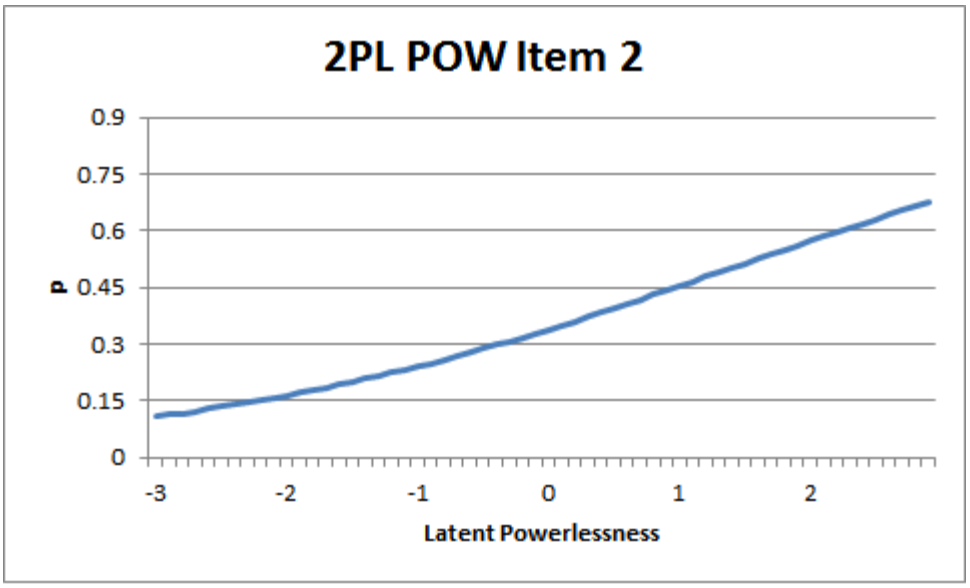


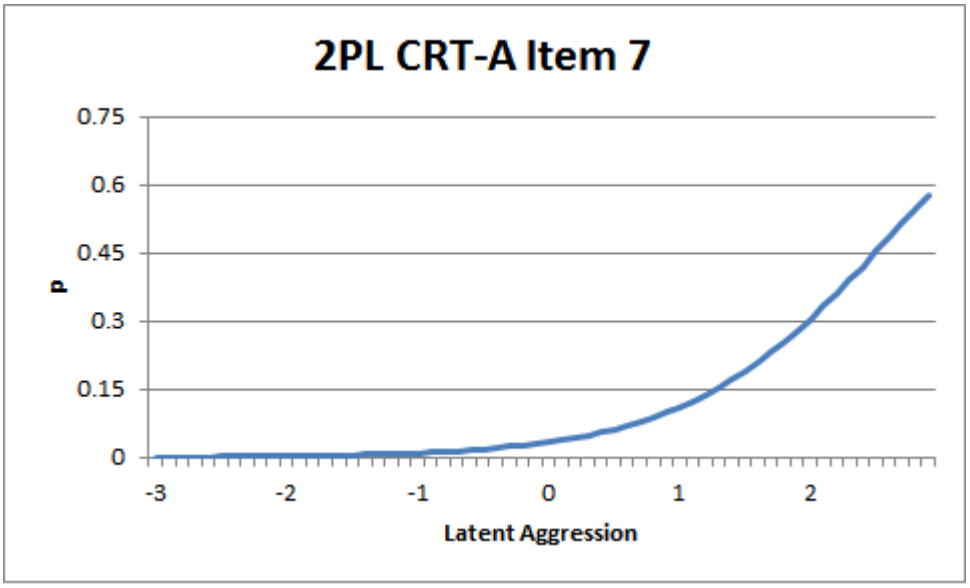
Figure 24. Two-Parameter Logistic Model item characteristic curve and information function for item 2 of the Powerlessness subscale of the Conditional Reasoning Test of Aggression.

Overall, item six has strong item characteristics. Responses to item six are well suited to indicate both latent aggression and latent powerlessness. There is some evidence of DIF, but it is constrained to difference in intercept, meaning that item six discriminates among student and applied respondents similarly.

Item 7

The seventh CRT-A item stem is "Girl Scouts and Boy Scouts teach young people a sense of discipline. They also teach respect for authority, neatness, dependability, and loyalty. Which of the following is the most logical prediction of what Scouts will be like when they grow up?" The aggressive logical response option is "They will be easily controlled by leaders." The nonaggressive logical response option is "They will be ready to take on responsibility." The two illogical options are "They will be reluctant to attend foreign films" and "They will be self-conscious about their height." Of the 5,511 responses, 6.21% chose the aggressive logical option, 93.49% chose the nonaggressive logical option, 0.27% chose one of the illogical options, and 0.04% of responses were missing. The item-total correlation for item seven and the dichotomously scored CRT-A is 0.12.

The 2PL slope estimate is 1.26 with a standard error of 0.11. The data are consistent with the hypothesis in that the slope coefficient is statistically greater than zero and monotonically increases with estimates of latent aggression. The 2PL intercept estimate is 2.66 with a standard error of 0.16. Information increases up to $\theta = 2.7$, after which it decreases through $\theta = 2.9$. This indicates that this item is best suited for individuals high on latent aggression. The 2PL ICC and information graph are presented in Figure 25.



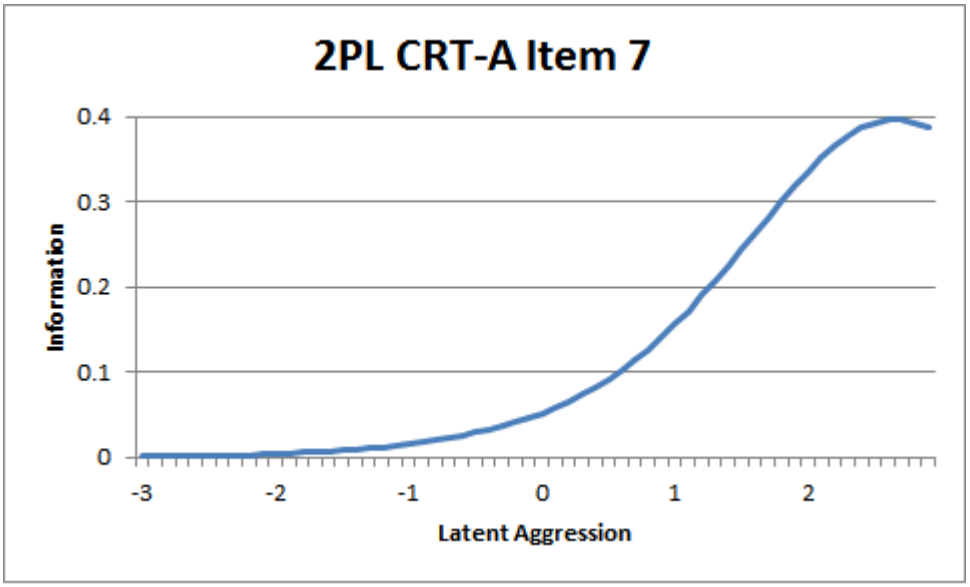
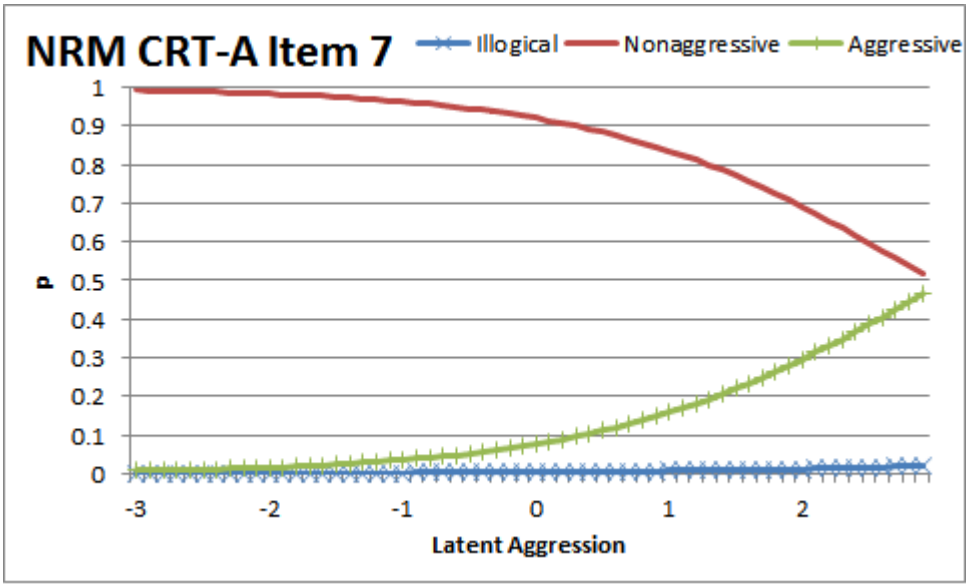


Figure 25. Two-Parameter Logistic Model item characteristic curve and information function for item 7 of the Conditional Reasoning Test of Aggression.

The linear NRM slope estimates are .071 for the aggressive logical response option, -.074 for the nonaggressive logical response option, and .003 for the combined illogical options. This is consistent with the expectation that the aggressive option will have a positive relationship with latent aggression while the nonaggressive option has a negative relationship with latent aggression. Information monotonically increases up to $\theta = 2.9$, indicating that this item is best suited for individuals high on latent aggression. The NRM ICC and information graph are presented in Figure 26.



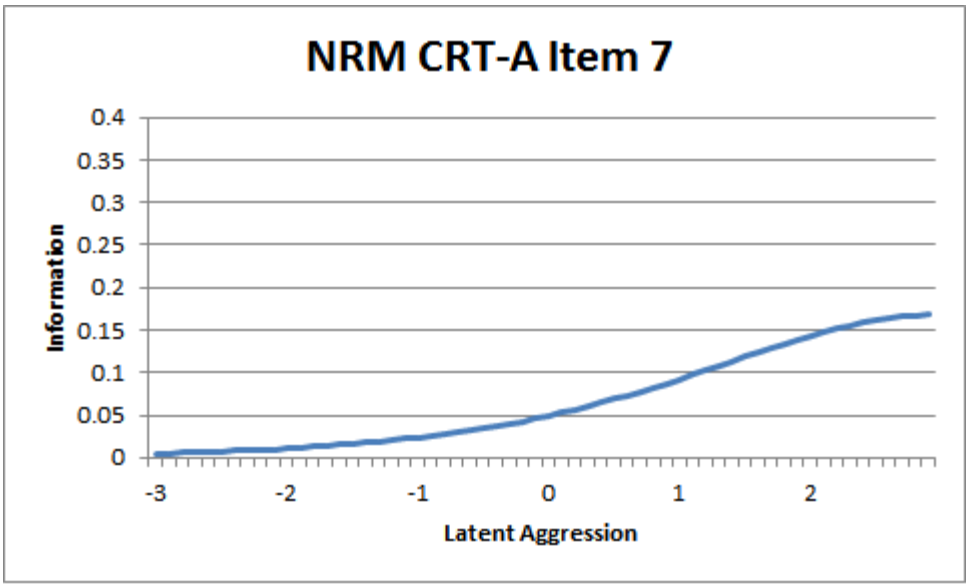


Figure 26. Nominal Response Model item characteristic curve and information function for item 7 of the Conditional Reasoning Test of Aggression.

Item seven displays evidence of DIF (X^2, $2df = 6.5$, $p < .05$). This chi-square value can be partitioned into slope (X^2, $1df = 0.1$, $p > .05$) and intercept (X^2, $1df = 6.3$, $p < .05$) components, indicating that student and applied samples statistically differ in terms of intercept, but not slope. A graph displaying the ICCs for each group can be found in Figure 27.

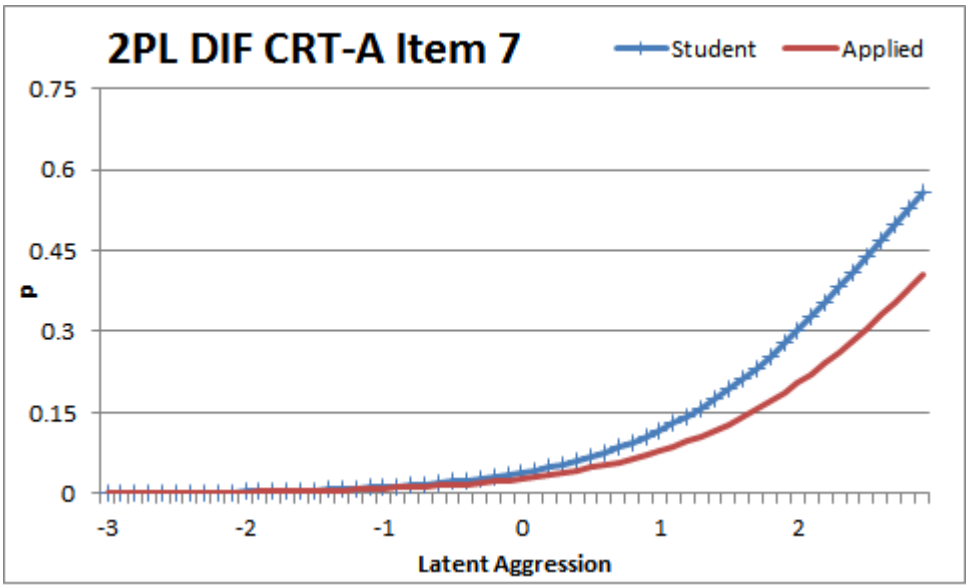


Figure 27. Two-Parameter Logistic Model differential item functioning item characteristic curve for item 7 of the Conditional Reasoning Test of Aggression.

The factor loadings for item seven are 0.449 for externalizing controls, 0.431 for internalizing controls, and 0.480 for powerlessness, indicating that item seven is primarily related to powerlessness, but also related to both externalizing and internalizing controls. The factor-specific 2PL slope estimate for item seven is 1.00 with a standard error of 0.16. The 2PL intercept estimate is 3.12 with a standard error of 0.37. This indicates that item seven works well as an indicator of latent powerlessness as well as latent aggression. The factor-specific 2PL ICC for item seven can be found in Figure 28.

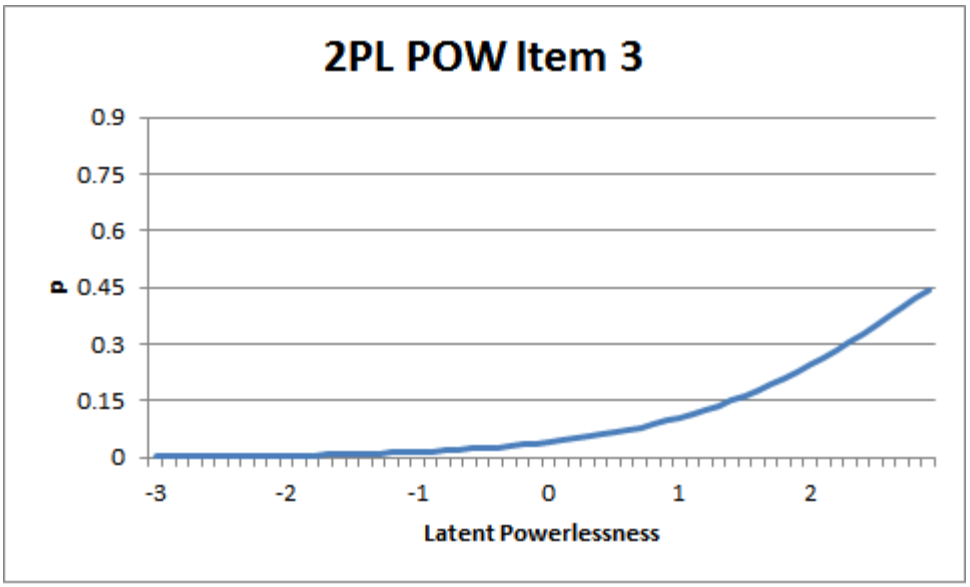


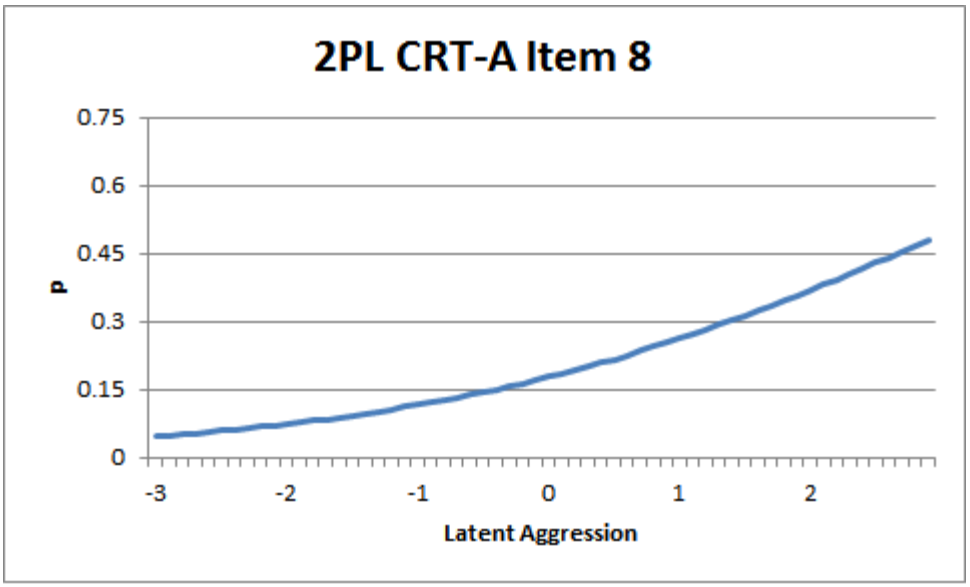
Figure 28. Two-Parameter Logistic Model item characteristic curve and information function for item 3 of the Powerlessness subscale of the Conditional Reasoning Test of Aggression.

Overall, item seven has strong item characteristics. Responses to item seven are well suited to indicate both latent aggression and latent powerlessness. There is some evidence of DIF, but it is constrained to difference in intercept, meaning that item seven discriminates among student and applied respondents similarly.

Item 8

The eighth CRT-A item stem is "People in a rich neighborhood in New York were pushed around for years by a homeless man. This man slept in alleys, stayed drunk or high on drugs, and cursed and threatened to hurt many of the residents. The police were called many times. But the homeless man always got a lawyer and returned to the neighborhood and caused trouble. Which of the following is the most logical conclusion regarding the people who lived in this neighborhood?" The aggressive logical response option is "They were afraid of the man, and would not fight back." The nonaggressive logical response option is "They did all that they could do within the law." The two illogical options are "They were used to dealing with the cold weather" and "They worked in New Jersey." Of the 5,511 responses, 18.91% chose the aggressive logical option, 80.51% chose the nonaggressive logical option, 0.51% chose one of the illogical options, and 0.07% of responses were missing. The item-total correlation for item eight and the dichotomously scored CRT-A is 0.12.

The 2PL slope estimate is 0.50 with a standard error of 0.06. The data are consistent with the hypothesis in that the slope coefficient is statistically greater than zero and monotonically increases with estimates of latent aggression. The 2PL intercept estimate is 3.08 with a standard error of 0.33. Information monotonically increases up to $\theta = 2.9$, indicating that this item is best suited for individuals high on latent aggression. The 2PL ICC and information graph are presented in Figure 29.



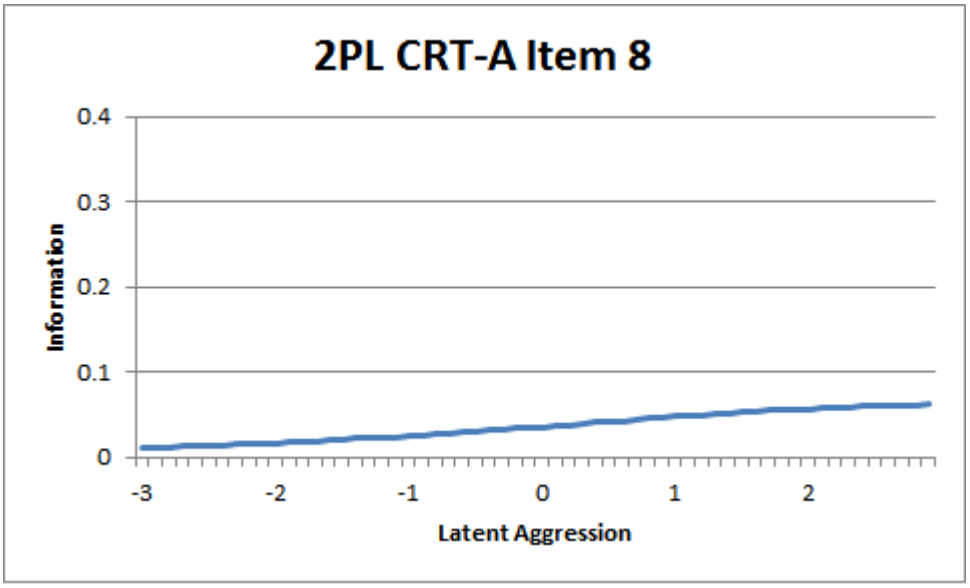


Figure 29. Two-Parameter Logistic Model item characteristic curve and information function for item 8 of the Conditional Reasoning Test of Aggression.

The linear NRM slope estimates are .110 for the aggressive logical response option, -.113 for the nonaggressive logical response option, and .003 for the combined illogical options. This is consistent with the expectation that the aggressive option will have a positive relationship with latent aggression while the nonaggressive option has a negative relationship with latent aggression. Information increases up to $\theta = 1.7$, after which it decreases through $\theta = 2.9$. This

indicates that this item is best suited for individuals moderately high to high on latent aggression. The NRM ICC and information graph are presented in Figure 30.

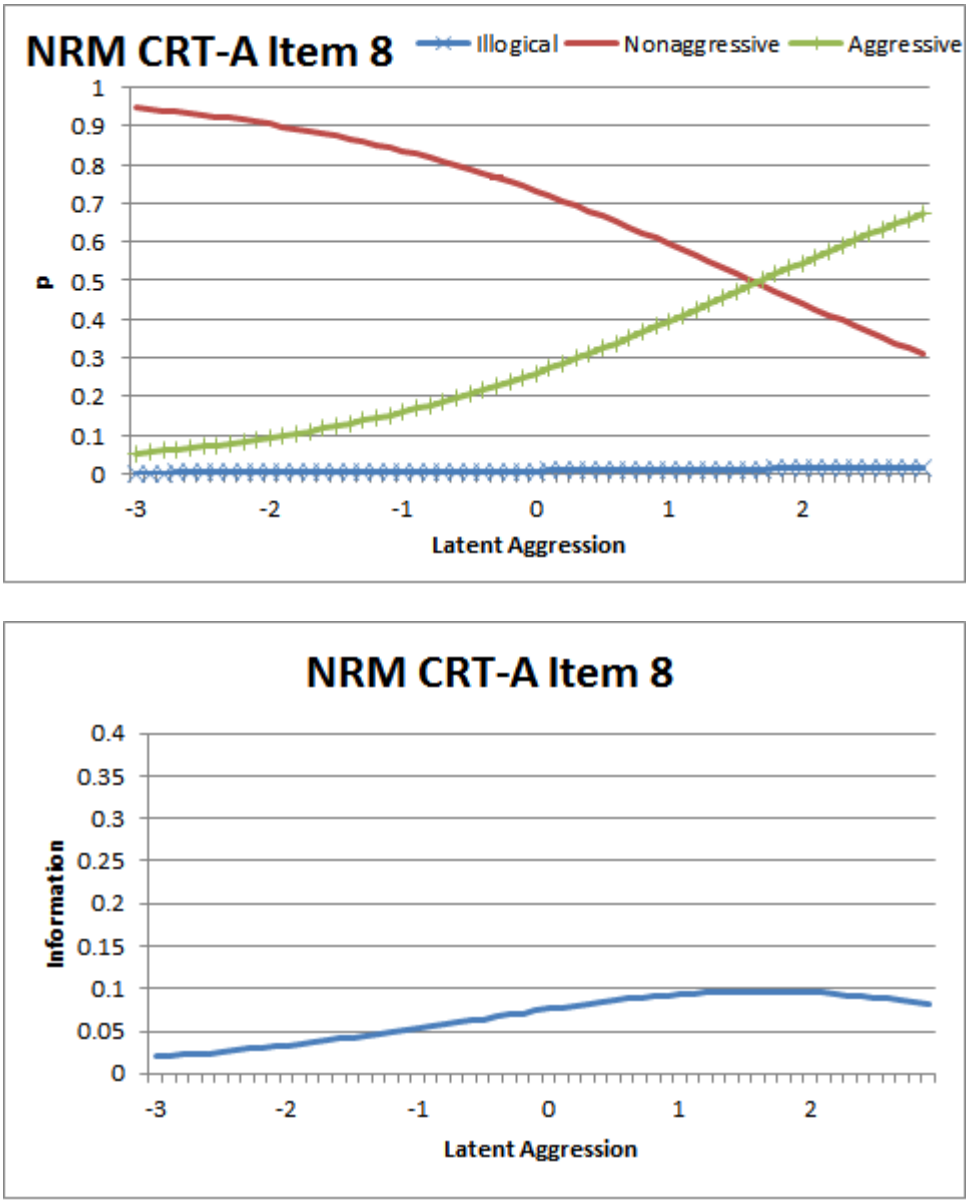


Figure 30. Nominal Response Model item characteristic curve and information function for item 8 of the Conditional Reasoning Test of Aggression.

Item eight displays evidence of DIF (X^2, $2df = 16.6$, $p < .05$). This chi-square value can be partitioned into slope (X^2, $1df = 2.1$, $p > .05$) and intercept (X^2, $1df = 14.5$, $p < .05$)

components, indicating that student and applied samples statistically differ in terms of intercept, but not slope. A graph displaying the ICCs for each group can be found in Figure 31.

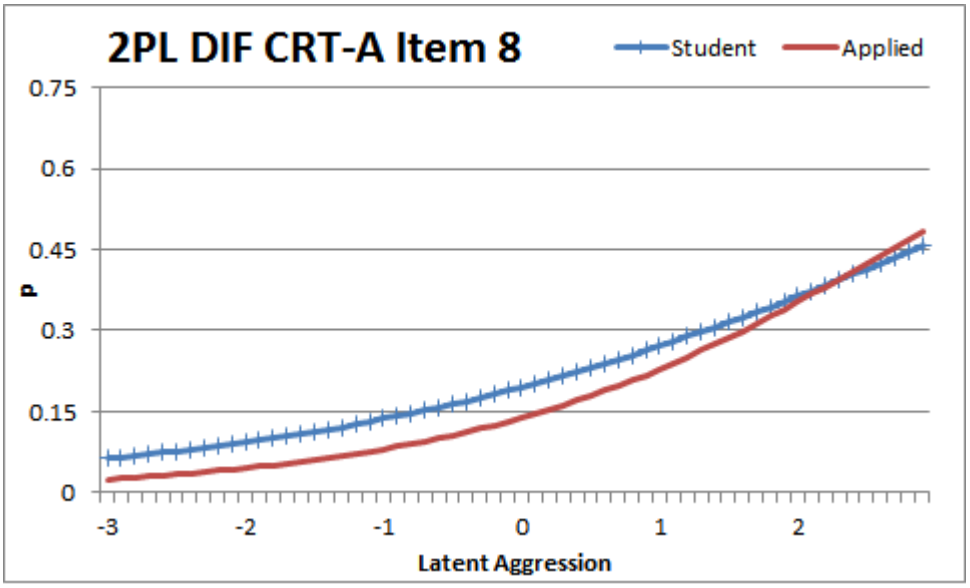


Figure 31. Two-Parameter Logistic Model differential item functioning item characteristic curve for item 8 of the Conditional Reasoning Test of Aggression.

The factor loadings for item eight are 0.319 for externalizing controls, 0.127 for internalizing controls, and 0.233 for powerlessness, indicating that item eight is primarily related to externalizing controls, but also related to internalizing controls and powerlessness. The factor-specific 2PL slope estimate for item eight is 0.62 with a standard error of 0.07. The 2PL intercept estimate is 2.54 with a standard error of 0.26. This indicates that item eight works well as an indicator of latent externalizing controls as well as latent aggression. The factor-specific 2PL ICC for item eight can be found in Figure 32.

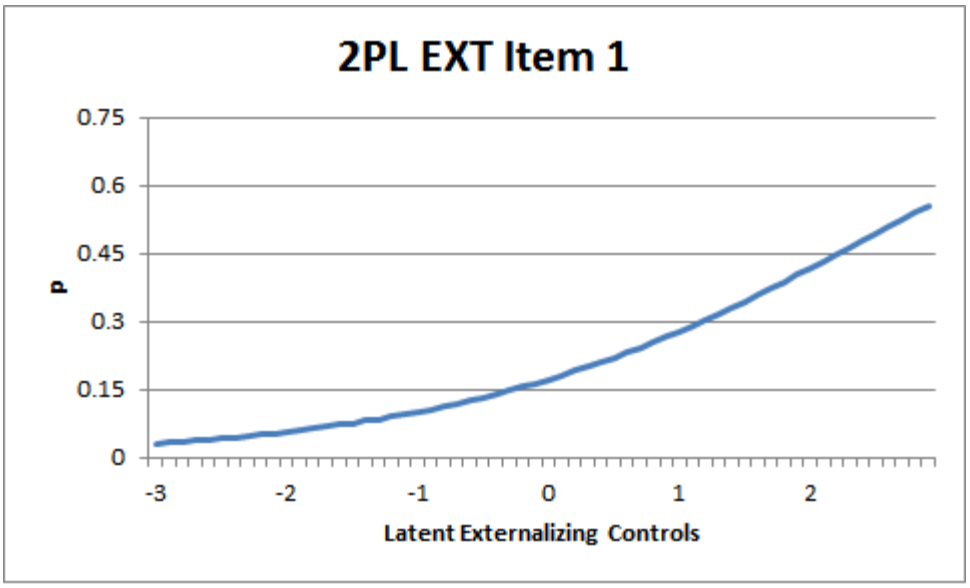


Figure 32. Two-Parameter Logistic Model item characteristic curve and information function for item 1 of the Externalizing Controls subscale of the Conditional Reasoning Test of Aggression.

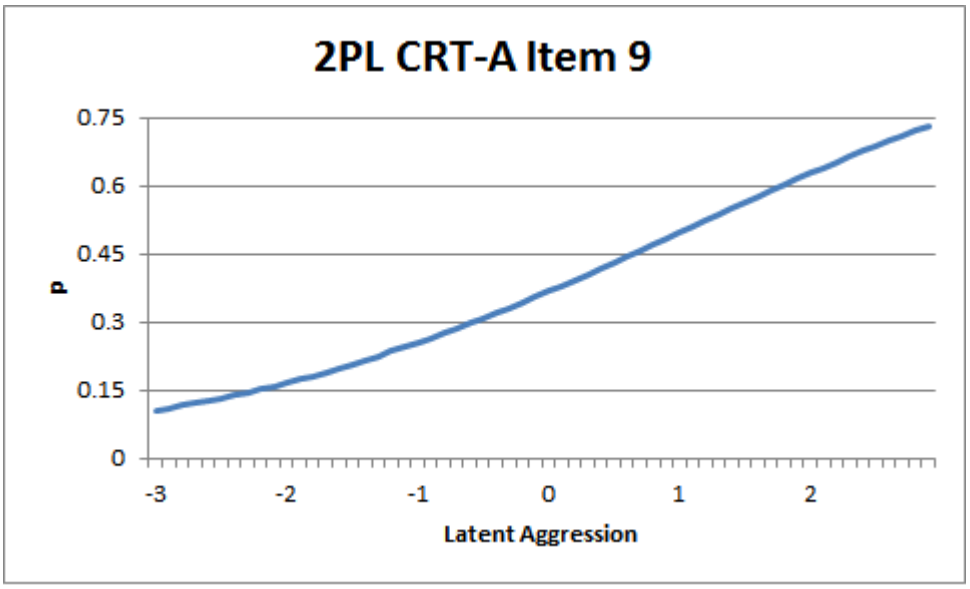
Overall, item eight has strong item characteristics. Responses to item eight are well suited to indicate both latent aggression and latent externalizing controls. There is some evidence of DIF, but it is constrained to difference in intercept, meaning that item eight discriminates among student and applied respondents similarly.

