

A PSYCHOMETRIC INVESTIGATION OF THE COMPREHENSIVE
ASSESSMENT OF SADISTIC TENDENCIES (CAST): EVIDENCE
FROM FACTOR ANALYSIS AND ITEM RESPONSE THEORY

BY

RACHEL LYNN AMRHEIN

THESIS

Submitted in partial fulfillment of the requirements
for the degree of Master of Science in Psychology
in the Graduate College of the
University of Illinois at Urbana-Champaign, 2018

Urbana, Illinois

Adviser:

Associate Professor Daniel A. Newman

Abstract

Sadism research has grown exponentially in recent years, establishing the trait as a key predictor of cruel behaviors (bullying, aggression, internet trolling). However, empirical validations of the most popular sadism measure are needed. To address this, we administered the comprehensive assessment of sadistic tendencies (CAST; Buckels & Paulhus, 2014) to 432 undergraduates to investigate the factor structure, item response theory item characteristics (Samejima graded response model and generalized graded unfolding model parameters), underlying response processes (dominance versus ideal point), and measurement invariance across gender using factor analytic methods. Factor analysis results validated the three-factor structure of sadism (physical, verbal, and vicarious), whereas item response theory models demonstrated that a dominance response process fit better than an ideal point response process. We observed that the CAST provides the most information in the high trait range (distinguishing moderately sadistic participants from those who were high in sadism). Results also indicate weak discriminant validity between the CAST and measures of Machiavellianism and Psychopathy. Data further support past research suggesting males are more sadistic than females across all CAST subscales. Further analyses of scalar measurement invariance across gender suggest a true gender gap in the underlying trait, rather than measurement bias. This study begins to answer how to best assess sadism by examining the construct's most popular measure, while highlighting a number of weaknesses to be addressed in future scholarship.

TABLE OF CONTENTS

CHAPTER 1: INTRODUCTION	1
CHAPTER 2: METHODS	11
CHAPTER 3: RESULTS	14
CHAPTER 4: DISCUSSION	42
REFERENCES	48
APPENDIX: ADDITIONAL TABLES AND FIGURES	65

CHAPTER 1: INTRODUCTION

Cruelty for its own sake is particularly insidious for the functioning of society, yet “everyday” sadism is a trait that manifests in normative populations (Chabrol, van Leeuwen, Rodgers, & Séjourné, 2009; Davies & O’Meara, 2007). Contemporary research on cruel behavior has turned to the predictive power of socially aversive personality traits; namely, the dark triad (Paulhus & Williams, 2002) of Psychopathy, Narcissism, and Machiavellianism (Baughman, Dearing, Giammarco, & Vernon, 2012; Furnham, Richards & Paulhus, 2013; Jonason & Tost, 2010; O’Boyle, Forsyth, Banks, & McDaniel, 2012). Recent recognition that the dark triad alone may fail to account for variance in such malevolence has prompted expanding the triad to a “dark tetrad” that includes sadism (Book et al., 2016; Chabrol et al., 2009; Paulhus & Jones, 2016).

Sadism is the tendency to derive pleasure from committing cruel acts upon others (Buckels & Paulhus, 2014; Paulhus, 2014). Unlike the other dark triad traits, sadism involves an “appetitive motivation” (Buckels, 2012) for cruelty that is separate from instrumental motivations (Balakrishnan, Plouffe, & Saklofske, 2017; Paulhus, 2014). The explosion of research on subclinical sadism is attributed to (a) the recent development of self-report instruments that measure the trait (Paulhus & Jones, 2016) and (b) evidence that scores from these instruments can incrementally predict behavior beyond other well-known traits (Buckels, Jones, & Paulhus, 2013; Buckels, Trapnell, & Paulhus, 2014; Greitemeyer & Sagiolou, 2017; Miller, 2017; Pfattheicher & Schindler, 2015; van Geel, Goemans, Toprak, & Vedder, 2017).

The rapid adoption of self-report instruments measuring sadism means that psychometric assessments of these measures have lagged behind (Book et al., 2016). The current paper

provides a psychometric investigation into the most popular, contemporary measure of sadistic personality: the Comprehensive Assessment of Sadistic Tendencies (CAST; Buckels & Paulhus, 2014). The current investigation draws from factor analytic and item response theory (IRT) frameworks, to assess the item characteristics, factor structure, response processes (dominance vs. ideal point), and measurement invariance of the CAST.

Sadism as a personality trait

This paper uses the term “sadism” to refer to the modern conceptualization of the personality trait (sometimes referred to as “everyday” sadism) (Buckels, 2012; Chabrol et al., 2009). Everyday sadism is the tendency to derive pleasure from inflicting or witnessing harm on others (Buckels et al., 2013). This definition differs substantially from historic uses of the word both inside and outside of psychology. Traditionally, sadism was specific to sexual pleasure; sadists were those who caused others pain during sexual intercourse (or fantasized about doing so; Krueger, 2010).

Sadism was first introduced to psychology by psychiatrists who were trying to define which behaviors and motivations fell within “normal” sexuality and which fell within “abnormal” perversion and pathology (Oosterhuis, 2012). Krafft-Ebing (1893) coined the term “sadism” to refer to violent sexual desires, after the French author Marquis de Sade, who depicted violent sexual fantasies. Ultimately, Krafft-Ebing concluded that sadism was an extreme behavior in the distribution of sexuality, but not pathological (Oosterhuis, 2012). Freud (1905) incorporated sadistic sexual pleasure into his theories of sexuality, which position aggression as a main driver of sexual impulses, stating: “that cruelty and sexual impulse are most intimately connected is beyond doubt taught by the history of civilization,” (p. 26). Freud discusses sadistic impulses as normal, but still regards sadistic sex as a perversion. Throughout

the 20th century, sadism researchers focused on the question of whether sadism is pathological or normative, with most work limiting sadism to the context of paraphilias residing within clinical and forensic populations (Heilbrun & Loftus, 1986; Meloy, 1997; Mokros, Osterheider, Hucker, & Nitschke, 2011; Stone, 2010).

Sadistic personality disorder was included in the Diagnostic and Statistical Manual of Mental Disorders (DSM) until the 4th edition (2000), when psychologists feared that the diagnosis might be used to exculpate violent offenders (O'Meara, Davies, & Hammond, 2011). After the removal of sadism from the inventory of clinical disorders, the definition of the psychological construct diversified. Forensic psychology retained the sexual nature of sadism and focused on predicting sexual crime (e.g., Marshall & Kennedy, 2003), while personality psychology began treating sadism as a personality trait (Baumeister & Campbell, 1999; Chabrol et al., 2009; Meloy, 1997).

In the last decade, studies focusing on sub-clinical or “everyday” presentations of sadism have skyrocketed. These studies have shown that motivations beyond the traditional dark triad can account for a number of deleterious behaviors, such as preferring violent media (Greitemeyer, 2015) and increasingly seeking it out over time (Greitemeyer & Sagioglou, 2017), destructive online behaviors like trolling (Buckels et al., 2014) and cyberbullying (van Geel, Goemans, Toprak, & Vedder, 2017), traditional bullying behaviors (Chabrol et al., 2009; van Geel et al., 2017), going out of one's way to inflict harm on others (Buckels et al., 2013; Pfattheicher & Schindler, 2015), and an increased propensity to engage in acts of unprovoked aggression (Reidy, Zeicher, & Seibert, 2011). One of the most popular measures of sadism is the comprehensive assessment of sadistic tendencies (CAST; Buckels & Paulhus, 2014).

The comprehensive assessment of sadistic tendencies (CAST)

The CAST (Buckels & Paulhus, 2014) measures sadism as a multi-faceted trait consisting of three factors: (1) physical sadism (use of force); (2) verbal sadism (use of insults); and (3) vicarious sadism (enjoying witnessing pain, suffering, or acts of cruelty committed by others). The CAST is not the only instrument that emerged following the removal of sadistic personality disorder from the DSM (see review by Meloy, 1997), but it is unique in several ways. First, it conceptualizes sadism as a broad range of motivated behaviors (Buckels, 2012; Reidy, Zeichner, & Seibert, 2011) that are not specific to sexual encounters or fantasies; this sets the CAST apart from earlier attempts to operationalize sadism as sexual attraction to distress (Heilbrun & Loftus, 1986) or as a sexual paraphilia (Krueger, 2010). Second, it treats sadism as a continuous latent trait, rather than a categorical diagnosis or taxonomic type (Millon, Millon, Davis, & Grossman, 2009). Third, it is not constrained to be unidimensional, like the Short Sadistic Impulses Scale (SSIS: O'Meara, Davies, & Hammond, 2011), which was purposefully constructed to measure a single latent trait and satisfy the criteria for analyzing responses using the Rasch (1966) psychometric model. The multi-faceted structure of the CAST is desirable because evidence suggests that scores for the individual factors (physical, verbal, and vicarious sadism) can differentially predict outcomes (Buckels, Trapnell, & Paulhus, 2014; Greitemeyer, 2015; Paulhus, Curtis, & Jones, 2018; Russell & King, 2016). Fourth, it expands the definition of sadism to include both direct and indirect forms of sadism. Direct sadism requires the individual to be responsible for causing harm to others, whereas indirect sadism requires the individual to enjoy witnessing pain and suffering that was caused by someone or something else. Empirical evidence shows that indirect, vicarious sadism can often predict cruelty and aggression as well as direct sadism can (Buckels, Jones, & Paulhus, 2013; Buckels, Trapnell, & Paulhus, 2014). The inclusion of indirect sadism items is a strength of the CAST, as this creates a more

comprehensive assessment that is applicable to normative populations (Paulhus & Jones, 2015) who might be less likely to have engaged in more extreme acts (e.g., physical assault) but still have a sadistic disposition. Finally, the CAST only includes behaviors that specify intrinsic enjoyment of inflicting physical or emotional pain on others. This distinguishes the CAST from other contemporary measures like the Assessment of Sadistic Personality (ASP; Plouffe, Saklofske, & Smith, 2017), which include aggressive behaviors that may be instrumental, rather than intrinsically pleasurable (e.g., “I control my friends through intimidation”). Although subjugation might be something sadists engage in, these items could also serve as indicators of other aversive personality traits and are not specific to sadism (Buckels & Paulhus, 2014).

Extant research has provided ample evidence for the criterion validity of the CAST, demonstrating its ability to predict sadistic and antisocial behaviors and attitudes (Buckels et al., 2014; Jonason, Zeigler-Hill & Okan, 2017). Yet to our knowledge, the factor structure, response processes, measurement invariance across gender, and item characteristics of this (still unpublished) measure have not been investigated. In the current paper, we combine factor analysis and item response theory (IRT) approaches to provide psychometric information about the CAST in a normative/nonclinical sample.

Structure of the CAST

The multifaceted structure of the CAST was driven primarily by theoretical concerns (including items that represented both direct and indirect sadistic tendencies) and predictive validity concerns (ensuring that the items can predict behaviors in normative populations; Buckels, 2012; Buckels et al., 2013, Buckels et al., 2014). The CAST is a revised version of the Varieties of sadistic tendencies scale (VAST), which included items measuring political sadism (e.g., “if our lives were threatened, I would be in favor of torturing a terrorist”, Buckels, 2012).

The authors note structural difficulties with the VAST (highly correlated factors), and the political sadism scale (the 4th factor in the VAST) was ultimately dropped in revision (Paulhus & Jones, 2016). Nonetheless, there is little work that replicates and validates the structure of sadism proposed by the CAST. Furthermore, the CAST has demonstrated that it can predict unique variance in behaviors, even after accounting for other dark triad traits (Buckels et al., 2014; Greitemeyer & Sagioulou, 2017), but there is little known about how the CAST fits with the other dark triad traits in a factor analytic model. To that end, the current paper estimates a series of factor analytic models.

Research question 1a: *For the CAST, does a three-factor model (verbal, physical, and vicarious sadism) fit the data better than a single factor or two factor (direct and indirect sadism) model? [This question involves both overall model fit and the magnitudes of the latent correlations among ostensibly different factors.]*

Research question 1b: *Does a model of all four dark tetrad traits (CAST sadism, Machiavellianism, Narcissism, and Psychopathy) fit the data well? [This question involves both overall model fit and the magnitudes of the latent correlations among ostensibly different factors.]*

Item Response Theory Models

If the field is proposing that researchers include sadism in personality assessment (Paulhus, 2014), then it is important to understand the properties of the items used to assess the trait. Item response theory (IRT) models offer a unique way to understand the information provided by items in psychological assessments. IRT analyses differ from classical test theory approaches because they do not assume that errors are evenly distributed across all levels or ignore specific response patterns in favor of total/mean scores (Hambleton, Swaminathan, & Rogers, 1991). The item characteristics derived from the IRT framework can provide insight into how well sadism items can differentiate between individuals, which items provide the most information, and which items are most likely to be endorsed at varying levels of the latent trait.

Additionally, IRT allows us to answer questions about the underlying response process participants use when responding to sadism measures. IRT models have been developed to model two distinct types of response processes—dominance and ideal point. By fitting sadism data to each, we not only identify which model more accurately represents the data, but can also test whether participants are treating the items like maximal performance thresholds (dominance process) or typical performance items requiring introspection (ideal point process; Drasgow et al., 2010; O’Brien & LaHuis, 2011).

Dominance models assume a monotonic relationship between the latent trait (θ) and the probability that an individual will agree with the item, such that the item response function forms an S-shaped (logistic ogive) curve (Coombs, 1964). In other words, the higher you are on the trait, the more likely you are to endorse said item. Items are typically explained in terms of “difficulty,” or the level of the underlying trait at which a respondent is more likely to get the question correct than incorrect; while assuming that individuals who endorse difficult items will likely also endorse moderate and easy items (Chernyshenko et al., 2007; Drasgow, Chernyshenko, & Stark, 2010). Thus, if sadism scores follow a dominance response process, we would assume that someone who endorses more extreme items (e.g., “I enjoy physically hurting people”) would also endorse less extreme behaviors (e.g., “in video games, I like the realistic blood spurts”).

In contrast to a *dominance model* response process, it is possible that responses to particular survey items might instead follow an *ideal point model* response process. Ideal point processes assume a non-monotonic relationship between θ and the probability of endorsement, such that the item response function forms a bell-shaped curve (Coombs, 1964; Roberts, Donoghue, & Laughlin, 2000). Individuals who possess a level of the trait near the ideal point of

an item (i.e., the peak of the bell-shaped item response function) will have the highest probability of endorsing the statement; and as a corollary, as a person's trait level gets farther from the ideal point in either direction, the probability of endorsing the item decreases. Thus, the ideal point response process requires introspection in a way that the dominance process does not (comparing whether one's own behavior deviates from that described in the item, in either direction; Drasgow et al., 2010). Participants must consider how much the trait is reflected in each statement and then compare the statement to their own typical behavior, deciding whether the item describes their level of the trait well.

The difference between the two response processes is most apparent for items reflecting a moderate level of the latent trait. Recall that a dominance process predicts that only people very low on the trait will not endorse a moderate item, whereas the ideal point process predicts that people who are either very low or very high on the trait will not be likely to endorse the item. If sadism scores follow an ideal point process, then we would assume that even individuals who are very sadistic might not score high on the CAST because they are not endorsing less extreme items that merely reflect a low or moderate level of sadistic tendencies. It is unclear whether a dominance or ideal point response process will best describe responses to the CAST. Ideal point models are theoretically best suited to measure the typical behaviors reflected in measures of attitudes, personality, and individual differences (Carter & Dalal, 2010; Chernyshenko et al., 2007; Tay, Drasgow, Rounds, & Williams, 2009); however, not all measures of personality traits fit an ideal point model better than a dominance model (Cho, Drasgow, Cao, 2015; LaPalme, Wang, Joseph, Saklofske, & Yan, 2015; Carter, Guan, Maples, Williamson, & Miller, 2015; Zampetakis, Lerakis, Kafetsios, & Moustakis, 2015). This discrepancy exists in part because ideal point models are being fit to items already developed according to dominance

approaches—e.g. factor analysis (Chernyshenko et al., 2007; Drasgow et al., 2010), which eliminates items with moderate or low ideal points (Brown & Maydeu-Olivares, 2010; Cao, Drasgow, & Cho, 2015; Chernyshenko, Stark, Chan, Drasgow, & Williams, 2001). Even if researchers were motivated to develop items representing the full range of the trait continuum, good intermediate items are often difficult to write (Credé, 2010; Huang & Mead, 2014; for instructions on how to write such items, see Cao, Drasgow & Cho., 2015). Given that sadism is a personality construct, but the CAST was likely developed according to traditional test design techniques, we do not have a specific hypothesis about which model will fit the data better.

Consequently the second goal of the current paper is to determine whether responses to the CAST fit a dominance or ideal point model better, and then to provide the item characteristics of the CAST using the more appropriate model. This information will enable researchers to draw conclusions about which items are likely to be endorsed by people who have very low, moderate, or high levels of sadism; the items that do the best job discriminating highly sadistic people; and the response process that best describes how participants evaluate CAST items.

Research question 2: Do responses to the CAST better fit a dominance or ideal point response model?

