Toward the Narrow Approach to Posttraumatic Stress Disorder Diagnostic Criteria: An Item Response Theory Analysis

by

Madison Wyn Silverstein

A dissertation submitted to the Graduate Faculty of
Auburn University
in partial fulfillment of the
requirements for the Degree of
Doctor of Philosophy

Auburn, Alabama
August 3, 2019

Keywords: item response theory, posttraumatic stress disorder, *DSM-5*, *ICD-11*

Approved by

Dr. Frank Weathers (Chair), Professor of Psychology
Dr. Joseph Bardeen, Assistant Professor of Psychology
Dr. Chris Correia, Professor of Psychology
Dr. Kathy Jo Ellison, Associate Professor of Nursing
Dr. Dan Svyantek, Professor of Psychology

Abstract

A crucial debate in the field of traumatic stress involves the question of whether posttraumatic stress disorder (PTSD) is better represented by a broad or a narrow approach to establishing the set of symptom criteria (Stein et al., 2014). The broad approach, exemplified by the *Diagnostic and Statistical Manual of Mental Disorders, 5th Edition* (*DSM-5*; APA, 2013), includes a wide range of trauma-related symptoms, regardless of whether they overlap with other disorders. Conversely, the narrow approach, exemplified by the *International Classification of Diseases, 11th Edition* (*ICD-11*; WHO, expected release 2018) retains only a limited set of what are argued to be core symptoms specific to PTSD (Brewin, 2013; Maercker et al., 2013; Resick & Miller, 2009).. Although the *ICD-11* workgroup narrowed symptom criteria using theory, empirical research, and clinical judgment, it remains empirically unclear whether the retained symptoms are truly the core PTSD symptoms. Item response theory (IRT), a statistical technique that examines each symptom's relative contribution to a construct, is a powerful tool that can inform PTSD symptom selection for the narrow approach. Although IRT studies on PTSD measures exist, no firm conclusions can be drawn about the core symptoms due to the over-restrictiveness of the models employed, variations in measures and populations examined, and the change from *DSM-IV-TR* (APA, 2000) to *DSM-5* (APA, 2013) criteria. To empirically address the question of which items represent the core PTSD symptoms, IRT was employed to examine item difficulty and item discrimination parameters. Undergraduates who experienced a *DSM-5* Criterion A event completed the LEC-5 and PCL-5. Physiological reactivity, internal avoidance, persistent negative emotional state, detachment from others, and concentration and sleep difficulties emerged as the most discriminating symptoms within each *DSM-5* symptom cluster. Importantly, this list only has one symptom in common with *ICD-11* PTSD criteria, suggesting that, in general, the symptoms retained for *ICD-11* are not in fact the most discriminating. Future research should employ IRT in a clinical population.

Acknowledgments

I am grateful for the support of Dr. Frank Weathers and my family.

Table of Contents

List of Tables

List of Figures

List of Abbreviations

PTSD	Posttraumatic stress disorder
DSM-5	*Diagnostic and Statistical Manual of Mental Disorders, 5th Edition*
ICD-11	*International Classification of Diseases, 11th Edition*
IRT	Item response theory
DSM-IV-TR	*Diagnostic and Statistical Manual of Mental Disorders, 4th Edition*
LEC-5	Life Events Checklist – 5
PCL-5	Posttraumatic Stress Disorder Checklist – 5
NACM	Negative alterations in cognition and mood
AAR	Alterations in arousal and reactivity
CTT	Classical test theory
IRF	Item response function
ICC	Item characteristic curve
CRC	Category response curve
Θ	Theta
TIF	Test information function
DIF	Differential item functioning
1PL	One-parameter logistic model
2PL	Two-parameter logistic model
χ^2	Chi-square test statistic

PCL-C	Posttraumatic Stress Disorder Checklist – Civilian
CAPS	Clinician Administered PTSD Scale
NCS-R	National Comorbidity Study-Replication
PCL-M	Posttraumatic Stress Disorder Checklist – Military
FIML	Full information maximum likelihood estimation
CFI	Comparative fit index
TLI	Tucker-Lewis index
RMSEA	Root mean square error of approximation
CFA	Confirmatory factor analysis
MGCFA	Multi-group confirmatory factor analysis
M-H	Mantel-Haenszel test
R	Reexperiencing
AV	Avoidance
ST	Sense of threat
Rex	Reexperiencing
Avoid	Avoidance
SD	Standard deviation
a	IRT discrimination parameter
b	IRT difficulty parameter
Cat	Category

Introduction

Currently, one of the most hotly debated topics in the field of traumatic stress centers around the issue of how the symptom criteria for posttraumatic stress disorder (PTSD) should be conceptualized and arranged. Key points of contention in this debate include questions such as the number of symptoms and clusters and which specific symptoms should be included. Although this is a longstanding debate, it has emerged with increased prominence in recent years because of its centrality to both the substantial revision of the PTSD criteria for *Diagnostic and Statistical Manual of Mental Disorders, 5th Edition* (*DSM-5*; American Psychiatric Association, 2013), and the parallel, even more dramatic reconceptualization of PTSD for the *International Classification of Diseases, 11th Edition* (*ICD-11*; World Health Organization, expected release 2018). Current conceptualizations of the PTSD criteria for posttraumatic stress disorder are increasingly polarized, with *DSM-5* following a broad, inclusive approach and *ICD-11* following a narrow, restrictive approaches (Stein et al., 2014). Whereas the broad approach asserts that all clinically significant symptoms of PTSD should be included in the criteria, the narrow approach asserts that only the core symptoms of PTSD are necessary (Stein et al., 2014). Considering theoretical and empirical evidence, it remains unclear which approach best reflects the essence of PTSD and which is more clinically useful.