Item 9

The ninth CRT-A item stem is "Businesses say they want to give customers a good product at a low price. To keep costs down, companies have cut back to the smallest workforce possible. And the pay for most workers does not buy as much as it used to. Which of the following is the most logical conclusion based on the above?" The aggressive logical response option is "As long as their prices are low, companies don't care about the quality of life of their employees." The nonaggressive logical response option is "Getting customers depends on keeping costs low." The two illogical options are "Many companies pay employees monthly" and "Companies usually raise prices to attract customers." Of the 5,511 responses, 37.51% chose the aggressive logical option, 60.08% chose the nonaggressive

logical option, 2.32% chose one of the illogical options, and 0.09% of responses were missing. The item-total correlation for item nine and the dichotomously scored CRT-A is 0.16.

The 2PL slope estimate is 0.53 with a standard error of 0.05. The data are consistent with the hypothesis in that the slope coefficient is statistically greater than zero and monotonically increases with estimates of latent aggression. The 2PL intercept estimate is 1.02 with a standard error of 0.10. Information increases up to $\theta = 1.3$, after which it decreases through $\theta = 2.9$. This indicates that this item is best suited for individuals moderately high to high on latent aggression. The 2PL ICC and information graph are presented in Figure 33.



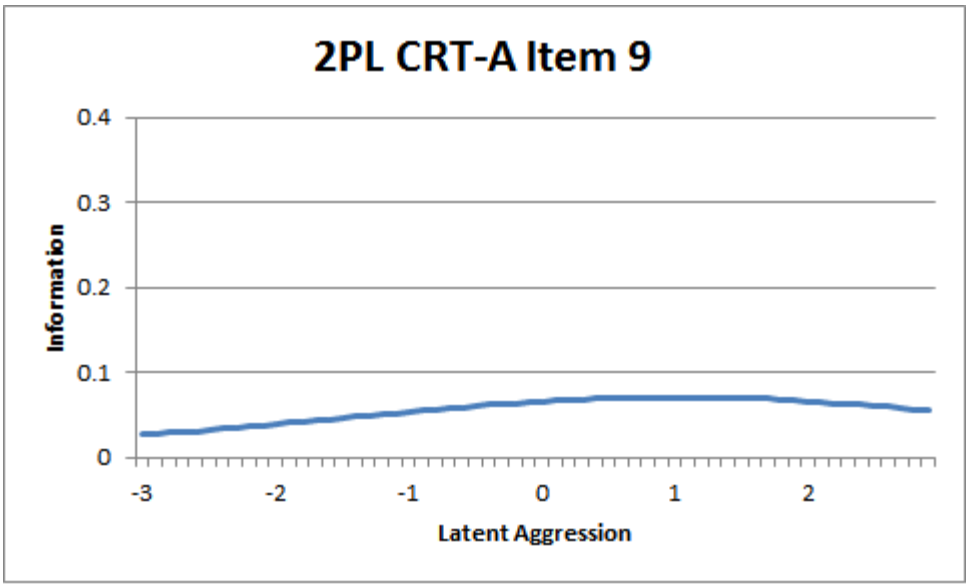
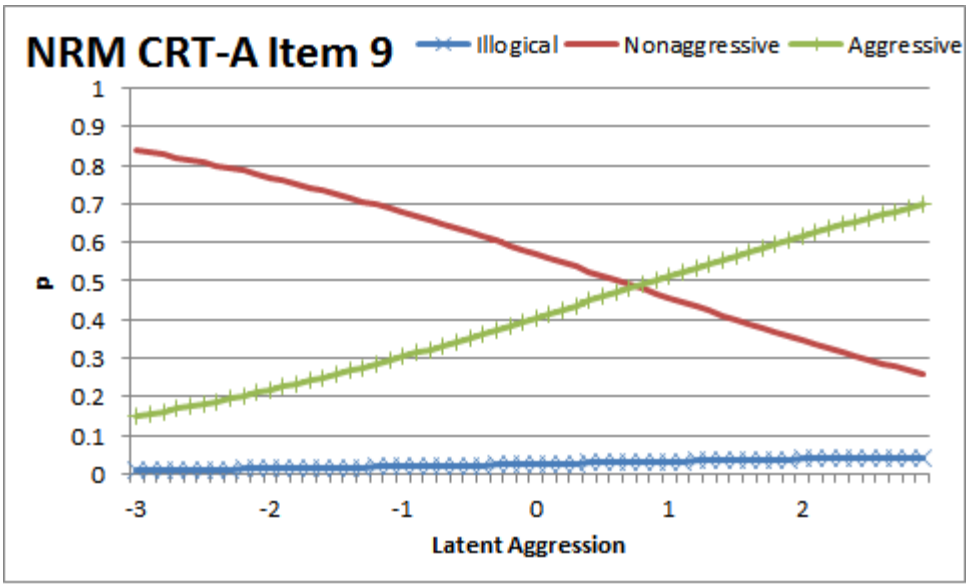


Figure 33. Two-Parameter Logistic Model item characteristic curve and information function for item 9 of the Conditional Reasoning Test of Aggression.

The linear NRM slope estimates are .097 for the aggressive logical response option, -.103 for the nonaggressive logical response option, and .006 for the combined illogical options. This is consistent with the expectation that the aggressive option will have a positive relationship with latent aggression while the nonaggressive option has a negative relationship with latent aggression. Information increases up to $\theta = 0.9$, after which it decreases through $\theta = 2.9$. This indicates that this item is best suited for individuals moderately high to high on latent aggression. The NRM ICC and information graph are presented in Figure 34.



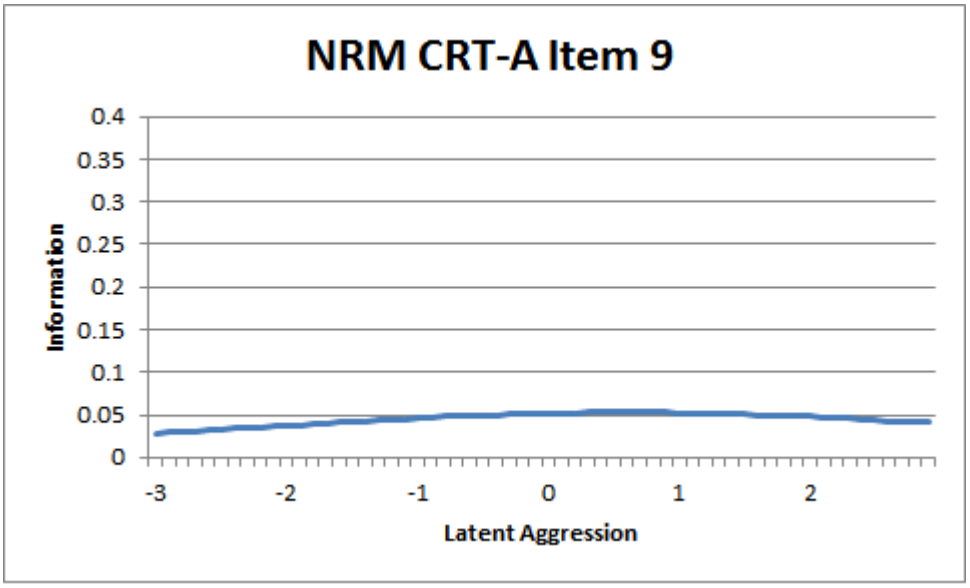


Figure 34. Nominal Response Model item characteristic curve and information function for item 9 of the Conditional Reasoning Test of Aggression.

Item nine displays no evidence of DIF (X^2, $2df = 1.7$, $p > .05$). This chi-square value can be partitioned into slope (X^2, $1df = 1.6$, $p > .05$) and intercept (X^2, $1df < 0.1$, $p > .05$) components, indicating that student and applied samples do not statistically differ in either slope or intercept. A graph displaying the ICCs for each group can be found in Figure 35.

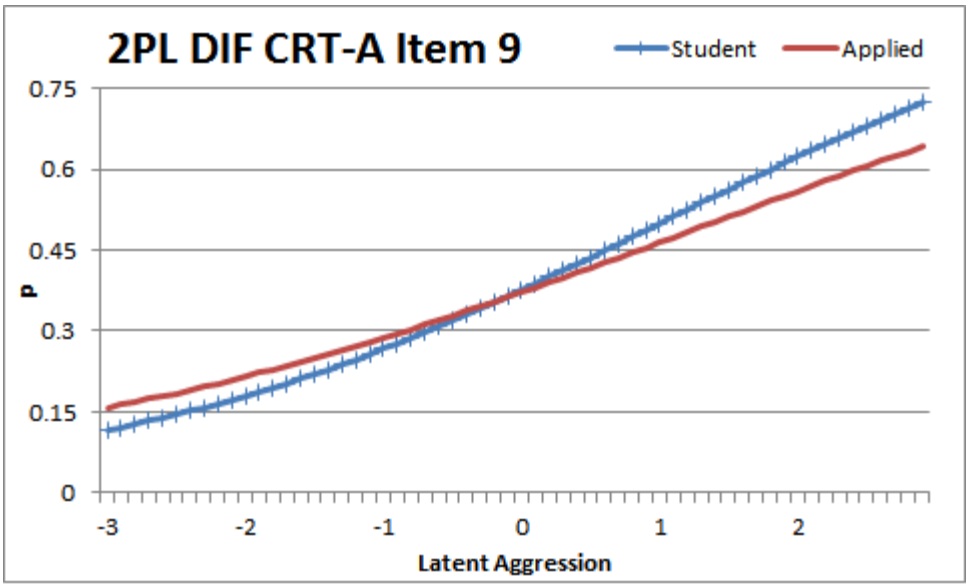


Figure 35. Two-Parameter Logistic Model differential item functioning item characteristic curve for item 9 of the Conditional Reasoning Test of Aggression.

The factor loadings for item nine are 0.128 for externalizing controls, 0.099 for internalizing controls, and 0.403 for powerlessness, indicating that item nine is primarily related to powerlessness, but also related to externalizing controls. The factor-specific 2PL slope estimate for item nine is 0.69 with a standard error of 0.10. The 2PL intercept estimate is 0.81 with a standard error of 0.10. This indicates that item nine works well as an indicator of latent powerlessness as well as latent aggression. The factor-specific 2PL ICC for item nine can be found in Figure 36.

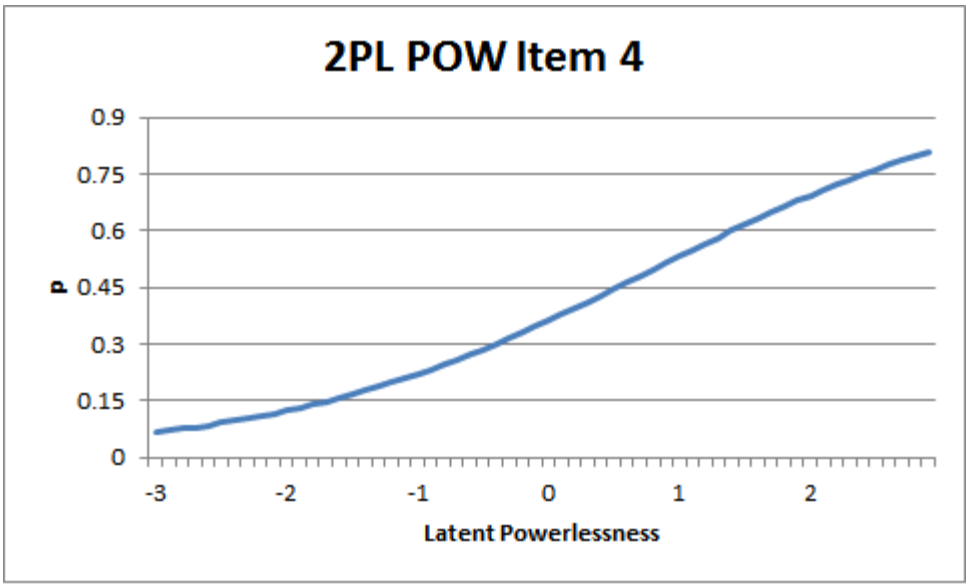


Figure 36. Two-Parameter Logistic Model item characteristic curve and information function for item 4 of the Powerlessness subscale of the Conditional Reasoning Test of Aggression.

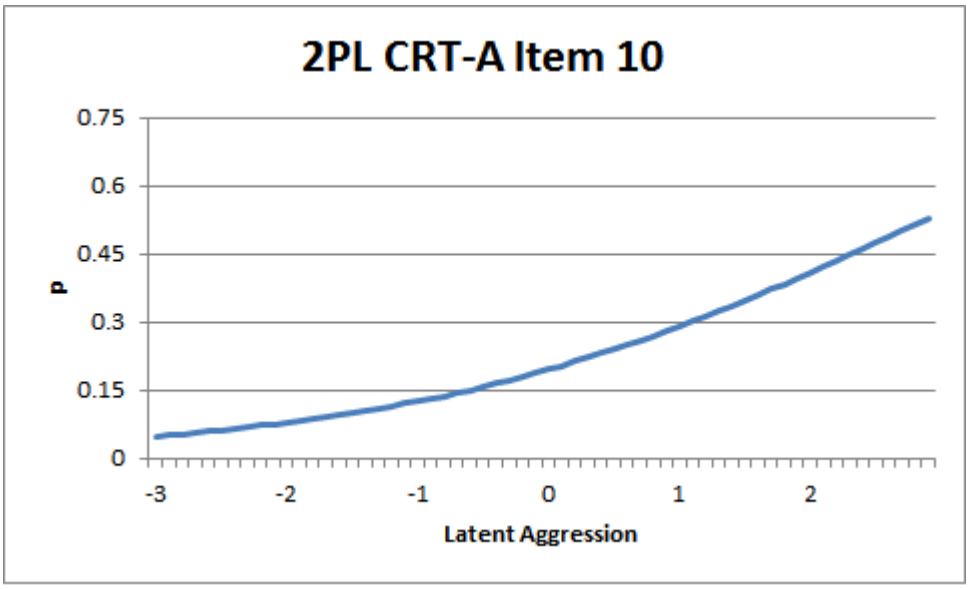
Overall, item nine has strong item characteristics. Responses to item nine are well suited to indicate both latent aggression and latent powerlessness. There is no evidence of DIF, meaning that item nine discriminates among student and applied respondents similarly.

Item 10

The tenth CRT-A item stem is "100 years ago, male college students often fought duels with swords. One or both fighters were cut. Some people argued that duels should be outlawed. Other people stood up for dueling. They said that duels were a good way to pick out leaders who were brave and strong. In those days, leaders in the military and business often had dueling scars. Ultimately, however, duels were outlawed. Which of the following is the most logical conclusion based on the above?" The aggressive logical response option is "Without duels, it became harder to identify good leaders." The nonaggressive logical response option is "Colleges wanted to be known as places of learning rather than fighting." The two illogical options are "Guns made duels less dangerous" and "People interested in business stopped attending college." Of the 5,511 responses, 20.79% chose the aggressive logical option, 78.01% chose the nonaggressive logical option, 1.05% chose one of the illogical options,

and 0.15% of responses were missing. The item-total correlation for item ten and the dichotomously scored CRT-A is 0.14.

The 2PL slope estimate is 0.52 with a standard error of 0.06. The data are consistent with the hypothesis in that the slope coefficient is statistically greater than zero and monotonically increases with estimates of latent aggression. The 2PL intercept estimate is 2.70 with a standard error of 0.27. Information monotonically increases up to $\theta = 2.9$, indicating that this item is best suited for individuals high on latent aggression. The 2PL ICC and information graph are presented in Figure 37.



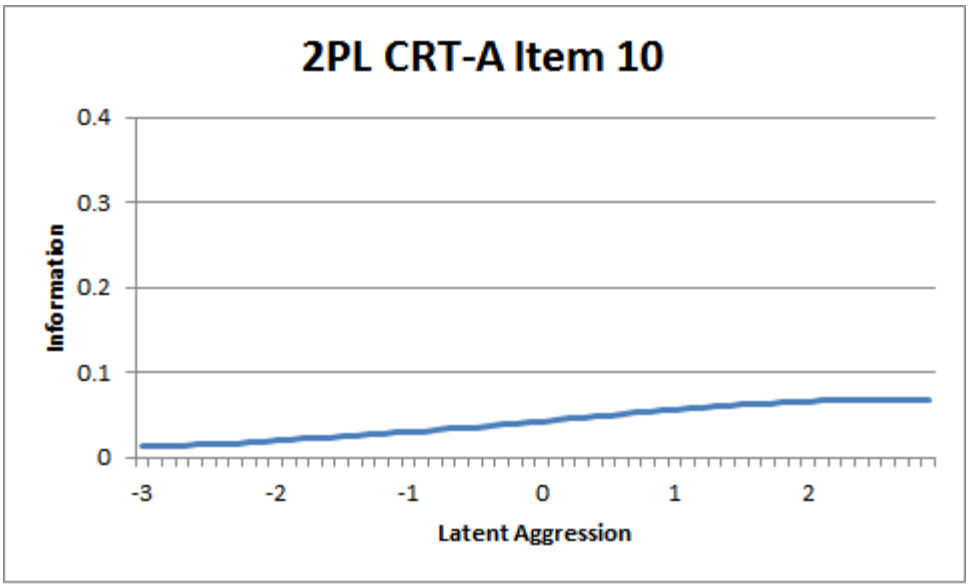
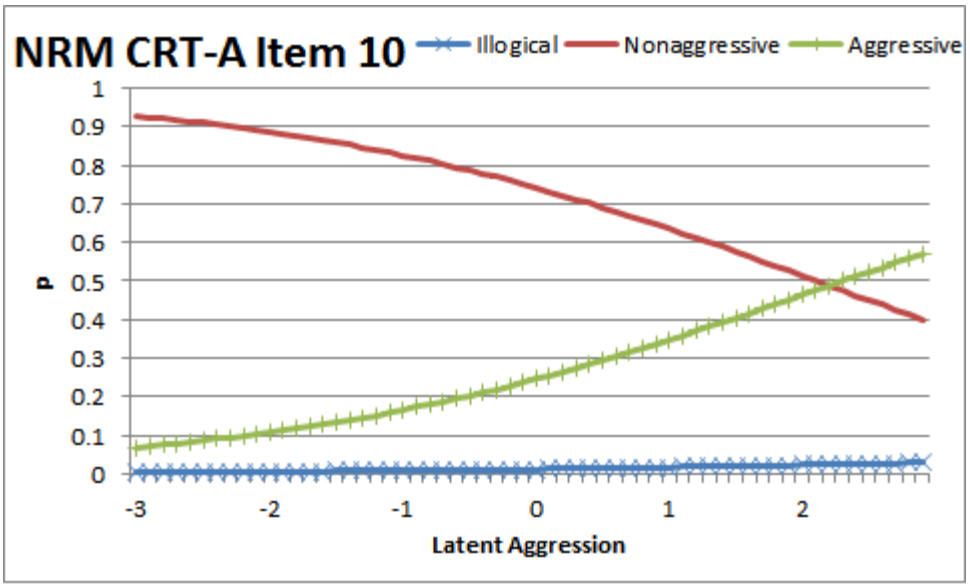


Figure 37. Two-Parameter Logistic Model item characteristic curve and information function for item 10 of the Conditional Reasoning Test of Aggression.

The linear NRM slope estimates are .087 for the aggressive logical response option, -.092 for the nonaggressive logical response option, and .004 for the combined illogical options. This is consistent with the expectation that the aggressive option will have a positive relationship with latent aggression while the nonaggressive option has a negative relationship with latent aggression. Information increases up to $\theta = 2.4$, after which it decreases through $\theta = 2.9$. This indicates that this item is best suited for individuals high on latent aggression. The NRM ICC and information graph are presented in Figure 38.



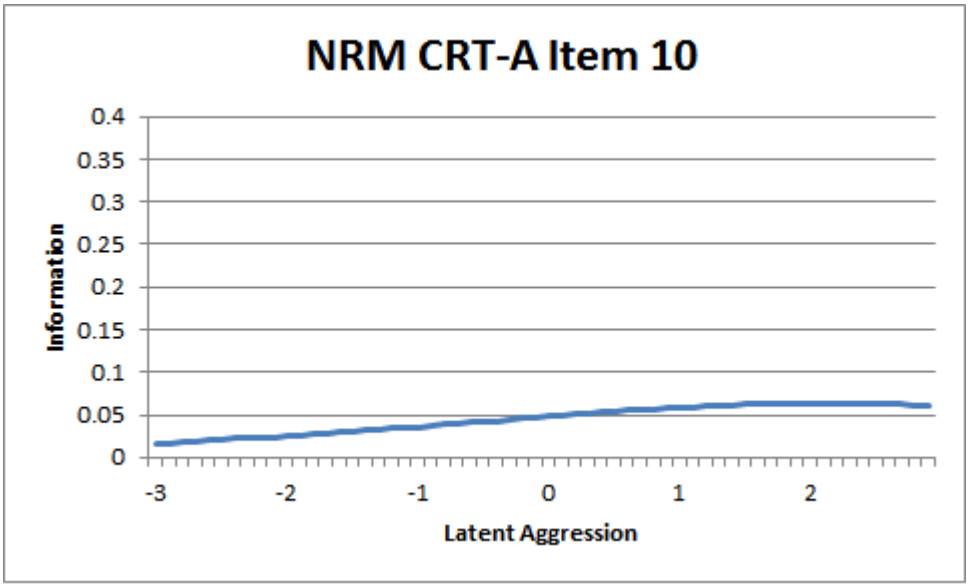


Figure 38. Nominal Response Model item characteristic curve and information function for item 10 of the Conditional Reasoning Test of Aggression.

Item ten displays no evidence of overall DIF (X^2, $2df = 4.4$, $p > .05$), although after partitioning the chi-square into slope (X^2, $1df < 0.1$, $p > .05$) and intercept (X^2, $1df = 4.4$, $p < .05$) components, results indicate that student and applied samples statistically differ in terms of intercept, but not slope. A graph displaying the ICCs for each group can be found in Figure 39.

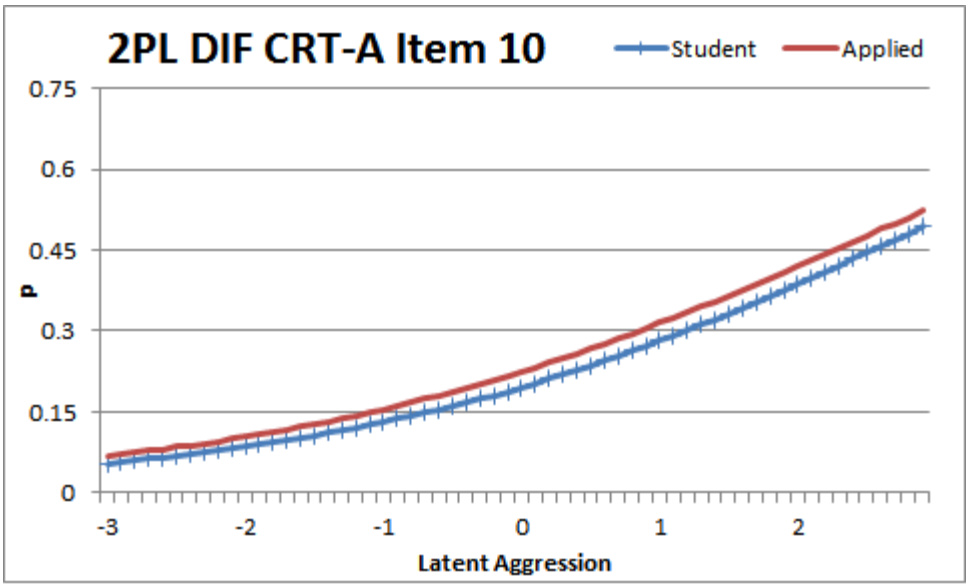


Figure 39. Two-Parameter Logistic Model differential item functioning item characteristic curve for item 10 of the Conditional Reasoning Test of Aggression.

The factor loadings for item ten are 0.114 for externalizing controls, 0.432 for internalizing controls, and 0.195 for powerlessness, indicating that item ten is primarily related to internalizing controls, but also related to externalizing controls and powerlessness. The factor-specific 2PL slope estimate for item ten is 0.90 with a standard error of 0.12. The 2PL intercept estimate is 1.72 with a standard error of 0.18. This indicates that item ten works well as an indicator of latent internalizing controls as well as latent aggression. The factor-specific 2PL ICC for item ten can be found in Figure 40.

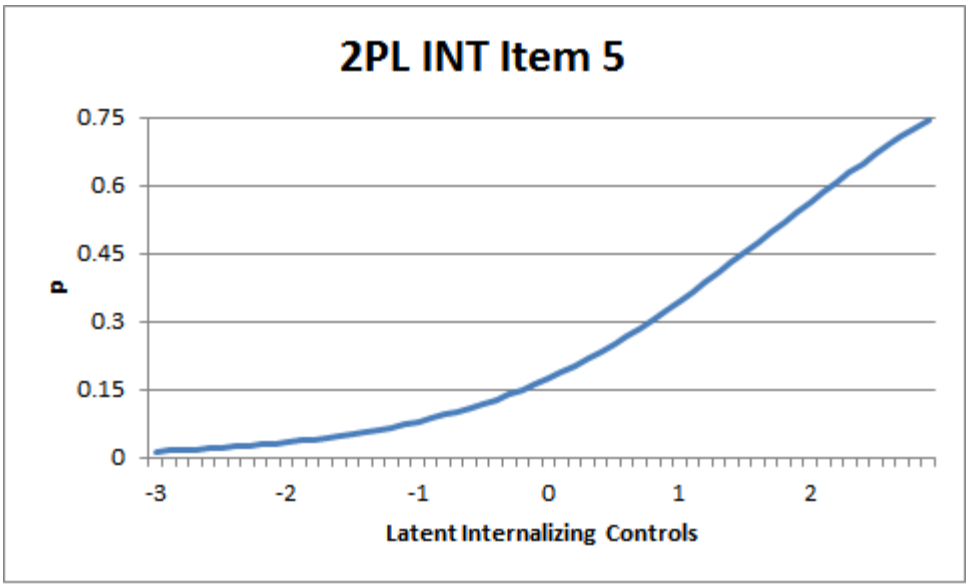


Figure 40. Two-Parameter Logistic Model item characteristic curve and information function for item 5 of the Internalizing Controls subscale of the Conditional Reasoning Test of Aggression.

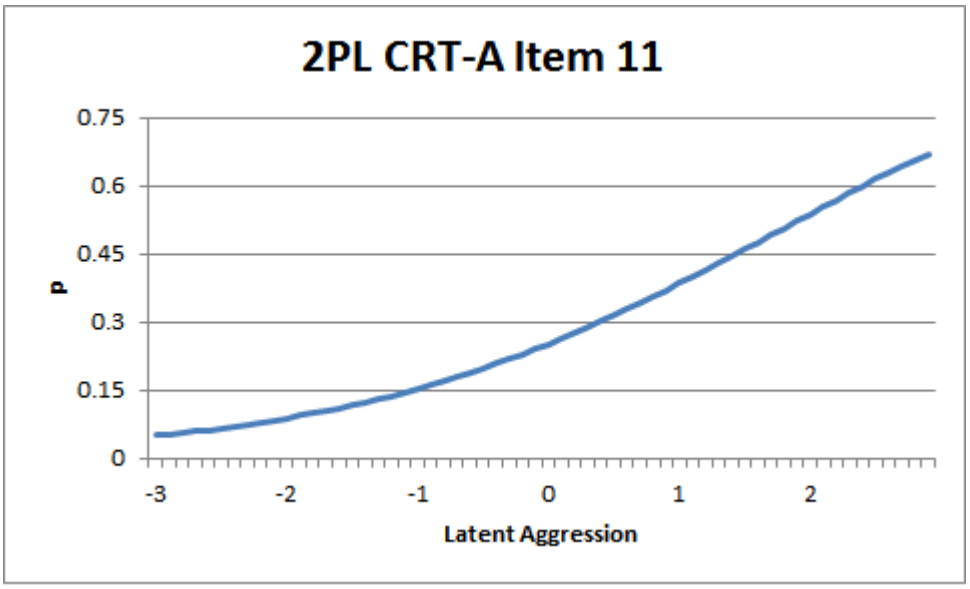
Overall, item ten has strong item characteristics. Responses to item ten are well suited to indicate both latent aggression and latent internalizing controls. There is some evidence of DIF, but it is constrained to difference in intercept, meaning that item ten discriminates among student and applied respondents similarly.

Item 11

The eleventh CRT-A item stem is "Doreen has noticed that a new girl at her high school has been looking at her from across the cafeteria. The new girl is like Doreen in many ways. She is pretty, wears nice clothes, cuts her hair short, and seems to get along with both girls and boys. Doreen notices that the new girl is checking out who Doreen's friends are and how Doreen acts around boys. Which of the following is the most logical conclusion based on the above?" The aggressive logical response option is "The new girl is checking Doreen out as a likely rival." The nonaggressive logical response option is "The new girl may become friends with Doreen." The two illogical options are "The new girl is planning on joining the soccer team" and "Doreen has algebra during second period." Of the 5,511 responses, 26.82% chose the aggressive logical option, 72.84% chose the nonaggressive logical

option, 0.25% chose one of the illogical options, and 0.09% of responses were missing. The item-total correlation for item eleven and the dichotomously scored CRT-A is 0.17.

The 2PL slope estimate is 0.62 with a standard error of 0.06. The data are consistent with the hypothesis in that the slope coefficient is statistically greater than zero and monotonically increases with estimates of latent aggression. The 2PL intercept estimate is 1.75 with a standard error of 0.14. Information increases up to $\theta = 1.9$, after which it decreases through $\theta = 2.9$. This indicates that this item is best suited for individuals moderately high to high on latent aggression. The 2PL ICC and information graph are presented in Figure 41.



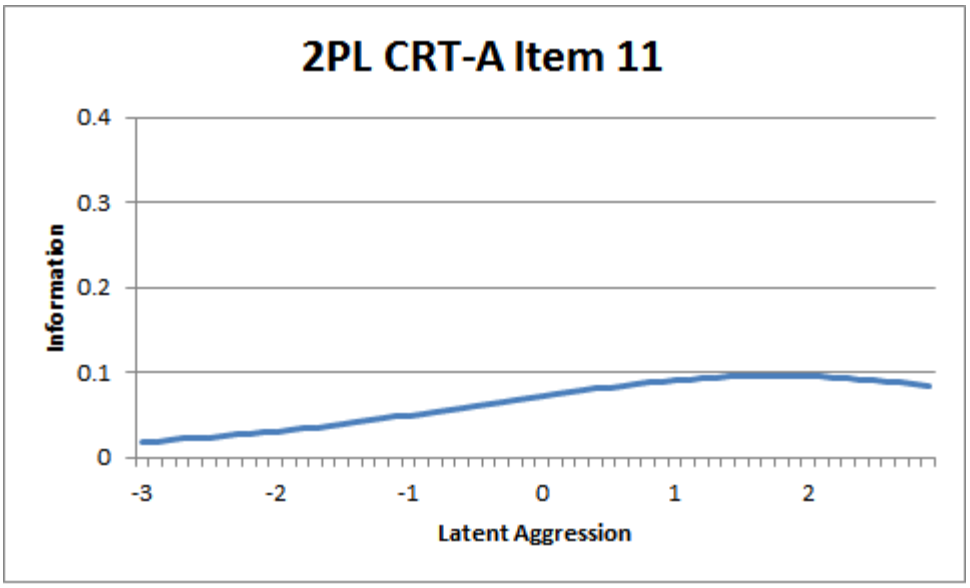
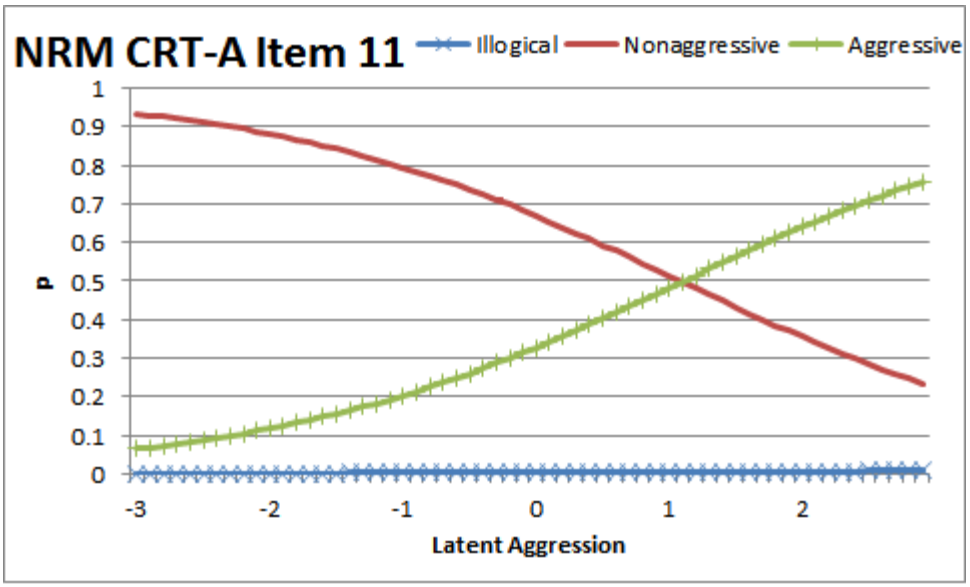


Figure 41. Two-Parameter Logistic Model item characteristic curve and information function for item 11 of the Conditional Reasoning Test of Aggression.