Gender and sadism

Because the CAST is an emerging instrument, there have been very few studies that have examined gender differences in its subscales. Studies reporting gender differences find CAST scores are higher for males than females on the overall scale, as well as on each subscale (Buckels et al., 2013; Jonason et al., 2017; Paulhus & Jones, 2016). While gender differences have not been the primary focus of sadism research, there has been considerable attention paid to gender in other dark personality traits (Grijalva et al., 2015; Miller, Watts, & Jones, 2011). The

interest in gender is partially attributable to the literature examining how the dark triad influences romantic relationships (Jonason & Kavanagh, 2010; Jonason, Luevano, & Adams, 2012), mating strategy (Lee, Ashton, Wiltshire, Bourdage, Visser & Gallucci, 2013; Carter, Campbell, & Muncer, 2014), and partner violence (Miller, Dir, Gentile, Wilson, Pryor & Campbell, 2010). Because males routinely score higher on antisocial variables, such as aggression (Archer, 2004; Bettencourt & Miller, 1996; Björkqvist, 1994), sensation seeking (Cross, Copping, & Campbell, 2011), and empathic deficits (Eisenberg & Lennon, 1983); it is not surprising that men score higher on traits within the dark triad literature (Furnham & Trickey, 2011; Jonason, Koenig, & Tost, 2010; Paulhus & Williams, 2002) and the emerging sadism literature (Buckels & Paulhus, 2014; Plouffe et al., 2017).

Given that (a) there are only a few studies that have evaluated gender differences and (b) the CAST includes items (vicarious sadism scale) about indirect aggression, which is more commonly expressed than is physical aggression by females; whereas males more commonly express physical aggression than indirect aggression (Osterman et al., 1998). Also, girls are more relationally aggressive than boys, whereas boys are more overtly aggressive than girls (Crick, Grotpeter, 1995). As such, there is value in continuing to investigate whether men are more sadistic than women, and whether this gender gap differs across sadism's facets.

Furthermore, if there are mean differences in sadism between men and women, it is unclear if interpreting these observed mean differences is appropriate without first establishing that the CAST has measurement invariance across both groups (i.e., Is the CAST calibrated the same way for men as for women?). As such, our final goal is to evaluate the measurement invariance of the CAST across gender groups.

Research question 3: Does the CAST demonstrate measurement invariance across gender?

CHAPTER 2: METHODS

Participants and Procedure

Participants (N=432) were recruited from the Psychology Subject Pool at a large mid-western University and received course credit for completing a one-hour laboratory study in the psychology building, during which they filled out a battery of self-report surveys. We report here only the measures of the dark tetrad traits: sadism, psychopathy, Machiavellianism, and narcissism. The sample was mostly female (70%) with a mean age of 19.15 years. Most participants reported their ethnicity as White/Caucasian (49%) or Asian (28%), but the sample also included participants who identified as Latino/Latina/Hispanic (8%), Black/African American (6%), or reported multiple ethnicities (7%).

Measures

Sadism. We measured sadism with the Comprehensive Assessment of Sadistic Tendencies (CAST: Buckels & Paulhus, 2014). The CAST is an unpublished 18 item measure that assesses “everyday sadism” in nonclinical samples and is a revised version of an earlier published measure called the Varieties of Sadistic Tendencies (VAST: Buckels et al., 2014; Paulhus & Jones, 2015). The revised CAST removed the political sadism subscale that is included in the VAST (Buckels, 2012; Paulhus & Jones, 2016). While the authors provide a number of filler items that can be used to balance the negative valence of the items, we only administered sadism relevant items. The CAST includes three subscales to measure both direct (verbal and physical) and indirect (vicarious) sadism (item text is displayed in Table 1). All items were rated on a 7 point Likert scale (1 = strongly disagree; 4 = neither agree nor disagree; 7 = strongly agree). Internal consistency reliability of the overall CAST scale ($\alpha = .85$), as well as the three subscales (verbal sadism: $\alpha = .81$; physical sadism: $\alpha = .77$; and vicarious sadism: $\alpha = .77$)

was sufficiently high. Item text and means are provided in Table 1.

Narcissism. Narcissism was measured with the short form narcissistic personality inventory (NPI- 16: Ames, Rose, & Anderson, 2006). Items use a forced choice format where narcissism-consistent responses are coded as 1 and narcissism inconsistent responses are coded as 0 ($\alpha = .85$).

Machiavellianism. Machiavellianism was measured with the Machiavellian Personality Scale (MPS; Dahling, Whitaker, & Levy, 2009). The measure contains 16 items rated on a 7 point Likert scale (1 = strongly disagree; 4 = neither agree nor disagree; 7 = strongly agree) that measure four subscales: amorality ($\alpha = .82$), desire for control ($\alpha = .74$), desire for status ($\alpha = .83$), and distrust of others ($\alpha = .70$).

Psychopathy. Psychopathy was measured with the 64 item Self-report Psychopathy Scale (SRP-III; Paulhus, Neumann, & Hare, 2009). All items are rated on a 7 point Likert scale (1 = strongly disagree; 4 = neither agree nor disagree; 7 = strongly agree). The SRP-III has four subscales intended to capture both factor 1 and factor 2 psychopathy (Harpur, Hare, & Hakstian, 1989): interpersonal manipulation ($\alpha = .79$) callous affect ($\alpha = .72$), erratic lifestyle ($\alpha = .80$), and antisocial behaviors ($\alpha = .69$).

Table 1

Comprehensive assessment of sadistic tendencies (CAST) item text, subscale, and means

Item	Item text	Subscale	Mean score
1	I was purposely mean to some people in high school.	Verbal	2.45
2	I enjoy making jokes at the expense of others.	Verbal	2.50
3	I have purposely tricked someone and laughed when they looked foolish.	Verbal	2.58
4	When making fun of someone, it is especially amusing if they realize what I'm doing.	Verbal	2.41
5	Perhaps I shouldn't have, but I never got tired of mocking certain classmates.	Verbal	2.25
6*	I would never purposely humiliate someone.	Verbal	3.4
7	I enjoy physically hurting people.	Physical	1.40
8	I enjoy tormenting people.	Physical	1.31
9	I have the right to push certain people around.	Physical	1.48
10	I have dominated others using fear.	Physical	1.55
11	I enjoy hurting my partner during sex (or pretending to).	Physical	1.51
12	In video games, I like the realistic blood spurts.	Vicarious	1.99
13	I love to watch YouTube clips of people fighting.	Vicarious	1.74
14	I enjoy watching cage fighting (or MMA), where there is no escape.	Vicarious	1.78
15	I sometimes replay my favorite scenes from gory slasher films.	Vicarious	1.44
16*	There's way too much violence in sports.	Vicarious	4.79
17	I enjoy playing the villain in games and torturing other characters.	Vicarious	1.85
18	In professional car-racing, it's the accidents that I enjoy most.	Vicarious	1.69

Note. Asterisks indicate that the item is reverse scored. Responses are on a Likert scale that ranges from 1 to 7.

CHAPTER 3: RESULTS

Factor structure

We assessed the factor structure of the CAST using confirmatory factor analysis (CFA), implemented in the R package lavaan (Rosseel, 2012). Item correlations are reported in Appendix. Researchers are often concerned that data on dark personality traits do not follow a normal distribution and therefore have biased standard errors and model fit statistics (Herve & Yuille, 2017; Wright, Pincus, & Lenzenweger, 2012). To address this potential concern, analyses were performed using maximum likelihood estimation (estimator = ‘MLM’), which provides a Satorra-Bentler rescaled chi-square statistic (Satorra & Bentler, 1994) that is robust to non-normality. We note, however, that the data do not have particularly extreme values (Curran, West, & Finch, 1996) of skewness (overall sadism = .94; verbal sadism = .51; physical sadism = 2.03; vicarious sadism = 1.29), but there is some kurtosis (overall sadism = 3.48; verbal sadism = 2.52; physical sadism = 7.01; vicarious sadism = 4.17). This suggests that the data do not deviate substantially from a normal distribution, even if the mean scores are very low (overall sadism $M = 2.12$; verbal sadism $M = 2.60$; physical sadism $M = 1.45$; vicarious sadism $M = 2.18$.)

Table 2 shows fit statistics for all models that tested the structure of sadism. To address Research Question 1a, we estimated both a unidimensional model in which all items loaded onto a single latent trait (Model A depicted in Figure 1) and a three-factor oblique model in which the items loaded onto the three factors originally proposed by the authors of the instrument (Buckels & Paulhus, 2014; Model B depicted in Figure 2). The three-factor oblique model demonstrated adequate model fit to the data ($\chi^2 = 428.54$ (132); RMSEA = .072; SRMR = .059; CFI = .90), in comparison to the unidimensional model, which demonstrated poor fit ($\chi^2 = 1263.98$ (135);

RMSEA = .139; SRMR = .103; CFI = .619). A χ^2 difference test comparing the two models indicated that the multi-dimensional model fits significantly better than the unidimensional model ($\Delta\chi^2_{(df=3)} = 835.44; p < .05$). All CAST items exhibited strong loadings ($\lambda > .40$) onto their respective latent factors, with the exception of two items (items #6 and #16: the reverse-worded items). The latent correlations between the three factors in Model B were moderately large (verbal-physical $\phi = .56$; verbal-vicarious $\phi = .46$; physical-vicarious $\phi = .50$). Altogether, these results validate the structure of the CAST as a measure of three correlated subscales of sadism, and indicate that we should indeed treat sadism as a multidimensional construct consisting of verbal, physical, and vicarious sadism.

We next tested a hierarchical factor model (Model C depicted in Figure 3), in which the three sadism factors all loaded onto a single, higher-order latent construct. Because there are only three sadism subscale factors, fit indices cannot mathematically distinguish between the hierarchical (Model C) and the oblique model (Model B) of sadism (i.e., Models B and C have identical fit, by design; Bollen, 1989). Although we cannot demonstrate whether the hierarchical factor model fits better than the oblique model, we prefer the hierarchical model because it allows researchers to draw inferences at both the facet level (verbal, physical, and vicarious sadism) and the overall higher-order construct level (sadism). Parameter estimates for all three sadism models (A, B, and C) are presented in Table 3.

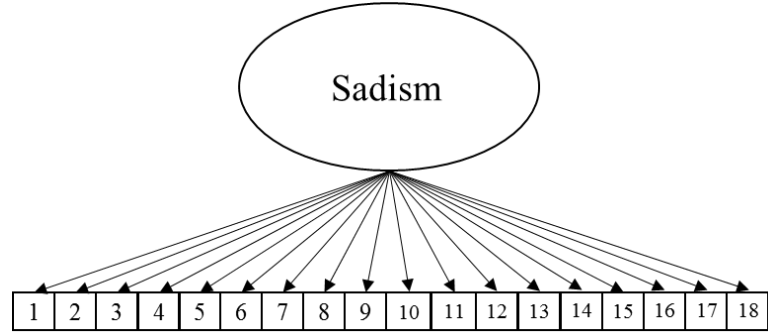


Figure 1. Model A: Unidimensional model of sadism.

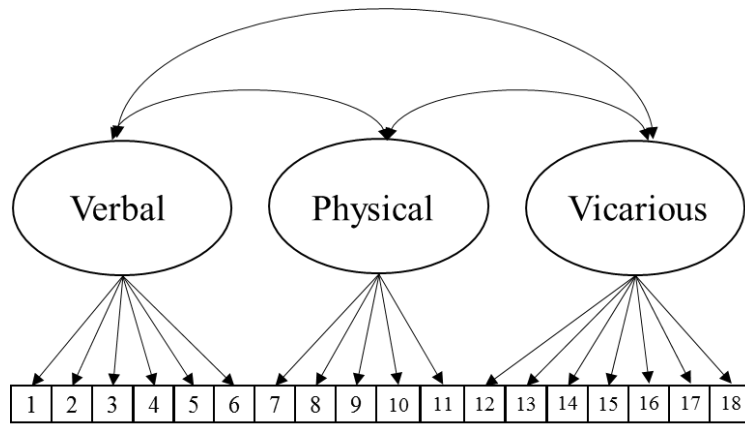


Figure 2. Model B: Three-factor oblique model of sadism.

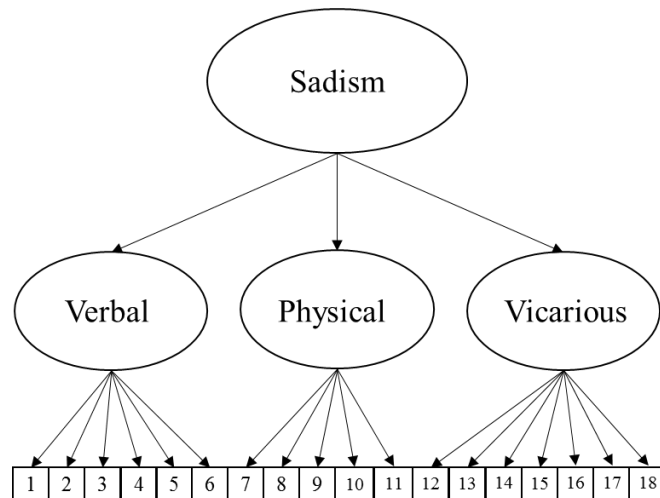


Figure 3. Model C: Three-factor hierarchical model of sadism

Table 2

Comparison of CFA fit statistics for sadism measured with the comprehensive assessment of sadistic tendencies (CAST)

Model		χ^2	CFI	TLI	AIC/BIC	RMSEA	SRMR	SB
A	Single factor	1263.9	0.62	0.57	25069.5/ 25289.2	.139	.098	1.54
		9 (135)				(0.132, 0.146)		
B	Three factor oblique	428.54	0.90	0.89	24240/ 24471.9	0.072	0.056	1.58
		(132)				(0.065, 0.080)		
C	Hierarchical	428.54	0.90	0.89	24240/ 24471.9	0.072	0.056	1.58
		(132)				(0.065, 0.080)		
D	Two factor oblique (direct and indirect sadism)	763.53	0.79	0.76	24571/ 24794.8	0.104	0.078	1.54
		(134)				(0.097, 0.112)		

Note: Parenthesis following χ^2 fit statistics indicate degrees of freedom; parentheses following RMSEA fit statistics provide the lower and upper bound of the RMSEA confidence interval (90%). CFI = comparative fit index; TLI = Tucker Lewis index; AIC = Akaike information criterion; BIC = Bayesian information criterion; RMSEA = root mean square error of approximation; SRMR = standardized root mean square residual; SB = Santorra Bentler scaling correction.

Table 3

Parameter estimates from sadism CFA models

	One-dim. (Model A)	Oblique Three Factor (Model B)			Hierarchical Model (Model C)		
Item	One Factor	Verb.	Phys.	Vic.	Verb.	Phys.	Vic.
1	0.54	0.61			0.61		
2	0.69	0.80			0.80		
3	0.69	0.82			0.82		
4	0.65	0.75			0.75		
5	0.68	0.73			0.73		
6	0.21	0.29			0.29		
7	0.52		0.69			0.69	
8	0.57		0.79			0.79	
9	0.62		0.75			0.75	
10	0.56		0.65			0.65	
11	0.44		0.42			0.42	
12	0.56			0.73			0.73
13	0.56			0.82			0.82
14	0.55			0.80			0.80
15	0.51			0.63			0.63
16	0.09			0.17			0.17
17	0.49			0.53			0.53
18	0.43			0.49			0.49
Factor Correlations							
Verbal		--					
Physical		0.56	--				
Vicarious		0.46	0.50	--			
Second order factor loadings					0.72	0.78	0.64

Note. All values in rows labeled 1 through 18 represent the factor loading of the corresponding CAST item. All factor loadings are standardized; verb. = verbal; phys. = physical; vic. = vicarious; one-dim = unidimensional model.

Although the CAST specifies items according to three subscale factors (physical, verbal, and vicarious), previous research has identified two types of sadism: direct sadism and indirect sadism (Buckels, 2012; Buckels, et al. 2013). To test this structure we estimated an oblique two factor model (Model D depicted in Figure 4). The direct sadism factor specifies as indicators all items from both the verbal and physical subscales; and the indirect sadism factor uses as indicators all items from the vicarious sadism subscale. The two-factor oblique model (Model D) demonstrated poor fit ($\chi^2 = 763.53$ (134); RMSEA = .104; SRMR = .082; CFI = .79) and performed much worse than the three-factor models (Models B and C). Distinguishing between physical and verbal sadism greatly improves the empirical fit of the factor analytic model.

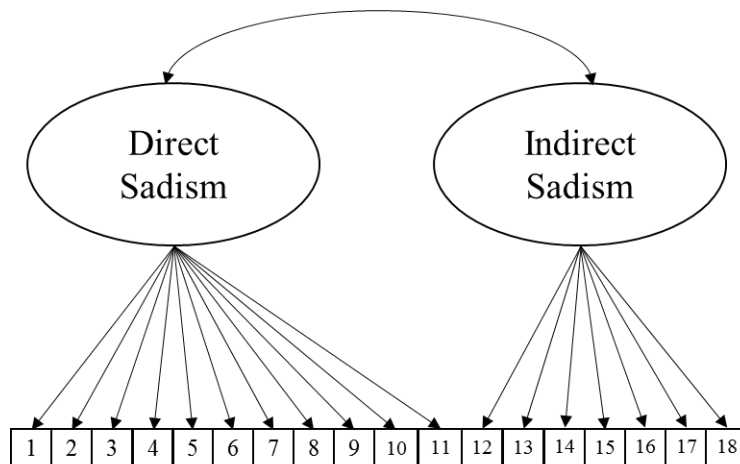


Figure 4. Model D: Two factor model of sadism (direct sadism combines verbal and physical items; indirect sadism includes vicarious sadism items)

Small factor loadings of reverse-worded items

A detailed look at the parameter estimates from the preferred three factor hierarchical model (Model C) shows that the two reverse scored items (item 6 and item 16) had considerably lower loadings ($\lambda_6 = .29$ and $\lambda_{16} = .17$) than the other items in the scale. Removal of such items can sometimes improve the structure of the scale by eliminating measurement artifacts (e.g., Idaszak & Drasgow, 1986). We thus estimated the fit of Model E (Figure 5), which is identical to Model B, except that the reverse scored items were removed, leaving only 16 indicators. Although Model E cannot be directly compared to Model B (because the two models are not nested), we nonetheless note that Model E displays adequate overall fit ($X^2 = 380.00$ (101); RMSEA = .08; SRMR = .07; CFI = .90).