The linear NRM slope estimates are .126 for the aggressive logical response option, -.127 for the nonaggressive logical response option, and .001 for the combined illogical options. This is consistent with the expectation that the aggressive option will have a positive relationship with latent aggression while the nonaggressive option has a negative relationship with latent aggression. Information increases up to $\theta = 1.2$, after which it decreases through $\theta = 2.9$. This indicates that this item is best suited for individuals moderately high to high on latent aggression. The NRM ICC and information graph are presented in Figure 42.



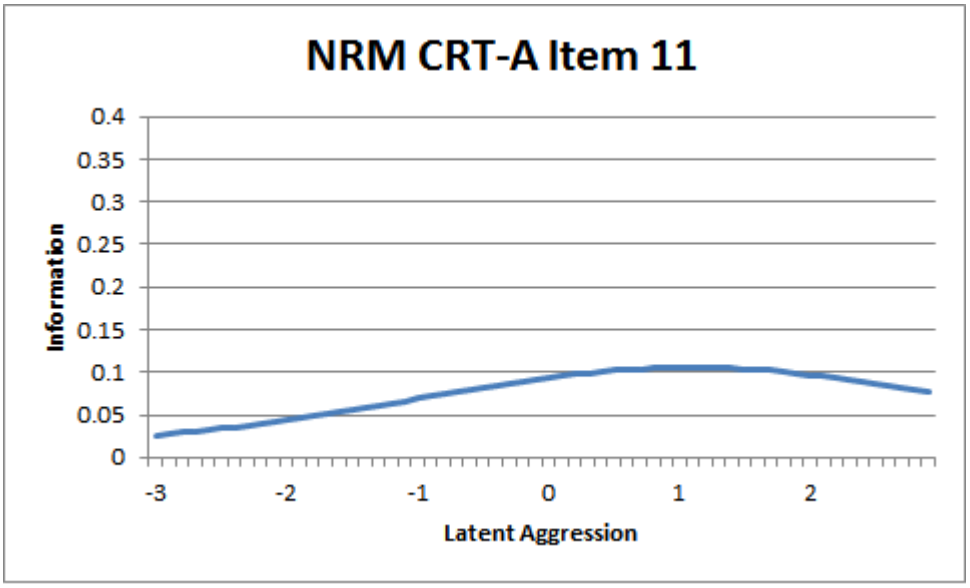


Figure 42. Nominal Response Model item characteristic curve and information function for item 11 of the Conditional Reasoning Test of Aggression.

Item eleven displays no evidence of DIF (X^2, $2df = 1.5$, $p > .05$). This chi-square value can be partitioned into slope (X^2, $1df = 1.2$, $p > .05$) and intercept (X^2, $1df = 0.3$, $p > .05$) components, indicating that student and applied samples do not statistically differ in either slope or intercept. A graph displaying the ICCs for each group can be found in Figure 43.

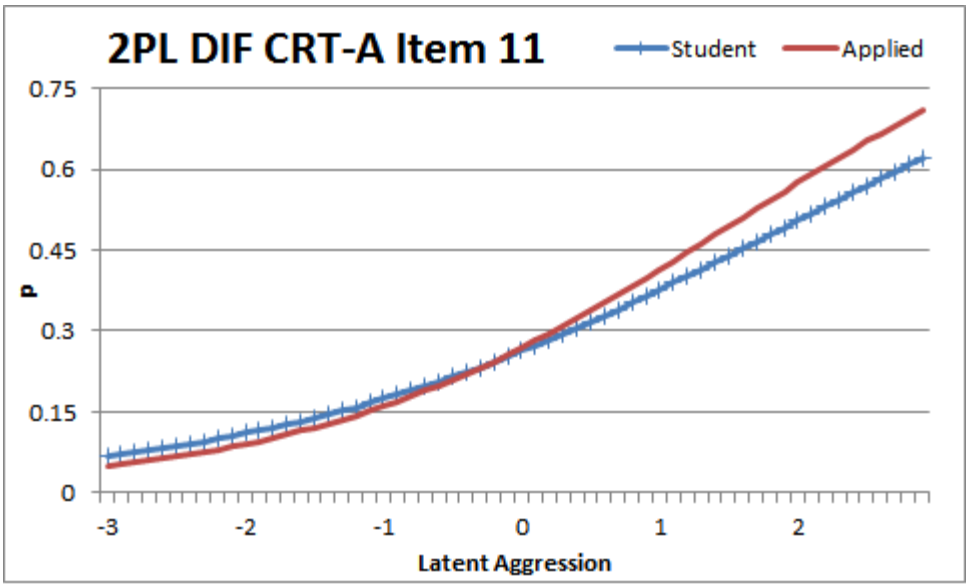


Figure 43. Two-Parameter Logistic Model differential item functioning item characteristic curve for item 11 of the Conditional Reasoning Test of Aggression.

The factor loadings for item eleven are 0.282 for externalizing controls, 0.193 for internalizing controls, and 0.292 for powerlessness, indicating that item eleven is primarily related to powerlessness, but also related to externalizing controls and internalizing controls. The factor-specific 2PL slope estimate for item eleven is 0.60 with a standard error of 0.09. The 2PL intercept estimate is 1.81 with a standard error of 0.24. This indicates that item eleven works well as an indicator of latent powerlessness as well as latent aggression. The factor-specific 2PL ICC for item eleven can be found in Figure 44.

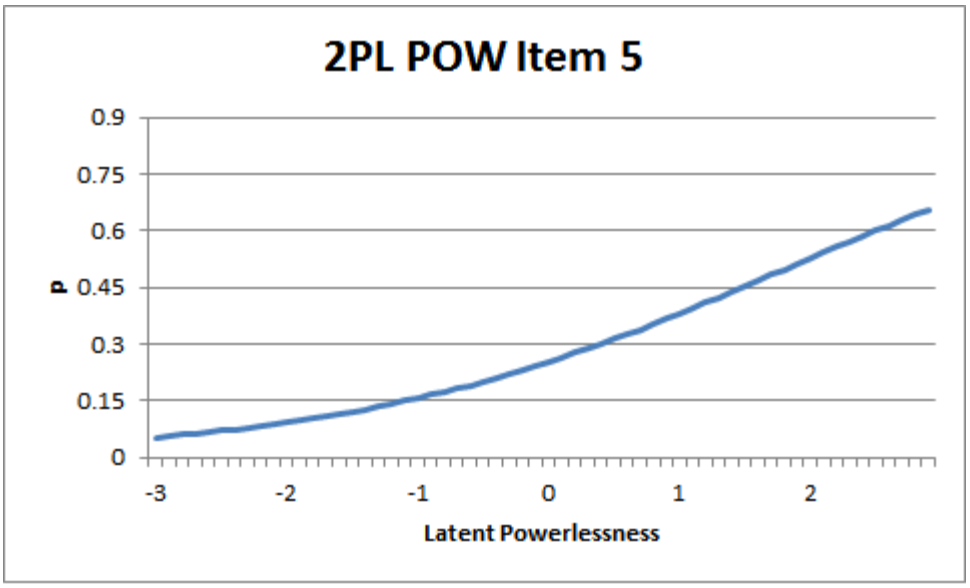


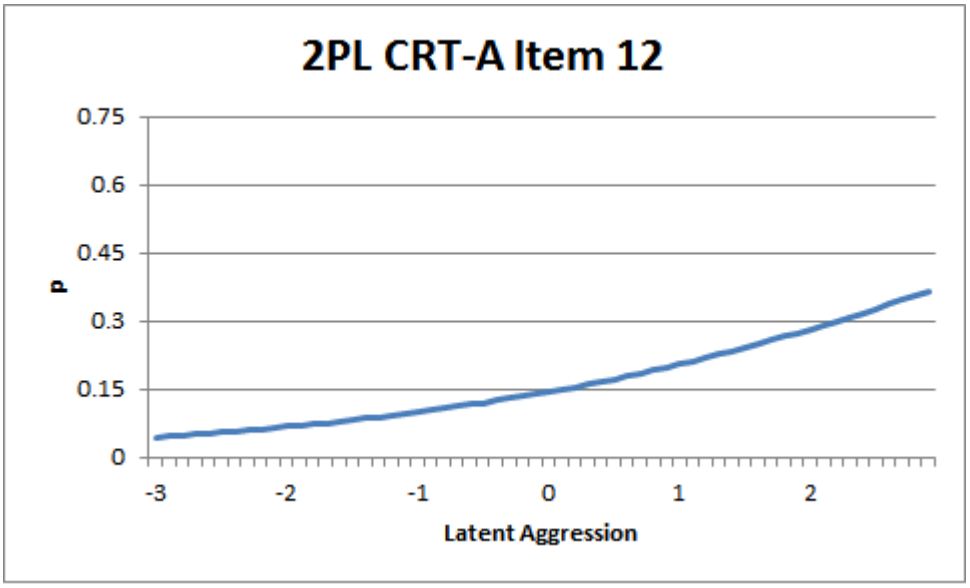
Figure 44. Two-Parameter Logistic Model item characteristic curve and information function for item 5 of the Powerlessness subscale of the Conditional Reasoning Test of Aggression.

Overall, item eleven has strong item characteristics. Responses to item eleven are well suited to indicate both latent aggression and latent powerlessness. There is no evidence of DIF, meaning that item eleven discriminates among student and applied respondents similarly.

Item 12

The twelfth CRT-A item stem is "More people are getting permits to carry guns. Most of these people say that they want to carry a gun to protect themselves. Which of the following is the most logical conclusion based on the above?" The aggressive logical response option is "These people would not mind shooting someone if threatened or attacked." The nonaggressive logical response option is "These people think they are less likely to be hurt if they have a gun." The two illogical options are "These people would gladly buy a new car" and "These people think they are less likely to be hurt if they have a gun." Of the 5,511 responses, 15.22% chose the aggressive logical option, 84.54% chose the nonaggressive logical option, 0.20% chose one of the illogical options, and 0.04% of responses were missing. The item-total correlation for item twelve and the dichotomously scored CRT-A is 0.10.

The 2PL slope estimate is 0.42 with a standard error of 0.06. The data are consistent with the hypothesis in that the slope coefficient is statistically greater than zero and monotonically increases with estimates of latent aggression. The 2PL intercept estimate is 4.21 with a standard error of 0.57. Information monotonically increases up to $\theta = 2.9$, indicating that this item is best suited for individuals high on latent aggression. The 2PL ICC and information graph are presented in Figure 45.



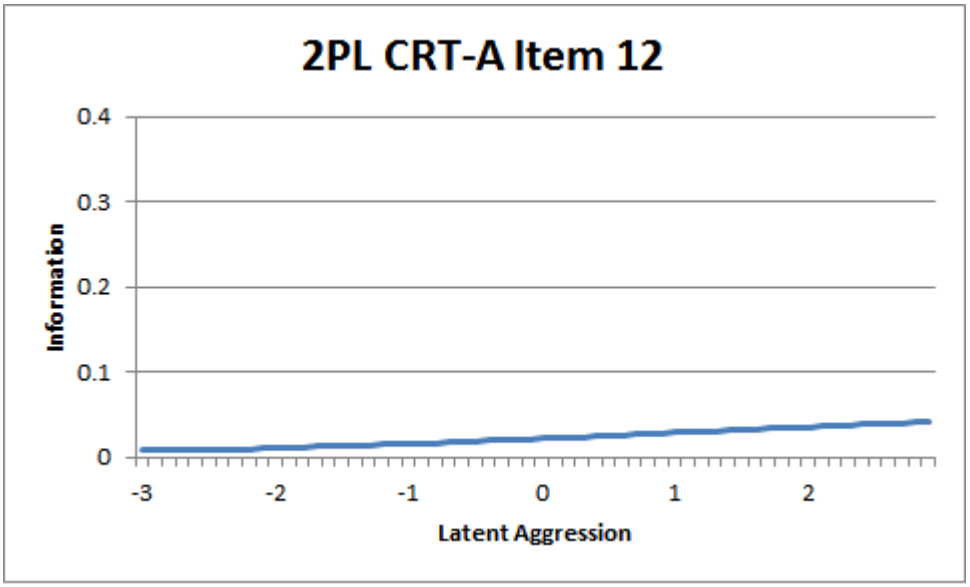
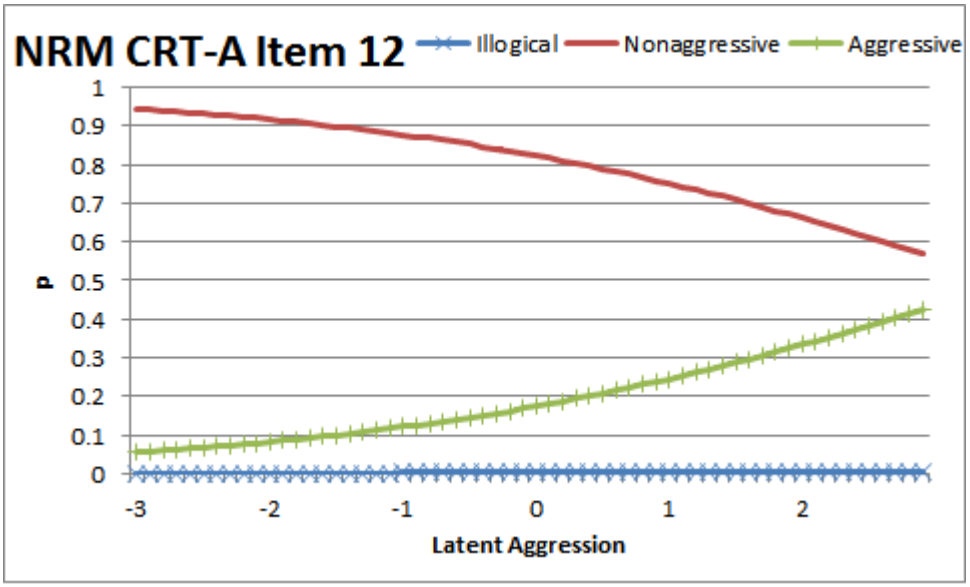


Figure 45. Two-Parameter Logistic Model item characteristic curve and information function for item 12 of the Conditional Reasoning Test of Aggression.

The linear NRM slope estimates are .062 for the aggressive logical response option, -.063 for the nonaggressive logical response option, and .001 for the combined illogical options. This is consistent with the expectation that the aggressive option will have a positive relationship with latent aggression while the nonaggressive option has a negative relationship with latent aggression. Information monotonically increases up to $\theta = 2.9$, indicating that this item is best suited for individuals high on latent aggression. The NRM ICC and information graph are presented in Figure 46.



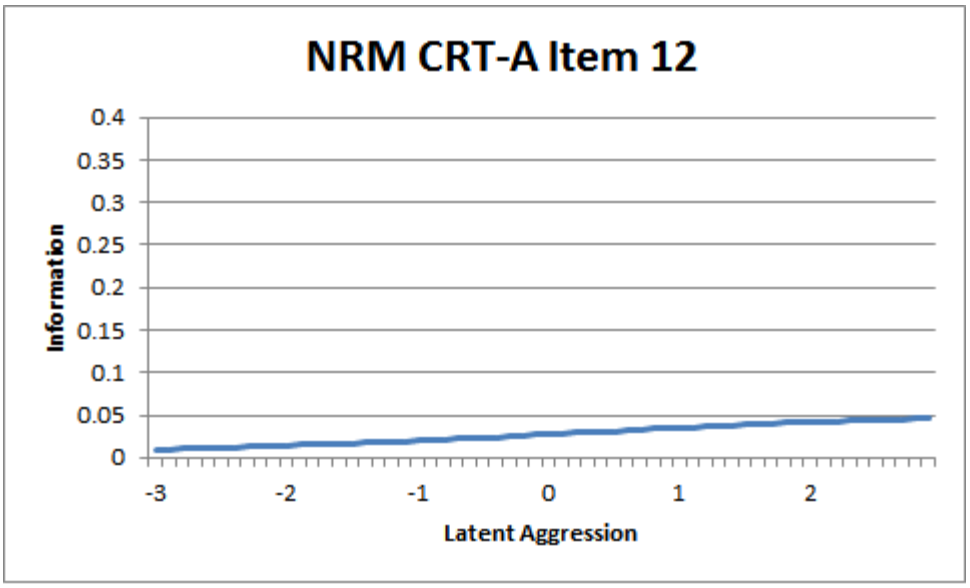


Figure 46. Nominal Response Model item characteristic curve and information function for item 12 of the Conditional Reasoning Test of Aggression.

Item twelve displays evidence of DIF (X^2, $2df = 10.6$, $p < .05$). This chi-square value can be partitioned into slope (X^2, $1df < 0.1$, $p > .05$) and intercept (X^2, $1df = 10.6$, $p < .05$) components, indicating that student and applied samples statistically differ in terms of intercept, but not slope. A graph displaying the ICCs for each group can be found in Figure 47.

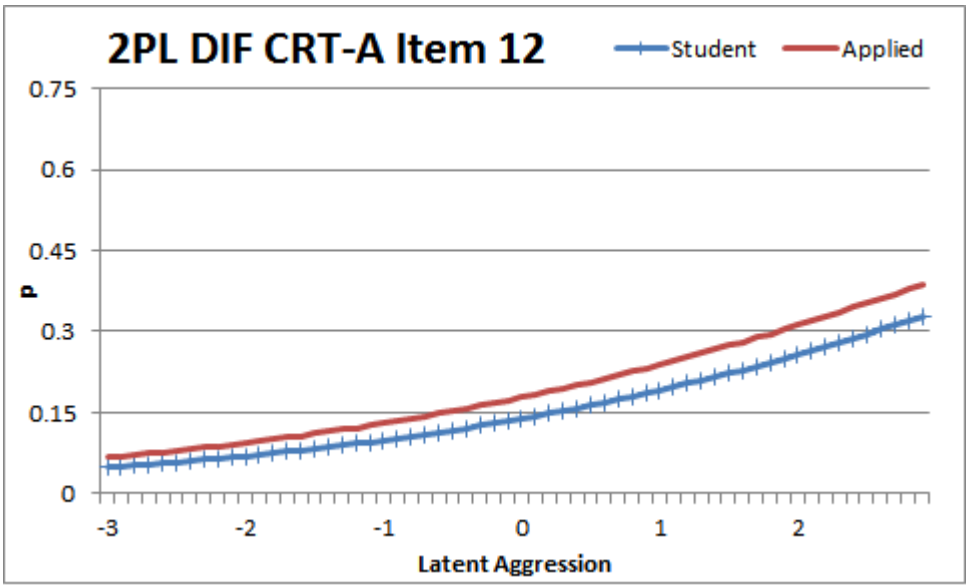


Figure 47. Two-Parameter Logistic Model differential item functioning item characteristic curve for item 12 of the Conditional Reasoning Test of Aggression.

The factor loadings for item twelve are 0.294 for externalizing controls, 0.222 for internalizing controls, and 0.059 for powerlessness, indicating that item twelve is primarily related to externalizing controls, but also related to internalizing controls. The factor-specific 2PL slope estimate for item twelve is 0.45 with a standard error of 0.07. The 2PL intercept estimate is 3.96 with a standard error of 0.58. This indicates that item twelve works well as an indicator of latent externalizing controls as well as latent aggression. The factor-specific 2PL ICC for item twelve can be found in Figure 48.

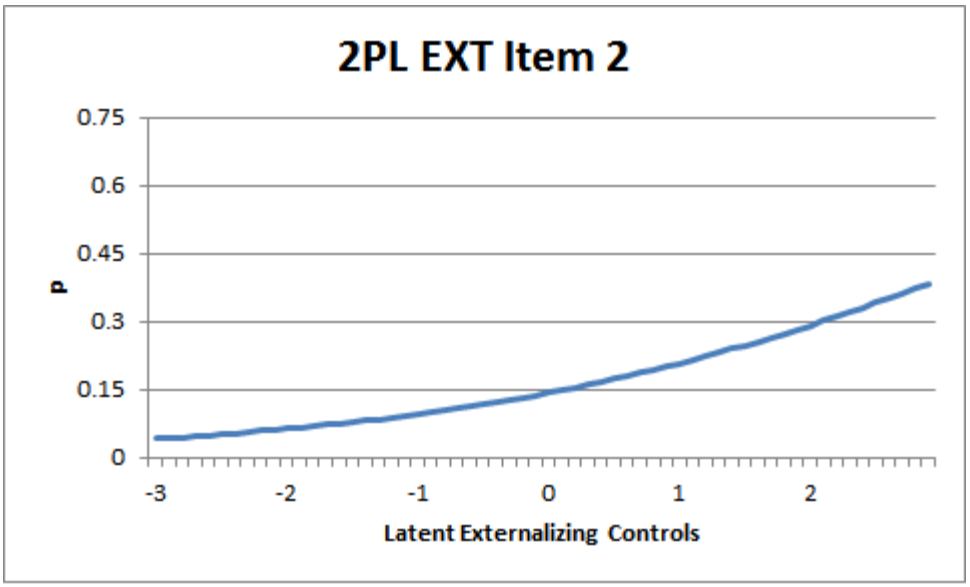


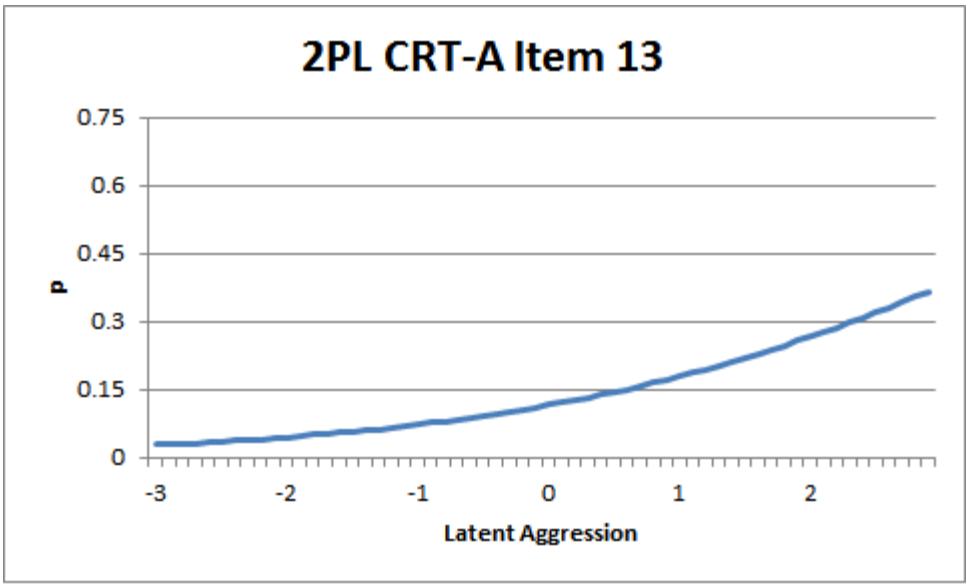
Figure 48. Two-Parameter Logistic Model item characteristic curve and information function for item 2 of the Externalizing Controls subscale of the Conditional Reasoning Test of Aggression.

Overall, item twelve has strong item characteristics. Responses to item twelve are well suited to indicate both latent aggression and latent externalizing controls. There is some evidence of DIF, but it is constrained to difference in intercept, meaning that item twelve discriminates among student and applied respondents similarly.

Item 13

The thirteenth CRT-A item stem is "American cars have gotten better in the last 15 years. American car makers started to build better cars when they began to lose business to the Japanese. Many American buyers thought that foreign cars were better made. Which of the following is the most logical conclusion based on the above?" The aggressive logical response option is "American car makers built cars to wear out 15 years ago, so they could make a lot of money selling parts." The nonaggressive logical response option is "The Japanese knew more than Americans about building good cars 15 years ago." The two illogical options are "America was the world's largest producer of airplanes 15 years ago" and "Swedish car makers lost business in America 15 years ago." Of the 5,511 responses, 12.59% chose the aggressive logical option, 85.88% chose the nonaggressive logical option, 1.43% chose one of the illogical options, and 0.09% of responses were missing. The item-total correlation for item thirteen and the dichotomously scored CRT-A is 0.10.

The 2PL slope estimate is 0.51 with a standard error of 0.07. The data are consistent with the hypothesis in that the slope coefficient is statistically greater than zero and monotonically increases with estimates of latent aggression. The 2PL intercept estimate is 3.98 with a standard error of 0.47. Information monotonically increases up to $\theta = 2.9$, indicating that this item is best suited for individuals high on latent aggression. The 2PL ICC and information graph are presented in Figure 49.



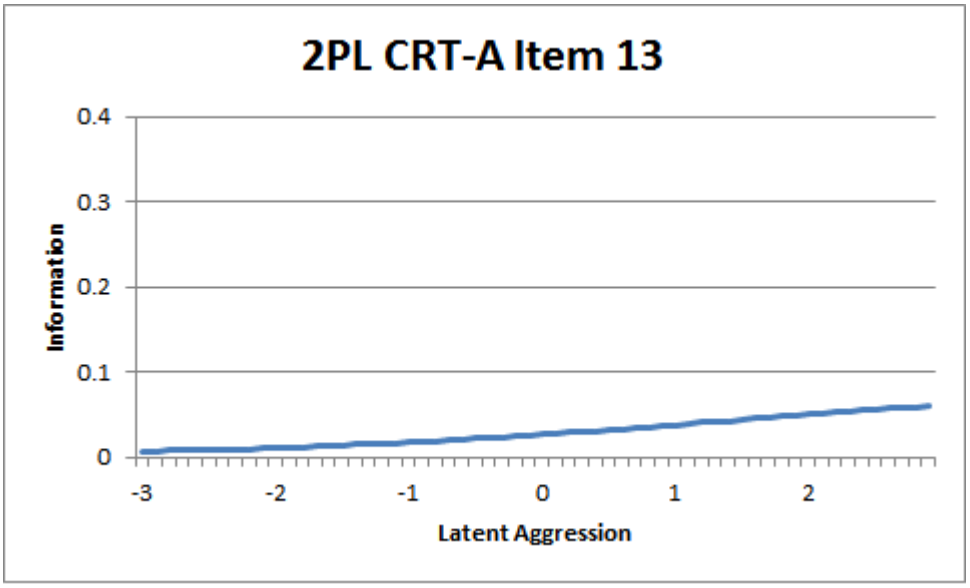


Figure 49. Two-Parameter Logistic Model item characteristic curve and information function for item 13 of the Conditional Reasoning Test of Aggression.

The linear NRM slope estimates are .071 for the aggressive logical response option, -.079 for the nonaggressive logical response option, and .008 for the combined illogical options. This is consistent with the expectation that the aggressive option will have a positive relationship with latent aggression while the nonaggressive option has a negative relationship with latent aggression. Information monotonically increases up to $\theta = 2.9$, indicating that this item is best

suited for individuals high on latent aggression. The NRM ICC and information graph are presented in Figure 50.

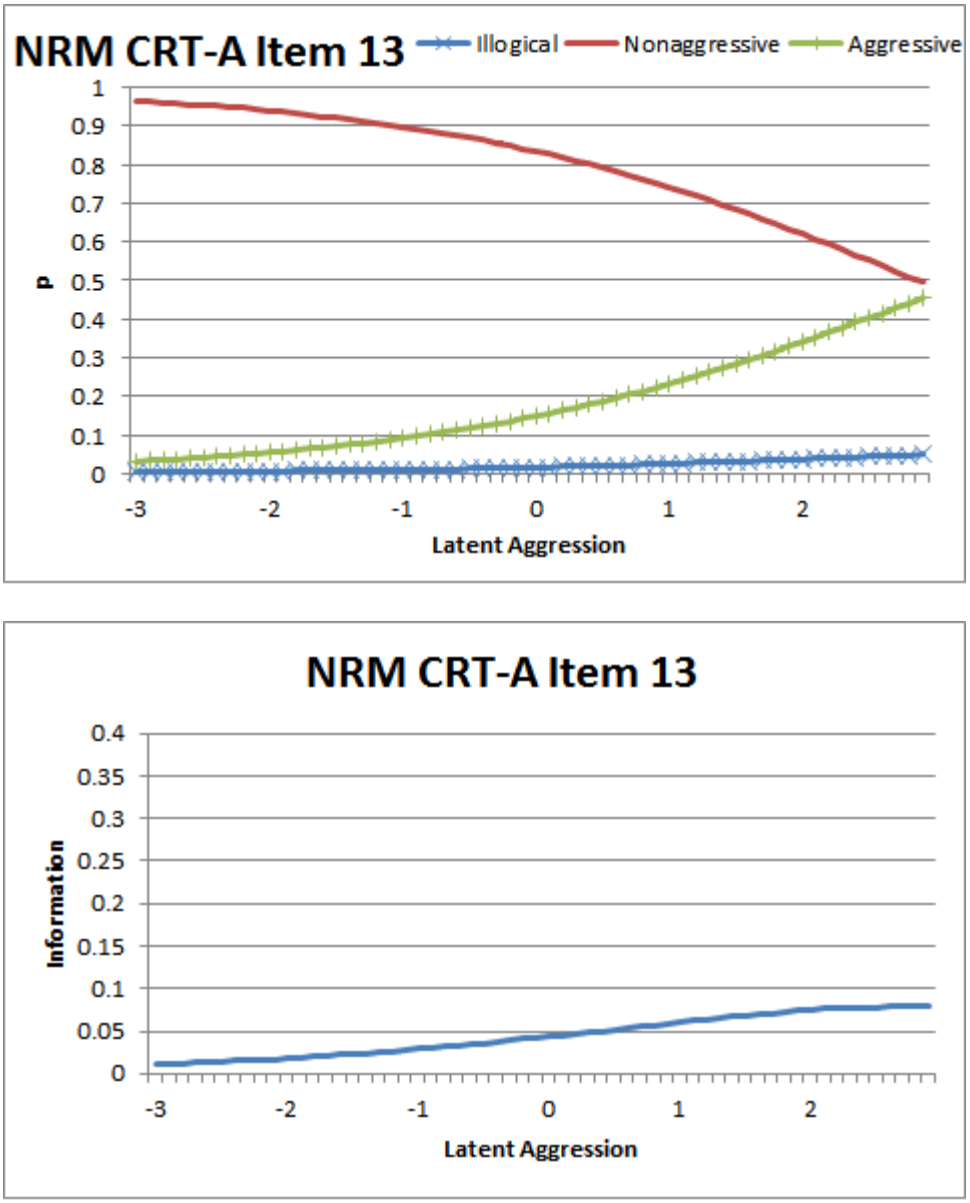


Figure 50. Nominal Response Model item characteristic curve and information function for item 13 of the Conditional Reasoning Test of Aggression.

Item thirteen displays no evidence of DIF (X^2, $2df = 0.3$, $p > .05$). This chi-square value can be partitioned into slope (X^2, $1df = 0.3$, $p > .05$) and intercept (X^2, $1df < 0.1$, $p > .05$)

components, indicating that student and applied samples do not statistically differ in either slope or intercept. A graph displaying the ICCs for each group can be found in Figure 51.

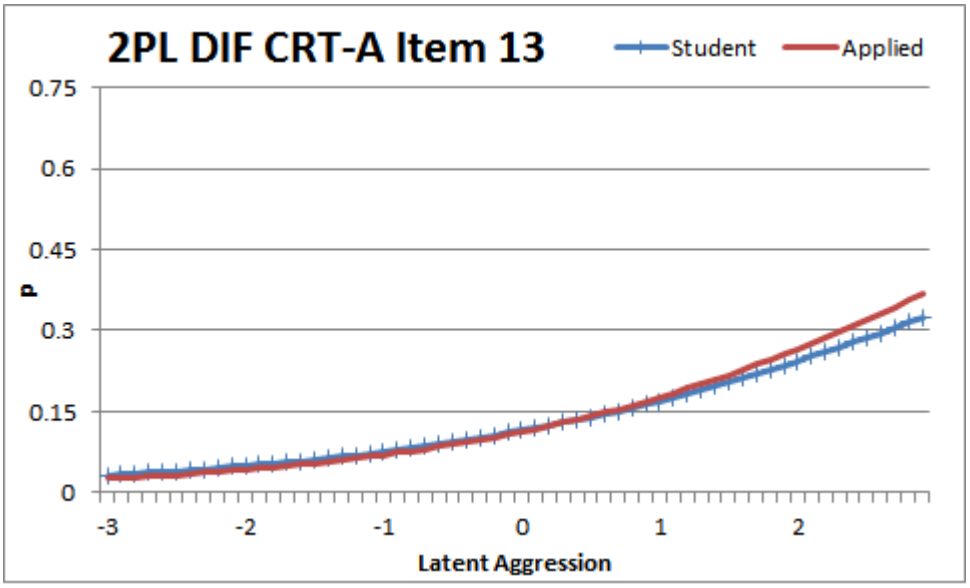


Figure 51. Two-Parameter Logistic Model differential item functioning item characteristic curve for item 13 of the Conditional Reasoning Test of Aggression.

The factor loadings for item thirteen are 0.481 for externalizing controls, 0.119 for internalizing controls, and 0.097 for powerlessness, indicating that item thirteen is primarily related to externalizing controls, but also related to internalizing controls. The factor-specific 2PL slope estimate for item thirteen is 0.80 with a standard error of 0.09. The 2PL intercept estimate is 2.70 with a standard error of 0.24. This indicates that item thirteen works well as an indicator of latent externalizing controls as well as latent aggression. The factor-specific 2PL ICC for item thirteen can be found in Figure 52.

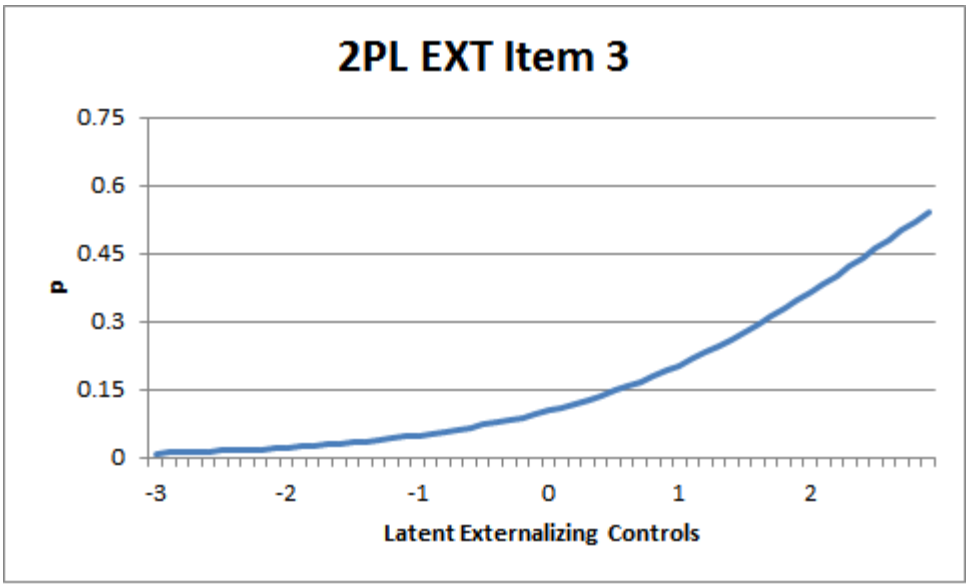


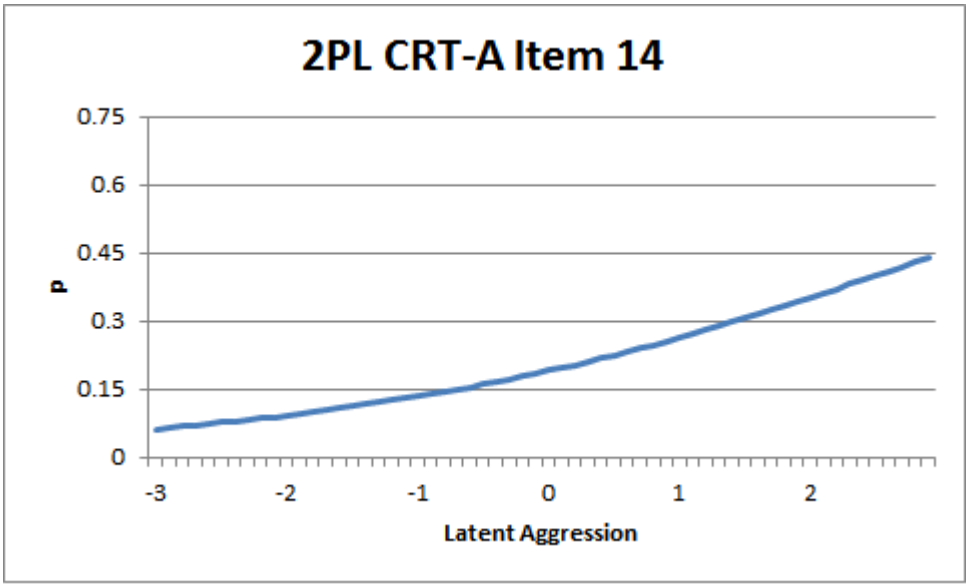
Figure 52. Two-Parameter Logistic Model item characteristic curve and information function for item 3 of the Externalizing Controls subscale of the Conditional Reasoning Test of Aggression.

Overall, item thirteen has strong item characteristics. Responses to item thirteen are well suited to indicate both latent aggression and latent externalizing controls. There is no evidence of DIF, meaning that item thirteen discriminates among student and applied respondents similarly.

Item 14

The fourteenth CRT-A item stem is "Store employees are told to watch out for people who look like shoplifters. If a customer looks like a shoplifter, then employees are supposed to watch the customer closely. Which of the following is the biggest problem with this practice?" The aggressive logical response option is "Abuse by store employees who use it as an excuse to bother people they don't like." The nonaggressive logical response option is "Many customers who look like shoplifters are honest and do not steal." The two illogical options are "Most retail stores don't open until 10:00 in the morning" and "Parking is getting harder to find in shopping malls." Of the 5,511 responses, 19.85% chose the aggressive logical option, 79.70% chose the nonaggressive logical option, 0.38% chose one of the illogical options, and 0.07% of responses were missing. The item-total correlation for item fourteen and the dichotomously scored CRT-A is 0.11.

The 2PL slope estimate is 0.42 with a standard error of 0.06. The data are consistent with the hypothesis in that the slope coefficient is statistically greater than zero and monotonically increases with estimates of latent aggression. The 2PL intercept estimate is 3.47 with a standard error of 0.44. Information monotonically increases up to $\theta = 2.9$, indicating that this item is best suited for individuals high on latent aggression. The 2PL ICC and information graph are presented in Figure 53.



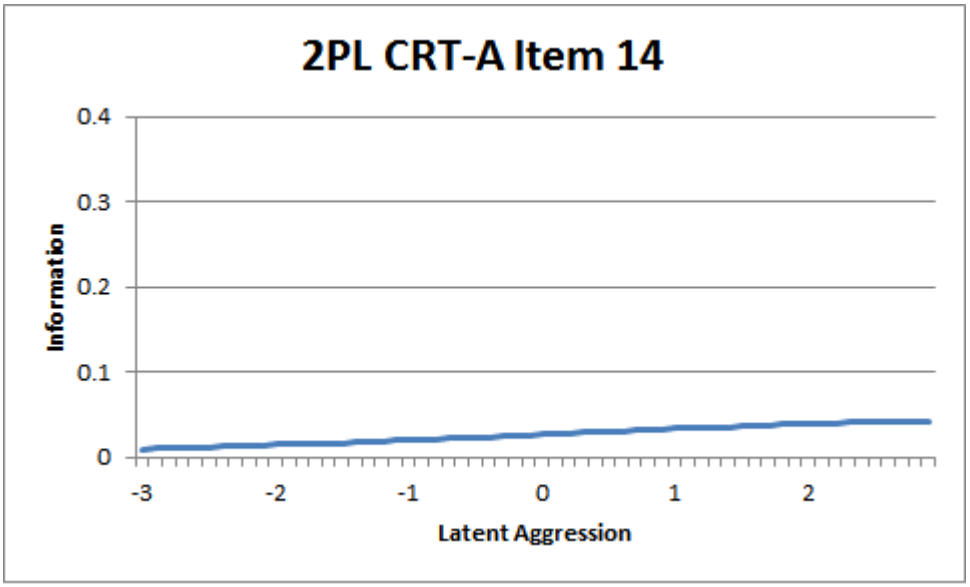
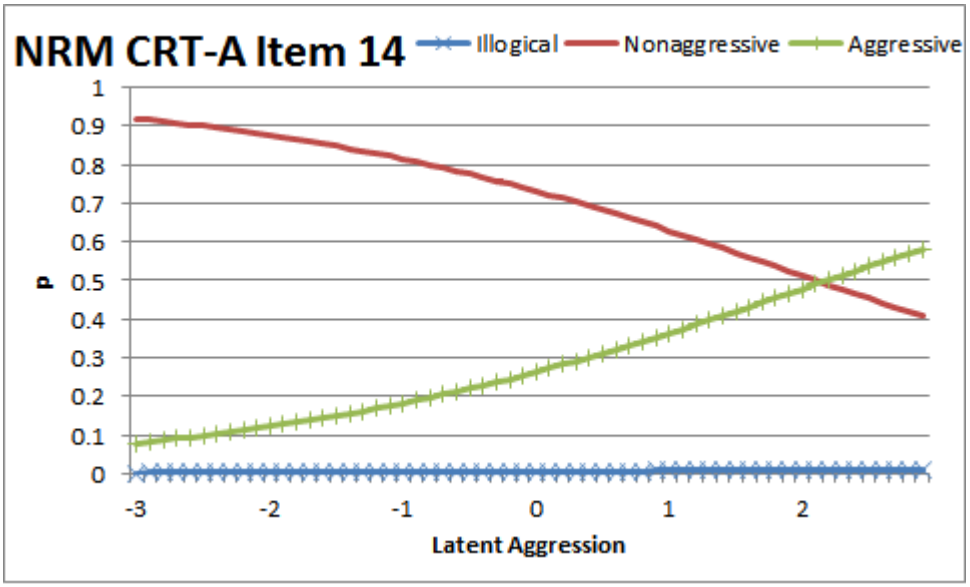


Figure 53. Two-Parameter Logistic Model item characteristic curve and information function for item 14 of the Conditional Reasoning Test of Aggression.

The linear NRM slope estimates are .087 for the aggressive logical response option, -.089 for the nonaggressive logical response option, and .002 for the combined illogical options. This is consistent with the expectation that the aggressive option will have a positive relationship with latent aggression while the nonaggressive option has a negative relationship with latent aggression. Information increases up to $\theta = 2.3$, after which it decreases through $\theta = 2.9$. This indicates that this item is best suited for individuals high on latent aggression. The NRM ICC and information graph are presented in Figure 54.



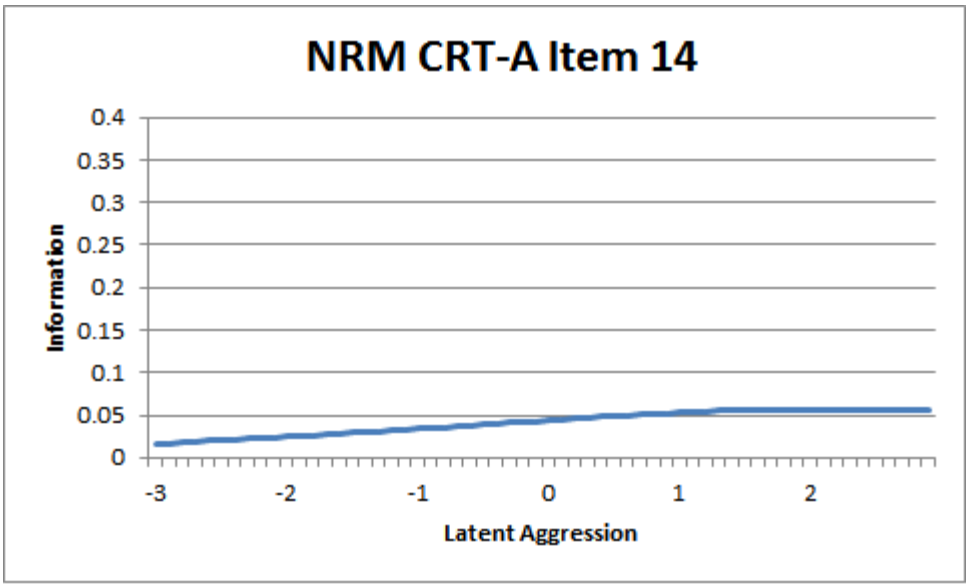


Figure 54. Nominal Response Model item characteristic curve and information function for item 14 of the Conditional Reasoning Test of Aggression.

Item fourteen displays evidence of DIF (X^2, $2df = 6.6$, $p < .05$). This chi-square value can be partitioned into slope (X^2, $1df = 0.4$, $p > .05$) and intercept (X^2, $1df = 6.3$, $p < .05$) components, indicating that student and applied samples statistically differ in terms of intercept, but not slope. A graph displaying the ICCs for each group can be found in Figure 55.

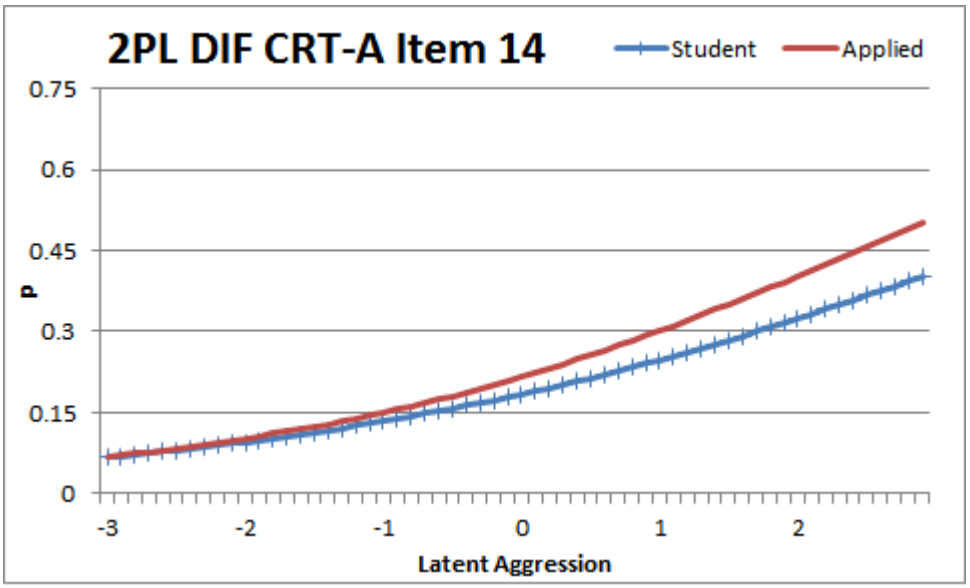


Figure 55. Two-Parameter Logistic Model differential item functioning item characteristic curve for item 14 of the Conditional Reasoning Test of Aggression.

The factor loadings for item fourteen are 0.294 for externalizing controls, 0.222 for internalizing controls, and -0.014 for powerlessness, indicating that item fourteen is primarily related to externalizing controls, but also related to internalizing controls. The factor-specific 2PL slope estimate for item fourteen is 0.49 with a standard error of 0.07. The 2PL intercept estimate is 3.01 with a standard error of 0.38. This indicates that item fourteen works well as an indicator of latent externalizing controls as well as latent aggression. The factor-specific 2PL ICC for item fourteen can be found in Figure 56.

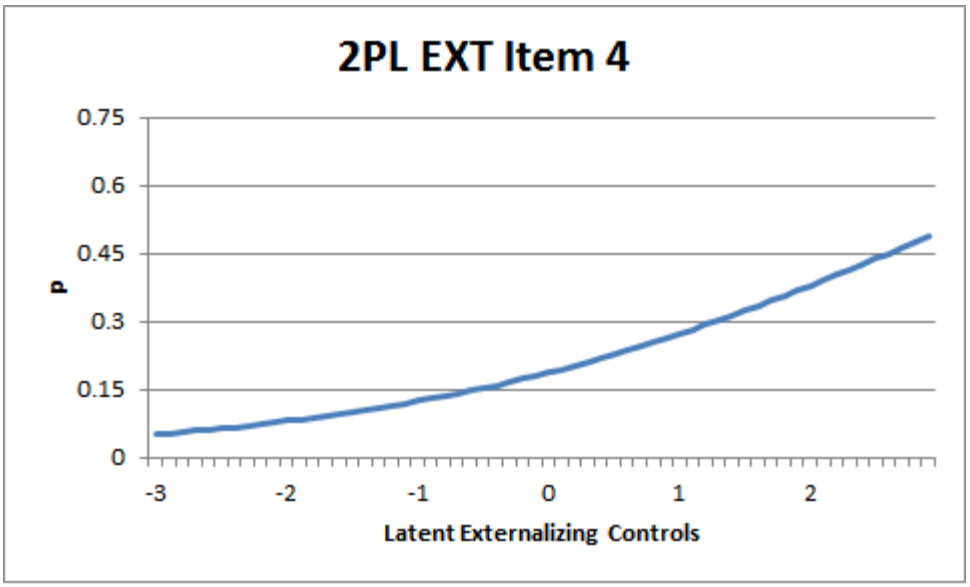


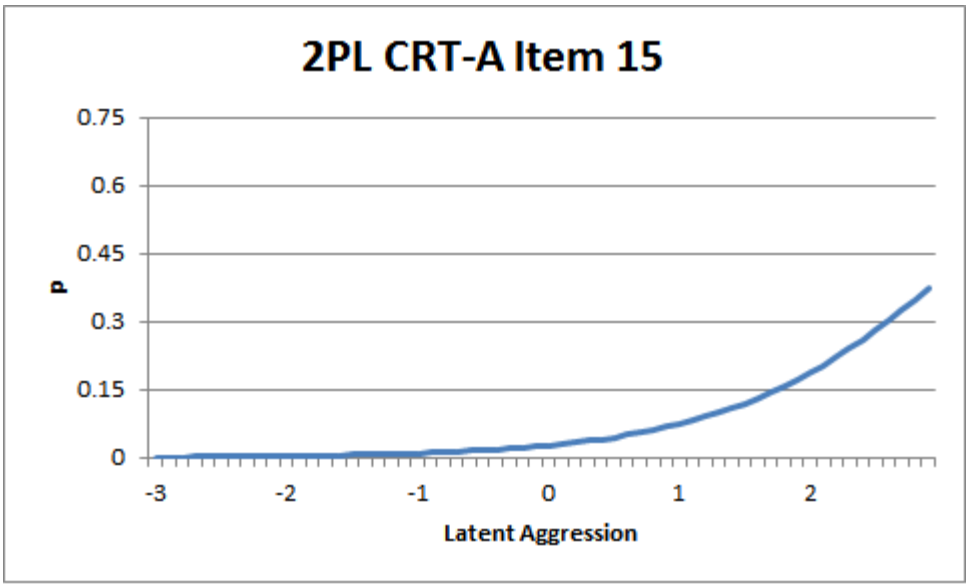
Figure 56. Two-Parameter Logistic Model item characteristic curve and information function for item 4 of the Externalizing Controls subscale of the Conditional Reasoning Test of Aggression.

Overall, item fourteen has strong item characteristics. Responses to item fourteen are well suited to indicate both latent aggression and latent externalizing controls. There is some evidence of DIF, but it is constrained to difference in intercept, meaning that item fourteen discriminates among student and applied respondents similarly.

Item 15

The fifteenth CRT-A item stem is "Many companies use bonuses to reward their employees. For example, salespeople are supposed to make a certain number of sales. If they sell more than they are supposed to, then they receive a bonus. Bonuses include extra pay and time off from work. Which of the following is the most logical explanation for why companies use bonuses?" The aggressive logical response option is "Bonuses give managers a way to have more control over their employees." The nonaggressive logical response option is "Bonuses give hard-working employees a way to earn extra money or time off." The two illogical options are "Bonuses give new employees a way to learn more about the business" and "Bonuses give customers a reward for being loyal." Of the 5,511 responses, 4.34% chose the aggressive logical option, 92.58% chose the nonaggressive logical option, 3.05% chose one of the illogical options, and 0.04% of responses were missing. The item-total correlation for item fifteen and the dichotomously scored CRT-A is 0.10.

The 2PL slope estimate is 1.05 with a standard error of 0.11. The data are consistent with the hypothesis in that the slope coefficient is statistically greater than zero and monotonically increases with estimates of latent aggression. The 2PL intercept estimate is 3.39 with a standard error of 0.28. Information monotonically increases up to $\theta = 2.9$, indicating that this item is best suited for individuals high on latent aggression. The 2PL ICC and information graph are presented in Figure 57.



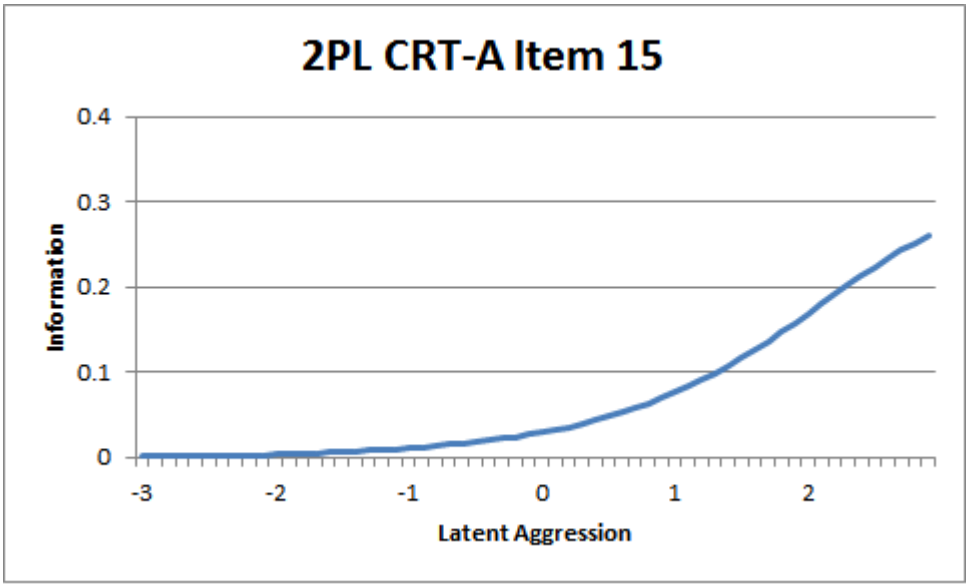


Figure 57. Two-Parameter Logistic Model item characteristic curve and information function for item 15 of the Conditional Reasoning Test of Aggression.

The linear NRM slope estimates are .061 for the aggressive logical response option, -.084 for the nonaggressive logical response option, and .022 for the combined illogical options. This is consistent with the expectation that the aggressive option will have a positive relationship with latent aggression while the nonaggressive option has a negative relationship with latent aggression. Information increases up to $\theta = 2.8$, after which it decreases through $\theta = 2.9$. This

indicates that this item is best suited for individuals high on latent aggression. The NRM ICC and information graph are presented in Figure 58.

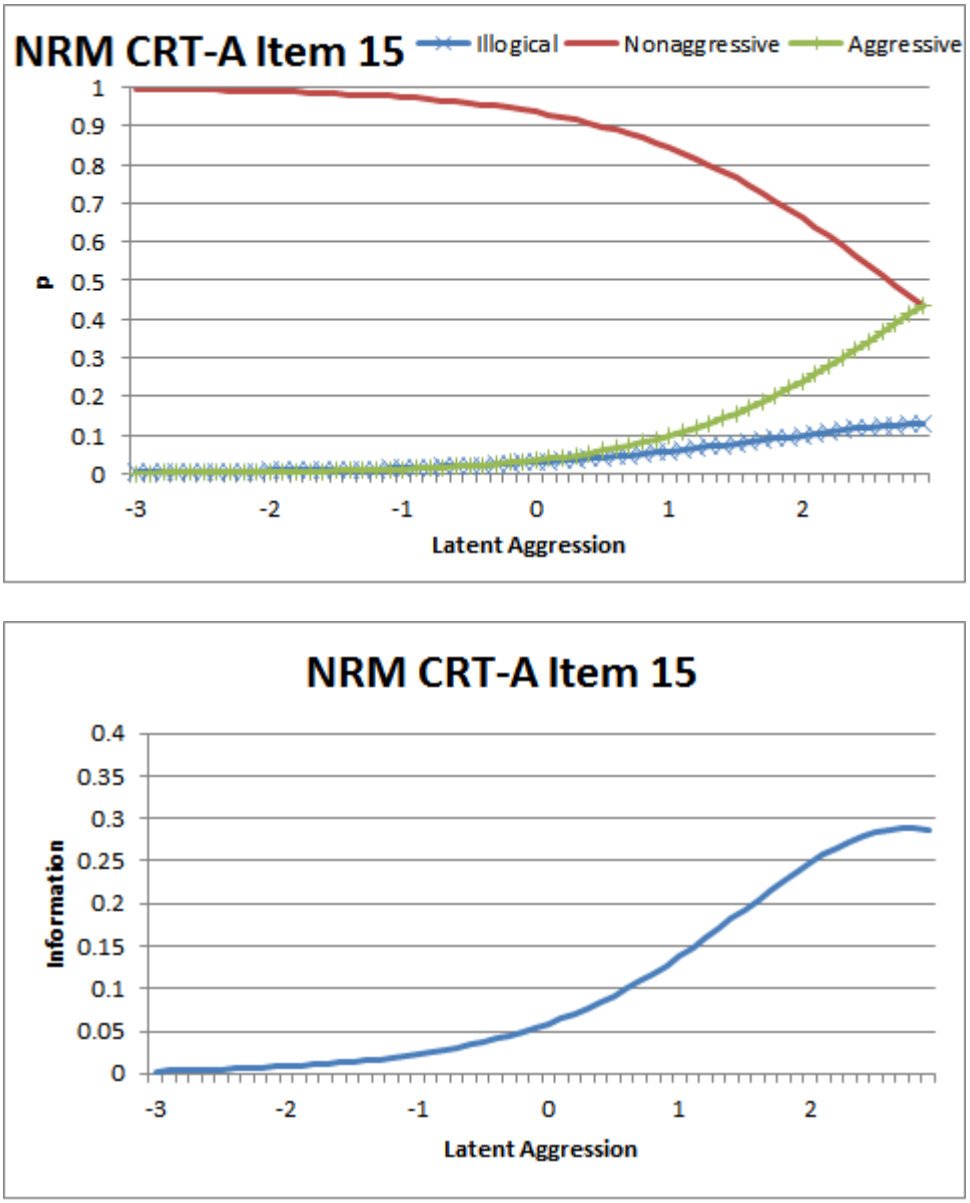


Figure 58. Nominal Response Model item characteristic curve and information function for item 15 of the Conditional Reasoning Test of Aggression.

Item fifteen displays evidence of DIF (X^2, $2df = 6.6$, $p < .05$). This chi-square value can be partitioned into slope (X^2, $1df = 1.8$, $p > .05$) and intercept (X^2, $1df = 4.8$, $p < .05$) components,

indicating that student and applied samples statistically differ in terms of intercept, but not slope. A graph displaying the ICCs for each group can be found in Figure 59.

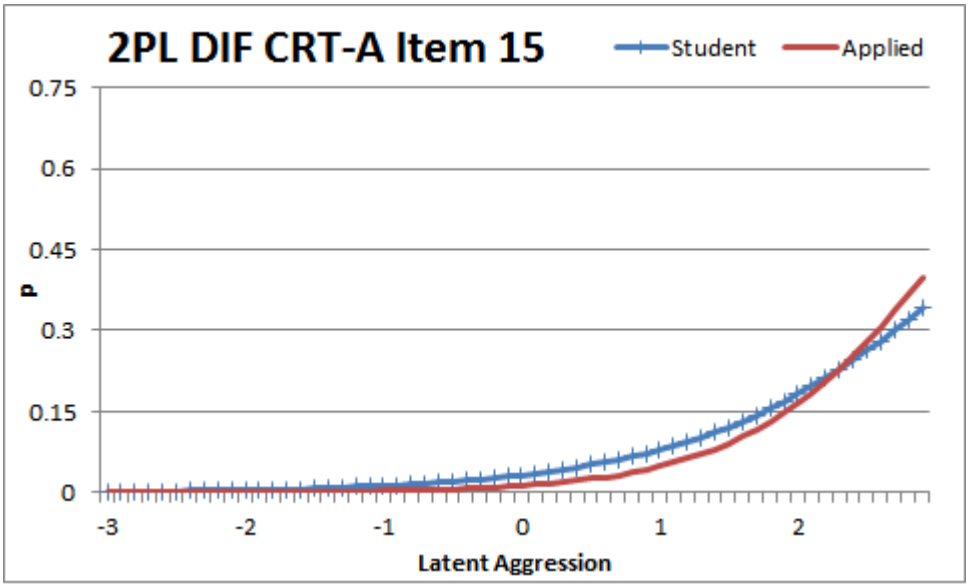


Figure 59. Two-Parameter Logistic Model differential item functioning item characteristic curve for item 15 of the Conditional Reasoning Test of Aggression.

The factor loadings for item fifteen are 0.560 for externalizing controls, 0.342 for internalizing controls, and 0.252 for powerlessness, indicating that item fifteen is primarily related to externalizing controls, but also related to internalizing controls and powerlessness. The factor-specific 2PL slope estimate for item fifteen is 0.94 with a standard error of 0.13. The 2PL intercept estimate is 3.71 with a standard error of 0.40. This indicates that item fifteen works well as an indicator of latent externalizing controls as well as latent aggression. The factor-specific 2PL ICC for item fifteen can be found in Figure 60.

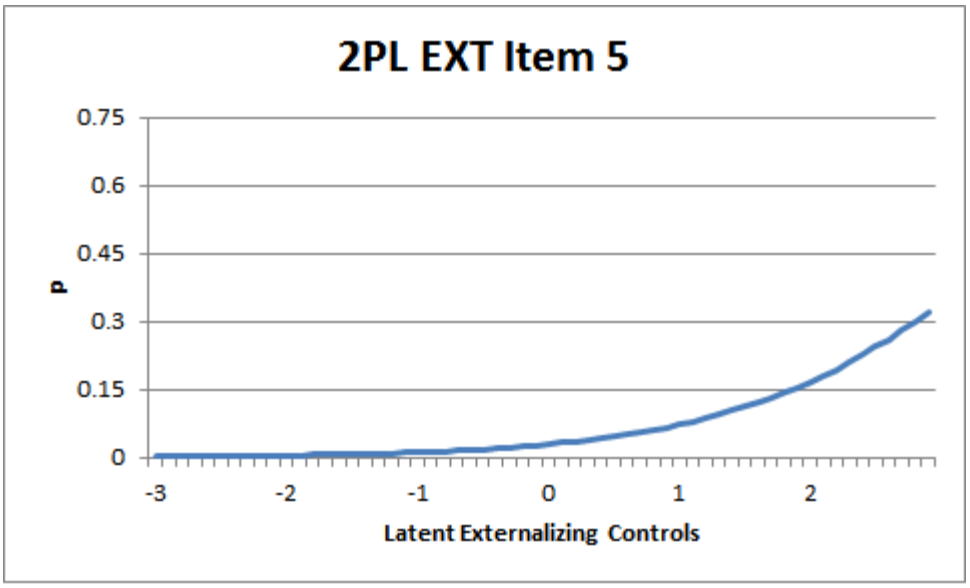


Figure 60. Two-Parameter Logistic Model item characteristic curve and information function for item 5 of the Externalizing Controls subscale of the Conditional Reasoning Test of Aggression.