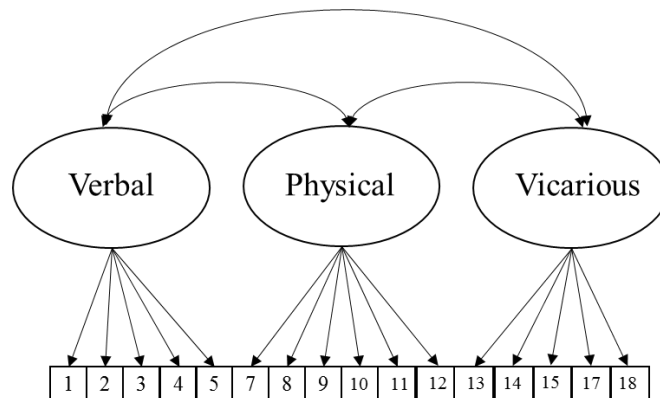


Figure 5. Model E: a three-factor model with the two reverse scored items (6 and 16) removed.

One possible explanation for the small factor loadings of items 6 and 16 is that they are reflections of both the item content and a methods factor related to negatively worded items. To test whether the reversed items (items 6 and 16) reflect such an artifact, we estimated Model F (Figure 6) in which the reverse scored items cross-loaded onto a separate “methods” factor [to

achieve model identification, we constrained the two method-factor loadings to be equal, and these method factor loadings were estimated to be *standardized* $\lambda_6 = .23$ and *standardized* $\lambda_{16} = .29$] ($\chi^2 = 426.40$ (131); RMSEA = .072; SRMR = .056; CFI = .90). Including a methods factor did not improve model fit compared to Model B ($\Delta\chi^2_{(df=1)} = 2.14$; $p > .05$; n.s.), but interpretation of this method factor is limited because there are only two observed indicators.

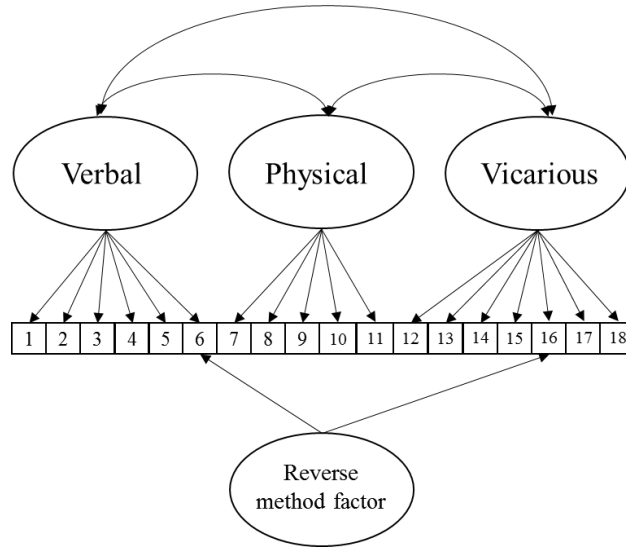


Figure 6. Model F: Three-factor model of sadism with the reverse scored items loading onto a separate methods factor.

To address this limitation of a potentially under identified reverse-wording method factor, we next estimated a model that included both sadism and psychopathy scores (Model G). This model allowed all sadism items to load onto their respective subscales (verbal, physical, and vicarious), all psychopathy items to load onto their respective subscales (interpersonal manipulation, antisocial behavior, erratic lifestyle, and callous affect), and for all 23 reverse-scored items across both measures (sadism and psychopathy) to load onto a methods factor. Although the RMSEA was acceptable, several other fit indices suggested that the fit was poor, which led us to conclude that the overall fit of Model G was poor ($\chi^2 = 6808.89$ (3188); CFI = .67; TLI = .66; RMSEA = .051; SRMR = .073). We caution against interpreting parameter

estimates from this model given that the CFI, TLI, and SRMR indices suggested poor model fit (see Appendix for parameter estimates). Even if we did interpret the model, the factor loadings of the items onto the reverse-wording method factor are small (average $\lambda = .28$). Notably, the two reverse scored CAST items do not load onto this reverse-wording method factor ($\lambda_6 = .08$ and $\lambda_{16} = -0.08$). This model does not provide strong evidence that these are poor items for reasons relating to a general negative wording artifact. CAST items 6 and 16 might have low factor loadings due to the specific content of the items, or (more plausibly) participant inattention when responding to items (Schmitt & Stults, 1985; cf. the CAST was administered after the SRP-III for all participants and was the last set of questions in a relatively lengthy study).

Table 4

Fit indices for CFA models investigating reverse scored CAST items

Model	Description	χ^2	CFI	TLI	RMSEA	SRMR	SB scaling correction
E	Three factor removing reversed items	380.00 (101)	0.90	0.89	0.080 (0.071, 0.089)	0.056	1.72
F	Three factor with reversed items cross- loading onto separate factor	414.79 (127)	0.90	0.89	0.072 (0.065, 0.080)	0.053	1.56
G	Reversed items from CAST and SRP-III loading onto a separate factor	6808.89 (3188)	0.67	0.66	0.051 (0.050, 0.053)	0.073	1.56

Note. CAST = comprehensive assessment of sadistic tendencies; SRP-III = self-report psychopathy scale; parenthesis following χ^2 fit statistics indicate degrees of freedom; parentheses following RMSEA fit statistics provide the lower and upper bound of the RMSEA confidence interval (90%). CFI = comparative fit index; TLI = Tucker Lewis index; RMSEA = root mean square error of approximation; SRMR = standardized root mean square residual.

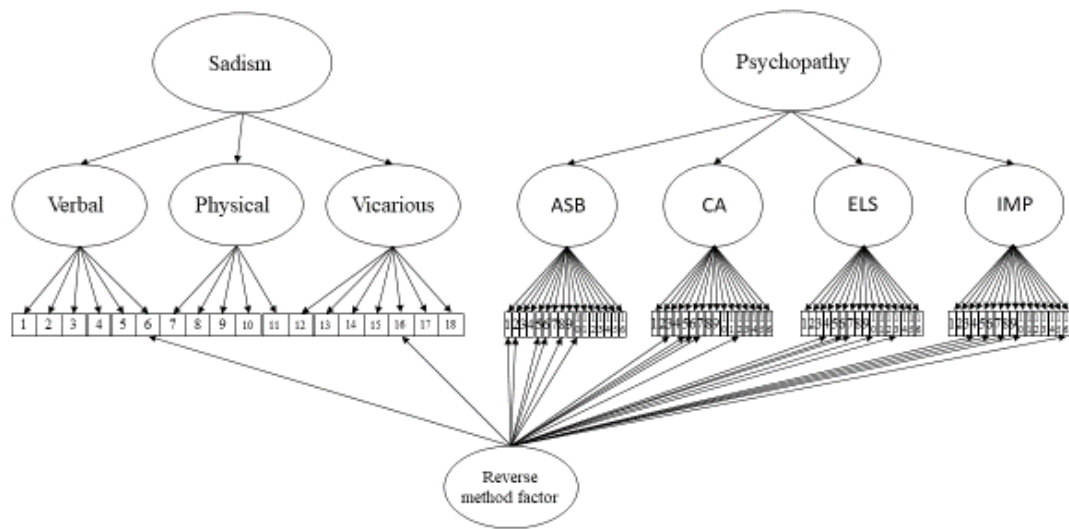


Figure 7. Model G: Items load onto latent traits (sadism and psychopathy) as well as reverse scored methods factor.

Sadism within the dark tetrad

The measurement models tested above provide information about the structure and item parameters of the CAST items when they are tested in models looking at sadism in isolation. However, sadism is theorized to be one component of a broader “dark tetrad” model of undesirable traits (Paulhus, 2014; Book et al., 2016). Therefore, we estimated a model of the complete dark tetrad (Model H: Figure 8) to address Research Question 2. This dark tetrad model uses the preferred hierarchical structure for modeling sadism (i.e., Model C) while also accounting for variance attributed to the other dark triad traits (Machiavellianism, narcissism, and psychopathy). The model did not converge when individual items were used as indicators of the dark triad traits, since our sample size ($N=432$) is small compared to the number of values in the covariance matrix (Marsh, Hau, Balla, & Gayson, 1998). In order to estimate the model, we used scale composite scores and item parcels as indicators for Machiavellianism, Narcissism, and Psychopathy (Bandalos & Finney, 2001; Kishton & Widaman, 1994; Little, Cunningham, Shahar, & Widaman, 2002). The indicators for Machiavellianism were created by calculating composite scores for each of the four subscales of the Machiavellian Personality Scale (subscale 1 = average of five items measuring amorality, subscale 2 = average of three items measuring desire for control, subscale 3 = average of three items measuring desire for status, and subscale 4 = average of five items measuring distrust of others). Similarly, the indicators for psychopathy were the scale composite scores for the four factors in the SRP-III: antisocial behaviors (16 items), callous affect (16 items), erratic lifestyle (16 items), and interpersonal manipulation (16 items). Since narcissism was measured with the short form NPI-16, which does not evaluate facet level scores, we created 3 parcels, by randomly selecting 3 sets of items (using the RAND function in Microsoft Excel). Parcel 1 included NPI-16 items 4, 5, 6, 9, and 13; parcel 2 included

NPI-16 items 2, 10, 11, 14, and 15; and parcel 3 included NPI-16 items 1, 3, 7, 8, 12, and 16.

Randomly selecting the items for each parcel makes it likely that each indicator is representative of the entire narcissism construct (i.e., the domain-representative approach—in which each parcel is formed by combining items across different facets; Williams, Vandenberg, & Edwards, 2009). The full model—in which the dark tetrad was estimated to include sadism, psychopathy, narcissism, and Machiavellianism (Figure XX; Model H)—demonstrated marginally adequate fit ($\chi^2 = 903.01$ (368); RMSEA = .058 ; SRMR = .056; CFI = .89; TLI = .88). SRMR and RMSEA are adequate, and CFI and TLI are marginally adequate.

The parameter estimates for the dark tetrad model (Model H: Figure 8) are provided in Table 5. Many of the latent correlations amongst the dark tetrad variables are very high in the current dataset (sadism-Machiavellianism $\phi = .74$; sadism-psychopathy $\phi = .85$; Machiavellianism-psychopathy $\phi = .86$; observed correlations reported in Appendix). Narcissism has a much smaller relationship to the other traits. Although these latent correlations do not reach unity (i.e., $\phi < 1.0$), these values suggest that sadism might lack discriminant validity from Machiavellianism and psychopathy in this sample. We performed Fornell & Larcker's (1981) less conservative test of discriminant validity, to further investigate whether these latent variable relationships are problematic. Previous studies have used this metric to provide evidence of discriminant validity using parameters estimated with factor analytic models (e.g., Joseph & Newman, 2010). Fornell and Larcker's (1981) test compares the latent factor score (ϕ) to the square root of the average indicator variance (calculated using sum of the squared factor loadings and the sum of the error variances). Evidence of discriminant validity is provided if this average indicator variance is larger than the latent correlation between constructs. The latent correlation for sadism-Machiavellianism (.74) was larger than the square root of the sadism-

Machiavellianism average indicator variance (.64), indicating only weak discriminant validity between sadism and Machiavellianism in the current sample. Similarly, the latent correlation for sadism-psychopathy (.85) was also larger than the square root of the sadism-psychopathy average indicator variance (.70), also indicating only weak discriminant validity between sadism and psychopathy in the current sample. This evidence suggests that, in our particular sample, sadism as measured by the CAST does not provide a strong unique addition to the dark triad (weak evidence of discriminant validity between sadism and other dark triad traits).

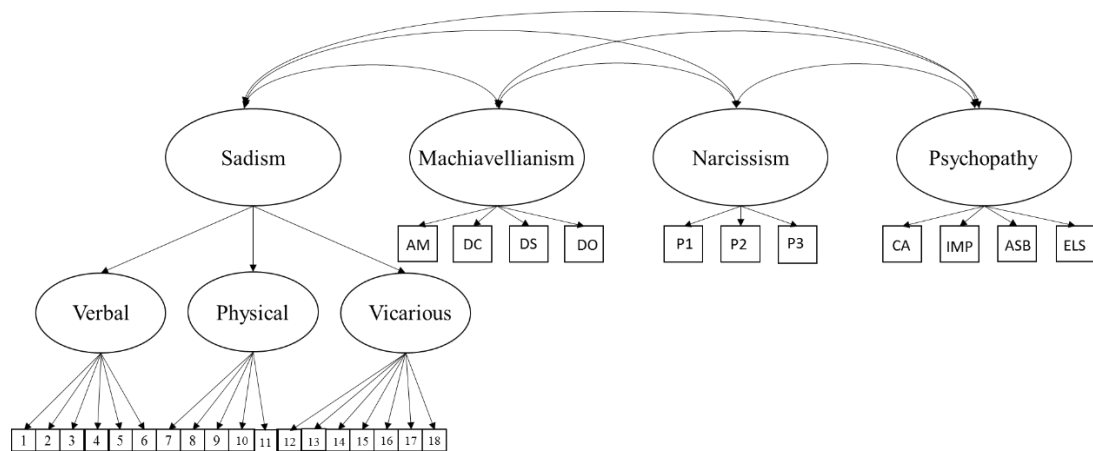


Figure 8. Model H: Dark tetrad oblique model that includes all four traits. Sadism indicators are the 18 CAST items; Machiavellianism indicators and Psychopathy indicators are the respective scale composite scores; Narcissism indicators are three randomly selected item parcels; AM = amorality; DC = desire for control; DS = desire for status; DO = distrust of others; IMP = interpersonal manipulation; CA = callous affect; ASB = antisocial behaviors; ELS = erratic lifestyle.

Table 5 (part 1 of 2)

Dark tetrad CFA model parameters

Indicator	Verb.	Phys.	Vic.	Sadism	Mach.	Narc.	Psych.
CAST 1	0.61						
CAST 2	0.81						
CAST 3	0.82						
CAST 4	0.74						
CAST 5	0.73						
CAST 6	0.29						
CAST 7		0.68					
CAST 8		0.77					
CAST 9		0.76					
CAST 10		0.67					
CAST 11		0.43					
CAST 12			0.73				
CAST 13			0.82				
CAST 14			0.81				
CAST 15			0.63				
CAST 16			0.17				
CAST 17			0.53				
CAST 18			0.49				
MPS Amorality					0.82		
MPS Desire for Control					0.55		
MPS Desire for Status					0.41		
MPS Distrust of Others					0.45		
NPI-16 parcel 1 (items 4, 5, 6, 9, and 13)						0.81	
NPI-16 parcel 2 (items 2, 10, 11, 14, and 15)						0.82	
NPI-16 parcel 3 (items 1, 3, 7, 8, 12, and 16)						0.67	
SRP-III Interpersonal Manipulation							0.82
SRP-III Erratic Lifestyle							0.65

Table 5 continued (part 2 of 2)

Dark tetrad CFA model parameters

SRP-III Callous Affect					0.77
SRP-III Antisocial Behavior					0.46
Factor Correlations					
Verbal sadism					
Physical sadism					
Vicarious sadism					
Sadism (overall)		--			
Machiavellianism	0.74	--			
Narcissism	0.11	0.17	--		
Psychopathy	0.85	0.86	0.21	--	
Second order factor loadings	0.72	0.66	0.77		

Note. CAST = comprehensive assessment of sadistic tendencies; MPS = Machiavellian Personality Scale; Verb. = verbal sadism; Phys. = physical sadism; Vic. = vicarious sadism; NPI-16 = 16 item Narcissistic personality inventory; SRP-III = self-report psychopathy scale; Mach = Machiavellianism; Narc. = Narcissism; Psych. = Psychopathy. Model fit: $\chi^2 = 903.01$ (368); RMSEA = .058 ; SRMR = .056; CFI = .89; TLI = .88.

Item response model estimation and comparison

Data Unidimensionality and Response Categories

After conducting psychometric analyses using factor analysis, we now turn to item response theory. Item response theory (IRT) models assume unidimensionality of measures, such that only one trait is being assessed (Hambleton et al., 1991). Therefore, we tested the unidimensionality of the CAST before proceeding to model estimation. Although the CFA results indicate that the CAST is clearly multidimensional, previous research has shown that it is reasonable to relax this assumption if there is a dominant first factor (Drasgow & Hulin, 1990; Drasgow & Parsons, 1983). We conducted a principal components analysis for the 18 CAST

items. The first eigenvalue was 5.93 and accounted for 33% of the variance, which exceeds the 20% typically recommended for demonstrating a dominant first factor (Reckase, 1979). Further, the ratio of the first-factor eigenvalue to the second-factor eigenvalue was large (ratio = 2.90), lending additional support for the unidimensionality assumption.

The CAST was administered using a 7-point Likert scale (Buckels & Paulhus, 2014). In order to conduct IRT analyses, we next collapsed the seven response categories into three categories, representing disagreement, agreement, or neutrality (0 = strongly disagree, disagree, or somewhat disagree; 1 = neither agree nor disagree; 2 = somewhat agree, agree, or strongly agree), which allows for more efficient item parameter estimation (Chernyshenko et al., 2007; Cho, Drasgow, Cao, 2015; Stark et al., 2006; Tay et al., 2009) and is desirable when datasets include items with response options that have very few endorsements, as is the case with our dataset (item 10, for example, was never endorsed with a ‘strongly agree’ response).

Dominance Model IRT

We first estimated item parameters of a dominance model using Samejima’s graded response model (SGRM) in the R software package *ltm* (Rizopoulos, 2006; Samejima, 1969). Table 6 includes the item parameters for each item, as well as classical test theory information (item means and item-test correlations). IRT item discrimination ranged from 0.14 to 1.49 ($M = 0.95$). Item discrimination quantifies how well an item can discriminant individuals at varying levels of the latent trait. It is typical for personality measures to have discrimination levels near 1.0 (Chernyshenko et al., 2007). The discrimination parameters for the CAST items are about 1.0, on average, which indicates that items are good at discriminating between different levels of sadism. The verbal subscale had the largest mean discrimination ($a = 1.12$), followed by physical ($a = 1.02$), and vicarious ($a = .77$). Overall, the discrimination parameters are sufficiently large

and in line with other personality item characteristics (Chernyshenko et al., 2007; Cho, Drasgow, & Cao, 2015).

The means of the two difficulty parameter estimates were 1.15 and 1.97, respectively. These estimates reflect the location of the inflection point of the item response function. There are 2 estimates to describe the inflection points for the two item information curves (the curve reflecting “agreement” responses and the curve reflecting “disagreement” responses). With the exception of item 16, the $b1$ threshold estimates are relatively large and positive, which indicates that the items are somewhat “difficult,” meaning only people with high trait levels of sadism are typically endorsing CAST items (i.e., agreeing with the item statement). We note that item 16 (a reverse-scored item) is the only item with negative difficulty parameters, indicating that even individuals with very low theta levels would be likely to endorse this item. Item 16 also has the lowest item-test correlation ($r = .21$) and a mean score (4.80) that is much higher than any other item, suggesting item 16 may not be a good indicator of sadism. Although the parameter estimates for the other reverse scored item, item 6, were not problematic, it did have the second lowest item-test correlation ($r = .34$) and the lowest discrimination ($a = .47$). These item characteristics are not inherently problematic, but consistent with CFA results that these items appear to be functioning differently and perhaps are not good indicators of sadism in the current sample.

Ideal-Point Model IRT

We then estimated item parameters using a generalized graded unfolding model (GGUM; Roberts, Donoghue, & Laughlin, 2000), an ideal point model, using GGUM2004 software (Roberts, Fang, Cui, & Wang, 2006). The discrimination (α) parameter estimates for the 18 items ranged from 0.60 to 6.31 ($M = 2.65$). These parameters provide information about the

discrimination of the item, and show that item discrimination in Physical sadism again had the highest mean item discrimination ($\alpha = 3.62$), followed by vicarious ($\alpha = 2.85$), and verbal ($\alpha = 1.54$). Several items had particularly large parameter α estimates (Items 7, 8, 9 13, and 14 all have discrimination parameters above 3). These discrimination parameters are generally very high, which means that CAST items are sufficiently discriminating between varying levels of the latent trait.

The location (δ) parameter estimates describe the item location or the theta level at which the probability of item endorsement is greatest (i.e., the peak of the bell curve). The δ estimates were mostly moderate, falling between 0 and 1; most item locations fall near zero ($M = 0.74$), with the exception of item 15. Following previous research guidelines (Carter & Dalal, 2010; Roberts & Shim, 2008), we examined how many items had delta parameters that fell between the 10th and 90th percentile of theta estimates. An unfolding process would fit the data best when the set of test items do not have extreme locations and include moderate items with delta parameters near zero. Therefore, a test with several items outside of this range is more likely to yield data that fits a dominance model better than an ideal point model. Seventeen out of 18 items fell within this range, which prevents us from ruling out the possibility that these items might demonstrate unfolding (i.e., an ideal point response process).

Overall the GGUM parameter estimates demonstrated high discrimination and moderately negative location, suggesting that the model is best for distinguishing individuals who fall just below mean sadism levels. However, given the items with unreasonably large δ parameter estimates, caution is warranted in interpreting these values and requires further examination of overall fit of the data to the GGUM.

Table 6

CAST item characteristics from classical test theory and item response theory models

Item	Classical test theory item discrimination (item-total correlation)	<u>SGRM (dominance)</u>			<u>GGUM (ideal point)</u>			
		<i>a</i>	<i>b1</i>	<i>b2</i>	α	δ	$\tau1$	$\tau2$
1	0.57	0.75	1.07	1.46	1.15	0.66	1.09	-1.19
2	0.71	1.33	0.79	1.30	2.27	0.69	00.0	-0.52
3	0.71	1.27	0.75	1.05	1.69	0.62	0.41	-0.91
4	0.66	0.98	0.97	1.61	1.74	0.74	0.09	-0.36
5	0.68	1.25	1.02	1.47	1.80	1.03	0.26	-0.9
6*	0.34	0.33	0.47	1.22	0.60	-0.21	2.45	-2.23
7	0.51	1.19	2.20	2.66	4.51	0.70	0.24	-0.07
8	0.54	1.49	2.15	2.75	4.46	0.75	0.33	-0.07
9	0.59	1.39	1.90	2.29	4.01	0.73	0.24	-0.04
10	0.55	1.01	1.95	2.47	3.78	0.75	0.30	-0.29
11	0.47	0.72	2.32	2.90	1.70	0.67	0.83	-0.27
12	0.62	0.88	1.25	1.93	3.49	0.4	-0.01	-0.28
13	0.62	1.11	1.46	2.06	6.31	0.41	-0.11	-0.17
14	0.61	1.02	1.53	1.90	4.97	0.37	0.00	-0.26
15	0.53	0.92	2.20	2.98	0.99	4.99	-2.73	-4.04
16*	0.21	0.14	-5.45	-0.3	0.84	0.73	-0.94	-1.48
17	0.52	0.77	1.72	2.51	1.47	0.47	0.96	-0.18
18	0.48	0.53	2.44	3.19	1.89	0.32	0.91	-0.39

Note/ SGRM = Samejima's graded response model GGUM = generalized graded unfolding model; Asterisks indicate reverse scored items.

Comparing Dominance vs. Ideal-Point IRT Models

In order to compare the SGRM (dominance model) and GGUM (ideal point model; Research Question 2), we assessed Model fit with sample-size adjusted ratios of chi-square to degrees of freedom (X^2/df) using the MODFIT program (Stark, 2001), where the sample size is adjusted to 3,000 and the test uses the noncentral X^2 distribution (Drasgow, Levine, Tsien, Williams, & Mead, 1995; Tay, Ali, Drasgow, & Williams, 2011). Chi-square tests compare the observed frequency of response options to the frequency expected from the item response function (Drasgow et al., 1995). Each model (SGRM and GGUM) yields different response functions for each item, which allows us to compare relative fit using the X^2/df statistics. Since single item chi-square statistics can be blind to certain types of misfit (Van den Wollenberg, 1982), we also obtained the chi-square statistics for item doublets and triplets, and we focus on these values for the purpose of comparing fit of the SGRM and GGUM. Lower values of X^2/df (less than 3) indicate better model fit (Drasgow et al., 1995; Tay et al., 2011). SGRM had more item doublets with $X^2/df < 3$ than the GGUM (15 compared to 8). The same was true for item triplets—SGRM had all but two triplet with a $X^2/df < 3$, while the GGUM only had two triplets with $X^2/df < 3$. Because the SGRM had low values of X^2/df for individual item doublets and triplets and had a lower mean X^2/df ($M = 0.69$, $SD = 0.76$) compared to the average X^2/df from the GGUM ($M = 46.44$, $SD = 67.89$), we conclude that *the dominance model fits the data better than the ideal point model*.

To graphically evaluate item fit, we examined the item fit plots for both models to determine how much discrepancy exists between the option response function (ORF) and observed responses (see appendix). Fit plots were somewhat mixed. The SGRM plots

demonstrated moderately good fit between the ORF and participant responses, and most items showed response patterns consistent with dominance models. Yet according to visual inspection some items did appear to show poor fit (items 6, 16 and 18). The items that fit poorly were the two reverse scored items and item 18, which had a limited range of responses compared to other items. Most of the GGUM plots also demonstrated moderately good fit between item response functions and responses, with the exception of items 6, 15, 16, and 18. There were no items that clearly fit the GGUM better than the SGRM, but some items showed poor fit with either model. The fit plots confirm that the reverse scored items were not well predicted by either a dominance or ideal point response curve (note, we refer to items 6 and 16 as “reverse scored” even though the responses are not technically recoded [reverse scored] for estimating the GGUM).

Item information and response functions

We estimated test information curves for both models (Figure 9-10). The SGRM accounts for more total information than the GGUM. The information curves showed that most items provide information when theta is above 0. We note that the GGUM information curve is bimodal, while the SGRM information curve is not. These differences are attributable to the way the item response functions are calculated in the dominance and ideal point models. Item information functions estimated within an ideal point model framework are bimodal because the information function is related to the slope of the item response function and the slope is zero at the item location (i.e., the peak of the bell-shaped curve; Andrich, 1996; Chernyshenko et al., 2007). Consequently, the most information is provided for the theta values immediately surrounding the item location. When the item information functions are combined to form the test information functions, they produce different curves and provide information about different ranges of the latent trait. Further inspection of each item’s information function shows that the

reverse scored items 6 and 16 provided the least amount of information for both models (combined they provide information statistics < 1 for both the SGRM and GGUM).

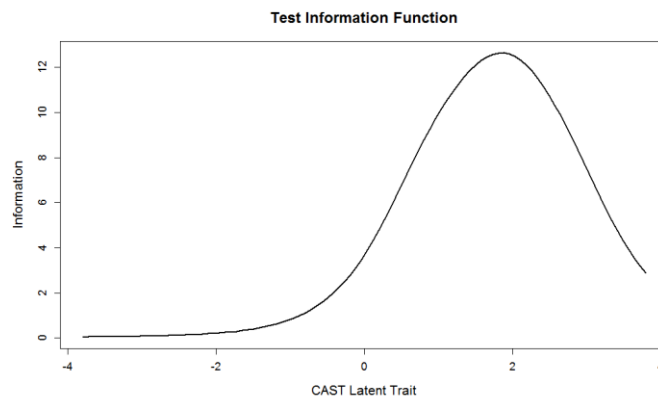


Figure 9 Test information function for the SGRM

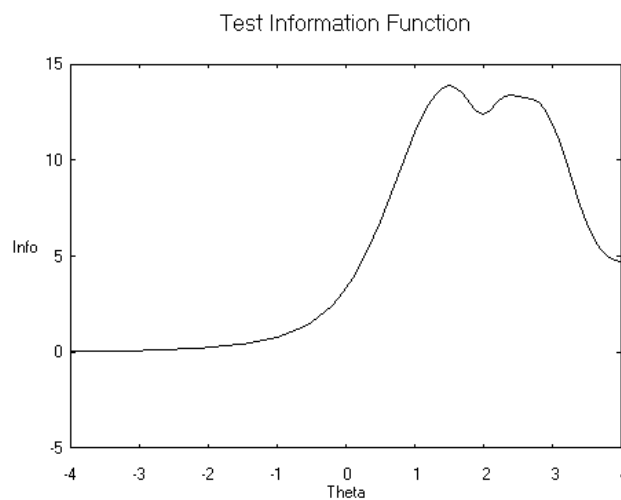


Figure 10. Test information function for the GGUM

To better illustrate how each item functions, we present the ORFs for the 18 CAST items (Figure 11). These response functions depict the probability of agreeing, disagreeing, or responding neutrally to the item at various levels of the underlying trait (theta). Given that the SGRM provides more information, has more reasonable item parameter estimates, and has better overall fit with the data than the GGUM, we will only interpret the plots from SGRM.

The most notable plots are those for the reverse scored items (6 and 16), which show

quite linear functions and are uncharacteristic of dominance model ORFs. In general, the plots illustrate the main finding from the item parameter estimates, which is that all items are quite “difficult.” With the exception of the reverse scored items, the probability of not endorsing an item (i.e., responding “disagree” or “strongly disagree”) is nearly certain until theta reaches at least zero (the population mean), which suggests the CAST is poor at differentiating among individuals who are low on the trait. Furthermore, there are several items (e.g., items 8 and 15) that require very high levels of the trait before the probability of endorsing the item reaches 50% (i.e., very difficult items).

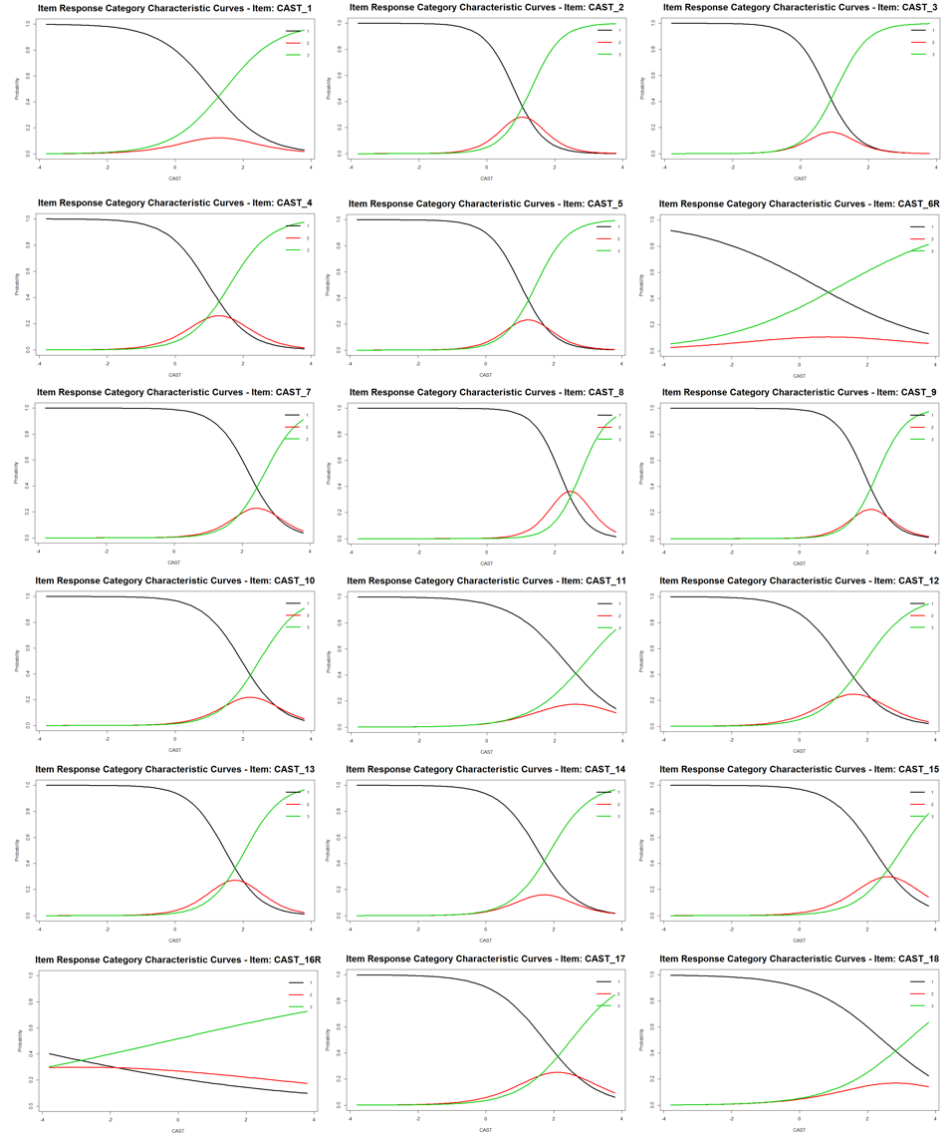


Figure 11. Option Response Functions for all CAST items (item parameters estimated with SGRM).

Measurement invariance across gender

Before interpreting any mean differences in CAST scores between men and women, we assessed measurement invariance (or equivalence) across gender. We used a mean and covariance structures (MACS) method (Cheung & Rensvold, 2000; Schmitt & Kuljanin, 2008) of measurement equivalence rather than an item response method, since the MACS method is preferred for multidimensional constructs and moderate sample sizes (Meade & Lautenschlager, 2004; Raju, Laffitte, & Byrne, 2002). This approach uses progressive model comparisons to test three types of measurement equivalence: configural, metric, and scalar (Vandenberg & Lance, 2000). Before testing these models, we estimated the fit of the oblique three-factor structure of sadism for the male ($N = 132$) ($\chi^2 = 218.83$ (132); RMSEA = .071; SRMR = .083; CFI = .89) and female ($N = 300$) ($\chi^2 = 398.48$ (132); RMSEA = .082; SRMR = .066; CFI = .86) subsamples separately. The fit for the two samples was marginally adequate. Although the ratio of χ^2 to degrees of freedom was reasonable, the RMSEA, SRMR, and CFI fit statistics were slightly outside the typical cutoffs of .08, .05, and .90, respectively (Cheung & Rensvold, 2002; Hu & Bentler, 1999; Van de Schoot, Lugtig, & Hox, 2012). We proceed with comparing the relative fit of the nested models (Table 7), but caution the underlying factor model only showed modest fit.