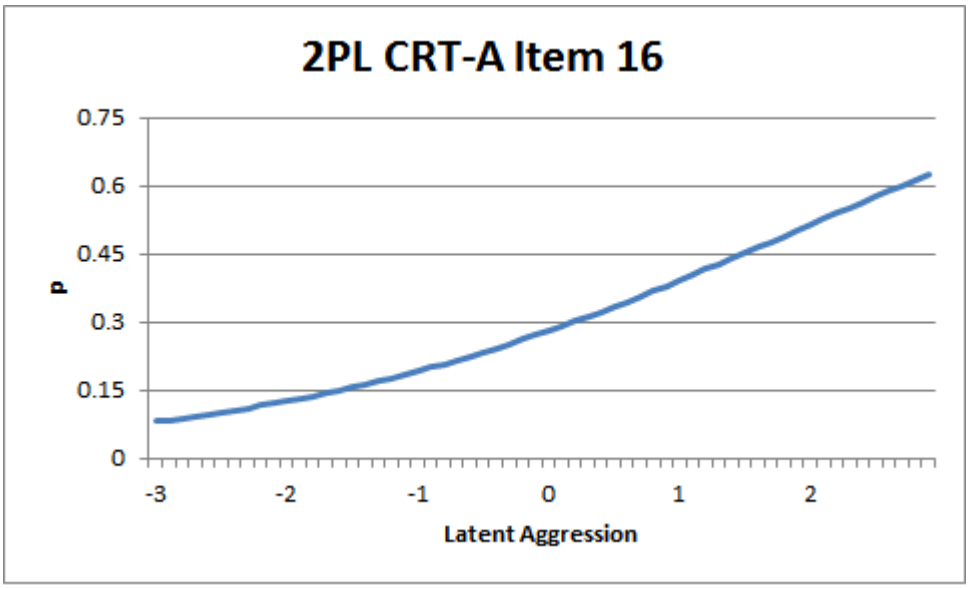
Overall, item fifteen has strong item characteristics. Responses to item fifteen are well suited to indicate both latent aggression and latent externalizing controls. There is some evidence of DIF, but it is constrained to difference in intercept, meaning that item fifteen discriminates among student and applied respondents similarly.

Item 16

The sixteenth CRT-A item stem is "People who work for restaurants often have their purses or bags searched. Managers search employees as they leave work. The reason given for the searches is that they reduce theft of food and equipment. Which of the following is the biggest problem with this reasoning?" The aggressive logical response option is "Employees who steal are too smart to be caught by this type of search." The nonaggressive logical response option is "Most restaurant employees are honest and feel embarrassed by the searches." The two illogical options are "Many restaurant employees receive tips from customers" and "More restaurants are opening up for lunch." Of the 5,511 responses, 29.12% chose the aggressive logical option, 66.74% chose the nonaggressive logical

option, 4.08% chose one of the illogical options, and 0.05% of responses were missing. The item-total correlation for item sixteen and the dichotomously scored CRT-A is 0.14.

The 2PL slope estimate is 0.50 with a standard error of 0.05. The data are consistent with the hypothesis in that the slope coefficient is statistically greater than zero and monotonically increases with estimates of latent aggression. The 2PL intercept estimate is 1.89 with a standard error of 0.18. Information increases up to $\theta = 2.1$, after which it decreases through $\theta = 2.9$. This indicates that this item is best suited for individuals high on latent aggression. The 2PL ICC and information graph are presented in Figure 61.



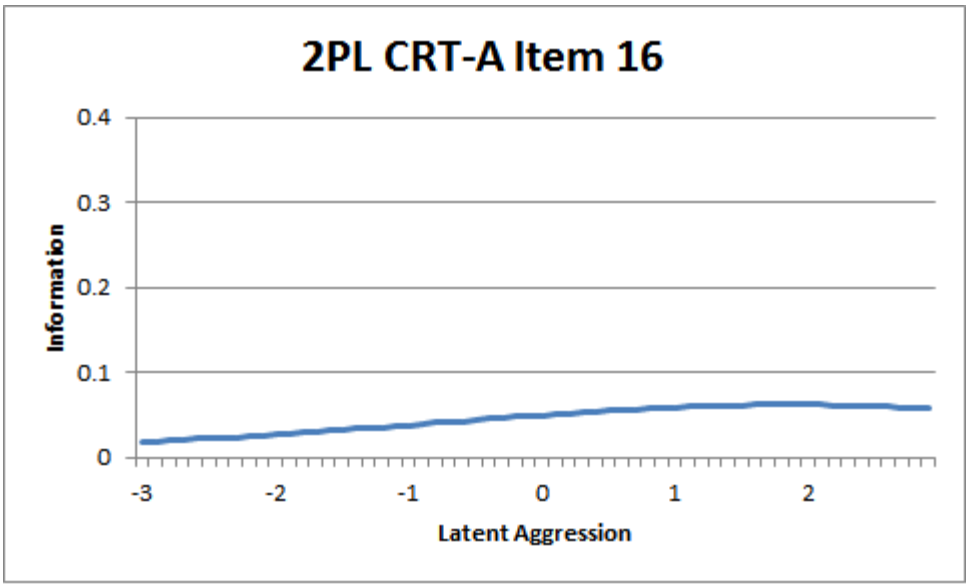
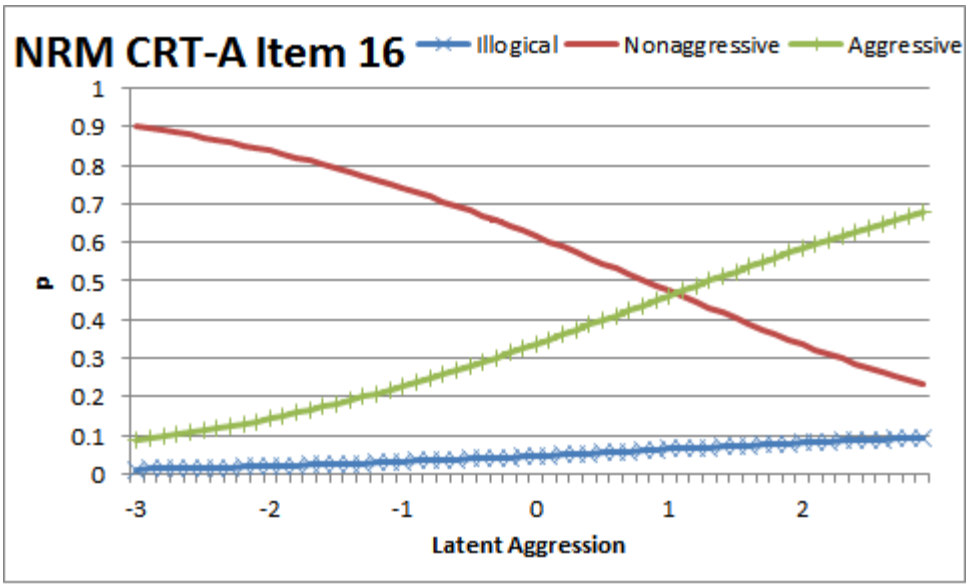


Figure 61. Two-Parameter Logistic Model item characteristic curve and information function for item 16 of the Conditional Reasoning Test of Aggression.

The linear NRM slope estimates are .107 for the aggressive logical response option, -.122 for the nonaggressive logical response option, and .015 for the combined illogical options. This is consistent with the expectation that the aggressive option will have a positive relationship with latent aggression while the nonaggressive option has a negative relationship with latent aggression. Information increases up to $\theta = 1.1$, after which it decreases through $\theta = 2.9$. This indicates that this item is best suited for individuals moderately high to high on latent aggression. The NRM ICC and information graph are presented in Figure 62.



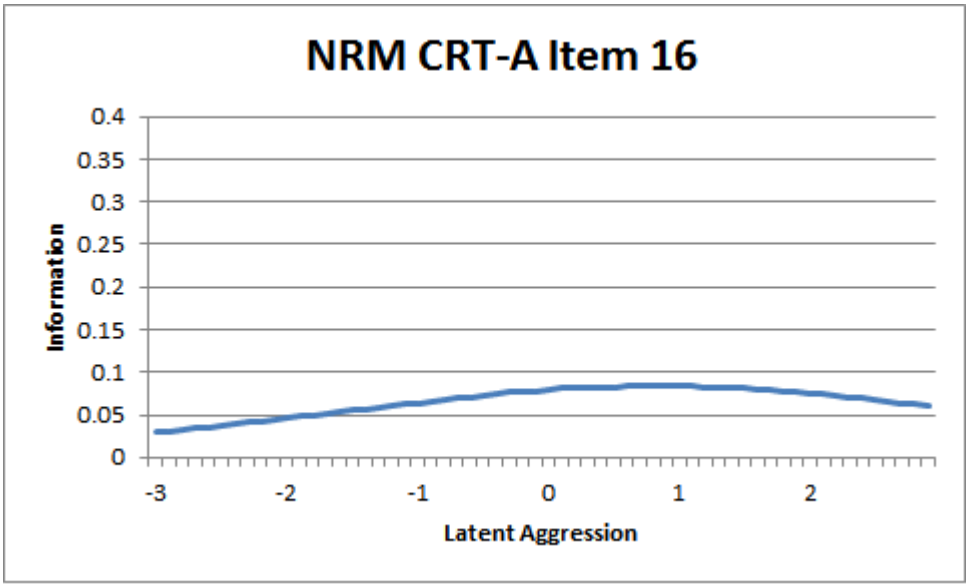


Figure 62. Nominal Response Model item characteristic curve and information function for item 16 of the Conditional Reasoning Test of Aggression.

Item sixteen displays no evidence of DIF (X^2, $2df = 1.3$, $p > .05$). This chi-square value can be partitioned into slope (X^2, $1df < 0.1$, $p > .05$) and intercept (X^2, $1df = 1.3$, $p > .05$) components, indicating that student and applied samples do not statistically differ in either slope or intercept. A graph displaying the ICCs for each group can be found in Figure 63.

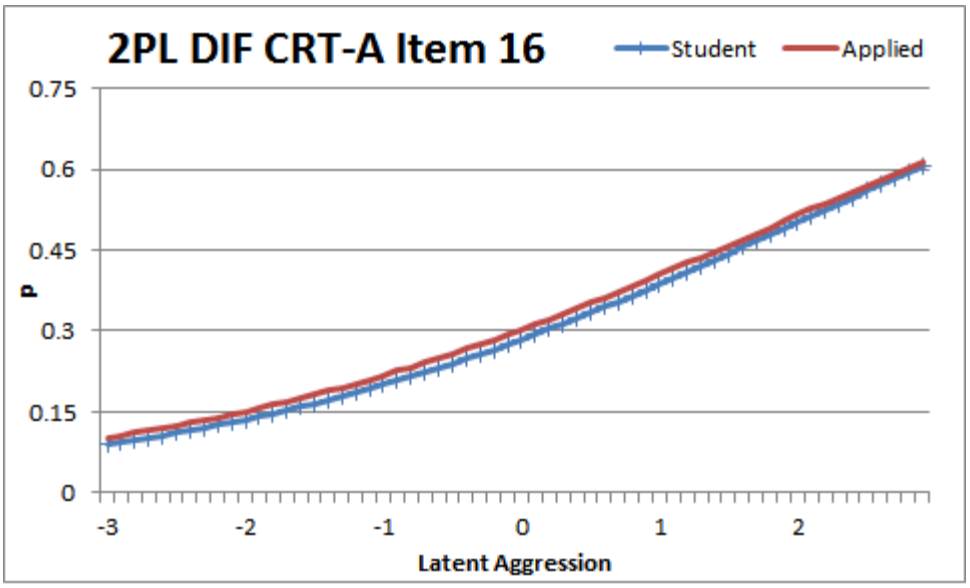


Figure 63. Two-Parameter Logistic Model differential item functioning item characteristic curve for item 16 of the Conditional Reasoning Test of Aggression.

The factor loadings for item sixteen are 0.254 for externalizing controls, 0.228 for internalizing controls, and 0.166 for powerlessness, indicating that item sixteen is primarily related to externalizing controls, but also related to internalizing controls and powerlessness. The factor-specific 2PL slope estimate for item sixteen is 0.40 with a standard error of 0.06. The 2PL intercept estimate is 2.29 with a standard error of 0.32. This indicates that item sixteen works well as an indicator of latent externalizing controls as well as latent aggression. The factor-specific 2PL ICC for item sixteen can be found in Figure 64.

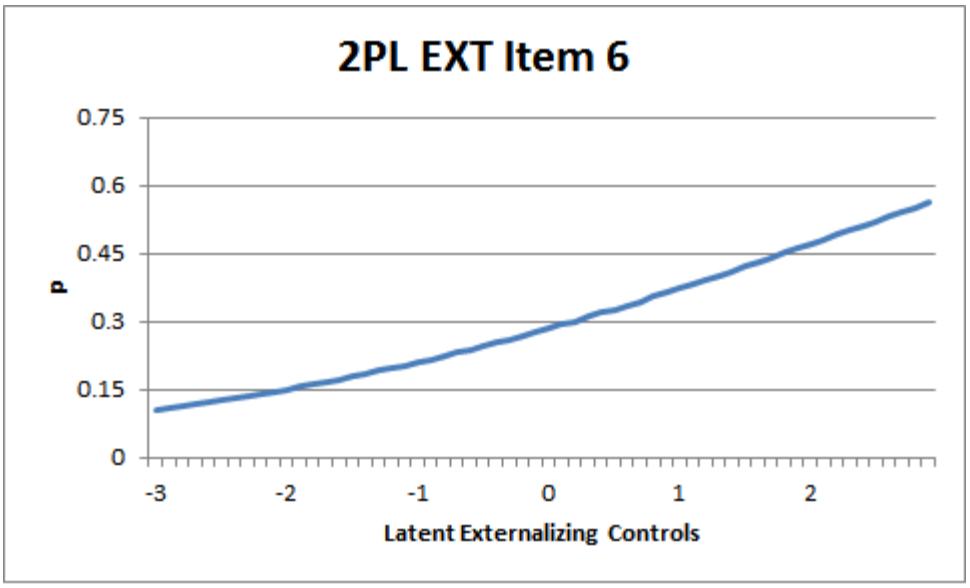


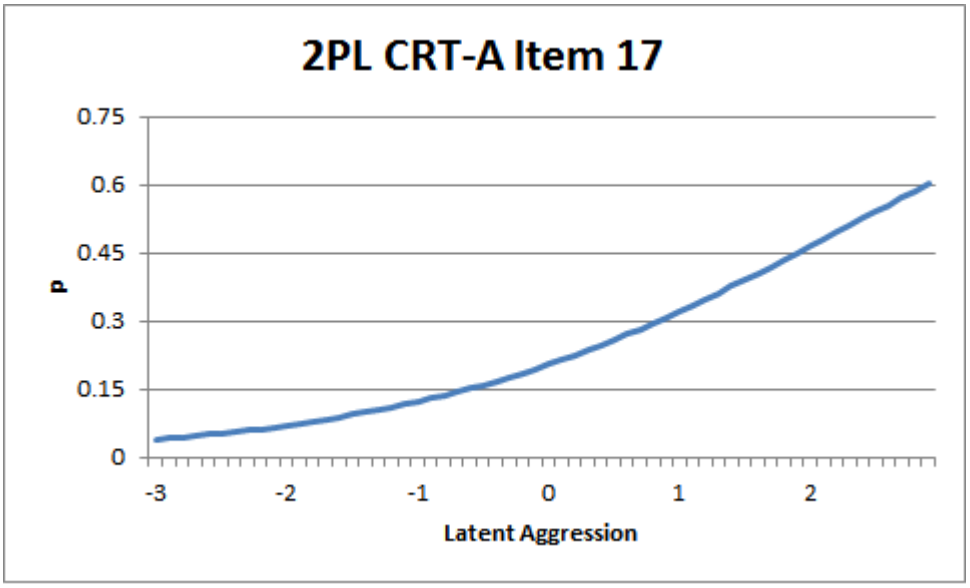
Figure 64. Two-Parameter Logistic Model item characteristic curve and information function for item 6 of the Externalizing Controls subscale of the Conditional Reasoning Test of Aggression.

Overall, item sixteen has strong item characteristics. Responses to item sixteen are well suited to indicate both latent aggression and latent externalizing controls. There is no evidence of DIF, meaning that item sixteen discriminates among student and applied respondents similarly.

Item 17

The seventeenth CRT-A item stem is "Gangs have formed in many large cities. Gangs often fight over territory, selling drugs, and insults. Gang members are often killed in these fights. Few murders of gang members are solved. Which of the following is the most logical conclusion based on the above?" The aggressive logical response option is "The police don't really care about the deaths of a few gang members." The nonaggressive logical response option is "Too many people are in gang fights to know who committed the murders." The two illogical options are "Gangs never use weapons in fights" and "Most police are trained in hand-to-hand combat." Of the 5,511 responses, 22.08% chose the aggressive logical option, 75.05% chose the nonaggressive logical option, 2.69% chose one of the illogical options, and 0.18% of responses were missing. The item-total correlation for item seventeen and the dichotomously scored CRT-A is 0.15.

The 2PL slope estimate is 0.61 with a standard error of 0.06. The data are consistent with the hypothesis in that the slope coefficient is statistically greater than zero and monotonically increases with estimates of latent aggression. The 2PL intercept estimate is 2.23 with a standard error of 0.19. Information increases up to $\theta = 2.5$, after which it decreases through $\theta = 2.9$. This indicates that this item is best suited for individuals high on latent aggression. The 2PL ICC and information graph are presented in Figure 65.



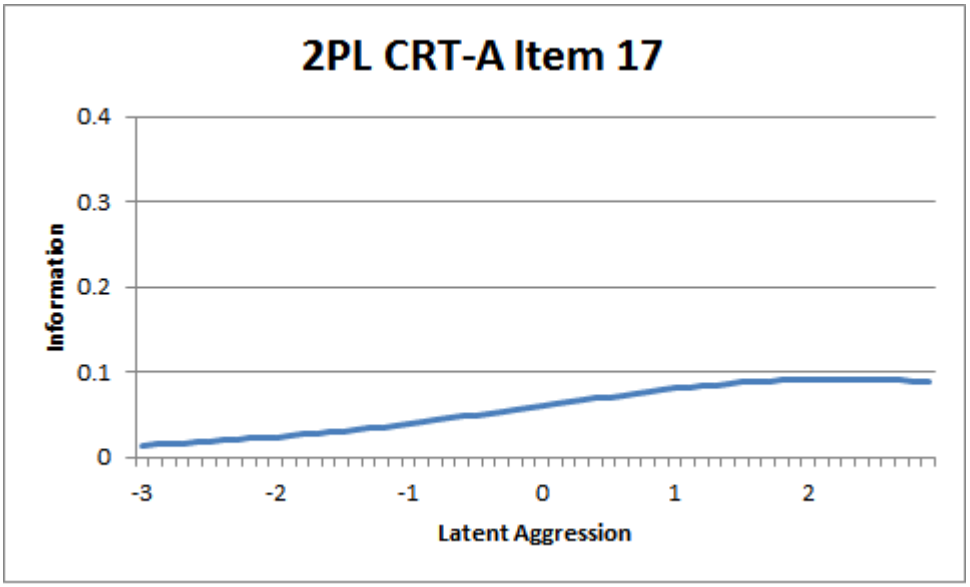
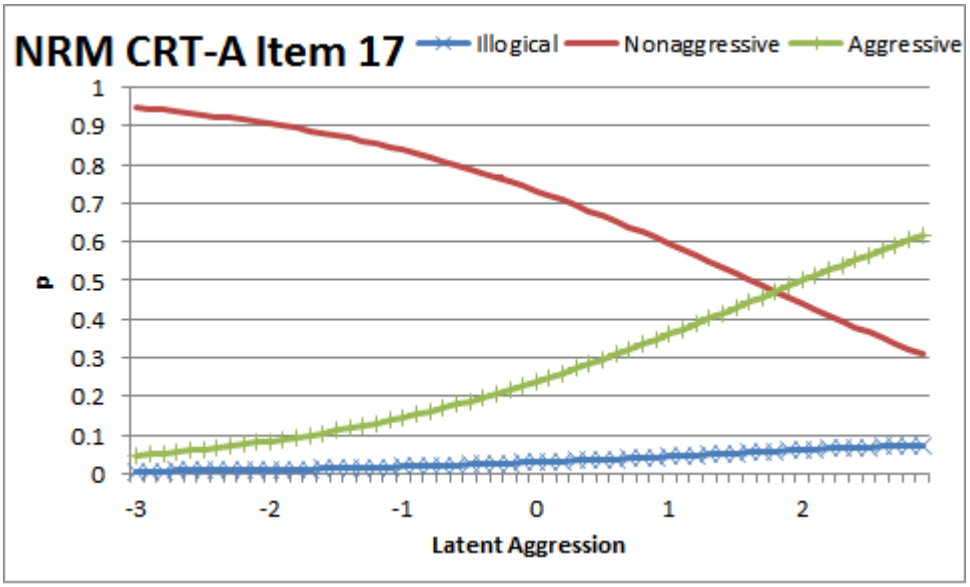


Figure 65. Two-Parameter Logistic Model item characteristic curve and information function for item 17 of the Conditional Reasoning Test of Aggression.

The linear NRM slope estimates are .101 for the aggressive logical response option, -.113 for the nonaggressive logical response option, and .012 for the combined illogical options. This is consistent with the expectation that the aggressive option will have a positive relationship with latent aggression while the nonaggressive option has a negative relationship with latent aggression. Information increases up to θ = 1.7, after which it decreases through θ = 2.9. This indicates that this item is best suited for individuals moderately high to high on latent aggression. The NRM ICC and information graph are presented in Figure 66.



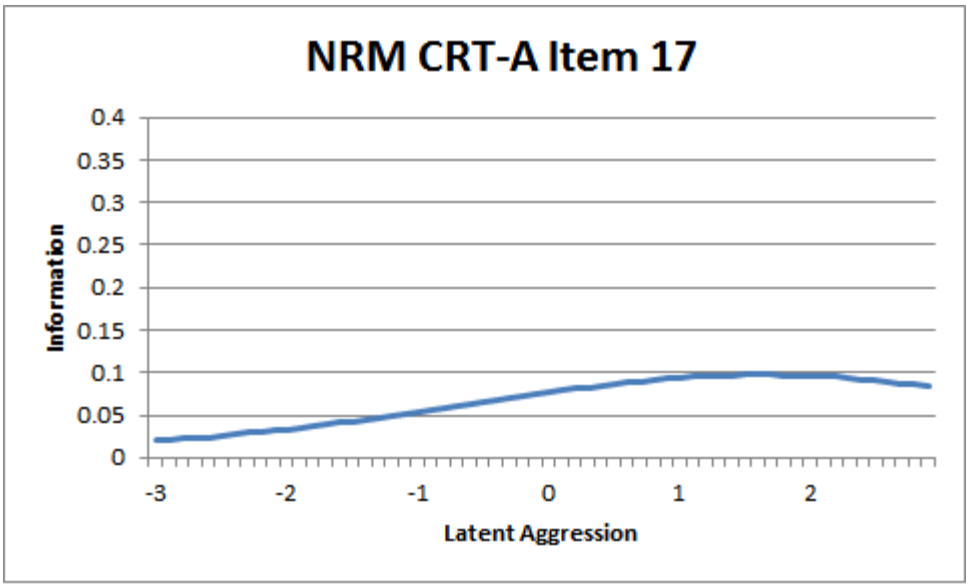


Figure 66. Nominal Response Model item characteristic curve and information function for item 17 of the Conditional Reasoning Test of Aggression.

Item seventeen displays evidence of DIF (X^2, $2df = 6.9$, $p < .05$). This chi-square value can be partitioned into slope (X^2, $1df = 1.8$, $p > .05$) and intercept (X^2, $1df = 5.1$, $p < .05$) components, indicating that student and applied samples statistically differ in terms of intercept, but not slope. A graph displaying the ICCs for each group can be found in Figure 67.

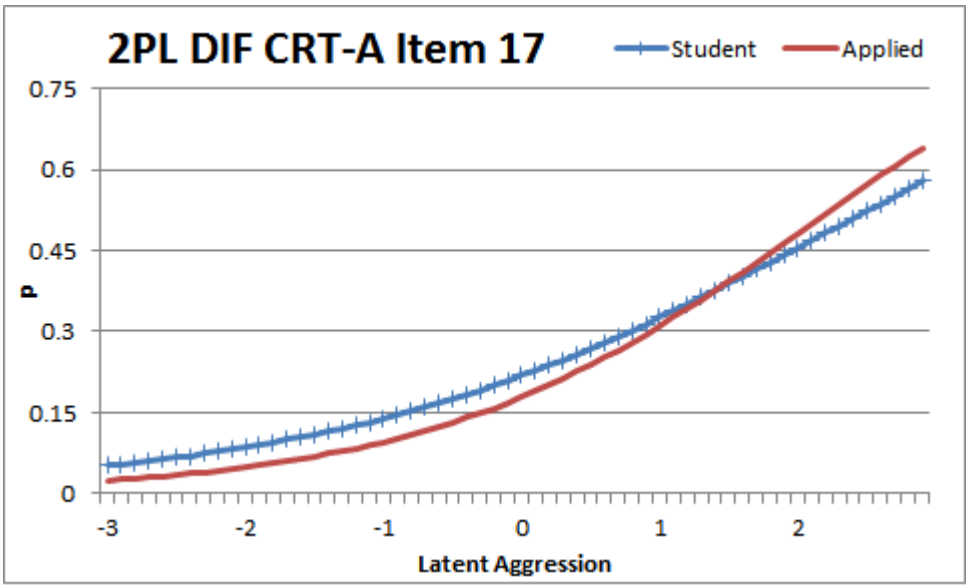


Figure 67. Two-Parameter Logistic Model differential item functioning item characteristic curve for item 17 of the Conditional Reasoning Test of Aggression.

The factor loadings for item seventeen are 0.358 for externalizing controls, 0.143 for internalizing controls, and 0.249 for powerlessness, indicating that item seventeen is primarily related to externalizing controls, but also related to internalizing controls and powerlessness. The factor-specific 2PL slope estimate for item seventeen is 0.70 with a standard error of 0.07. The 2PL intercept estimate is 1.98 with a standard error of 0.18. This indicates that item seventeen works well as an indicator of latent externalizing controls as well as latent aggression. The factor-specific 2PL ICC for item seventeen can be found in Figure 68.

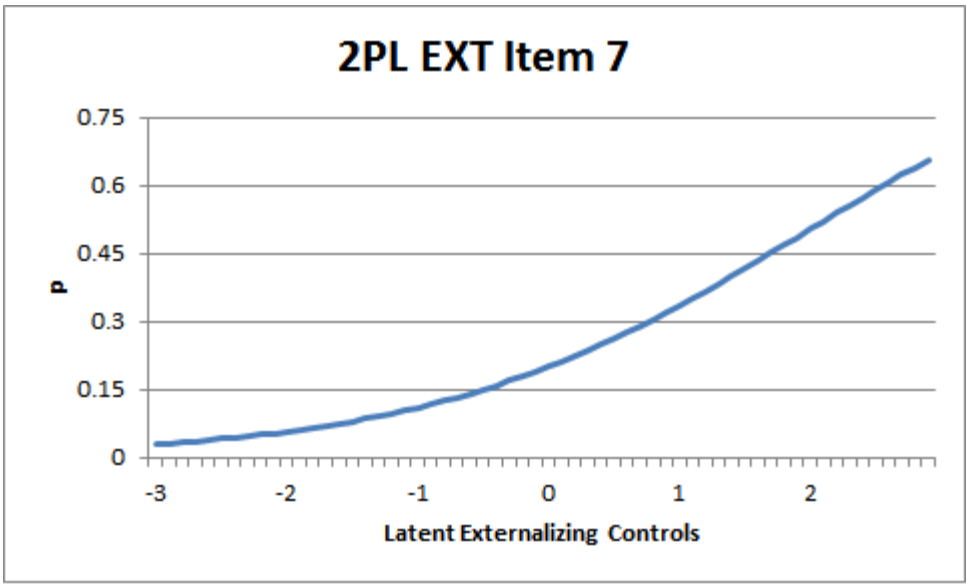


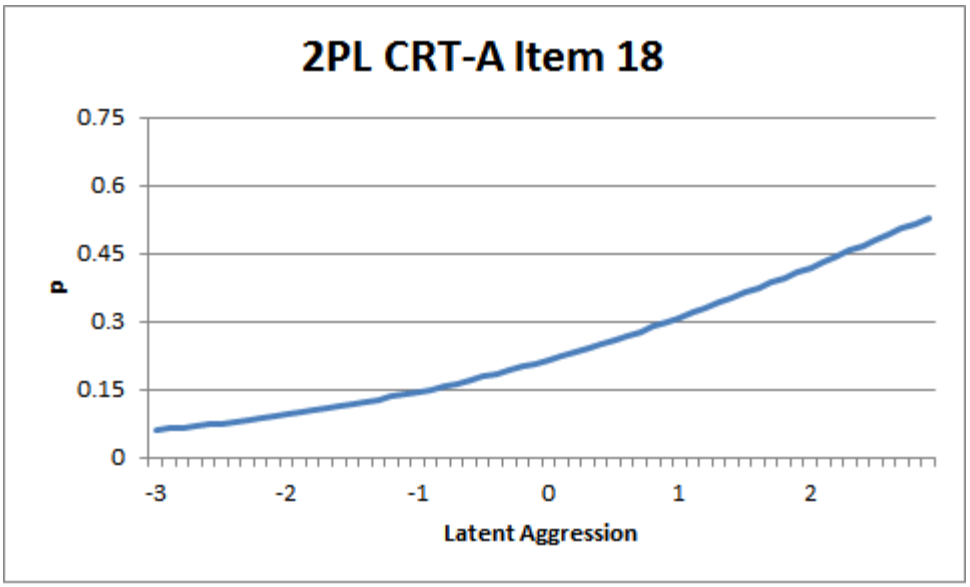
Figure 68. Two-Parameter Logistic Model item characteristic curve and information function for item 7 of the Externalizing Controls subscale of the Conditional Reasoning Test of Aggression.

Overall, item seventeen has strong item characteristics. Responses to item seventeen are well suited to indicate both latent aggression and latent externalizing controls. There is some evidence of DIF, but it is constrained to difference in intercept, meaning that item seventeen discriminates among student and applied respondents similarly.

Item 18

The eighteenth CRT-A item stem is "Wild animals often fight to see who will breed. This ensures that only the strongest animals reproduce. When strong animals reproduce, their young tend to grow into strong and powerful animals. Unlike animals, people who are not strong often reproduce. Which of the following is the most logical conclusion based on the above?" The aggressive logical response option is "Humans are becoming physically weaker." The nonaggressive logical response option is "People who are not strong can be successful." The two illogical options are "Animals breed most often in the Fall" and "The study of biology is getting less popular." Of the 5,511 responses, 22.61% chose the aggressive logical option, 75.05% chose the nonaggressive logical option, 2.09% chose one of the illogical options, and 0.25% of responses were missing. The item-total correlation for item eighteen and the dichotomously scored CRT-A is 0.12.

The 2PL slope estimate is 0.48 with a standard error of 0.05. The data are consistent with the hypothesis in that the slope coefficient is statistically greater than zero and monotonically increases with estimates of latent aggression. The 2PL intercept estimate is 2.67 with a standard error of 0.28. Information monotonically increases up to $\theta = 2.9$, indicating that this item is best suited for individuals high on latent aggression. The 2PL ICC and information graph are presented in Figure 69.