Configural equivalence was assessed by constraining the factor structure (i.e., pattern of factor loadings) to be equivalent between the male and female samples. The baseline model, configural equivalence, exhibited marginally adequate fit ($\chi^2 = 617.73$ (264); RMSEA = .079; CFI = .87, SRMR = .068). Again, the CFI is slightly below the typically applied .90 cutoff, but the RMSEA demonstrates adequate fit. We next tested metric equivalence, by constraining the factor loadings of the items to be equal in magnitude across gender groups ($\chi^2 = 641.73$ (279); RMSEA = .078; CFI = .87). The metric equivalence model (i.e., constraining factor loadings to be equal across genders) does not demonstrate worse fit than the configural model ($\Delta \text{CFI} < .01$),

which is interpreted as support for metric equivalence (Cheung & Rensvold, 2002). As such, we proceeded to test for scalar invariance by constraining the models to have equal intercepts across samples. This model of strong measurement invariance had adequate fit ($\chi^2 = 669.41$ (294); RMSEA = .077; CFI = .86) and did not fit worse than the metric invariant model (Δ CFI < .01), supporting the conclusion of scalar equivalence between genders. For completeness, we proceeded to estimate a model of strict measurement invariance in which the error variances (item uniquenesses) were constrained to be equal across gender groups, but the model fit was poor ($\chi^2 = 749.12$ (297); RMSEA = .084; CFI = .83). However, this level of measurement invariance is not necessary for interpreting mean differences between groups. As such, results support scalar measurement invariance between genders for the CAST.

Table 7

Nested Model Fit Statistics for measurement invariance tests

Model	Chi Square	df	CFI	RMSEA	SRMR	TLI	BIC	Δ CFI
<u>Configural</u> (same pattern of loadings)	617.30	264	0.870	0.079	0.068	0.85	24083.5	--
<u>Metric</u> (configural + same magnitude of loadings)	641.73	279	0.867	0.078	0.071	0.85	24016.9	.003
<u>Scalar</u> (metric + same item intercepts)	669.31	294	0.862	0.077	0.076	0.857	23953.5	.005
<u>Invariant Uniquenesses</u> (scalar + same item uniquenesses)	749.12	297	0.834	0.084	0.103	0.709	24015.1	.028

Mean gender differences

Men had higher mean sadism scores than women overall ($t = -8.68$, $p < .01$), and men also scored higher than women on all three factors of sadism (physical, verbal, and vicarious

(verbal, $t = -4.93$, $p < .01$; physical, $t = -3.70$, $p < .01$; vicarious, $t = -10.29$, $p < .01$; see Table 8).

Table 8

CAST gender score differences

	Female		Male			t-test	
	Mean	Standard Deviation	Mean	Standard Deviation	d	Statistic	P value
CAST	1.91	0.64	2.60	0.81	0.90	-8.67	0.0000
Verbal	2.41	1.15	3.05	1.15	0.52	-4.93	0.0000
Physical	1.36	0.67	1.67	0.67	0.42	-3.70	0.0003
Vicarious	1.88	0.65	2.89	0.65	1.10	49.85	0.0000

Note: $N = 440$; Female $N = 300$; Male $N = 140$; male coded 1, female coded 2.

Chapter 4: Discussion

The current study presented a psychometric investigation of the most popular self-report measure of sadism, the Comprehensive Assessment of Sadistic Tendencies (CAST: Buckels & Paulhus, 2014). The sudden increase in the popularity of sadism research means that researchers have rapidly incorporated the CAST into their measures without knowing much about the psychometric properties of the instrument. We aimed to shed light on these properties by providing empirical estimates of the factor structure, item characteristics, measurement invariance, and response processes underlying sadism. We provide evidence confirming the three-factor structure of everyday sadism (verbal, physical, and vicarious), an underlying dominance response process (as opposed to an ideal point response process), and measurement invariance across gender.

We first note that scores on the CAST are very low (for both the overall measure and verbal, physical, and vicarious facet scales)—most people give low levels of endorsement to the items. With the relatively low endorsement of sadism, it is still noteworthy that there exist some individuals who endorse many items. The CAST, consistent with other “dark triad” measures (Jakobwitz & Egan, 2006; Paulhus et al., 2018), is able to identify people high on the trait, even in non-clinical samples.

The three-factor structure of physical, verbal, and vicarious sadism proposed by Buckels and Paulhus (2014) was supported, as it demonstrated superior fit compared to a unidimensional model or a two-factor model distinguishing between direct and indirect sadism. Factor analytic models fit better when facets of sadism distinguish between the type of pain being enjoyed (i.e., verbal and physical). We also assessed the structure of the entire dark tetrad (sadism, Machiavellianism, narcissism, and psychopathy) and found that the model demonstrated

marginally good fit, but the relationships between the latent constructs in the dark tetrad model do not support the discriminant validity of sadism in the current dataset. Machiavellianism and psychopathy, specifically, are highly correlated with sadism in our sample. We encourage future researchers to continue to measure sadism, but to also account for the variance explained by the other dark triad traits (Machiavellianism, psychopathy, and narcissism). Although we did not measure honesty-humility in the current study, other researchers have found evidence that sadism is redundant with this “H” factor (Book et al., 2016). More research should be done to demonstrate that sadism is unique from other personality traits.

Items 6 and 16 (the two reverse scored items) had particularly low factor loadings, low item-test correlations, unusual IRT item parameters and item characteristic curves, and provide very little information about the latent trait of the individual. While it is clear that these items did not behave similarly to other items, it is unclear why. Perhaps there is a “method” factor influencing the observed responses, in which participants either do not recognize the reverse wording and erroneously endorse an item, or the cognitive process of appraising their tendency to *not* do something caused participants to answer in a manner not reflective of their trait level of sadism. The measures we administered did not include attention check items, making it difficult to evaluate whether these responses are due to inattentive responding. It could also be that the specific content of these items is less sadistic. These items qualify the intentions of the participant. Item 6 asks if participants agree that they “would never purposely humiliate someone,” but it is unclear how participants would respond if they have humiliated someone unintentionally. Item 6 also uses an extreme frequency qualifier (Nye, Newman, & Joseph, 2010), specifying that the participant would “never” do this behavior; but it is unclear how participants would respond if they rarely engaged in humiliation. Item 16 similarly asks if there

is “too much violence in sports,” which might be difficult for participants to answer if they believe there is some, but not “too much,” violence. Violence in sports might also be accepted as a societal norm, and consequently not a good indicator of sadism. It is less clear why the content of item 6 would be a poor indicator, although sadism may allow individuals to infer reduced culpability (Trémolière & Djeriouat, 2016), making it plausible that those high on the trait do not recognize their purposeful role in inflicting pain even if they derive pleasure from the behavior. We conclude that the CAST functions well as an instrument of self-report sadism with the proposed 18 items, but that researchers should pay particular attention to the reverse-worded items (6 and 16) in the future and attempt to include items that flag inattentive responding patterns, so that future psychometric evaluations can be carried out on the subset of respondents who passed attention check items. It might be appropriate to remove items 6 and 16 from the scale if future studies demonstrate that the poor item characteristics are not due to inattentive responding.

In order to gain a better understanding of the participants’ response process regarding the CAST, we estimated and compared two distinct classes of IRT models: a dominance model and ideal point model. Although the response process underlying the ideal point model is theoretically better suited for some personality traits (Dragow et al., 2010), results from X^2/df tests and graphic fit plots suggest that a dominance model had superior fit for the CAST. This is not entirely unexpected, given that the CAST was likely developed with a factor analytic approach, which eliminates items with moderate locations on the theta continuum. The fit plots showed patterns indicative of a dominance response process, but estimates had relatively large standard errors. Nonetheless, the CAST appears to contain the most test information in the high trait range (at higher theta levels). We encourage future researchers to cross-validate our findings

before any of the parameter estimates are incorporated into estimating sadism scores for individuals.

Sadism was conceived as an insidious trait that characterizes individuals within normative, functional populations in our society. The item parameters obtained from the better fitting dominance model reveal that most items are difficult, meaning respondents need to be very sadistic in order to endorse most items. This mirrors the descriptive statistics that show a response distribution with a very low mean score. Furthermore, the test information function shows that the CAST provides the most information for individuals who are above average on the latent trait. This is likely a good test attribute, given that the instrument is intended to differentiate sadistic individuals. It is not particularly important for the CAST to be able to provide information about people who are very low on the trait. Given that relatively few studies have actively compared different measurement tools for sadism, it is unclear if other measures (O'Meara et al., 2011; Plouffe et al., 2017) show a greater distribution of scores or provide more information across the entire trait continuum.

Finally, we investigated differences in sadism between men and women. Before examining the mean score differences, we conducted a measurement invariance analysis to ensure that the factor structure, factor loadings, and intercepts were equivalent for both groups. The model fit statistics for the nested CFAs were only moderately good, since the CFI and RMSEA values mostly fell just outside typical cutoffs. Nonetheless, relative comparisons of models demonstrated evidence of metric invariance across gender. This means that mean differences between men and women represent real differences in levels of the latent trait and are not due to measurement bias. Consistent with previous results (Buckels & Paulhus, 2014), men scored significantly higher than women on the overall CAST as well as the three subscales, with

the largest gender difference occurring on the vicarious sadism scale. Given that the content of the items on this scale deals heavily with sports and videogames, we suspect that the difference may simply be driven by items focusing on stereotypically male-specific behaviors (Greenberg, Sherry, Lachlan, Lucas, & Holmstrom, 2010; Lucas & Sherry, 2004).

Limitations

While we believe this study marks a good first step in psychometrically investigating the CAST as an assessment of sadism, there are a number of methodological and theoretical limitations to note that require follow-up. First, we used a college subject-pool sample, which is consistent with the stated goal of the CAST as an assessment of sadism in the general population; however, it goes without saying that sadism is highly relevant to questions posed by clinical (Međedović, 2017; Zeigler-Hill & Vonk, 2015) and forensic psychologists (Chabrol et al., 2009; Reidy et al., 2011; Russell & King, 2016; Stone, 2010; Trémolière & Djeriouat, 2016). It is unclear whether the CAST is a good instrument for these populations or whether it should only be used in normative samples. Second, although our sample size is sufficiently large for the analyses we conducted, IRT item parameter estimates and model comparison can be unstable in sample sizes less than 1000 (Tay et al., 2011). Our male sample, in particular, was relatively small, which may explain the moderately poor fit statistics of the underlying models in the measurement invariance tests.

Conclusion

The Comprehensive Assessment of Sadistic Tendencies Scales (CAST) has been instrumental in providing an assessment tool for identifying trait sadism in normative populations (Buckels & Paulhus, 2014). The growth in sadism research without documented validation for the CAST is problematic. The current paper aimed to begin alleviating these issues

by providing psychometric information about the test and its items using factor analytic and item response theory approaches. We confirmed the three-factor oblique structure proposed by Buckels and Paulhus (2014) that includes measures of three facets of sadism: verbal, physical, and vicarious. Item response theory analyses revealed that the response process for the CAST follows a dominance process rather than an ideal point response process and that the CAST provides the most information for those high on the latent trait. Additionally, we provided evidence of scalar measurement invariance across gender, which allowed us to meaningfully compare mean differences in sadism between men and women. Men scored higher than women on the overall scale as well as each subscale (verbal, physical, vicarious). We do note a few causes for concern that researchers might want to consider when they use the CAST. First, there may not be sufficient evidence of discriminant validity between Machiavellianism and Psychopathy. Second, the two reversed items (item 6: “I would never purposefully humiliate someone” and item 16: “There is way too much violent in sports”) have unusual psychometric properties. We recommend that researchers include attention checks in future administrations of the CAST in order to rule out insufficient attention as the cause of these atypical response patterns.

References

- Ames, D. R., Rose, P., & Anderson, C. P. (2006). The NPI-16 as a short measure of narcissism. *Journal of research in personality*, 40(4), 440-450.
<https://doi.org/10.1016/j.jrp.2005.03.002>
- Andrich, D. (1996). A hyperbolic cosine latent trait model for unfolding polytomous responses: Reconciling Thurstone and Likert methodologies. *British Journal of Mathematical and Statistical Psychology*, 49(2), 347-365. 10.1111/j.2044-8317.1996.tb01093.x
- Archer, J. (2004). Sex differences in aggression in real-world settings: a meta-analytic review. *Review of general Psychology*, 8(4), 291-322. <http://dx.doi.org/10.1037/1089-2680.8.4.291>
- Balakrishnan, A., Plouffe, R. A., & Saklofske, D. H. (2017). What do sadists value? Is honesty-humility an intermediary? Replicating and extending findings on the link between values and “dark” personalities. *Personality and Individual Differences*, 109, 142-147.
<https://doi.org/10.1016/j.paid.2016.12.055>
- Bandalos, D. L., & Finney, S. J. (2001). Item parceling issues in structural equation modeling. *New developments and techniques in structural equation modeling*, 269, V296.
- Baughman, H. M., Dearing, S., Giammarco, E., & Vernon, P. A. (2012). Relationships between bullying behaviours and the Dark Triad: A study with adults. *Personality and Individual Differences*, 52(5), 571-575. <https://doi.org/10.1016/j.paid.2011.11.020>

- Baumeister, R. F., & Campbell, W. K. (1999). The intrinsic appeal of evil: Sadism, sensational thrills, and threatened egotism. *Personality and Social Psychology Review*, 3(3), 210-221. https://doi.org/10.1207/s15327957pspr0303_4
- Bettencourt, B., & Miller, N. (1996). Gender differences in aggression as a function of provocation: a meta-analysis. *Psychological bulletin*, 119(3), 422-447. <http://dx.doi.org/10.1037/0033-2909.119.3.422>
- Björkqvist, K. (1994). Sex differences in physical, verbal, and indirect aggression: A review of recent research. *Sex roles*, 30(3), 177-188. doi:10.1007/BF01420988
- Book, A., Visser, B. A., Blais, J., Hosker-Field, A., Methot-Jones, T., Gauthier, N. Y., Volk, A., Holden, R.R., & D'Agata, M. T. (2016). Unpacking more “evil”: What is at the core of the dark tetrad?. *Personality and Individual Differences*, 90, 269-272. <https://doi.org/10.1016/j.paid.2015.11.009>
- Brown, A., & Maydeu-Olivares, A. (2010). Issues that should not be overlooked in the dominance versus ideal point controversy. *Industrial and Organizational Psychology*, 3(4), 489-493. doi:10.1111/j.1754-9434.2010.01277.x
- Buckels, E. E. (2012). *The pleasures of hurting others: Behavioral evidence for everyday sadism* (Doctoral dissertation, University of British Columbia). doi:10.14288/1.0073016
- Buckels, E. E., & Paulhus, D. L. (2014). Comprehensive assessment of sadistic tendencies (CAST). Unpublished instrument, Vancouver, Canada: University of British Columbia.
- Buckels, E. E., Jones, D. N., & Paulhus, D. L. (2013). Behavioral confirmation of everyday sadism. *Psychological Science*, 24(11), 2201-2219. doi:10.1177/0956797613490749

- Buckels, E. E., Trapnell, P. D., & Paulhus, D. L. (2014). Trolls just want to have fun. *Personality and Individual Differences*, 67, 97-102. <https://doi.org/10.1016/j.paid.2014.01.016>
- Cao, M., Drasgow, F., & Cho, S. (2015). Developing ideal intermediate personality items for the ideal point model. *Organizational Research Methods*, 18(2), 252-275.
doi:10.1177/1094428114555993
- Carter, G. L., Campbell, A. C., & Muncer, S. (2014). The dark triad personality: Attractiveness to women. *Personality and Individual Differences*, 56, 57-61.
<https://doi.org/10.1016/j.paid.2013.08.021>
- Carter, N. T., & Dalal, D. K. (2010). An ideal point account of the JDI Work satisfaction scale. *Personality and Individual Differences*, 49(7), 743-748.
<https://doi.org/10.1016/j.paid.2010.06.019>
- Carter, N. T., Guan, L., Maples, J. L., Williamson, R. L., & Miller, J. D. (2015). The downsides of extreme conscientiousness for psychological well-being: The role of obsessive compulsive tendencies. *Journal of personality*, 84(4), 510-522. doi:10.1111/jopy.12177
- Chabrol, H., Van Leeuwen, N., Rodgers, R., & Séjourné, N. (2009). Contributions of psychopathic, narcissistic, Machiavellian, and sadistic personality traits to juvenile delinquency. *Personality and Individual Differences*, 47(7), 734-739.
<https://doi.org/10.1016/j.paid.2009.06.020>
- Chernyshenko, O. S., Stark, S., Chan, K. Y., Drasgow, F., & Williams, B. (2001). Fitting item response theory models to two personality inventories: Issues and insights. *Multivariate Behavioral Research*, 36(4), 523-562. http://dx.doi.org/10.1207/S15327906MBR3604_03

- Chernyshenko, O. S., Stark, S., Drasgow, F., & Roberts, B. W. (2007). Constructing personality scales under the assumptions of an ideal point response process: Toward increasing the flexibility of personality measures. *Psychological Assessment*, 19(1), 88-106.
<http://dx.doi.org/10.1037/1040-3590.19.1.88>
- Cheung, G. W., & Rensvold, R. B. (2000). Assessing extreme and acquiescence response sets in cross-cultural research using structural equations modeling. *Journal of Cross-Cultural Psychology*, 31(2), 187-212. doi:[10.1177/0022022100031002003](https://doi.org/10.1177/0022022100031002003)
- Cheung, G. W., & Rensvold, R. B. (2002). Evaluating goodness-of-fit indexes for testing measurement invariance. *Structural equation modeling*, 9(2), 233-255.
http://dx.doi.org/10.1207/S15328007SEM0902_5
- Cho, S., Drasgow, F., & Cao, M. (2015). An investigation of emotional intelligence measures using item response theory. *Psychological assessment*, 27(4), 1241-1252.
<http://dx.doi.org/10.1037/pas0000132>
- Coombs, C. H. (1964). *A theory of data*. New York, NY: Wiley.
- Credé, M. (2010). Two caveats for the use of ideal point items: Discrepancies and bivariate constructs. *Industrial and Organizational Psychology*, 3(4), 494-497. doi:[10.1111/j.1754-9434.2010.01278.x](https://doi.org/10.1111/j.1754-9434.2010.01278.x)
- Crick, N. R., & Grotpeter, J. K. (1995). Relational aggression, gender, and social-psychological adjustment. *Child development*, 66(3), 710-722. [10.1111/j.1467-8624.1995.tb00900.x](https://doi.org/10.1111/j.1467-8624.1995.tb00900.x)
- Cross, C. P., Copping, L. T., & Campbell, A. (2011). Sex differences in impulsivity: a meta-analysis. *Psychological bulletin*, 137(1), 97-130. <http://dx.doi.org/10.1037/a0021591>

- Curran, P. J., West, S. G., & Finch, J. F. (1996). The robustness of test statistics to nonnormality and specification error in confirmatory factor analysis. *Psychological methods*, 1(1), 16-29
- Dahling, J. J., Whitaker, B. G., & Levy, P. E. (2009). The development and validation of a new Machiavellianism scale. *Journal of management*, 35(2), 219-257.
<https://doi.org/10.1177/0149206308318618>
- Davies, J., & O'Meara, A. (2007). 'I consider myself sadistic': a qualitative analysis of sadistic endorsement in a group of Irish undergraduates. *The British Journal of Forensic Practice*, 9(1), 24-30. <https://doi.org/10.1108/14636646200700005>
- Drasgow, F., Chernyshenko, O. S., & Stark, S. (2010). 75 years after Likert: Thurstone was right!. *Industrial and Organizational Psychology*, 3(4), 465-476. doi:10.1111/j.1754-9434.2010.01273.x
- Drasgow, F., & Hulin, C. L. (1990). Item response theory, In M. D. Dunnette & L. M. Hough (Eds.), *Handbook of industrial organizational psychology* (pp. 577-636). Palo Alto, CA.: Consulting Psychologist Press.
- Drasgow, F., Levine, M. V., Tsien, S., Williams, B., & Mead, A. D. (1995). Fitting polytomous item response theory models to multiple-choice tests. *Applied Psychological Measurement*, 19(2), 143-166. doi:[10.1177/014662169501900203](https://doi.org/10.1177/014662169501900203)
- Drasgow, F., & Parsons, C. K. (1983). Application of unidimensional item response theory models to multidimensional data. *Applied Psychological Measurement*, 7(2), 189-199. doi:[10.1177/014662168300700207](https://doi.org/10.1177/014662168300700207)

- Eisenberg, N., & Lennon, R. (1983). Sex differences in empathy and related capacities. *Psychological Bulletin*, 94(1), 100. <http://dx.doi.org/10.1037/0033-2909.94.1.100>
- Fornell, C., & Larcker, D. F. (1981). Evaluating structural equation models with unobservable variables and measurement error. *Journal of marketing research*, 39-50. DOI: 10.2307/3151312
- Freud, S. (1905). The sexual aberrations. *Three Essays on the Theory of Sexuality*, 1-38.
- Furnham, A., Richards, S. C., & Paulhus, D. L. (2013). The Dark Triad of personality: A 10 year review. *Social and Personality Psychology Compass*, 7(3), 199-216.
doi:10.1111/spc3.12018
- Furnham, A., & Trickey, G. (2011). Sex differences in the dark side traits. *Personality and Individual Differences*, 50(4), 517-522. <https://doi.org/10.1016/j.paid.2010.11.021>
- Greenberg, B. S., Sherry, J., Lachlan, K., Lucas, K., & Holmstrom, A. (2010). Orientations to video games among gender and age groups. *Simulation & Gaming*, 41(2), 238-259.
doi:[10.1177/1046878108319930](https://doi.org/10.1177/1046878108319930)
- Greitemeyer, T. (2015). Everyday sadism predicts violent video game preferences. *Personality and Individual Differences*, 75, 19-23. <https://doi.org/10.1016/j.paid.2014.10.049>
- Greitemeyer, T., & Sagioglou, C. (2017). The longitudinal relationship between everyday sadism and the amount of violent video game play. *Personality and Individual Differences*, 104, 238-242. <https://doi.org/10.1016/j.paid.2016.08.021>

- Grijalva, E., Newman, D. A., Tay, L., Donnellan, M. B., Harms, P. D., Robins, R. W., & Yan, T. (2015). Gender differences in narcissism: A meta-analytic review. *Psychological bulletin*, 141(2), 261-310. <http://dx.doi.org/10.1037/a0038231>
- Hambleton, R. K., Swaminathan, H., & Rogers, H. J. (1991). *Fundamentals of item response theory*. London, England: Sage.
- Harpur, T. J., Hare, R. D., & Hakstian, A. R. (1989). Two-factor conceptualization of psychopathy: Construct validity and assessment implications. *Psychological Assessment: A Journal of consulting and clinical Psychology*, 1(1), 6-17. <http://dx.doi.org/10.1037/1040-3590.1.1.6>
- Heilbrun Jr, A. B., & Loftus, M. P. (1986). The role of sadism and peer pressure in the sexual aggression of male college students. *Journal of Sex Research*, 22(3), 320-332. <http://dx.doi.org/10.1080/00224498609551312>
- Hervé, H., & Yuille, J. C. (Eds.). (2017). *The psychopath: Theory, research, and practice*. Routledge.
- Hu, L. T., & Bentler, P. M. (1999). Cutoff criteria for fit indexes in covariance structure analysis: Conventional criteria versus new alternatives. *Structural equation modeling: a multidisciplinary journal*, 6(1), 1-55. <http://dx.doi.org/10.1080/10705519909540118>
- Huang, J., & Mead, A. D. (2014). Effect of personality item writing on psychometric properties of ideal-point and likert scales. *Psychological assessment*, 26(4), 1162-1172. <http://dx.doi.org/10.1037/a0037273>

- Idaszak, J. R., & Drasgow, F. (1987). A revision of the Job Diagnostic Survey: Elimination of a measurement artifact. *Journal of Applied Psychology*, 72(1), 69-74.
<http://dx.doi.org.proxy2.library.illinois.edu/10.1037/0021-9010.72.1.69>
- Jakobwitz, S., & Egan, V. (2006). The dark triad and normal personality traits. *Personality and Individual Differences*, 40(2), 331-339. <https://doi.org/10.1016/j.paid.2005.07.006>
- Jonason, P. K., & Kavanagh, P. (2010). The dark side of love: Love styles and the Dark Triad. *Personality and Individual Differences*, 49(6), 606-610.
<https://doi.org/10.1016/j.paid.2010.05.030>
- Jonason, P. K., Koenig, B. L., & Tost, J. (2010). Living a fast life. *Human Nature*, 21(4), 428-442. doi:10.1007/s12110-010-9102-4
- Jonason, P. K., Luevano, V. X., & Adams, H. M. (2012). How the Dark Triad traits predict relationship choices. *Personality and Individual Differences*, 53(3), 180-184.
<https://doi.org/10.1016/j.paid.2012.03.007>
- Jonason, P. K., & Tost, J. (2010). I just cannot control myself: The Dark Triad and self-control. *Personality and Individual Differences*, 49(6), 611-615.
<https://doi.org/10.1016/j.paid.2010.05.031>
- Jonason, P. K., Zeigler-Hill, V., & Okan, C. (2017). Good v. evil: Predicting sinning with dark personality traits and moral foundations. *Personality and Individual Differences*, 104, 180-185. <https://doi.org/10.1016/j.paid.2016.08.002>

- Kishton, J. M., & Widaman, K. F. (1994). Unidimensional versus domain representative parceling of questionnaire items: An empirical example. *Educational and Psychological Measurement*, 54(3), 757-765. <https://doi.org/10.1177/0013164494054003022>
- Krafft-Ebing, R., & Chaddock, C. G. (1893). *Psychopathia sexualis: With especial reference to contrary sexual instinct: A medico-legal study*. Philadelphia, PA: FA Davis.
- Krueger, R. B. (2010). The DSM diagnostic criteria for sexual sadism. *Archives of sexual behavior*, 39(2), 325-345. <https://doi.org/10.1007/s10508-009-9586-3>
- LaPalme, M. L., Wang, W., Joseph, D. L., Saklofske, D. H., & Yan, G. (2016). Measurement equivalence of the Wong and Law Emotional Intelligence Scale across cultures: An item response theory approach. *Personality and Individual Differences*, 90, 190-198. <https://doi.org/10.1016/j.paid.2015.10.045>
- Lee, K., Ashton, M. C., Wiltshire, J., Bourdage, J. S., Visser, B. A., & Gallucci, A. (2013). Sex, power, and money: Prediction from the Dark Triad and Honesty–Humility. *European Journal of Personality*, 27(2), 169-184. doi:10.1002/per.1860
- Little, T. D., Cunningham, W. A., Shahar, G., & Widaman, K. F. (2002). To parcel or not to parcel: Exploring the question, weighing the merits. *Structural Equation Modeling*, 9(2), 151-173. http://dx.doi.org/10.1207/S15328007SEM0902_1
- Lucas, K., & Sherry, J. L. (2004). Sex differences in video game play: A communication-based explanation. *Communication research*, 31(5), 499-523. doi:[10.1177/0093650204267930](https://doi.org/10.1177/0093650204267930)

- Marsh, H. W., Hau, K. T., Balla, J. R., & Grayson, D. (1998). Is more ever too much? The number of indicators per factor in confirmatory factor analysis. *Multivariate behavioral research*, 33(2), 181-220. https://doi.org/10.1207/s15327906mbr3302_1
- Marshall, W. L., & Kennedy, P. (2003). Sexual sadism in sexual offenders: An elusive diagnosis. *Aggression and Violent Behavior*, 8(1), 1-22. [https://doi.org/10.1016/S1359-1789\(01\)00052-0](https://doi.org/10.1016/S1359-1789(01)00052-0)
- Meade, A. W., & Lautenschlager, G. J. (2004). A comparison of item response theory and confirmatory factor analytic methodologies for establishing measurement equivalence/invariance. *Organizational Research Methods*, 7(4), 361-388. doi:[10.1177/1094428104268027](https://doi.org/10.1177/1094428104268027)
- Međedović, J. (2017). Aberrations in emotional processing of violence-dependent stimuli are the core features of sadism. *Motivation and Emotion*, 41(2), 273-283. doi:10.1007/s11031-016-9596-0
- Meloy, J. R. (1997). The psychology of wickedness: Psychopathy and sadism. *Psychiatric Annals*, 27(9), 630-633. doi:10.3928/0048-5713-19970901-10
- Miller, M. J. (2017). The Influence of Dark Personality on Counterproductive Work Behaviors. *Computer and Information Technology*, 52, 18-44.
- Miller, J. D., Watts, A., & Jones, S. E. (2011). Does psychopathy manifest divergent relations with components of its nomological network depending on gender? *Personality and Individual Differences*, 50(5), 564-569. <https://doi.org/10.1016/j.paid.2010.11.028>

- Miller, J. D., Dir, A., Gentile, B., Wilson, L., Pryor, L. R., & Campbell, W. K. (2010). Searching for a vulnerable dark triad: Comparing factor 2 psychopathy, vulnerable narcissism, and borderline personality disorder. *Journal of personality*, 78(5), 1529-1564.
doi:10.1111/j.1467-6494.2010.00660.x
- Millon, T., Millon, C., Davis, R. D., & Grossman, S. (2009). *Millon clinical multiaxial inventory-III (MCMI-III): Manual*. Pearson/PsychCorp.
- Mokros, A., Osterheider, M., Hucker, S. J., & Nitschke, J. (2011). Psychopathy and sexual sadism. *Law and human behavior*, 35(3), 188-199. doi:10.1007/s10979-010-9221-9
- Nye, C. D., Newman, D. A., & Joseph, D. L. (2010). Never say “always”? Extreme item wording effects on scalar invariance and item response curves. *Organizational Research Methods*, 13(4), 806-830. <https://doi.org/10.1177/1094428109349512>
- O'Boyle Jr, E. H., Forsyth, D. R., Banks, G. C., & McDaniel, M. A. (2012). A meta-analysis of the dark triad and work behavior: A social exchange perspective. *Journal of Applied Psychology*, 97(3), 557-579. doi:10.1037/a0025679
- O'Brien, E., & LaHuis, D. M. (2011). Do applicants and incumbents respond to personality items similarly? A comparison of dominance and ideal point response models. *International Journal of Selection and Assessment*, 19(2), 109-118. doi:10.1111/j.1468-2389.2011.00539.x
- O'Meara, A., Davies, J., & Hammond, S. (2011). The psychometric properties and utility of the Short Sadistic Impulse Scale (SSIS). *Psychological assessment*, 23(2), 523-531.
<http://dx.doi.org/10.1037/a0022400>

- Oosterhuis, H. (2012). Sexual modernity in the works of Richard von Krafft-Ebing and Albert Moll. *Medical history*, 56(2), 133-155. <https://doi.org/10.1017/mdh.2011.30>
- Österman, K., Björkqvist, K., Lagerspetz, K. M., Kaukiainen, A., Landau, S. F., Frączek, A., & Caprara, G. V. (1998). Cross-cultural evidence of female indirect aggression. *Aggressive behavior*, 24(1), 1-8. 10.1002/(SICI)1098-2337
- Paulhus, D. L. (2014). Toward a taxonomy of dark personalities. *Current Directions in Psychological Science*, 23(6), 421-426. doi:[10.1177/0963721414547737](https://doi.org/10.1177/0963721414547737)
- Paulhus, D. L., Curtis, S. R., & Jones, D. N. (2018). Aggression as a Trait: The Dark Tetrad Alternative. *Current Opinion in Psychology*, 19, 88-92. <https://doi.org/10.1016/j.copsyc.2017.04.007>
- Paulhus, D. L., & Jones, D. N. (2015). Measures of dark personalities. In G.J. Boyle, D.H. Saklofske, G. Mathews (Eds.), *Measures of personality and social psychological constructs*, (pp. 562-594). San Diego, CA: Academic Press.
- Paulhus, D. L., & Jones, D. N. (2016). Individual Differences in Aggression. In B.J. Bushman (Ed.), *Aggression and Violence: A Social Psychological Perspective* (pp. 105-118). New York, NY: Taylor & Francis.
- Paulhus, D. L., Neumann, C. S., & Hare, R. D. (2009). Manual for the self-report psychopathy scale. *Toronto: Multi-health systems*.
- Paulhus, D. L., & Williams, K. M. (2002). The dark triad of personality: Narcissism, Machiavellianism, and psychopathy. *Journal of research in personality*, 36(6), 556-563. doi:[10.1016/S0092-6566\(02\)00505-6](https://doi.org/10.1016/S0092-6566(02)00505-6)

- Pfattheicher, S., & Schindler, S. (2015). Understanding the dark side of costly punishment: The impact of individual differences in everyday sadism and existential threat. *European Journal of Personality*, 29(4), 498-505. doi:10.1002/per.2003
- Plouffe, R. A., Saklofske, D. H., & Smith, M. M. (2017). The assessment of sadistic personality: Preliminary psychometric evidence for a new measure..*Personality and Individual Differences*, 104, 166-171.
doi:<http://dx.doi.org.proxy2.library.illinois.edu/10.1016/j.paid.2016.07.043>
- Raju, N. S., Laffitte, L. J., & Byrne, B. M. (2002). Measurement equivalence: A comparison of methods based on confirmatory factor analysis and item response theory. *Journal of Applied Psychology*, 87(3), 517-529. <http://dx.doi.org/10.1037/0021-9010.87.3.517>
- Rasch, G. (1966). An item analysis which takes individual differences into account. *British journal of mathematical and statistical psychology*, 19(1), 49-57. 10.1111/j.2044-8317.1966.tb00354.x
- Reckase, M. D. (1979). Unifactor latent trait models applied to multifactor tests: Results and implications. *Journal of educational statistics*, 4(3), 207-230.
doi:[10.3102/10769986004003207](http://dx.doi.org/10.3102/10769986004003207)
- Reidy, D. E., Zeichner, A., & Seibert, L. A. (2011). Unprovoked aggression: Effects of psychopathic traits and sadism. *Journal of personality*, 79(1), 75-100.
doi:10.1111/j.1467-6494.2010.00691.x
- Rizopoulos, D. (2006). ltm: An R package for Latent Variable Modelling and Item Response Theory Analyses, *Journal of Statistical Software*, 17 (5), 1-25.
URL <http://www.jstatsoft.org/v17/i05/>

- Roberts, J. S., Donoghue, J. R., & Laughlin, J. E. (2000). A general item response theory model for unfolding unidimensional polytomous responses. *Applied Psychological Measurement*, 24(1), 3-32. doi:[10.1177/01466216000241001](https://doi.org/10.1177/01466216000241001)
- Roberts, J. S., Fang, H. R., Cui, W., & Wang, Y. (2006). GGUM2004: A Windows-Based Program to Estimate Parameters in the Generalized Graded Unfolding Model. *Applied Psychological Measurement*, 30(1), 64-65. <http://dx.doi.org/10.1177/0146621605280141>
- Roberts, J. S., & Shim, H. S. (2008). GGUM2004 technical reference manual. Atlanta, GA: Georgia Institute of Technology. Retrieved from <http://prdlab.gatech.edu/unfolding/freesoftware/>
- Rosseel, Y. (2012). lavaan: An R Package for Structural Equation Modeling. *Journal of Statistical Software*, 48(2), 1-36. URL <http://www.jstatsoft.org/v48/i02/>.
- Russell, T. D., & King, A. R. (2016). Anxious, hostile, and sadistic: Maternal attachment and everyday sadism predict hostile masculine beliefs and male sexual violence. *Personality and Individual Differences*, 99, 340-345. <https://doi.org/10.1016/j.paid.2016.05.029>
- Russell, T. D., & King, A. R. (2017). Mean girls: PID-5 personality traits and everyday sadism predict hostile femininity. *Personality and Individual Differences*, 104, 252-257. <https://doi.org/10.1016/j.paid.2016.08.020>
- Samejima F. (1969). Estimation of latent ability using a response pattern of graded scores. *Psychometrika Monograph Supplement*, No. 17. Richmond, VA: Psychometric Society.
- Satorra, A., & Bentler, P. M. (1994). Corrections to test statistics and standard errors in covariance structure analysis. In A. von Eye & C. C. Clogg (Eds.), *Latent variables*

analysis: Applications for developmental research (pp. 399–419). Thousand Oaks, CA: Sage.