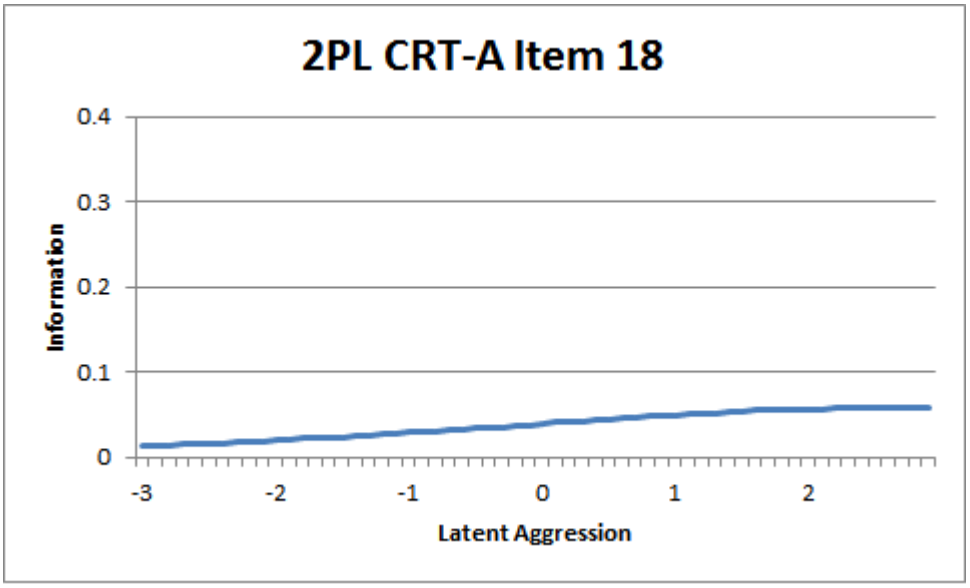


Figure 69. Two-Parameter Logistic Model item characteristic curve and information function for item 18 of the Conditional Reasoning Test of Aggression.

The linear NRM slope estimates are .084 for the aggressive logical response option, -.091 for the nonaggressive logical response option, and .008 for the combined illogical options. This is consistent with the expectation that the aggressive option will have a positive relationship with latent aggression while the nonaggressive option has a negative relationship with latent aggression. Information increases up to $\theta = 2.2$, after which it decreases through $\theta = 2.9$. This

indicates that this item is best suited for individuals high on latent aggression. The NRM ICC and information graph are presented in Figure 70.

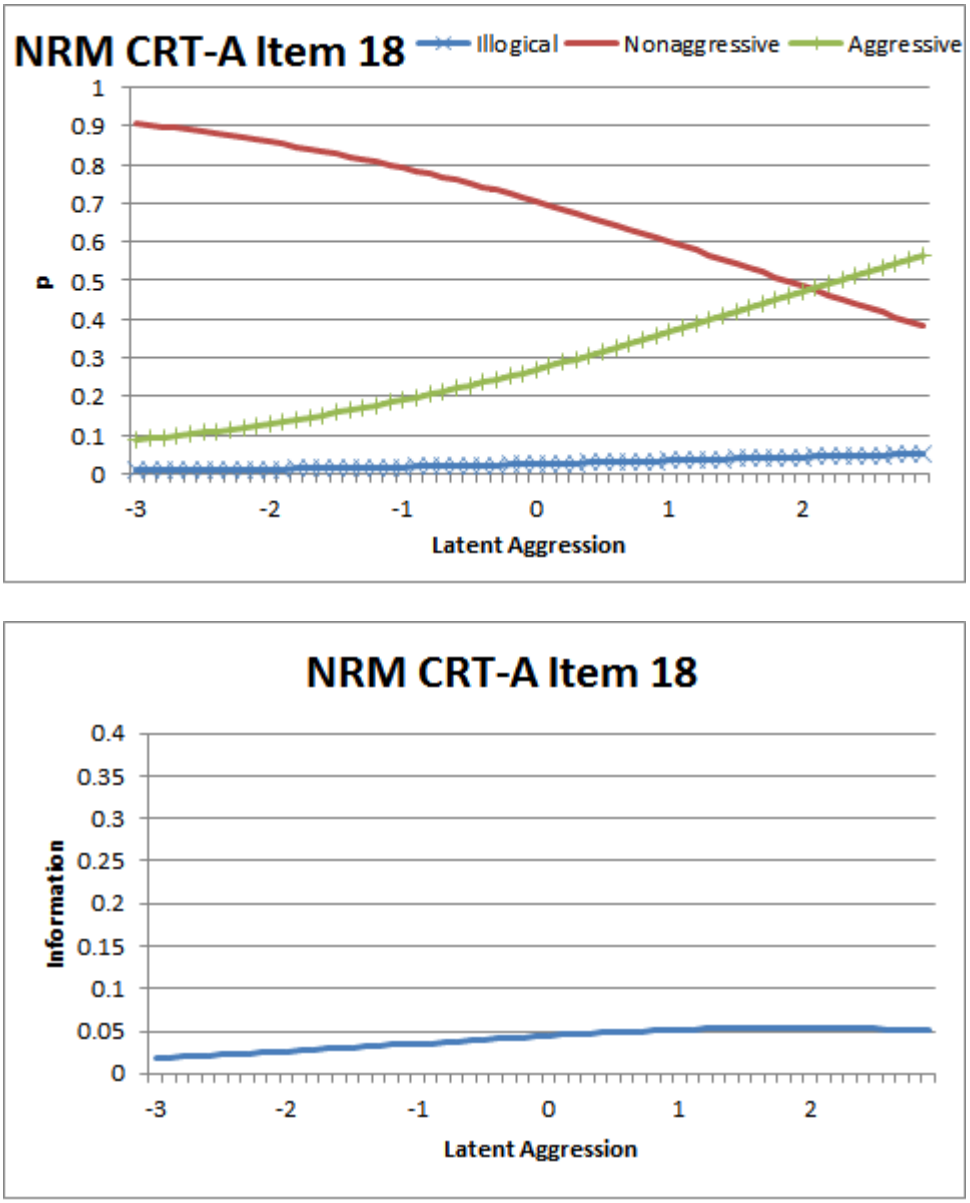


Figure 70. Nominal Response Model item characteristic curve and information function for item 18 of the Conditional Reasoning Test of Aggression.

Item eighteen displays no evidence of DIF (X^2, $2df = 2.7$, $p > .05$). This chi-square value can be partitioned into slope (X^2, $1df = 0.1$, $p > .05$) and intercept (X^2, $1df = 2.6$, $p > .05$)

components, indicating that student and applied samples do not statistically differ in either slope or intercept. A graph displaying the ICCs for each group can be found in Figure 71.

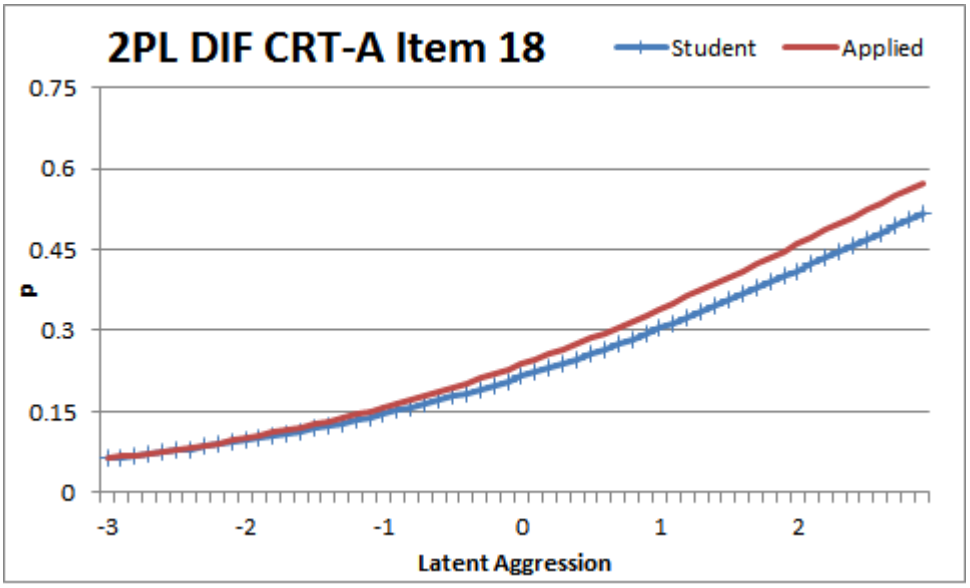


Figure 71. Two-Parameter Logistic Model differential item functioning item characteristic curve for item 18 of the Conditional Reasoning Test of Aggression.

The factor loadings for item eighteen are 0.113 for externalizing controls, 0.345 for internalizing controls, and 0.177 for powerlessness, indicating that item eighteen is primarily related to internalizing controls, but also related to externalizing controls and powerlessness. The factor-specific 2PL slope estimate for item eighteen is 0.81 with a standard error of 0.11. The 2PL intercept estimate is 1.71 with a standard error of 0.19. This indicates that item eighteen works well as an indicator of latent internalizing controls as well as latent aggression. The factor-specific 2PL ICC for item eighteen can be found in Figure 72.

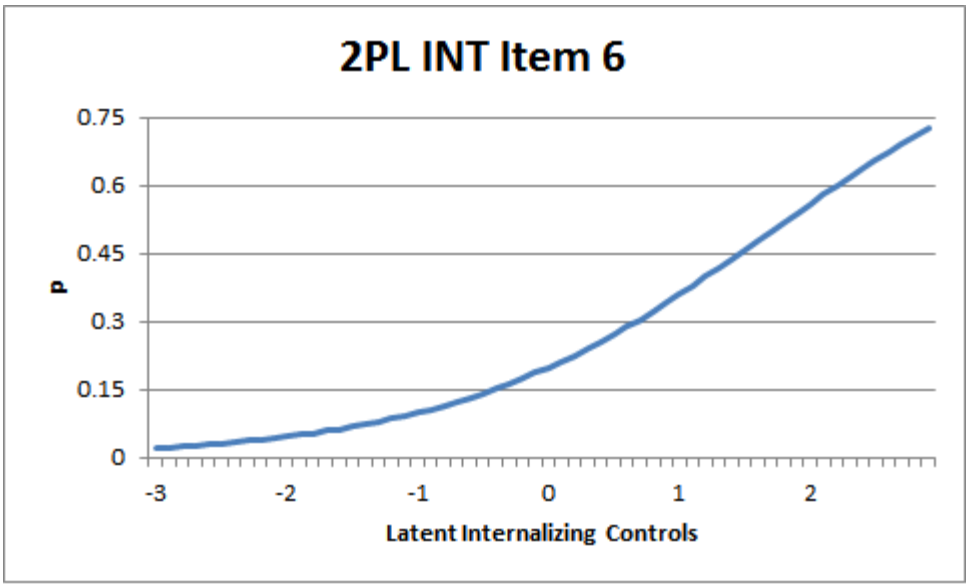


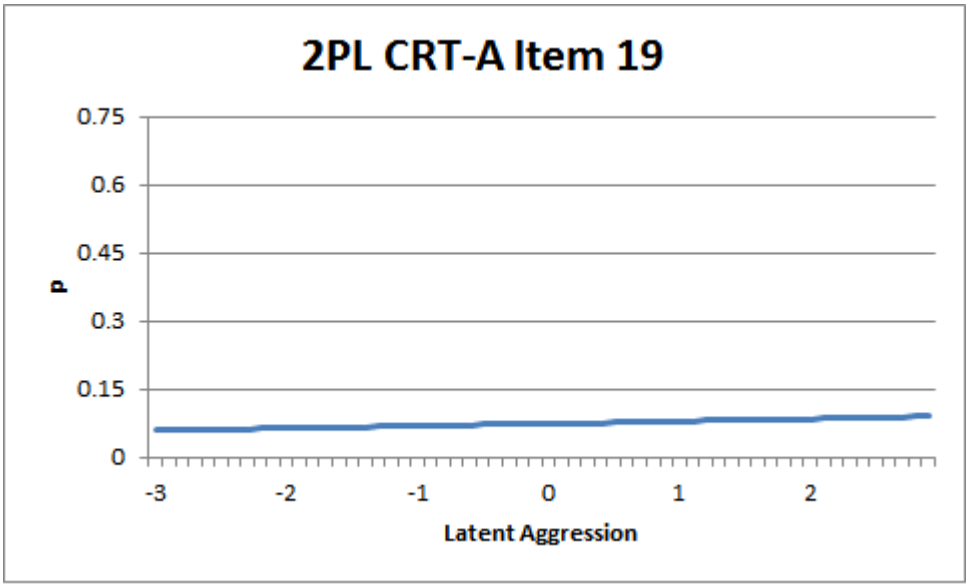
Figure 72. Two-Parameter Logistic Model item characteristic curve and information function for item 6 of the Internalizing Controls subscale of the Conditional Reasoning Test of Aggression.

Overall, item eighteen has strong item characteristics. Responses to item eighteen are well suited to indicate both latent aggression and latent internalizing controls. There is no evidence of DIF, meaning that item eighteen discriminates among student and applied respondents similarly.

Item 19

The nineteenth CRT-A item stem is "Many hold-ups take place on city streets. Hold-up victims are usually not hurt if they do everything a robber wants. Which of the following is the most logical conclusion regarding hold-up victims who do get hurt?" The aggressive logical response option is "They met a robber with a taste for violence." The nonaggressive logical response option is "They resisted, refused to turn over money, or started a fight." The two illogical options are "They were held up during the day rather than at night" and "They were able to outrun their attacker." Of the 5,511 responses, 7.42% chose the aggressive logical option, 88.86% chose the nonaggressive logical option, 3.57% chose one of the illogical options, and 0.15% of responses were missing. The item-total correlation for item nineteen and the dichotomously scored CRT-A is less than 0.01.

The 2PL slope estimate is 0.08 with a standard error of 0.08. The data for this item are not consistent with the hypothesis. Although the item has a positive slope, it is not large enough in magnitude to be statistically differentiated from zero. The 2PL intercept estimate is 33.61 with a standard error of 34.88. The information function is negligible and does not seem to change at different levels of latent aggression. The 2PL ICC and information graph are presented in Figure 73.



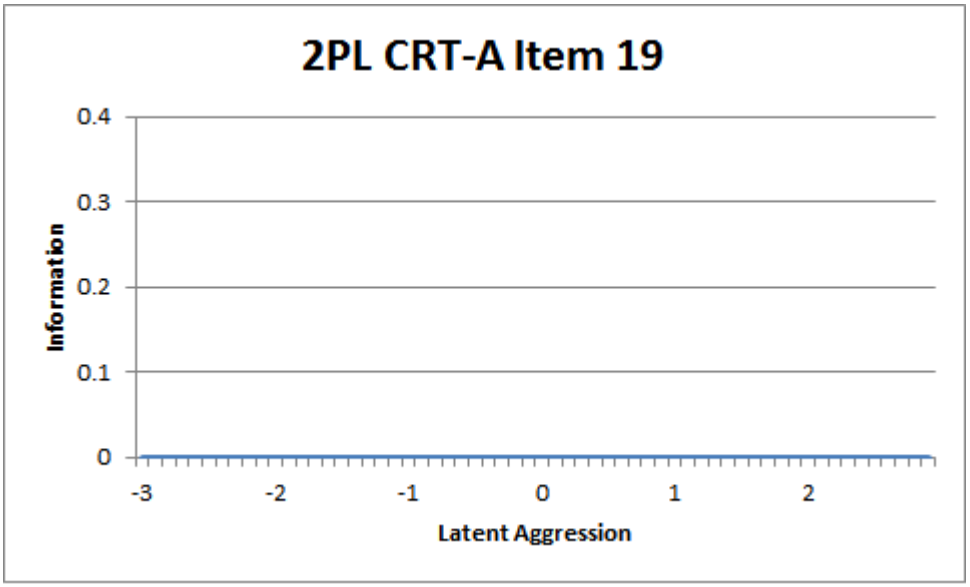
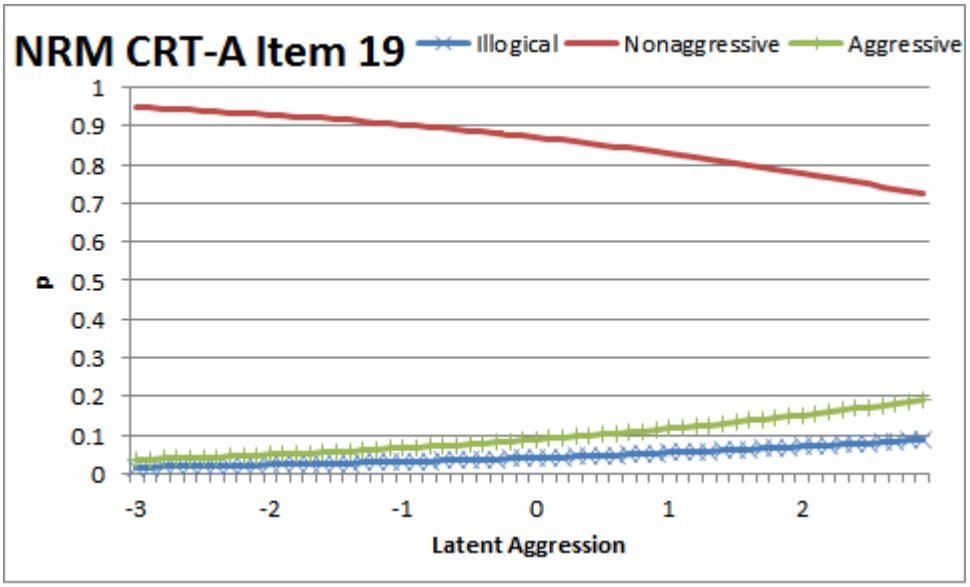


Figure 73. Two-Parameter Logistic Model item characteristic curve and information function for item 19 of the Conditional Reasoning Test of Aggression.

The linear NRM slope estimates are .026 for the aggressive logical response option, -.037 for the nonaggressive logical response option, and .012 for the combined illogical options. This is consistent with the expectation that the aggressive option will have a positive relationship with latent aggression while the nonaggressive option has a negative relationship with latent aggression. Information monotonically increases up to $\theta = 2.9$, indicating that this item is best suited for individuals high on latent aggression. The NRM ICC and information graph are presented in Figure 74.



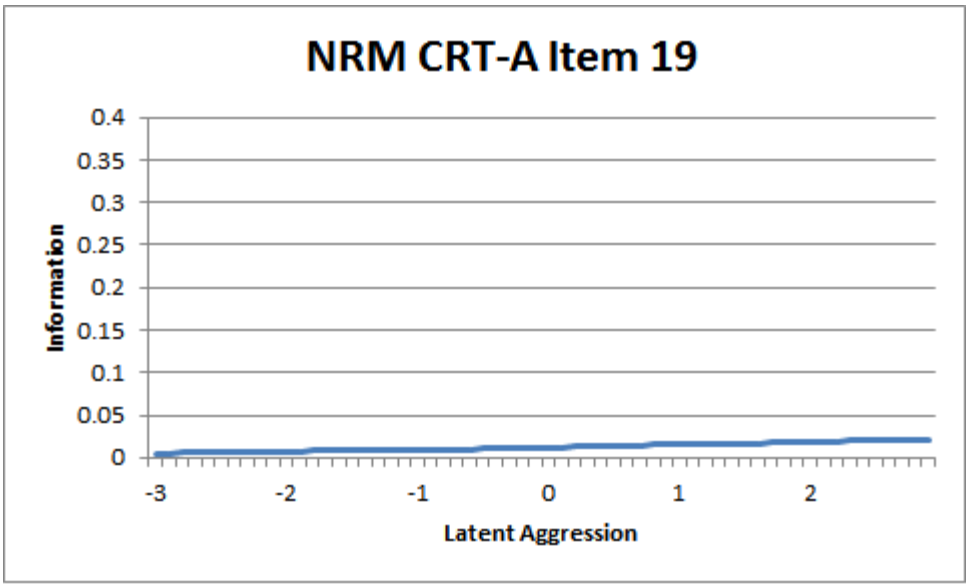


Figure 74. Nominal Response Model item characteristic curve and information function for item 19 of the Conditional Reasoning Test of Aggression.

Item nineteen displays no evidence of DIF (X^2, $2df = 1.9$, $p > .05$). This chi-square value can be partitioned into slope (X^2, $1df = 0.1$, $p > .05$) and intercept (X^2, $1df = 1.8$, $p > .05$) components, indicating that student and applied samples do not statistically differ in either slope or intercept. A graph displaying the ICCs for each group can be found in Figure 75.

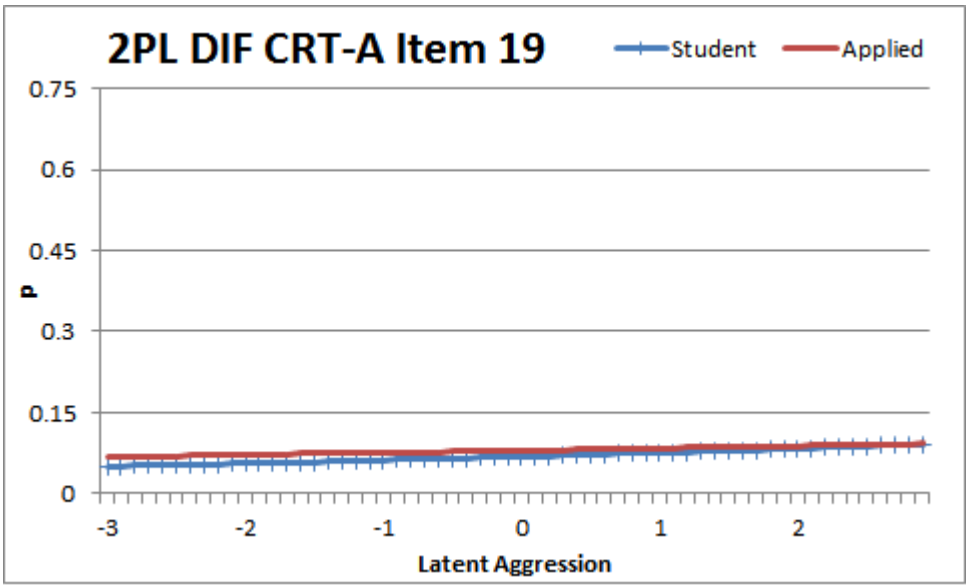


Figure 75. Two-Parameter Logistic Model differential item functioning item characteristic curve for item 19 of the Conditional Reasoning Test of Aggression.

The factor loadings for item nineteen are 0.435 for externalizing controls, 0.177 for internalizing controls, and -0.053 for powerlessness, indicating that item nineteen is primarily related to externalizing controls, but also related to internalizing controls. The factor-specific 2PL slope estimate for item nineteen is 0.34 with a standard error of 0.09. The 2PL intercept estimate is 7.61 with a standard error of 1.97. This indicates that item nineteen works better as an indicator of latent externalizing controls than an indicator of latent aggression. The factor-specific 2PL ICC for item nineteen can be found in Figure 76.

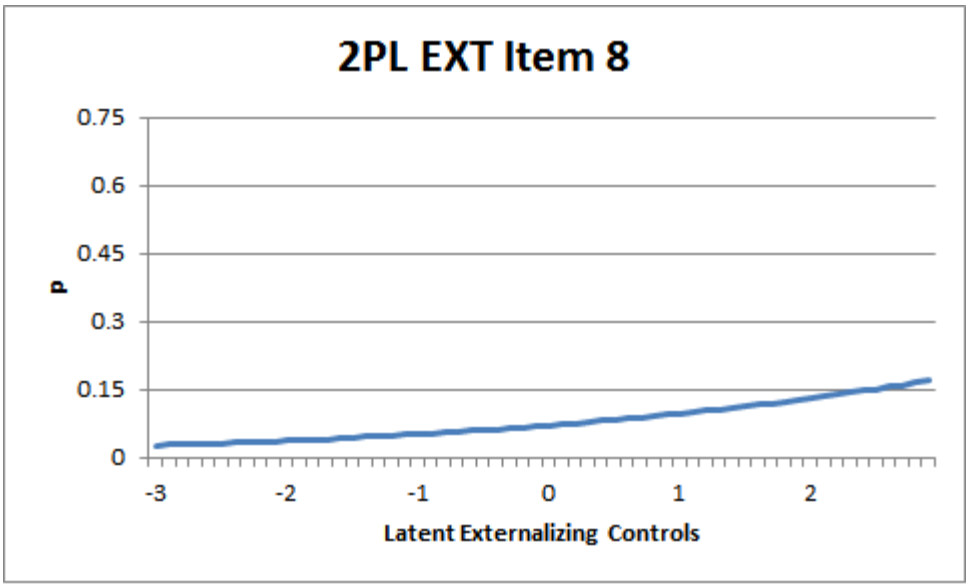


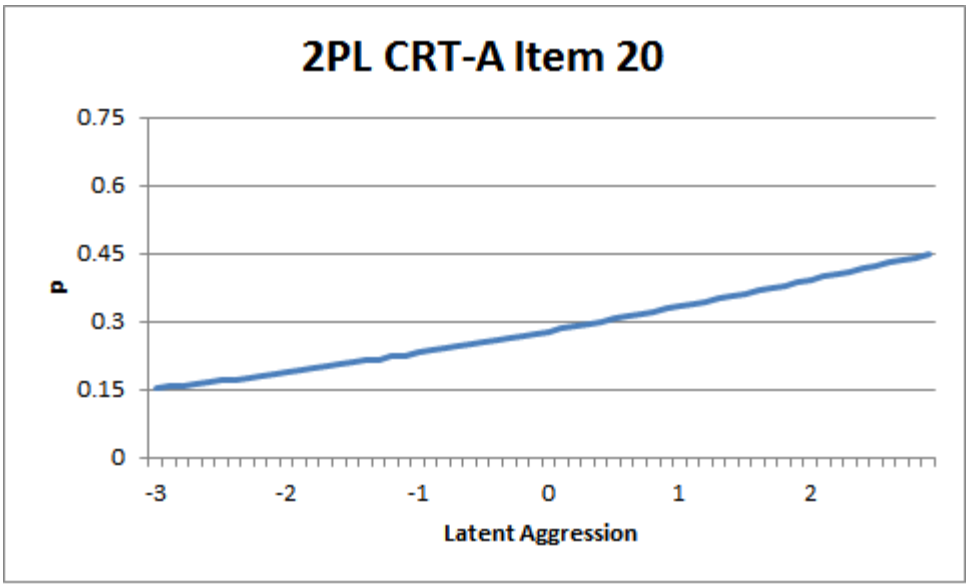
Figure 76. Two-Parameter Logistic Model item characteristic curve and information function for item 8 of the Externalizing Controls subscale of the Conditional Reasoning Test of Aggression.

Overall, item nineteen has weak item characteristics. Responses to item nineteen are well suited to indicate latent externalizing controls, but not well suited to indicate latent aggression. There is no evidence of DIF, meaning that item nineteen discriminates among student and applied respondents similarly.

Item 20

The twentieth CRT-A item stem is "Half of all marriages end in divorce. One reason for the large number of divorces is that getting a divorce is quick and easy. If a couple can agree on how to split their property fairly, then they can get a divorce simply by filling out forms and taking them to court. They do not need lawyers. Which of the following is the most logical conclusion based on the above?" The aggressive logical response option is "If one's husband or wife hires a lawyer, then he or she is not planning to play fair." The nonaggressive logical response option is "Couples might get back together if getting a divorce took longer." The two illogical options are "People are older when they get married" and "More men than women get divorced." Of the 5,511 responses, 28.11% chose the aggressive logical option, 67.34% chose the nonaggressive logical option, 4.21% chose one of the illogical options, and 0.34% of responses were missing. The item-total correlation for item twenty and the dichotomously scored CRT-A is 0.06.

The 2PL slope estimate is 0.26 with a standard error of 0.05. The data are consistent with the hypothesis in that the slope coefficient is statistically greater than zero and monotonically increases with estimates of latent aggression. The 2PL intercept estimate is 3.71 with a standard error of 0.67. Information monotonically increases up to $\theta = 2.9$, indicating that this item is best suited for individuals high on latent aggression. The 2PL ICC and information graph are presented in Figure 77.



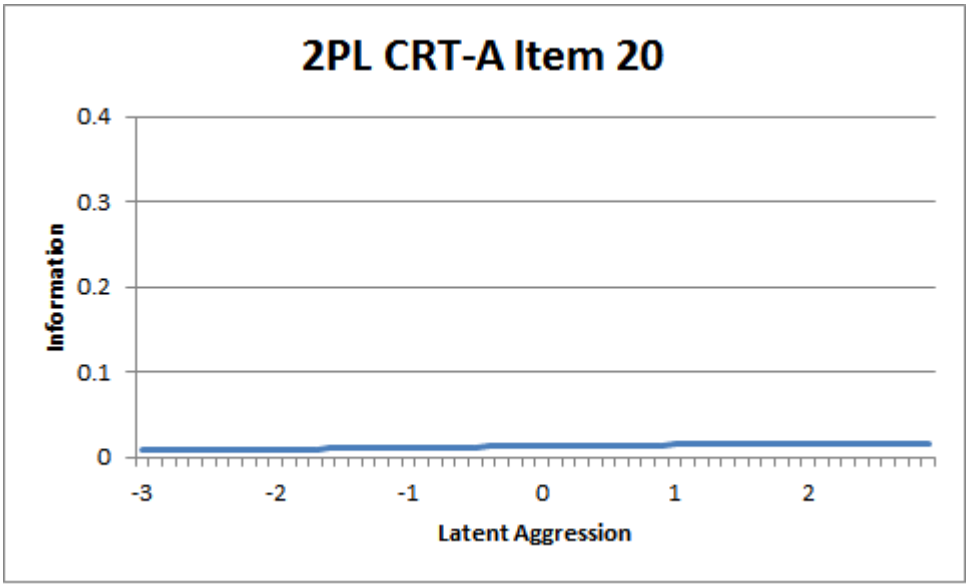


Figure 77. Two-Parameter Logistic Model item characteristic curve and information function for item 20 of the Conditional Reasoning Test of Aggression.

The linear NRM slope estimates are .061 for the aggressive logical response option, -.071 for the nonaggressive logical response option, and .009 for the combined illogical options. This is consistent with the expectation that the aggressive option will have a positive relationship with latent aggression while the nonaggressive option has a negative relationship with latent aggression. Information increases up to $\theta = 2.7$, after which it decreases through $\theta = 2.9$. This

indicates that this item is best suited for individuals high on latent aggression. The NRM ICC and information graph are presented in Figure 78.