Schmitt, N., & Kuljanin, G. (2008). Measurement invariance: Review of practice and implications. *Human Resource Management Review*, 18(4), 210-222.
<https://doi.org/10.1016/j.hrmr.2008.03.003>

Schmitt, N., & Stults, D. M. (1985). Factors defined by negatively keyed items: The result of careless respondents. *Applied Psychological Measurement*, 9, 367-373.

Stark, S. (2001). MODFIT: A computer program for model-data fit. Unpublished manuscript. University of Illinois at Urbana-Champaign.

Stark, S., Chernyshenko, O. S., & Drasgow, F. (2006). Detecting differential item functioning with confirmatory factor analysis and item response theory: Toward a unified strategy. *Journal of Applied Psychology*, 91(6), 1292-1306.
<http://dx.doi.org/10.1037/0021-9010.91.6.1292>

Stone, M. H. (2010). Sexual sadism: A portrait of evil. *Journal of The American Academy of Psychoanalysis and Dynamic Psychiatry*, 38(1), 133-157.
<https://doi.org/10.1521/jaap.2010.38.1.133>

Tay, L., Ali, U. S., Drasgow, F., & Williams, B. (2011). Fitting irt models to dichotomous and polytomous data: Assessing the relative model–data fit of ideal point and dominance models. *Applied Psychological Measurement*, 35(4), 280-295.
doi:10.1177/0146621610390674

- Tay, L., Drasgow, F., Rounds, J., & Williams, B. A. (2009). Fitting measurement models to vocational interest data: Are dominance models ideal?. *Journal of Applied Psychology*, 94(5), 1287-1304. <http://dx.doi.org/10.1037/a0015899>
- Trémolière, B., & Djeriouat, H. (2016). The sadistic trait predicts minimization of intention and causal responsibility in moral judgment. *Cognition*, 146, 158-171.
<https://doi.org/10.1016/j.cognition.2015.09.014>
- Van de Schoot, R., Lugtig, P., & Hox, J. (2012). A checklist for testing measurement invariance. *European Journal of Developmental Psychology*, 9(4), 486-492.
<http://dx.doi.org/10.1080/17405629.2012.686740>
- Van den Wollenberg, A. L. (1982). Two new test statistics for the Rasch Model. *Psychometrika*, 47(2), 123-140. doi:10.1007/BF02296270
- van Geel, M., Goemans, A., Toprak, F., & Vedder, P. (2017). Which personality traits are related to traditional bullying and cyberbullying? A study with the Big Five, Dark Triad and sadism. *Personality and Individual Differences*, 106, 231-235.
<https://doi.org/10.1016/j.paid.2016.10.063>
- Vandenberg, R. J., & Lance, C. E. (2000). A review and synthesis of the measurement invariance literature: Suggestions, practices, and recommendations for organizational research. *Organizational research methods*, 3(1), 4-70. doi:[10.1177/109442810031002](https://doi.org/10.1177/109442810031002)
- Williams, L. J., Vandenberg, R. J., & Edwards, J. R. (2009). Structural equation modeling in management research: A guide for improved analysis. *The Academy of Management Annals*, 3 (1), 543-604.

- Wright, A. G., Pincus, A. L., & Lenzenweger, M. F. (2012). An empirical examination of distributional assumptions underlying the relationship between personality disorder symptoms and personality traits. *Journal of abnormal psychology, 121*(3), 699-706.
<http://dx.doi.org/10.1037/a0029042>
- Zampetakis, L. A., Lerakis, M., Kafetsios, K., & Moustakis, V. (2015). Using item response theory to investigate the structure of anticipated affect: do self-reports about future affective reactions conform to typical or maximal models?. *Frontiers in psychology, 6*, 1438. doi:[10.3389/fpsyg.2015.01438](https://doi.org/10.3389/fpsyg.2015.01438)
- Zeigler-Hill, V., & Vonk, J. (2015). Dark personality features and emotion dysregulation. *Journal of Social and Clinical Psychology, 34*(8), 692-704.
<https://doi.org/10.1521/jscp.2015.34.8.692>

APPENDIX: Additional tables and figures

Table 9

CAST item correlations (part 1 of 2)

	1	2	3	4	5	6	7	8	9
CAST 1	--								
CAST 2	0.57	--							
CAST 3	0.48	0.66	--						
CAST 4	0.36	0.54	0.68	--					
CAST 5	0.49	0.59	0.55	0.56	--				
CAST 6	0.18	0.2	0.25	0.25	0.23	--			
CAST 7	0.2	0.25	0.26	0.27	0.31	0.05	--		
CAST 8	0.2	0.28	0.29	0.32	0.32	0.05	0.67	--	
CAST 9	0.31	0.36	0.33	0.39	0.44	0.1	0.45	0.58	--
CAST 10	0.29	0.33	0.33	0.32	0.35	0.14	0.38	0.45	0.57
CAST 11	0.3	0.26	0.24	0.21	0.27	0.09	0.2	0.29	0.33
CAST 12	0.2	0.31	0.29	0.24	0.29	0.03	0.23	0.24	0.25
CAST 13	0.22	0.32	0.29	0.27	0.25	0.04	0.26	0.24	0.27
CAST 14	0.19	0.31	0.28	0.26	0.29	0	0.29	0.31	0.23
CAST 15	0.14	0.18	0.23	0.22	0.23	-0.01	0.28	0.38	0.3
CAST 16	0	0.07	0.13	0.06	0.02	0.08	0	0.01	-0.03
CAST 17	0.16	0.26	0.22	0.21	0.23	0.01	0.33	0.29	0.32
CAST 18	0.11	0.24	0.26	0.22	0.22	0.04	0.22	0.2	0.23

Table 9

CAST item correlations (part 2 of 2)

	10	11	12	13	14	15	16	17	18
CAST 1									
CAST 2									
CAST 3									
CAST 4									
CAST 5									
CAST 6									
CAST 7									
CAST 8									
CAST 9									
CAST 10	--								
CAST 11	0.36	--							
CAST 12	0.25	0.29	--						
CAST 13	0.21	0.24	0.58	--					
CAST 14	0.17	0.2	0.55	0.72	--				
CAST 15	0.33	0.25	0.49	0.47	0.51	--			
CAST 16	-0.02	-0.02	0.08	0.18	0.2	0.06	--		
CAST 17	0.28	0.28	0.48	0.37	0.35	0.37	-0.07	--	
CAST 18	0.19	0.15	0.37	0.39	0.32	0.32	0.06	0.4	--

Table 10

Parameter estimates from Model G (part 1 of 3)

Indicator	<u>Sadism (CAST)</u>			ASB	<u>Psychopathy</u>			<u>Reverse</u>
	Verbal	Physical	Vic.		CA	ELS	IMP	
CAST 1	0.61							
CAST 2	0.81							
CAST 3	0.82							
CAST 4	0.74							
CAST 5	0.73							
CAST 6	0.29							0.05
CAST 7		0.68						
CAST 8		0.77						
CAST 9		0.76						
CAST 10		0.67						
CAST 11		0.43						
CAST 12			0.73					
CAST 13			0.82					
CAST 14			0.81					
CAST 15			0.63					
CAST 16			0.17					-0.03
CAST 17			0.53					
CAST 18			0.49					
SRP 5				0.27				0.73
SRP 6				0.23				0.90
SRP 10				0.40				
SRP 12				0.46				
SRP 18				0.17				0.31
SRP 21				0.19				0.13
SRP 29				0.53				
SRP 34				0.30				0.24
SRP 43				0.35				
SRP 46				0.17				0.13
SRP 49				0.56				
SRP 52				0.43				
SRP 57				0.60				
SRP 62				0.49				
SRP 63				0.71				
SRP 64				0.65				
SRP 2					0.30			
SRP 7					0.48			
SRP 11					0.15			0.08

Table 10 continued

Parameter estimates from Model G (part 2 of 3)

Indicator	<u>Sadism</u>		Vicarious	ASB	<u>Psychopathy</u>			<u>Reverse</u>
	Verbal	Physical			CA	ELS	IMP	
SRP 15					0.61			
SRP 19					0.32			0.02
SRP 23					0.11			0.07
SRP 26					0.32			0.07
SRP 30					0.32			
SRP 33					0.24			
SRP 37					0.55			
SRP 40					0.52			
SRP 44					0.40			0.02
SRP 48					0.42			
SRP 53					0.48			
SRP 56					0.37			
SRP 60					0.46			
SRP 1						0.63		
SRP 4						0.45		
SRP 9						0.71		
SRP 14						0.23		0.09
SRP 17						0.53		
SRP 22						0.25		0.03
SRP 25						0.39		0.02
SRP 28						0.71		
SRP 32						0.61		
SRP 36						0.29		0.09
SRP 39						0.32		
SRP 42						0.50		
SRP 47						0.59		0.17
SRP 51						0.39		
SRP 55						0.22		
SRP 59						0.45		
SRP 3							0.44	
SRP 8							0.47	
SRP 13							0.53	
SRP 16							0.33	0.33
SRP 20							0.57	
SRP 24							0.23	0.08
SRP 27							0.62	

Table 10 continued

Parameter estimates from Model G (part 3 of 3)

Indicator	<u>Sadism</u>		Vicarious	ASB	<u>Psychopathy</u>		IMP	<u>Reverse</u>
	Verbal	Physical			CA	ELS		
SRP 31							0.32	0.11
SRP 38							0.31	0.10
SRP 41							0.63	
SRP 45							0.36	
SRP 50							0.22	
SRP 54							0.48	
SRP 58							0.47	
SRP 61							0.39	0.09
Second order factor loadings								
Verbal	0.74							
Physical		0.72						
Vicarious			0.68					
ASB				0.58				
CA					0.97			
ELS						0.63		
IMP							0.89	

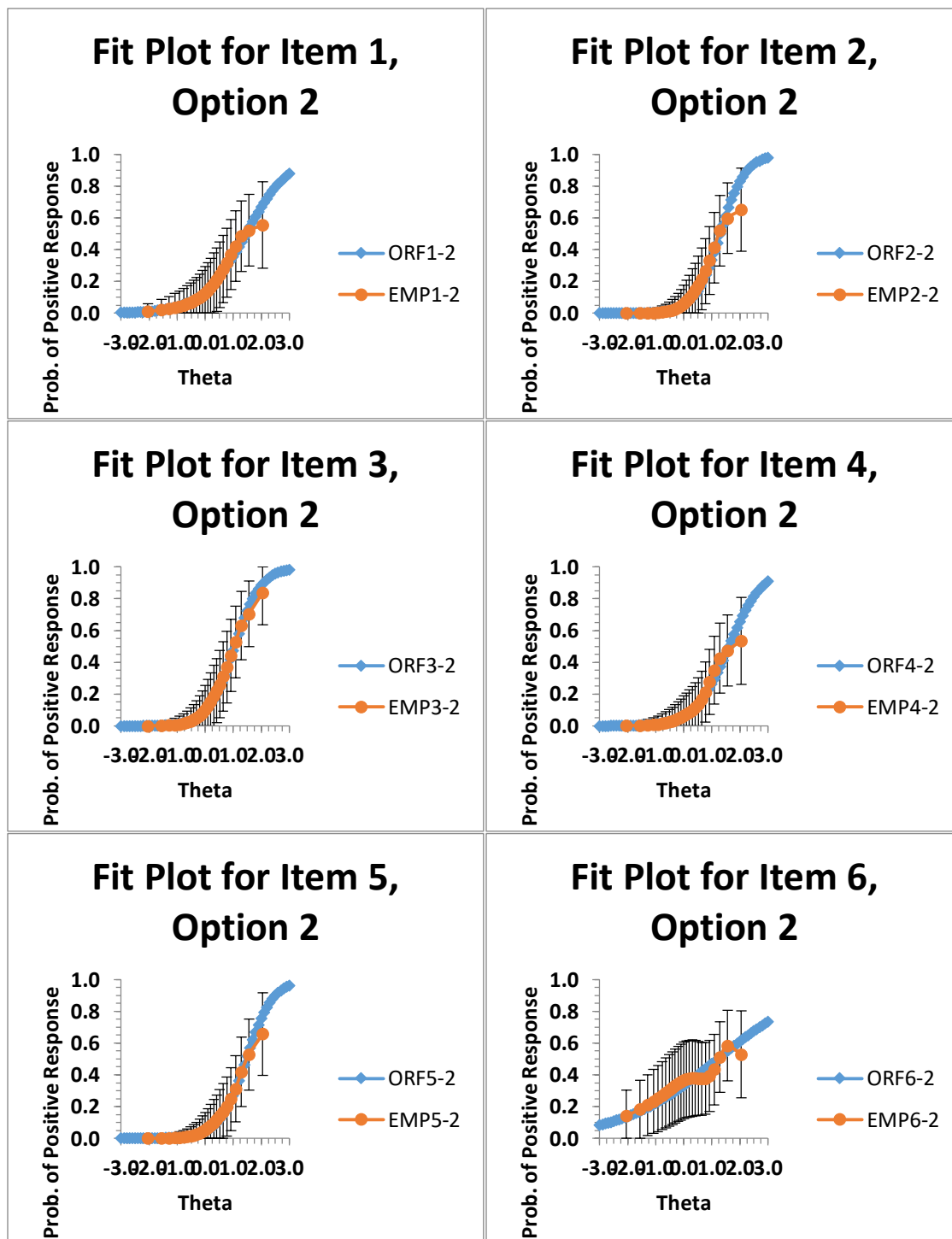
Note. CAST = comprehensive assessment of sadistic tendencies; SRP = self-report psychopathy scale; ASB = antisocial behavior; CA = callous affect; ELS = erratic lifestyle; IMP = interpersonal manipulation; all values are standardized.

Table 11

Observed correlations among dark tetrad traits

	1	2	3	4
1. Sadism (CAST)				
2. Narcissism (NPI-16)	0.16			
3. Machiavellianism (MPS)	0.45	0.13		
4. Psychopathy (SRP-III)	0.65	0.16	0.57	0.65

Note. $N = 432$.



Figures 12. Fit plots for Samejima graded response model. Empirical data based on endorsement.