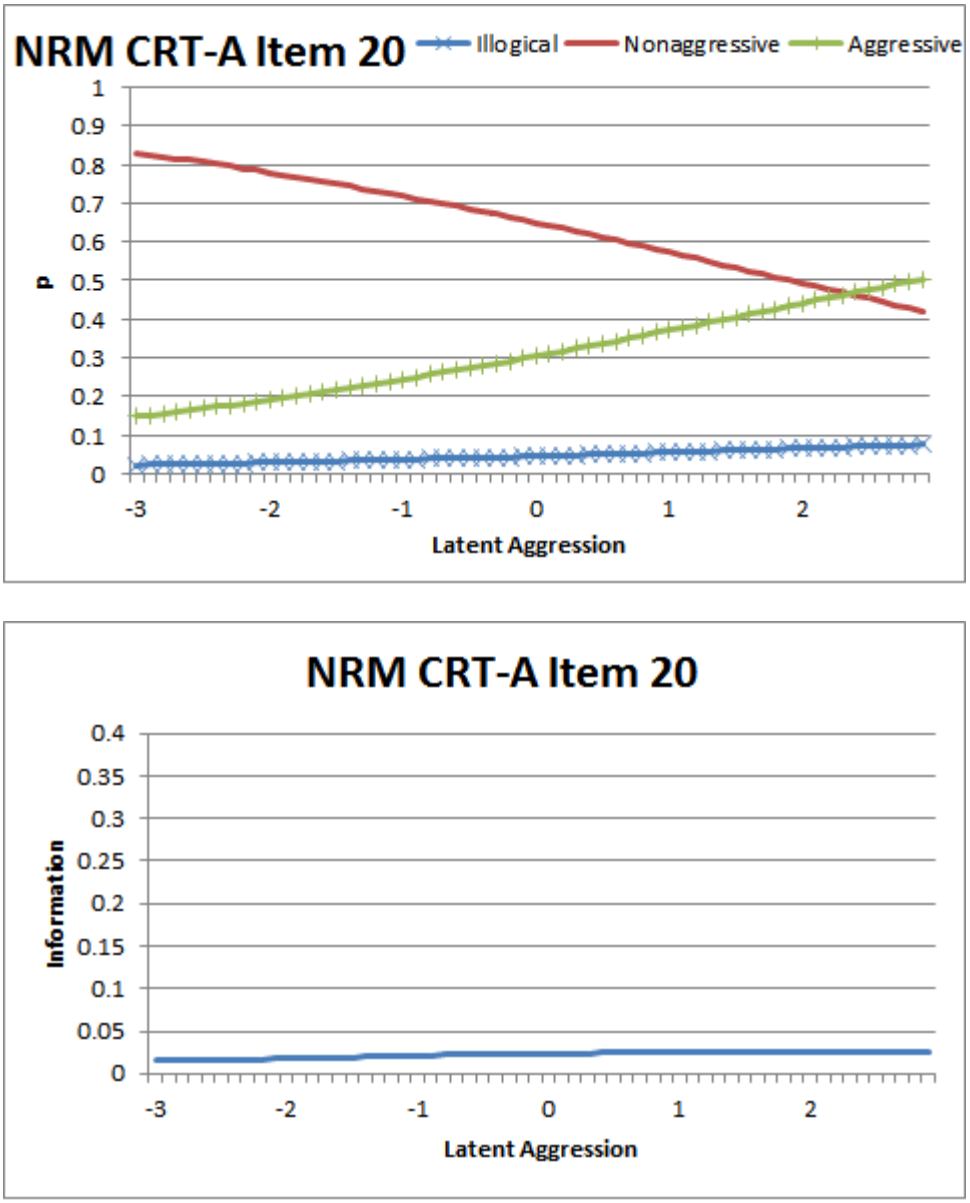


Figure 78. Nominal Response Model item characteristic curve and information function for item 20 of the Conditional Reasoning Test of Aggression.

Item twenty displays no evidence of DIF (X^2, $2df = 2.8$, $p > .05$). This chi-square value can be partitioned into slope (X^2, $1df = 2.2$, $p > .05$) and intercept (X^2, $1df = 0.6$, $p > .05$)

components, indicating that student and applied samples do not statistically differ in either slope or intercept. A graph displaying the ICCs for each group can be found in Figure 79.

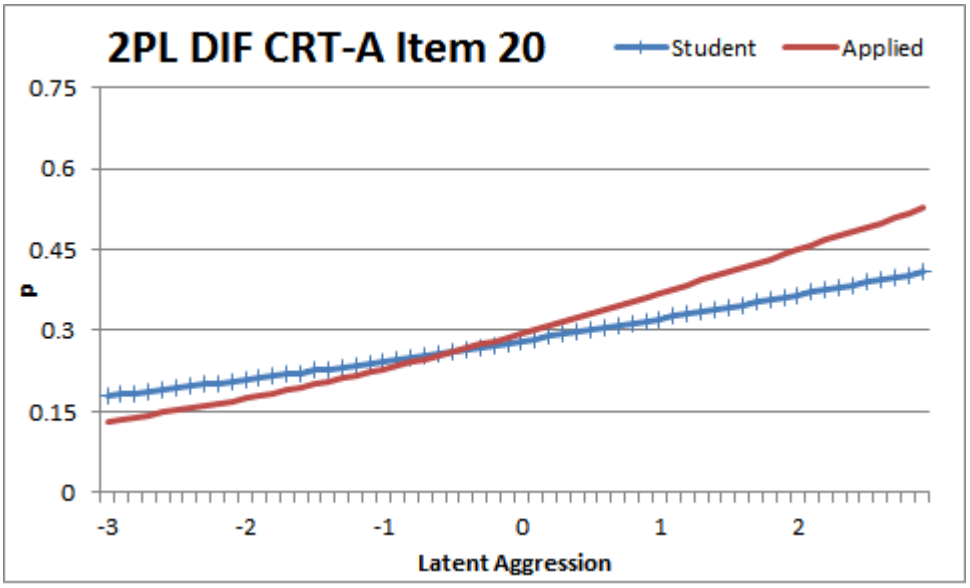


Figure 79. Two-Parameter Logistic Model differential item functioning item characteristic curve for item 20 of the Conditional Reasoning Test of Aggression.

The factor loadings for item twenty are 0.238 for externalizing controls, 0.173 for internalizing controls, and -0.085 for powerlessness, indicating that item twenty is primarily related to externalizing controls, but also related to internalizing controls. The factor-specific 2PL slope estimate for item twenty is 0.30 with a standard error of 0.06. The 2PL intercept estimate is 3.21 with a standard error of 0.59. This indicates that item twenty works well as an indicator of latent externalizing controls as well as latent aggression. The factor-specific 2PL ICC for item twenty can be found in Figure 80.

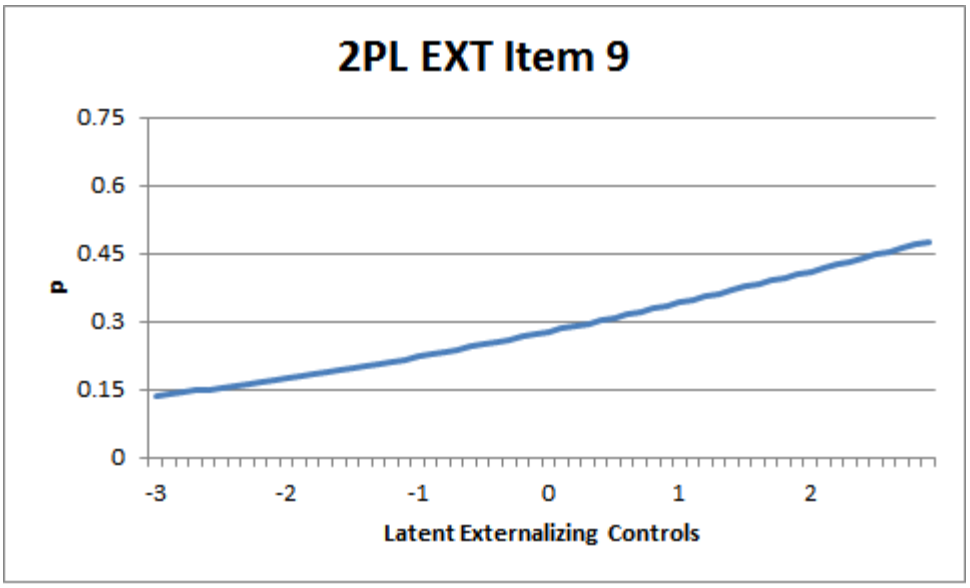


Figure 80. Two-Parameter Logistic Model item characteristic curve and information function for item 9 of the Externalizing Controls subscale of the Conditional Reasoning Test of Aggression.

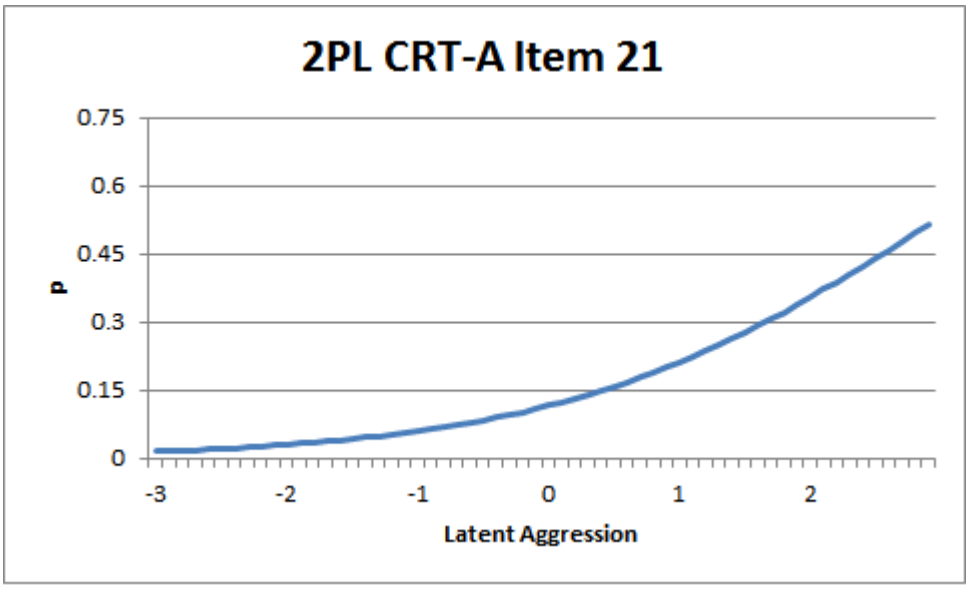
Overall, item twenty has strong item characteristics. Responses to item twenty are well suited to indicate both latent aggression and latent externalizing controls. There is no evidence of DIF, meaning that item twenty discriminates among student and applied respondents similarly.

Item 21

The twenty-one CRT-A item stem is "Some companies treat employees badly. For example, some companies lay people off and then expect one person to do the work of two people. Managers get big raises in some companies, but employees get only small increases. To get even, some employees have damaged company equipment, slacked off on the job, or faked being sick. However, most employees do not act in these ways. Which of the following is the most logical conclusion based on the above?" The aggressive logical response option is "Most employees are afraid of being caught." The nonaggressive logical response option is "Most employees value good behavior at work." The two illogical options are "Most employees never get sick" and "Most employees drive to work rather than walk." Of the 5,511 responses, 13.52% chose the aggressive logical option, 83.34% chose the nonaggressive logical option, 2.74% chose one of the illogical options, and 0.40% of responses

were missing. The item-total correlation for item twenty-one and the dichotomously scored CRT-A is 0.15.

The 2PL slope estimate is 0.72 with a standard error of 0.07. The data are consistent with the hypothesis in that the slope coefficient is statistically greater than zero and monotonically increases with estimates of latent aggression. The 2PL intercept estimate is 2.83 with a standard error of 0.23. Information monotonically increases up to $\theta = 2.9$, indicating that this item is best suited for individuals high on latent aggression. The 2PL ICC and information graph are presented in Figure 81.



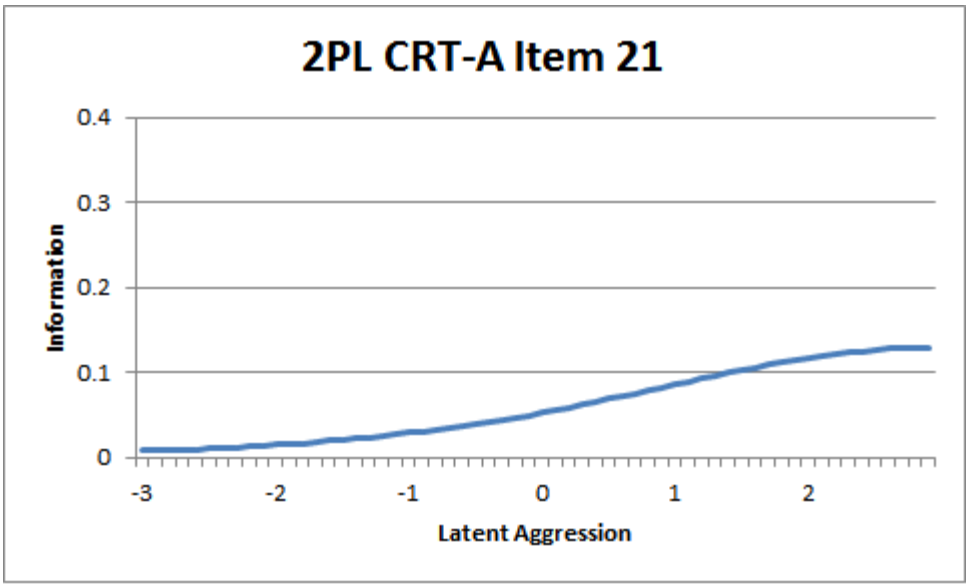
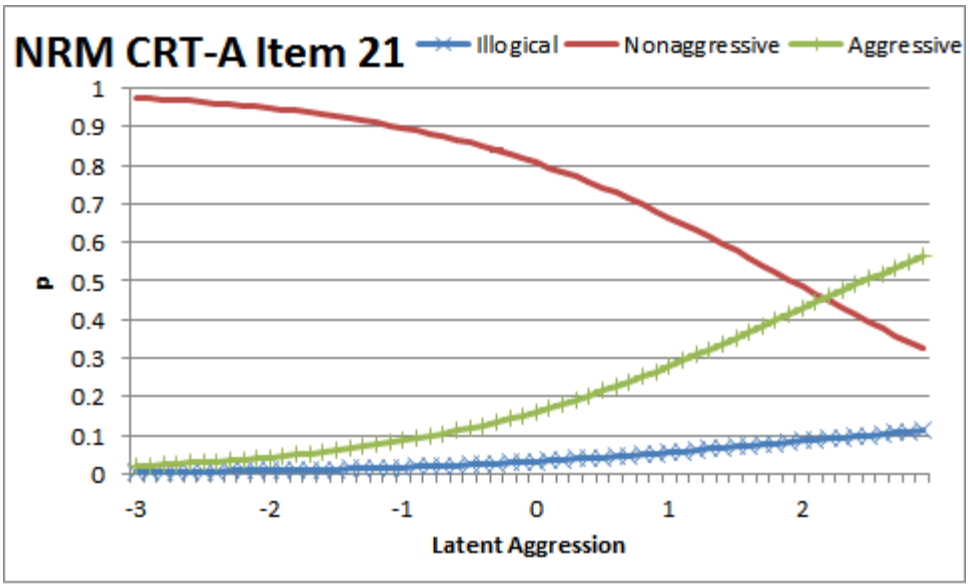


Figure 81. Two-Parameter Logistic Model item characteristic curve and information function for item 21 of the Conditional Reasoning Test of Aggression.

The linear NRM slope estimates are .094 for the aggressive logical response option, -.113 for the nonaggressive logical response option, and .019 for the combined illogical options. This is consistent with the expectation that the aggressive option will have a positive relationship with latent aggression while the nonaggressive option has a negative relationship with latent aggression. Information increases up to $\theta = 2.1$, after which it decreases through $\theta = 2.9$. This indicates that this item is best suited for individuals high on latent aggression. The NRM ICC and information graph are presented in Figure 82.



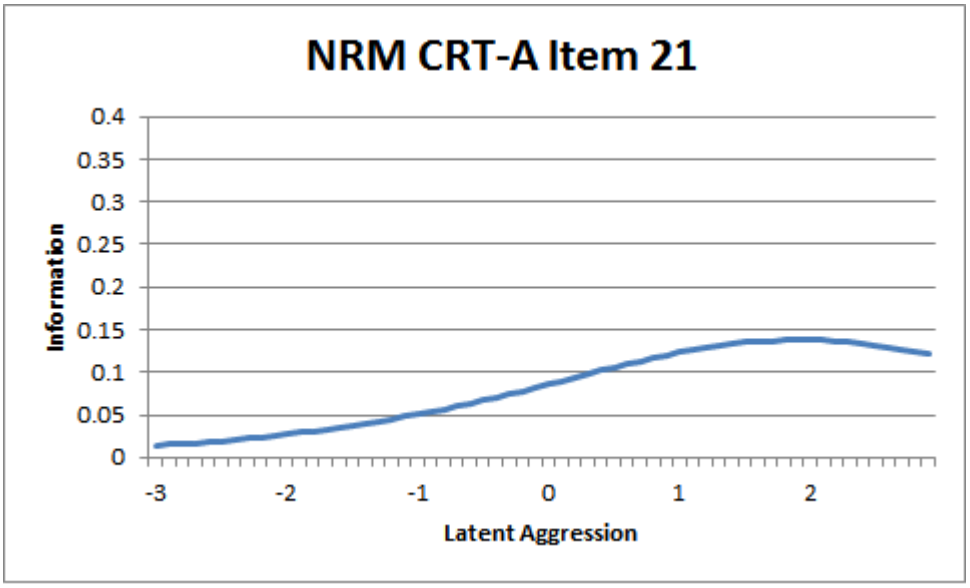


Figure 82. Nominal Response Model item characteristic curve and information function for item 21 of the Conditional Reasoning Test of Aggression.

Item twenty-one displays no evidence of overall DIF (X^2, $2df = 4.7$, $p > .05$), although after partitioning the chi-square into slope (X^2, $1df = 4.3$, $p < .05$) and intercept (X^2, $1df = 0.4$, $p > .05$) components, results indicate that student and applied samples statistically differ in terms of slope, but not intercept. A graph displaying the ICCs for each group can be found in Figure 83.

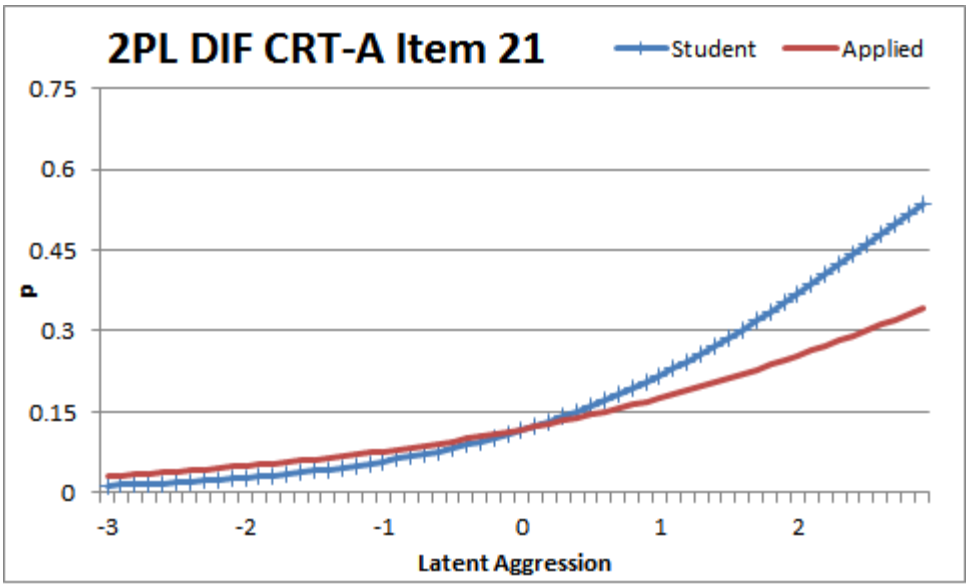


Figure 83. Two-Parameter Logistic Model differential item functioning item characteristic curve for item 21 of the Conditional Reasoning Test of Aggression.

The factor loadings for item twenty-one are 0.530 for externalizing controls, 0.296 for internalizing controls, and 0.130 for powerlessness, indicating that item twenty-one is primarily related to externalizing controls, but also related to internalizing controls and powerlessness. The factor-specific 2PL slope estimate for item twenty-one is 0.84 with a standard error of 0.09. The 2PL intercept estimate is 2.49 with a standard error of 0.21. This indicates that item twenty-one works well as an indicator of latent externalizing controls as well as latent aggression. The factor-specific 2PL ICC for item twenty-one can be found in Figure 84.

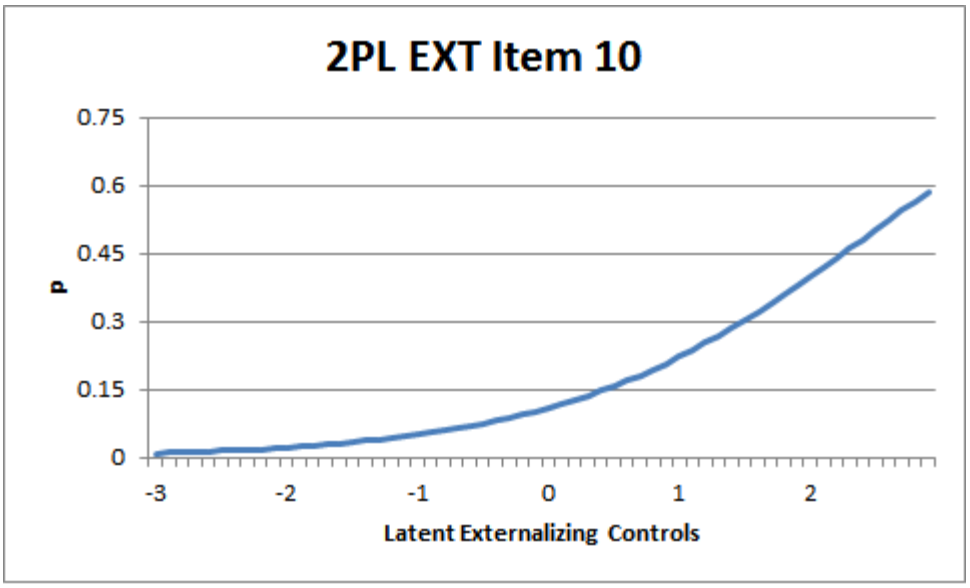


Figure 84. Two-Parameter Logistic Model item characteristic curve and information function for item 10 of the Externalizing Controls subscale of the Conditional Reasoning Test of Aggression.

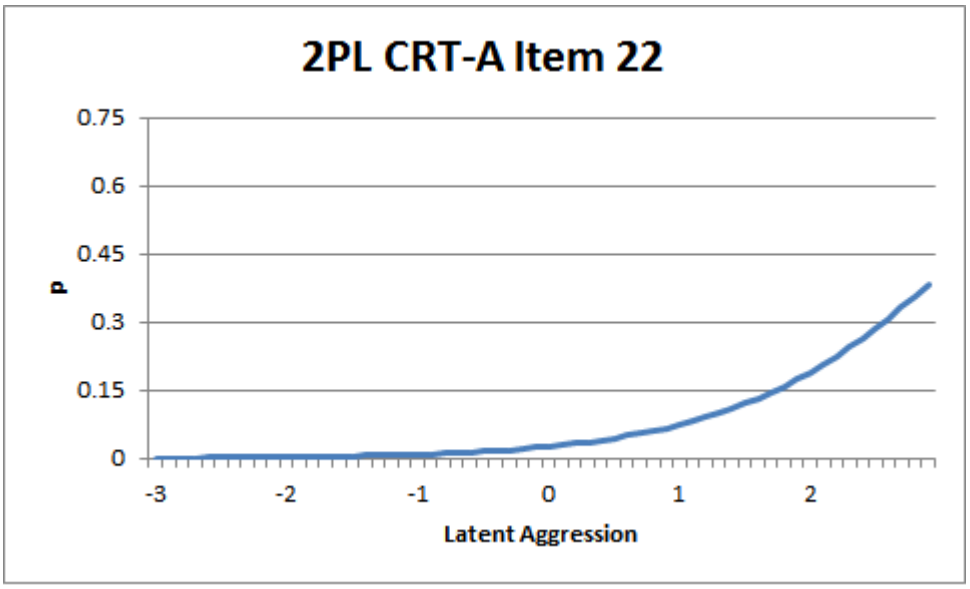
Overall, item twenty-one has strong item characteristics. Responses to item twenty-one are well suited to indicate both latent aggression and latent externalizing controls. There is some evidence of DIF, indicating that item twenty-one may discriminate among student respondents differently than it discriminates among applied respondents.

Item 22

The twenty-two CRT-A item stem is "Germany took over many small countries before World War II. Other countries thought that they could stop Germany. They had Germany sign agreements promising not to attack again. Germany broke these promises many times. Which of the following is the most logical conclusion based on the above?" The aggressive logical response option is "Only weak countries follow agreements." The nonaggressive logical response option is "Signing agreements works best when all countries can be trusted." The two illogical options are "England should not have invaded France" and "Small countries are always more powerful than large countries." Of the 5,511 responses, 4.28% chose the aggressive logical option, 93.40% chose the nonaggressive logical

option, 1.67% chose one of the illogical options, and 0.65% of responses were missing. The item-total correlation for item twenty-two and the dichotomously scored CRT-A is 0.14.

The 2PL slope estimate is 1.08 with a standard error of 0.12. The data are consistent with the hypothesis in that the slope coefficient is statistically greater than zero and monotonically increases with estimates of latent aggression. The 2PL intercept estimate is 3.35 with a standard error of 0.27. Information monotonically increases up to $\theta = 2.9$, indicating that this item is best suited for individuals high on latent aggression. The 2PL ICC and information graph are presented in Figure 85.



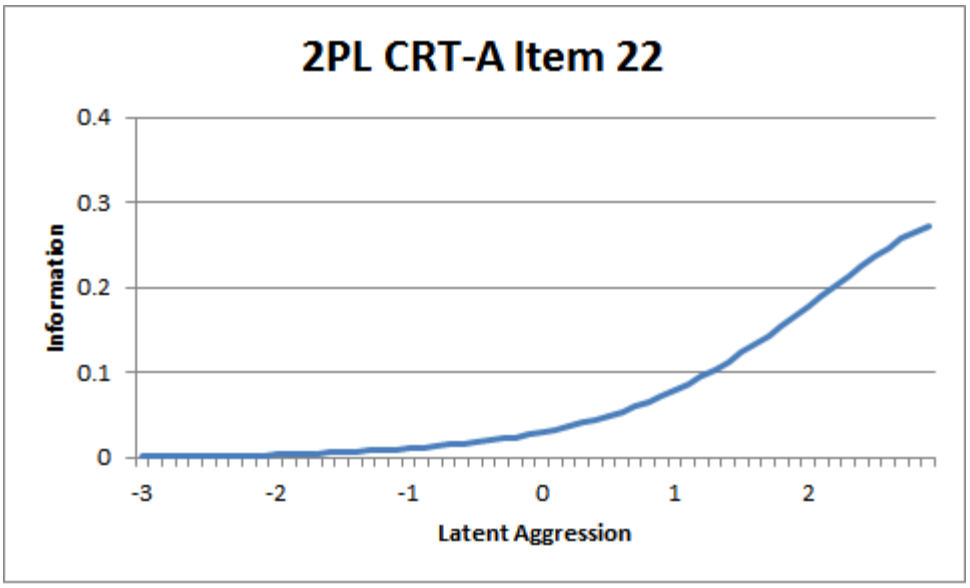
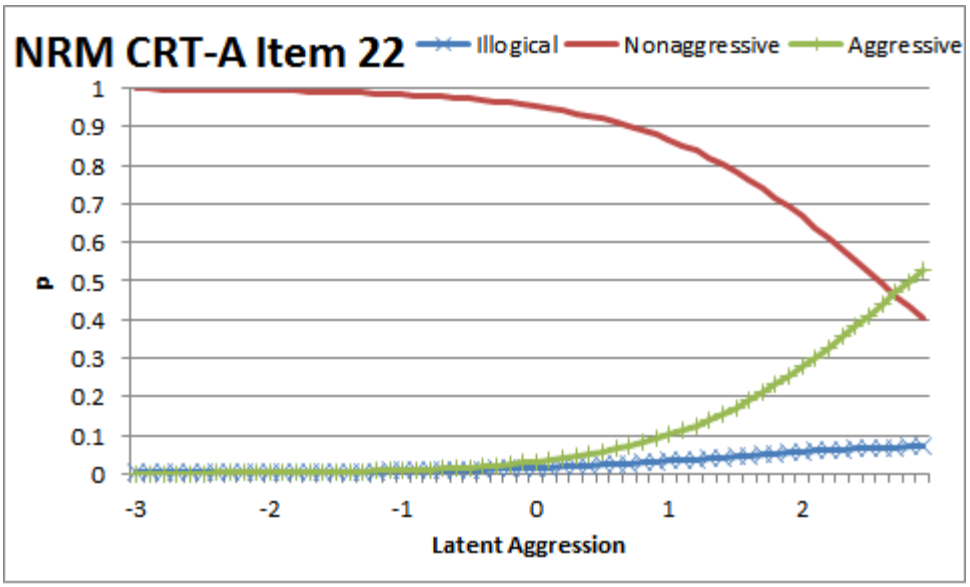


Figure 85. Two-Parameter Logistic Model item characteristic curve and information function for Item 22 of the Conditional Reasoning Test of Aggression.

The linear NRM slope estimates are .072 for the aggressive logical response option, -.085 for the nonaggressive logical response option, and .013 for the combined illogical options. This is consistent with the expectation that the aggressive option will have a positive relationship with latent aggression while the nonaggressive option has a negative relationship with latent aggression. Information increases up to $\theta = 2.7$, after which it decreases through $\theta = 2.9$. This indicates that this item is best suited for individuals high on latent aggression. The NRM ICC and information graph are presented in Figure 86.



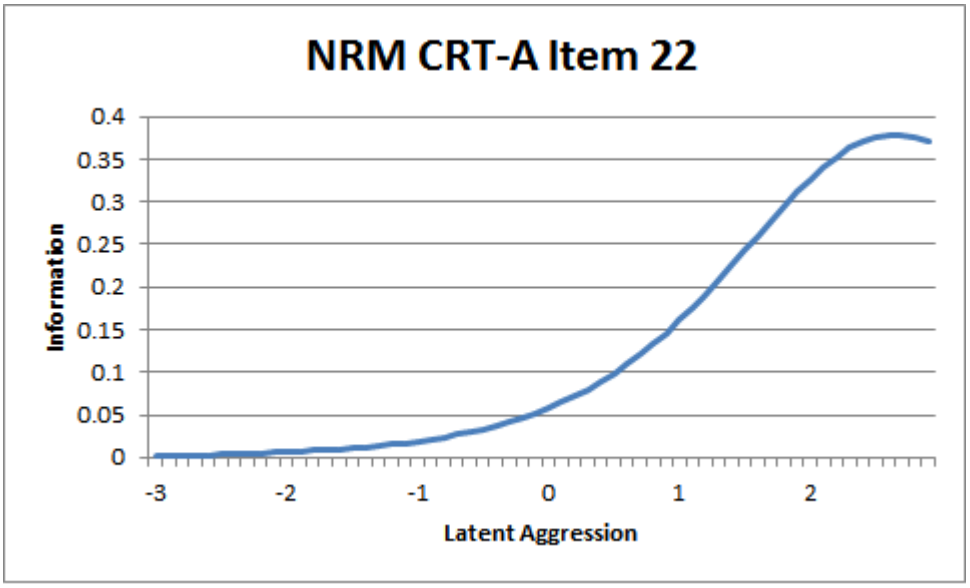


Figure 86. Nominal Response Model item characteristic curve and information function for Item 22 of the Conditional Reasoning Test of Aggression.

Item twenty-two displays no evidence of DIF (X^2, $2df = 0.4$, $p > .05$). This chi-square value can be partitioned into slope (X^2, $1df = 0.4$, $p > .05$) and intercept (X^2, $1df < 0.1$, $p > .05$) components, indicating that student and applied samples do not statistically differ in either slope or intercept. A graph displaying the ICCs for each group can be found in Figure 87.

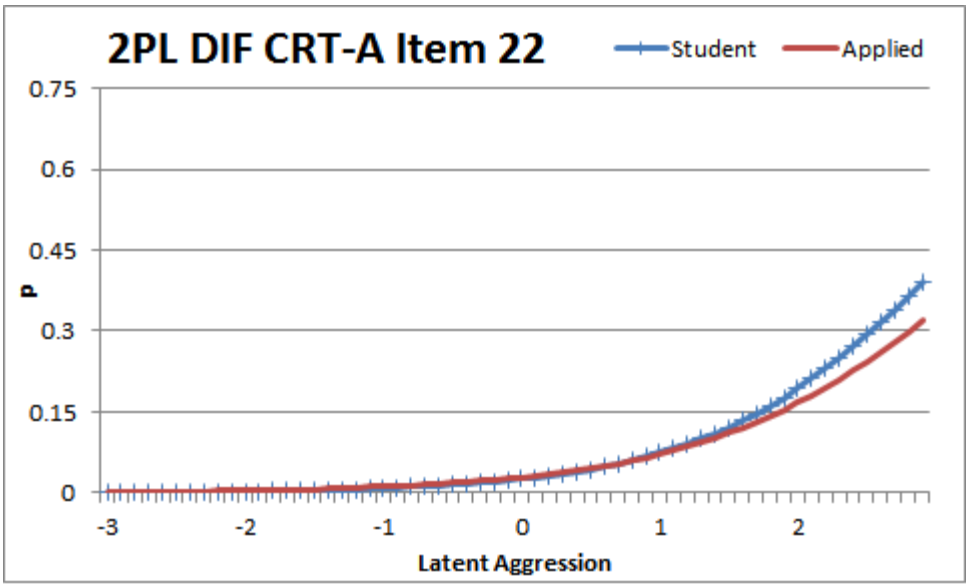


Figure 87. Two-Parameter Logistic Model differential item functioning item characteristic curve for Item 22 of the Conditional Reasoning Test of Aggression.