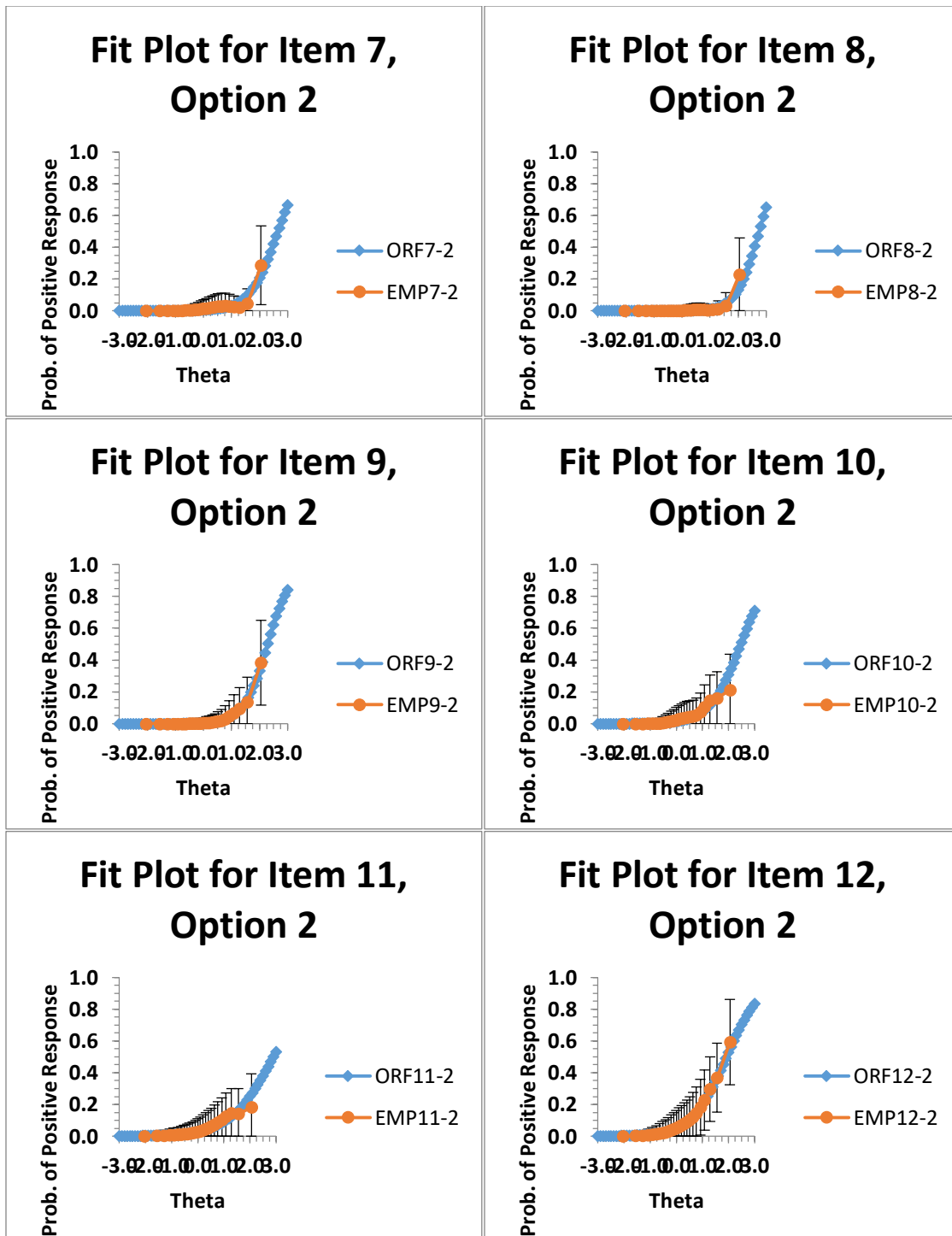


Figure 12 (continued). Fit plots for Samejima graded response model. Empirical data based on endorsement.

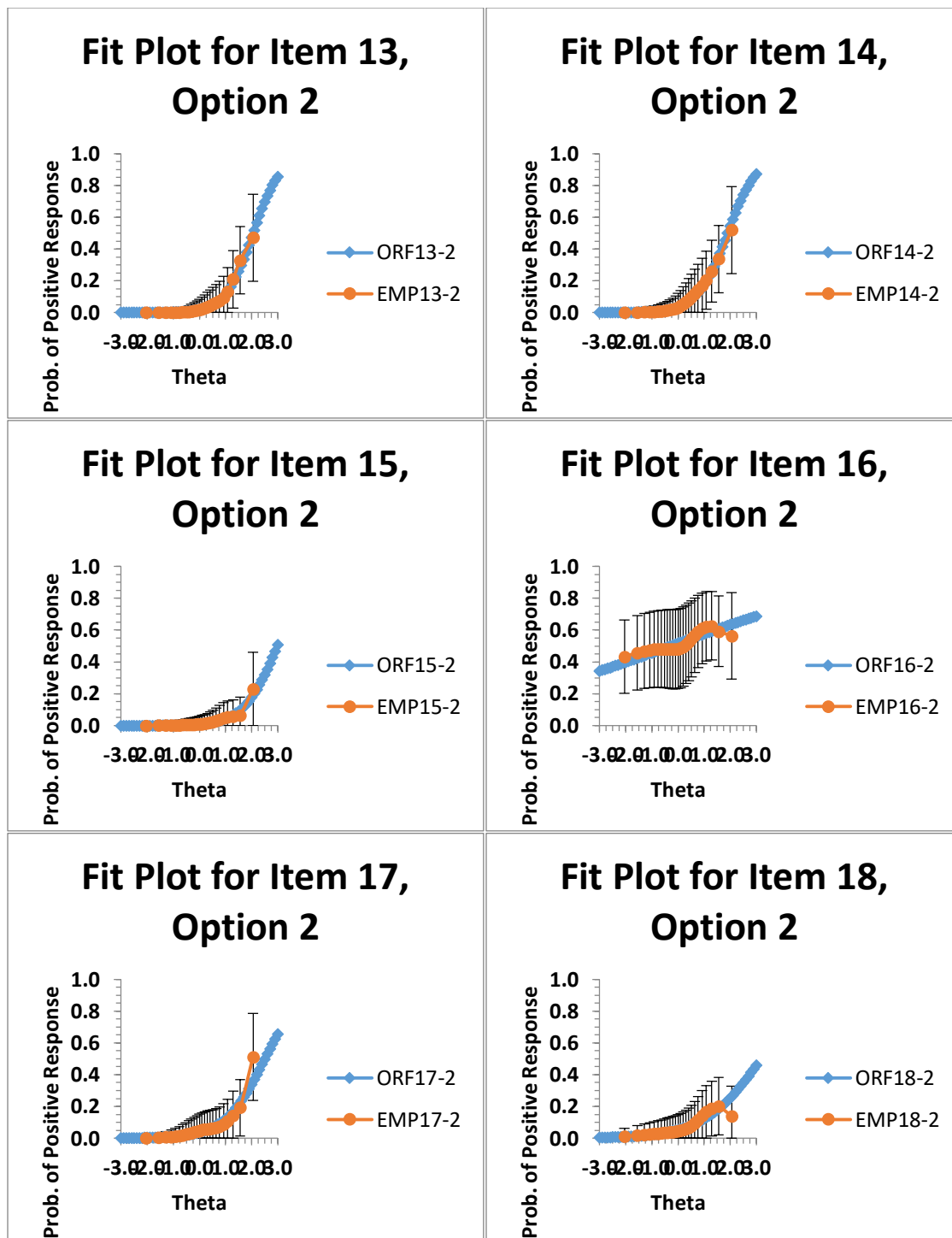
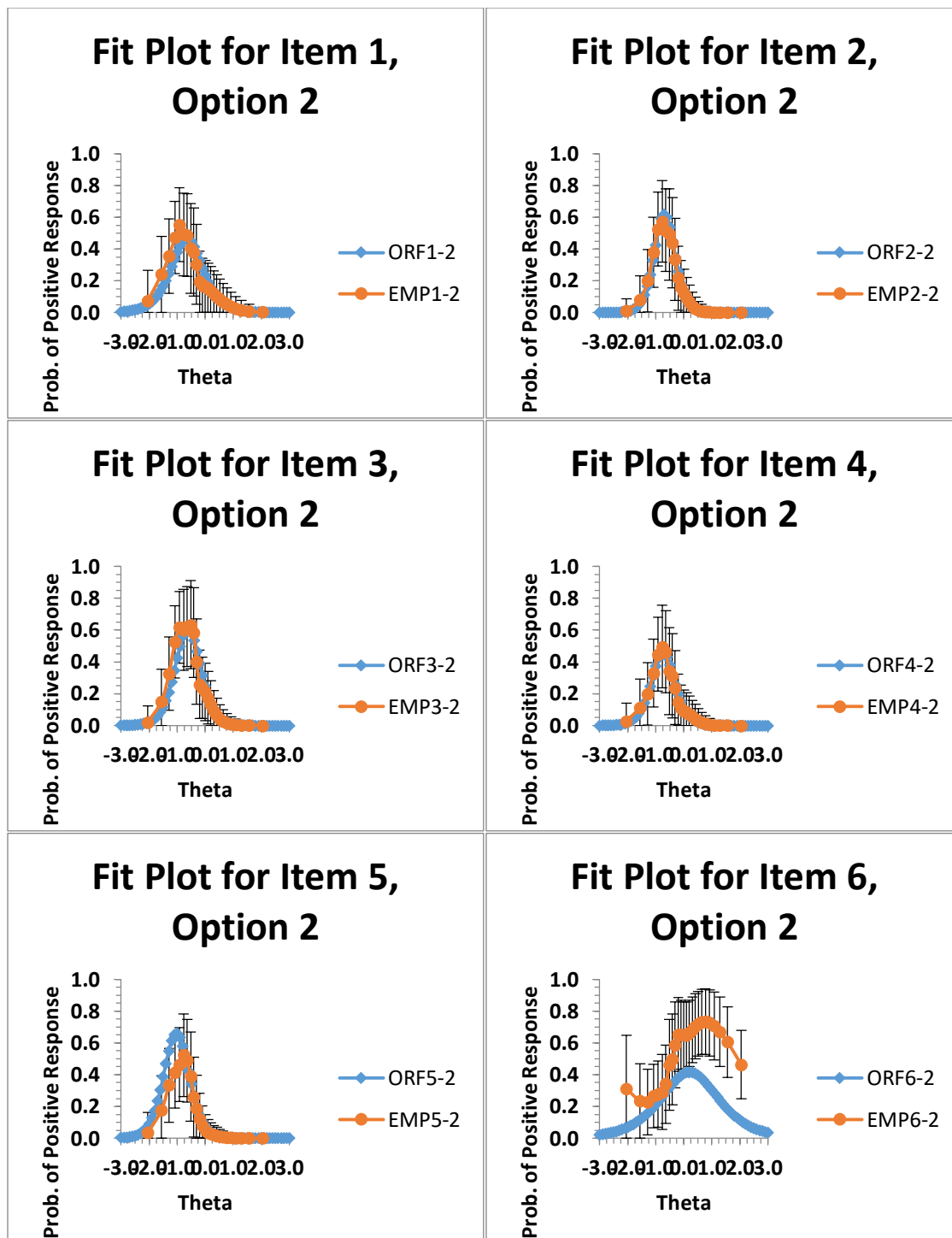
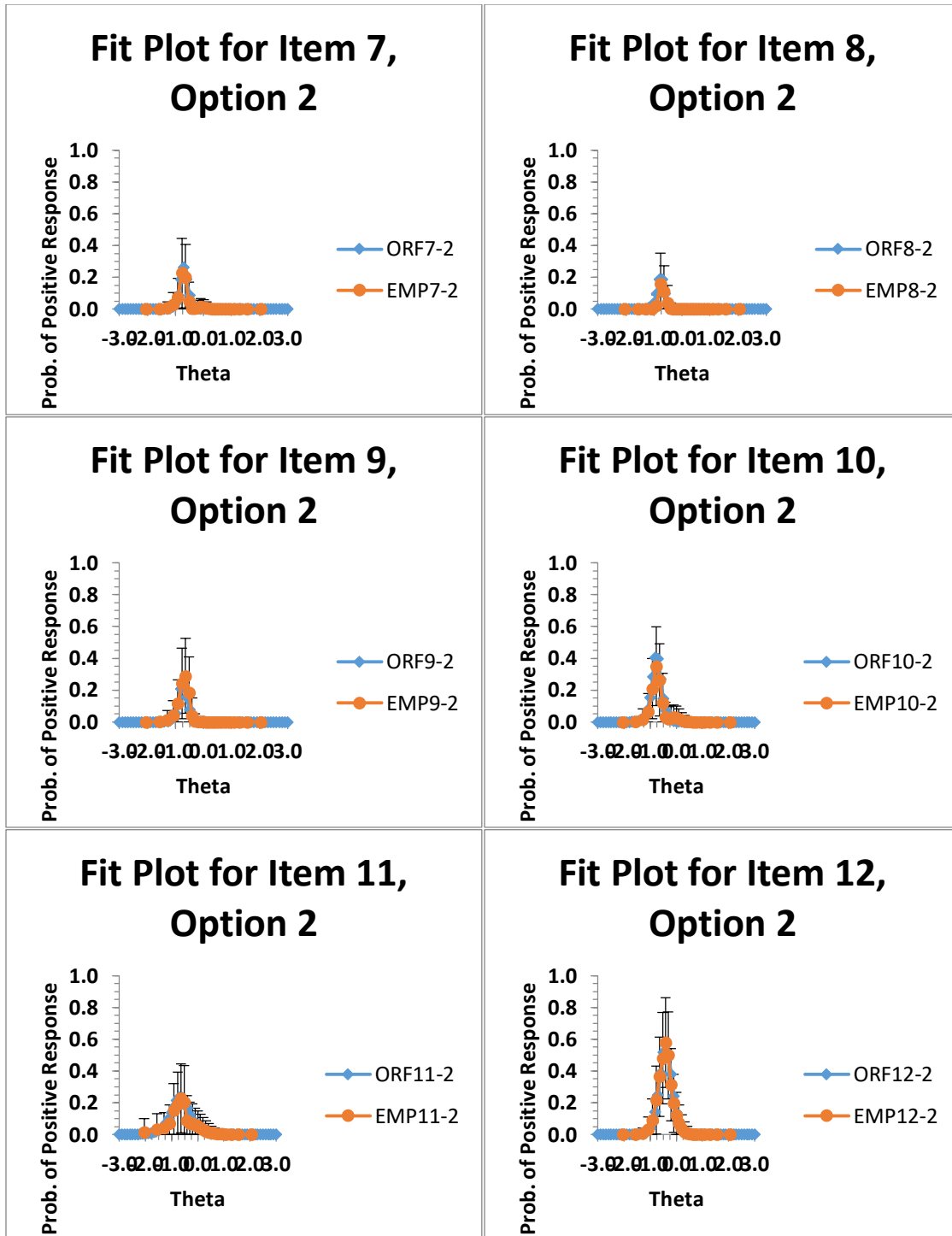


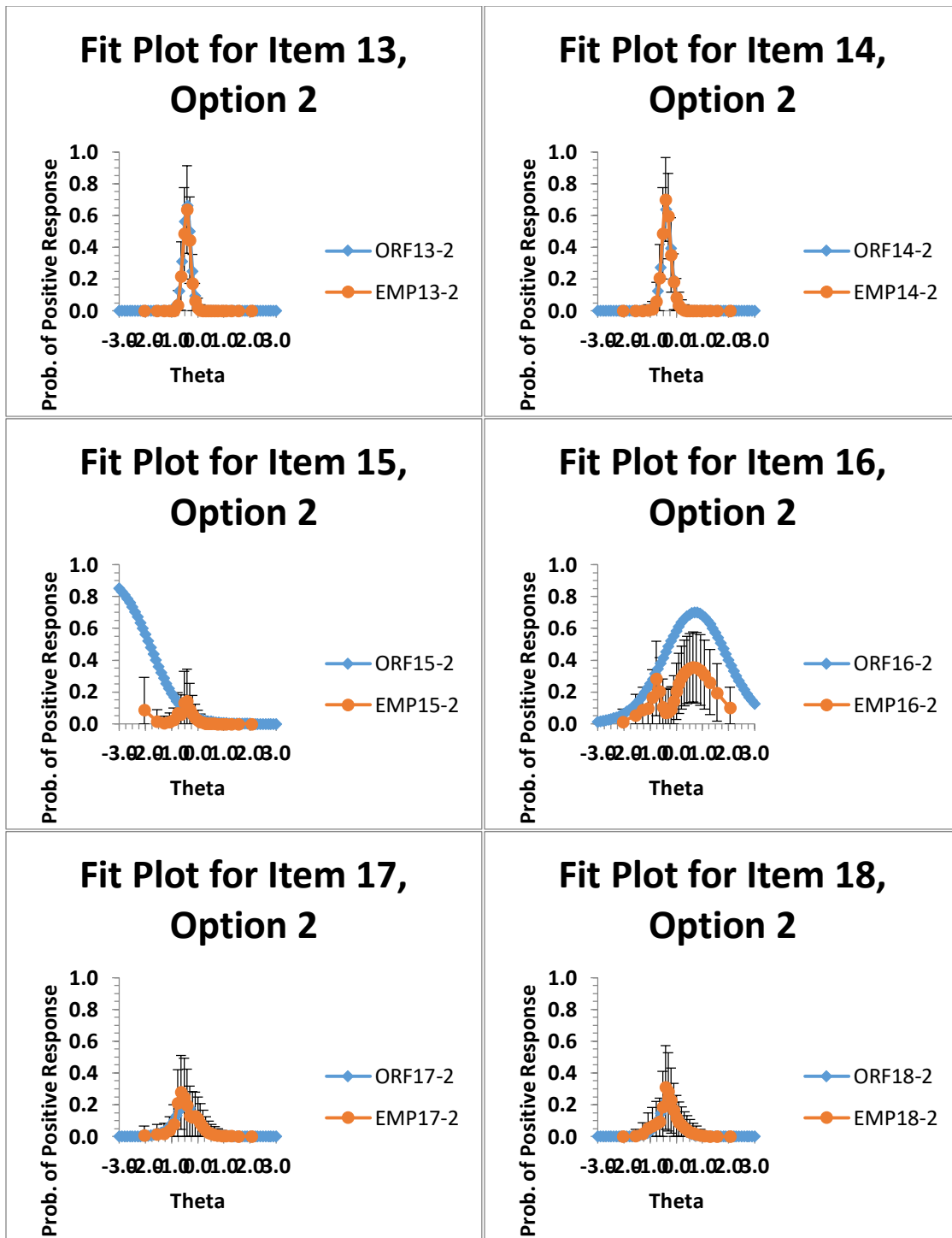
Figure 12 (Continued). Fit plots for Samejima graded response model. Empirical data based on endorsement.



Figures 13. Fit plots for generalized graded unfolding model. Empirical data based on endorsement.



Figures 13 (continued). Fit plots for generalized graded unfolding model. Empirical data based on endorsement.



Figures 13 (continued). Fit plots for generalized graded unfolding model. Empirical data based on endorsement.