The factor loadings for item twenty-two are 0.588 for externalizing controls, 0.395 for internalizing controls, and 0.107 for powerlessness, indicating that item twenty-two is primarily related to externalizing controls, but also related to internalizing controls and powerlessness. The factor-specific 2PL slope estimate for item twenty-two is 1.10 with a standard error of 0.14. The 2PL intercept estimate is 3.30 with a standard error of 0.31.This indicates that item twenty-two works well as an indicator of latent externalizing controls as well as latent aggression. The factor-specific 2PL ICC for item twenty-two can be found in Figure 88.

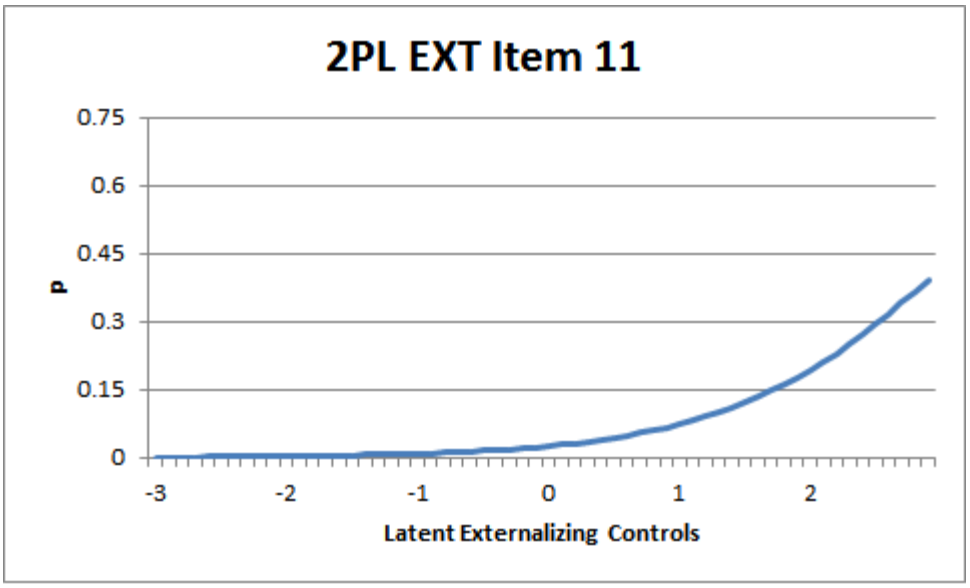


Figure 88. Two-Parameter Logistic Model item characteristic curve and information function for item 11 of the Externalizing Controls subscale of the Conditional Reasoning Test of Aggression.

Overall, item twenty-two has strong item characteristics. Responses to item twenty-two are well suited to indicate both latent aggression and latent externalizing controls. There is no evidence of DIF, meaning that item twenty-two discriminates among student and applied respondents similarly.

Conditional Reasoning Test of Aggression Total Information

The total information function for the CRT-A increases monotonically up to $\theta = 2.9$. This indicates that the CRT-A is best suited for the assessment of individuals high in latent aggression. The total information and standard error for the CRT-A are displayed in Figure 89.

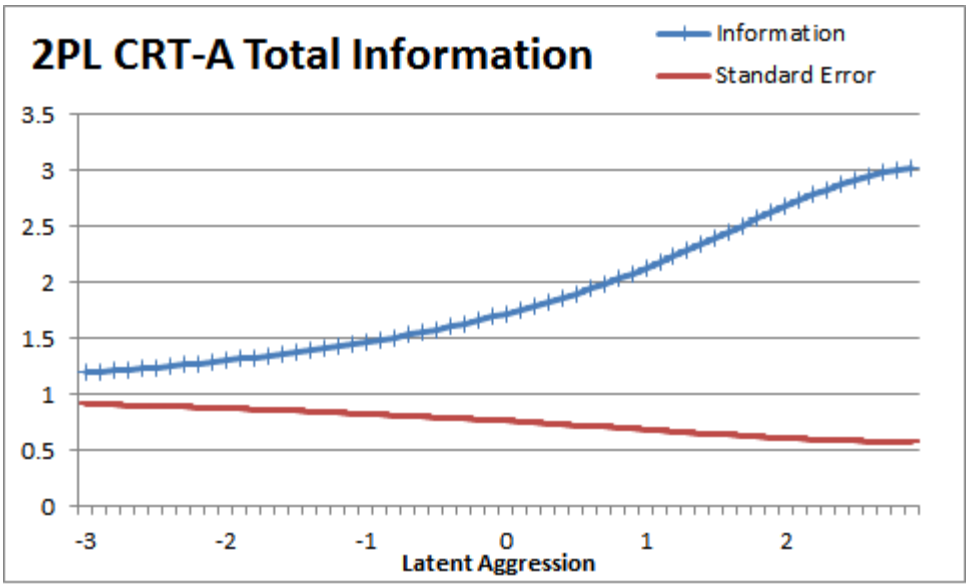


Figure 89. Information function for the Conditional Reasoning Test of Aggression.

Differential Item Functioning Analysis

Twelve items indicated a statistical difference between intercept coefficients. One item indicated a statistical difference between slope coefficients. Although these items differed in the magnitude of item characteristics, none reflected differences in direction. This may serve as an indication that observed differences between student and applied samples are minor and do not reflect fundamental differences in how the CRT-A measures latent aggression in the two groups. Further evidence for this assertion is the high correlation between the intercept and slope coefficients of the two groups. Intercept coefficients are correlated 0.98 and slope coefficients are correlated 0.85.

Person Characteristics and Empirical Reliability

One final analysis is the estimation of theta scores for each of the 5,511 individuals measured in this analysis. The theta estimates range from -1.121 to 2.901 with a mean of 0.00

and a standard deviation of 0.654. The mean square error[3] is 0.572 and the empirical reliability[4] is 0.428.

[3] Computed by summing the squared individual standard error estimates for theta.

[4] Computed using the formula $\frac{s_\theta^2}{s_\theta^2 + Mean\ Square\ Error_\theta}$.

CHAPTER 7

DISCUSSION

Summary

The 2PL model is a useful analytic tool for exploration of the CRT-A. Collapsing the illogical and nonaggressive logical response options provides a dichotomous scoring system that parallels the scoring system recommended by the CRT-A test manual. Given the possible unidimensional nature of the CRT-A, the 2PL model appropriately assesses the likelihood that a respondent will select the aggressive logical response option as a function of latent aggression.

If a polytomous model is desired for the purposes of analyzing the nonaggressive response options, the NRM is better suited to analyze the CRT-A than the GRM or the GPCM, but only slightly. Model comparison results indicate that the three polytomous models are comparable, but the NRM has a slightly lower AIC and BIC. This supports the expectation that the GRM and GPCM are less appropriate since CRT-A response options cannot be ordered on a single dimension. As a result, it is best to treat CRT-A response options as nominal categories as opposed to an ordered progression.

A detailed item analysis of the 22 items on the CRT-A reveals that most items work as intended. With the exception of item 19, each item slope is positive, indicating that individuals are more likely to select the logical aggressive response option as latent aggression increases. This is consistent with the stated hypothesis and provides evidence that the CRT-A works as theoretically intended.

In addition to its relatively flat slope, item 19 has the lowest item-total correlation calculated using this sample. There are a number of potential reasons for the lack of effectiveness

of item 19. The topic of this question involves the potentially violent situation of a robbery. As such, aggression may be primed in respondents due to the nature of the item stem. Also, the logical responses are somewhat ambiguous with respect to one of the JMs for aggression. The aggressive logical response option may appeal to those who have a hostile attribution bias toward the robber (the robber was violent). The nonaggressive logical response option, on the other hand, may appeal to those who have a hostile attribution bias toward the victim (the victim started a fight). Finally, the nonaggressive logical response option may be considered more consistent with the stem (victims are rarely hurt if they comply, so a plausible reason that they would get hurt is if they refuse to comply).

Future research should attempt to modify item 19 to reduce the likelihood of these potential problems. The aggressive logical response option could be changed to comply with the stem to the same degree as the aggressive logical response option (e.g. "The only thing the robber wanted was to hurt the victim."). Additionally, hostile attribution bias may be removed from the nonaggressive logical response option by removing the phrase "or started a fight." If these suggestions do not improve the item, then item 19 should be rewritten entirely or removed from the test.

It should be noted that item 19 had a statistically significant positive slope when related to the factor-specific scale of externalizing controls. Item 19 had a relatively small slope compared to the other ten items in the externalizing controls scale, but the evidence suggest that even though item 19 is a poor indicator of latent aggression, it can be used as a relatively weak indicator of externalizing controls.

As expected, each of the 22 items demonstrated evidence of a negative slope for the nonaggressive logical response option. Given the low response rate for illogical items, this is

consistent with the positive slopes for the aggressive logical response options. The slopes for the combined illogical options were small, but positive. Interestingly, the illogical slopes were slightly higher on the second half of the test.

DIF analyses revealed that twelve items showed evidence of student and applied samples differing in intercept, while only one item (item 21) showed evidence of the two groups differing in slope. This demonstrates that the items differentiate similarly across the two groups, although item "difficulty" (which is related to the base rate of aggressive logical responses) differs between groups in about half of the items. As noted above, high correlations between the intercept and slope coefficients in the two groups suggest that the magnitude of these differences is small.

An examination of the ICCs for item 21 reveals that student and applied samples are quite similar until $\theta \approx 0.5$, at which point the student ICC becomes much steeper than the applied ICC. The stem for item 21 takes place in a generic business setting ("some companies treat employees badly"), but eight other stems also mention businesses and two are set in schools. Unlike the other business-related items, the logical response options for item 21 reflect the opposing moral standpoints of avoidance of punishment and internalizing social norms. Future research should attempt to determine if there is some specific element of the college environment that allows item 21 to better differentiate students with high levels of latent aggression than applied samples with similar levels.

The information function for the CRT-A is a positive and monotonically increasing function with rising values of latent aggression. Most individual items have similar information functions, though some reach a maximum prior to $\theta = 2.9$. Excluding item 19, all CRT-A item information functions reach their maximum value above $\theta = 0$. This indicates that the CRT-A (as

well as most individual items) is best suited to assess latent aggression in respondents high on the trait.

The vast majority of existing research on the CRT-A examines normal populations, or populations that are not known to be prone to aggressive behavior. Since the CRT-A works optimally at high levels of latent aggression, it would be advisable for future research to examine the efficacy of the CRT-A in populations with higher aggression. The CRT-A may be useful in clinical or correctional samples where the prediction of aggressive behavior may prove beneficial to outcomes such as therapy efficacy or prediction of recidivism.

Analysis of the three factor subscales indicates that each of the 22 items appropriately reflects its respective factor. All slopes were positive and monotonically increasing, indicating that the items work as intended. Currently, the subscales are somewhat short (five to eleven items), but future development and expansion may yield stand-alone tests capable of assessing the specific types of rationale used by individuals to justify aggressive thoughts or behaviors.

Limitations and Future Research

Future research on the CRT-A should attempt to address the limitations of this dissertation as well as the questions raised by the results. For example, item 19 should be modified and reanalyzed using the suggestions mentioned above. Reasons for slope differences in item 21 should also be examined.

Although the sample is large, CRT-A scores tend to display positive skew. Individuals high on latent aggression may be underrepresented. As a result, many of the information functions may not have reached their true maximum value in the current analysis. As mentioned above, future research should attempt to conduct research using samples higher in latent aggression as the CRT-A seems best suited for use in those populations.

While the CRT-A displays minimal DIF when comparing student and applied respondents, other forms of DIF may be worth examining in future research. A comparison of gender, age, and cultural groups may be of interest to researchers interested in aggression or aging. A related line of research should seek to expand the subscales of the CRT-A into full-length tests. By doing so, the above-mentioned DIF analyses may better able to describe the reasons behind any group differences in aggressive behavior.

The data demonstrate that CRT-A items vary in both difficulty and discrimination, supporting the decision to select IRT models that account for two parameters. As discussed above, there are theoretical reasons to believe that modeling guessing may be fruitless or counterproductive due to the nature of the test. An empirical evaluation of the merit of three-parameter IRT models would provide stronger evidence for or against their use.

No model used in this dissertation was developed specifically for use with Conditional Reasoning Tests. IRT models were designed for use with multiple-choice and Likert-type items used in the direct measurement of traits. The CRT-A employs indirect measurement to assess implicit aggression. The nature of the CRT-A may necessitate the formulation of a new type of IRT model. While this dissertation demonstrates that both the 2PL model and NRM can provide adequate analysis of the CRT-A, there is a possibility that a new model may be better able to estimate CRT-A item parameters.

Future research should also examine potential reasons for the slightly higher slope of illogical responses in items on the second half of the CRT-A. Although this trend is not statistically verified, it is interesting enough to merit investigation. It is plausible that aggressive individuals are more likely to decide to sabotage their results (or simply stop trying) near the end

of the test. An investigation of the cognitions and behaviors of aggressive individuals during reasoning tests could determine if this trend is worth further examination.

In summary, the CRT-A has long been known to be a valid and reliable predictor of aggression. This item analysis reveals that 21 of the 22 items work as intended. Very little difference was found in the ICCs of student and applied samples, indicating that the two groups respond similarly to CRT-A items. Finally, the CRT-A is best suited for the assessment of latent aggression in individuals high in this trait. It would be wise to expand future research into different areas to better take advantage of the capabilities of the CRT-A.

REFERENCES

Andrich, D. (1978). A rating formulation for ordered response categories. *Psychometrika, 43,* 561-573.

Baker, F. B. (1993). Equating tests under the nominal response model. *Applied Psychological Measurement, 17,* 239-251.

Bergman, S. M., McIntyre, M. D., James, L. R. (2004). Identifying the aggressive personality. *Journal of Emotional Abuse, 4*, 81-93.

Berry, C. M., Sackett, P. R., & Tobares V. (2010). A meta-analysis of conditional reasoning tests of aggression. *Personnel Psychology, 63,* 361-384.

Bock, R. D. (1972). Estimating item parameters and latent ability when responses are scored in two or more nominal categories. *Psychometrika, 37,* 29-51.

De Ayala, R. J. (1989). A comparison of the nominal response model and the three-parameter logistic model in computerized adaptive testing. *Educational and Psychological Measurement, 49,* 789-805.

De Ayala, R. J. (1992). The nominal response model in computerized adaptive testing. *Applied Psychological Measurement, 16,* 327-343.

De Ayala, R. J. (1993). An introduction to polytomous item response theory models. *Measurement and Evaluation in Counseling and Development, 25,* 172-189.

De Ayala, R. J. & Sava-Bolesta, M. (1999). Item parameter recovery for the nominal response model. *Applied Psychological Measurement, 23,* 3-19.

DeMars, C. E. (2003). Sample size and the recovery of nominal response model item parameters. *Applied Psychological Measurement, 27,* 275-288.

DeSimone, J. A. & Cookson, R. L. (2011). *Predicting Corrective Action Procedure Requests in Nuclear Power Plant Employees*. Poster presented at the 26th Annual Convention of the Society for Industrial and Organizational Psychology, Chicago, IL.

DeSimone, J. A. & James, L. R. (2011). *Psychometric Properties of the Conditional Reasoning Test of Aggression*. Poster presented at the 119th Annual Convention of the American Psychological Association, Washington, DC.

DeSimone, J. A., Lee, H. J., & James, L. R. (2012). *Moderated Meta-Analysis of the Conditional Reasoning Test of Aggression*. Poster presented at the 27th Annual Convention of the Society for Industrial and Organizational Psychology, San Diego, CA.

Dunning, D., Heath, C., & Suls, J. M. (2004). Flawed self-assessment: Implications for health, education, and the workplace. *Psychological Sciences and the Public Interest*, *5*, 69-106.

Ellingson, J. E., Sackett, P. R., & Hough, L. M. (1999). Social desirability corrections in personality measurement: Issues of applicant comparison and construct validity. *Journal of Applied Psychology*, *84*, 155-166.

Embretson, S. E. & Riese, S. P. (2000). *Item Response Theory for Psychologists*. Mahwah, NJ: Lawrence Erlbaum.

Entwisle, D. R. (1972). To dispel fantasies about fantasy-based measures of achievement motivation. *Psychological Bulletin, 77*, 377-391.

Gulliksen, H. (1950). *Theory of mental tests*. New York: John Wiley & Sons, Inc.

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

James, L. R. (1998). Measurement of personality via conditional reasoning. *Organizational Research Methods*, *2*, 131-163.

James, L. R. & LeBreton, J. M. (2012). *Assessing the Implicit Personality Through Conditional Reasoning*. Washington, DC: APA.

James, L. R. & Mazerolle, M. D. (2002). *Personality in Work Organizations.* Thousand Oaks, CA: Sage Publications, Inc.

James, L. R. & McIntyre, M. D. (2000). *Conditional Reasoning Test of Aggression Test Manual.* Knoxville, TN: Innovative Assessment Technology.

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). A conditional reasoning measure for aggression. *Organizational Research Methods*, *8*, 69-99.

Kang, T., Cohen, A. S., & Sung, H. (2009). Model selection indices for polyomous items. *Applied Psychological Measurement, 22,* 499-518.

Kieftenbeld, V. & Natesan, P. (2012). Recovery of Graded Response Model parameters: A comparison of Maximum Likelihood and Markov Chain Monte Carlo estimation. *Applied Psychological Measurement, 36,* 399-419.

Kinslinger, H. J. (1966). Application of projective techniques in personnel psychology since 1940. *Psychological Bulletin*, *66*, 134-149.

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.

Leary, M. R. & Kowalski, R. M. (1990). Impression management: A literature review and two-component model. *Psychological Bulletin*, *107*, 34-47.

Lord, F. M. & Novick, M. R. (1968). *Statistical theories of mental test scores*. Reading, MA: Addison-Wesley.

Masters, G. N. (1982). A Rasch model for partial credit scoring. *Psychometrika, 47,* 149-174.

Meade, A. W. (2004). Psychometric problems and issues involved with creating and using ipsative measures for selection. *Journal of Occupational and Organizational Psychology, 77,* 531-552.

Mellenbergh, G. J. (1995). Conceptual notes on models for discrete polytomous item responses. *Applied Psychological Measurement,* 91-100.

Morgan, C. D. & Murray, H. A. (1935). A method for investigating fantasies. *Archives of Neurology and Psychiatry*, *34*, 289-306.

Morgeson, F. P., Campion, M. A., Dipboye, R. L., Hollenbeck, J. R., Murphy, K., & Schmitt, N. (2007). Reconsidering the use of personality tests in personnel selection contexts. *Personnel Psychology*, *60*, 683-729.

Muraki, E. (1992). A generalized partial credit model: Application of an EMPLOYMENT logarithm. *Applied Psychological Measurement, 17,* 159-176.

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

Reise, S. P., & Yu, J. (1990). Parameter recovery in the graded response model using MULTILOG. *Journal of Educational Measurement, 27,* 133-144.

Rorschach, H. (1921). *Psychodiagnostics.* Berne: Haber.

Rosse, J. G., Stecher, M. D., Miller, J. L., & Levin, R. A. (1998). The impact of response distortion on preemployment personality testing and hiring decisions. *Journal of Applied Psychology*, *83*, 634-644.

Samejima, F. (1969). Estimation of latent ability using a response pattern of graded scores. *Psychometrika Monograph No. 17, 34,* (4, Pt. 2).

Samejima, F. (1979). *A new family of models for the multiple choice item*. (Research Report No. 79-4). Knoxville, TN: University of Tennessee.

Snell, A. F., Sydell, E. J., & Lueke, S. B. (1999). Towards a theory of applicant faking: Integrating studies of deception. *Human Resources Management Review*, *9*, 219-242.

Suh, Y., Bolt, D. M. (2010). Nested logit models for multiple-choice item response data. *Psychometrika, 75,* 454-473.

Thissen, D. (1993). Repeating rules that no longer apply to psychological measurement. In N. Frederiksen, R. J. Mislevy, & I. I. Behar (Eds.), *Test theory for a new generation of tests* (pp. 79-97). Hillsdale, NJ: Lawrence Erlbaum Associates.

Thissen, D. Cai, L., Bock, R. D. (2010). The nominal categories item response model. In M. L. Nering & R. Ostini (Eds.), *Handbook of polytmous item response theory models* (pp. 43-75). New York, NY: Routledge/Taylor & Francis Group.

Thissen, D. & Stienberg, L. (1984). A response model for multiple choice items. *Psychometrika, 49,* 501-519.

Thissen, D. & Steinberg, L. (1986). A taxonomy of item response models. *Psychometrika, 51,* 567-577.

Wollack, J. A., Bolt, D. M., Cohen, A. S., & Lee, Y. (2002). Recovery of item parameters in the Nominal Response Model: A comparison of Marginal Maximum Likelihood Estimation and Markov Chain Monte Carlo Estimation. *Applied Psychological Measurement, 26,* 339-352.

Zickar, M. J. & Robie, C. (1999). Modeling faking good on personality items: An item-level analysis. *Journal of Applied Psychology*, *84*, 551-563.

VITA

Justin A. DeSimone

EDUCATION

The Georgia Institute of Technology

Ph.D. in Industrial-Organizational Psychology (in progress) — December, 2012
Minor in Quantitative Methods (requirements complete) — (Projected)

M.S. in Industrial-Organizational Psychology — May, 2010
Thesis Title: Using the Conditional Reasoning Test for Aggression to Predict Corrective Action Requests in a Sample of Nuclear Power Plant Employees

Duke University

B.A. in Psychology — May, 2003
Concentration in Personality/Social
Concentration in Developmental

AWARDS AND HONORS

North Carolina Athletic Association Graduate Scholar — 2010

Dean's List — Spring, 2003
Fall, 2002
Spring, 2002
Fall, 2001

TEACHING EXPERIENCE

The Georgia Institute of Technology

Instructor for "Management Statistics" — Fall, 2012, Fall, 2011

Duties: Taught classes, developed and keyed homework assignments, developed and graded tests, held office hours

Teaching Assistant for "Statistics II: Regression" — Spring, 2012, Spring, 2010, Spring, 2009, Spring, 2008, Spring, 2007

Professor: Dr. Lawrence James

Duties: Held office hours, developed and keyed homework assignments and tests, graded homework assignments and tests, held review sessions, occasionally taught SPSS lab sessions

Teaching Assistant for "History and Systems of Psychology" — Spring, 2012, Fall, 2005

Professor: Dr. Phillip Ackerman

Duties: Held office hours

Teaching Assistant for "Research Design" — Fall, 2011

Duties: Wrote test questions, keyed and graded tests, held office hours

Instructor for "General Psychology" — Summer, 2011

Duties: Taught classes, developed, keyed, and graded tests, held office hours

Instructor for "Psychological Statistics" — Spring, 2011, Fall, 2010

Duties: Taught classes, developed and keyed homework assignments, developed and graded tests, held office hours

Teaching Assistant for "Hierarchical Linear Modeling" — Spring, 2010

Professor: Dr. Lawrence James

Duties: Held office hours, developed and keyed homework assignments

Laboratory Instructor for "Research Methods" Fall, 2009

Professor: James Broadway

Duties: Taught a three-hour session each week, held office hours, developed homework and quizzes, graded homework and quizzes

Teaching Assistant for "Personality Theory" Fall, 2008

Professor: Heather McIntyre

Duties: Held office hours and graded tests

Teaching Assistant for "Statistics I: ANOVA" Fall, 2007

Professor: Dr. James Roberts

Duties: Held office hours, developed and keyed homework assignments, graded homework assignments and tests, and advised the instructor and lab instructor of problematic topic areas

Teaching Assistant for "Multivariate Statistics" Fall, 2007

Professor: Dr. James Roberts

Duties: Held office hours, developed and keyed homework assignments, and graded homework assignments and tests

Teaching Assistant for "Abnormal Psychology" Fall, 2006

Professor: Dr. Dianne Leader

Duties: Held office hours, developed and keyed quizzes, and graded quizzes and tests

RELATED EXPERIENCE

Intern — May, 2010 to August, 2010

Company: Human Resources Research Organization, Alexandria, VA

Duties: Job analysis, certification test item review, facilitation of item-writing workshops, facilitation of KSA identification workshops, selection test design and development, behavior-based interview design and development, literature review, data entry, data analysis (both quantitative and qualitative)

Consultant — August, 2009 to present

Company: Leadership Freedom, Atlanta, GA

Duties: Validation of various measures of personality, statistical analysis including exploratory and confirmatory factor analysis, psychometric analysis including convergent validity and internal consistency reliability

Training and Organizational Development Specialist (Internship) — May, 2006 to August, 2007

Company: The Coca-Cola Company, Atlanta, GA

Duties: Literature search, selection test design and validation, organization of selection and development procedure for hourly employees, design of a reward and recognition system for hourly employees, organization of departmental long-term focus, development or training materials, design of a departmental documentation system, and presentations on various topics

Staff Specialist — May, 2006 to August, 2007

Company: Office of Grant Support, Durham, NC

Duties: Literature search, grant preparation, grant review and critique, website maintenance, database design, data entry, data analysis, and various clerical duties

Staff Specialist — January, 2004 to January, 2005

Company: Center for the Study of Medical Ethics and Humanities, Durham, NC

Duties: Literature search, manuscript preparation, Institutional Review Board submission and institutional transfer, interviewing study participants, database design, data entry, and various clerical duties

Research Assistant — August, 2003 to January, 2004

Company: Cognitive-Behavioral Research and Treatment Program, Durham, NC

Duties: Literature search, manuscript preparation, interviewing study participants, preparation of course materials, database design, data entry, and various clerical duties

PUBLICATIONS AND PAPERS

James, L. R., LeBreton, J. M., Mitchell, T. R., Smith, D. R., DeSimone, J. A., Cookson, R., & Lee, H. L. (accepted). Use of conditional reasoning to measure the power motive. To appear in R. Landis & J. M. Cortina (Editors), Advances in Methodology. SIOP Frontier Series. San Francisco: Jossey-Bass.

Schoen, J. L., DeSimone, J. A., James, L. R. (2011). Exploring joint variance between independent variables and a criterion: Meaning, effect, and size. *Organizational Research Methods, 14*, 674-695.

Received thanks for my role in:
Sugarman, J. (2004). The future of empirical research in bioethics. *The Journal of Law, Medicine, and Bioethics, 32,* 332-231.

CONFERENCE PRESENTATIONS

DeSimone, J. A., Lee, H. J., & James, L. R. (2012). *Moderated Meta-Analysis of the Conditional Reasoning Test for Aggression*. Poster presented at the 27th Annual Convention of the Society for Industrial and Organizational Psychology, San Diego, CA.

Schnure, K. & DeSimone, J. A. (2012). *Narcissism, Aggression, and Achievement Motivation: How Do They Measure Up?* Poster presented at the 27th Annual Convention of the Society for Industrial and Organizational Psychology, San Diego, CA.

DeSimone, J. A. & James, L. R. (2011). *Psychometric Properties of the Conditional Reasoning Test for Aggression*. Poster presented at the 119th Annual Convention of the American Psychological Association, Washington, DC.

Oliver, J., DeSimone, J. A., McCloy, R. A., & Whetzel, D. L. (2011). *Review of realistic job previews and interest inventories*. Paper presented at the annual conference of the International Personnel Assessment Council (IPAC), Washington, DC.

DeSimone, J. A. & Cookson, R. L. (2011). *Predicting Corrective Action Procedure Requests in Nuclear Power Plant Employees*. Poster presented at the 26th Annual Convention of the Society for Industrial and Organizational Psychology, Chicago, IL.

DeSimone, J. A. (2010). *Using the Conditional Reasoning Test for Aggression to Predict Corrective Action Requests in a Sample of Nuclear Power Plant Employees*. Poster presented at the Georgia Tech Research and Innovation Conference, Atlanta, GA.

Kelly, T. L., DeSimone, J. A., Lee, H. J., McNiel, P., & James, L. R. (2010). *Dissociative Studies of Implicit and Explicit Measures of Aggression*. Poster presented at the 118th Annual Convention of the American Psychological Association, San Diego, CA.

DeSimone, J. A., Lee, H. K., Ko, E. C., & James, L. R. (2009). *A Meta-Analysis of the Conditional Reasoning Test for Aggression*. Poster presented at the Georgia Tech Graduate Research Symposium, Atlanta, GA.

Kelly, T. L, DeSimone, J. A., Lee, H. J., & McNiel, P. (2009). *The Differential Prediction of Implicit and Explicit Measures of Aggression*. Poster presented at the Georgia Tech Graduate Research Symposium, Atlanta, GA.

Minton, M. & DeSimone, J. A. (2009). *Psychometric properties of the Conditional Reasoning Test for Aggression*. Presentation at the 21st Annual Convention of the Association for Psychological Science, San Francisco, CA.

TECHNICAL REPORTS

DeSimone, J. A., Massie, H., Scott, R., & Pocklington, C. (2011). *DNA Behavior International Communication DNA® Profile (CDNA)* [*Client name: Financial DNA*]. Atlanta, GA: Financial DNA.

DeSimone, J. A., Massie, H., Scott, R., & Pocklington, C. (2010). *DNA Behavior International Communication DNA® Profile (CDNA)* [*Client name: Financial DNA*]. Atlanta, GA: Financial DNA.

DeSimone, J. A., Waugh, G., Geimer, J., Khanna, C., (2010). *Job analysis report* [*Client name: US Marshals Service*] (FR-10-42). Alexandria, VA: Human Resources Research Organization.

Geimer, J.L., Hardoin, M.M., Khanna, C., DeSimone, J.A., Adeniyi, M., & Koch, A. (2010). *Development of a writing skills test for the 2010 U.S. Marshals Service criminal investigator merit promotion exam: Final report* (FR-10-48). Alexandria, VA: Human Resources Research Organization.

Oliver, J., Whetzel, D. L., McCloy, R. A., DeSimone, J. A. (2010). *Air Force realistic job preview/job exploration feasibility analysis* (FR-10-49). Alexandria, VA: Human Resources Research Organization.

Thibodeaux, C., Geimer, J., & DeSimone, J.A. (in progress). *Development and validation of a structured panel interview for entry-level selection* [*client name and job title are protected by non-disclosure agreement*]. Alexandria, VA: Human Resources Research Organization.

DeSimone, J. A., Ellis, L., Massie, H., & Pockington, C. (2009). *Financial DNA® Core Life Motivations Profile (CORELIFE)* [*Client name: Financial DNA*]. Atlanta, GA: Financial DNA.

GRADUATE COURSEWORK

Statistics and Research Methodology

Advanced Research Methods
Hierarchical Linear Modeling
Item Response Theory
Multivariate Statistics
Psychological Testing
Psychometric Modeling
Psychometric Theory
Research Design
Statistical Analysis I: ANOVA
Statistical Analysis II: Regression
Structural Equation Modeling

Industrial-Organizational Psychology

Employee Selection
Industrial/Organizational Psychology
Motivation and Job Attitudes

Other

Developmental Psychology
History and Systems of Psychology
Human Abilities
Personality Theories
Social Psychology
Teaching Practicum

REFERENCES

Dr. Lawrence James
Georgia Institute of Technology
(404) 385-4621
lawrence.james@psych.gatech.edu

Dr. Susan Embretson
Georgia Institute of Technology
(404) 385-0501
susan.embretson@psych.gatech.edu

Dr. Gregory Corso
Morehead State University
(606) 783-2981
g.corso@moreheadstate.edu

LANGUAGES

English – Native Language

MEMBERSHIPS

American Psychological Association
Society for Industrial and Organizational Psychology