Supporters of the broad approach claim that including all clinically significant symptoms better captures the complexity of PTSD (Friedman, Resick, Bryant, & Brewin, 2011; Resick & Miller, 2009). As responses to trauma are diverse, maintaining PTSD's heterogeneous symptom criteria ensures that a variety of related symptom profiles can be captured, thus increasing diagnostic coverage of the PTSD diagnosis (Friedman, 2013; Friedman et al., 2011). Further, broad approach supporters believe that including symptoms that overlap with other disorders is necessary because they also characterize PTSD (Friedman, 2013). Friedman (2013) points out that, in medical diagnostics, symptoms such as fever or pain are not excluded from the symptom criteria just because they are symptoms of multiple diseases. By applying this logic to the PTSD diagnosis, symptoms that also appear in disorders such as depression or anxiety should be retained because they represent important treatment targets and provide important information about the problems an individual is experiencing (Friedman, 2013).

The broad approach is exemplified by the *DSM-5* criteria for PTSD because it includes a high number of diverse symptoms categorized into multiple domains. In fact, the criteria for PTSD were modified substantially in the newest edition to further reflect the broad approach. Specifically, the *DSM-5* workgroup split the numbing and avoidance symptom clusters into avoidance and negative alterations in cognition and mood (NACM), reworded many of the existing symptoms, and added three additional symptoms to the criteria (Weathers et al., 2014).

Although the modifications to the *DSM-5* PTSD criteria were based on extensive empirical research (Friedman, 2013), critics argue that a disorder with 20 symptom criteria is too heterogeneous and that including non-specific symptoms leads to excessive comorbidity (Brewin, Lanius, Novac, Schnyder, & Galea, 2009; Galatzer-Levy & Bryant, 2013; Spitzer, First, & Wakefield, 2007; Young, Lareau, & Pierre, 2014). Specifically, critics of the broad approach claim that the polythetic nature of the PTSD diagnostic category and the high number of symptoms has led to too much heterogeneity in the PTSD diagnosis (Olbert, Gala, & Tupler, 2014; Young et al., 2014). In fact, with the addition of three symptoms, there are now 636,120 symptom combinations that meet PTSD diagnostic criteria (Galatzer-Levy & Bryant, 2013). This extraordinary number of symptom combinations leads to a high prevalence of disjoint pairs, meaning that two individuals can meet criteria for PTSD but have no overlapping symptoms

(Hickling, Barnett, & Gibbons, 2013; Olbert et al., 2014). Latent class analyses also indicate that there are up to four different categories of individuals diagnosed with PTSD (see Olbert et al., 2014 for a review). This heterogeneity within the diagnostic category of PTSD is clinically problematic because a diagnosis does not provide much information about symptoms the individual is experiencing (Olbert et al., 2014).

Heterogeneity within the PTSD diagnosis also has implications for research. For example, treatment effect studies often have small to large effect sizes, indicating that a treatment is more effective in some studies than others. The variation in effect sizes might be due to the heterogeneity of the symptom profiles within each sample because some symptom profiles are more responsive to treatment than others (Galatzer-Levy & Bryant, 2013). Therefore, determining which treatments are most effective for individuals diagnosed with PTSD will be difficult because so many different types of symptom profiles are subsumed under the PTSD category. Further, different symptom profiles are associated with differences across a variety of factors (e.g., genetic, physiological, clinical, social factors, personality, comorbidity; Nandi, Beard, & Galea, 2009; Olbert et al., 2014). Using a common diagnosis of PTSD despite many of these potentially important differential associations overshadows real differences within the individuals who meet criteria for PTSD. This problem becomes even more important as the National Institute of Mental Health moves towards the Research Domain Criteria to examine biological markers of disorders (Cuthbert, 2014). As Olbert and colleagues (2014) noted, the chances of finding biological markers in a heterogeneous population are slim “given the possibility that symptom heterogeneity may reflect etiological heterogeneity” (Olbert et al., 2014; p. 459).

Increased comorbidity is a distinct but closely related issue to increased heterogeneity in the PTSD diagnosis. Astonishingly, there are over one quintillion ways a PTSD diagnosis in *DSM-5* can be comorbid (Young et al., 2014). This comorbidity is mainly an effect of the many PTSD symptoms that overlap with other disorders (Brewin et al., 2009; Spitzer et al., 2007). Researchers argue that including these overlapping symptoms weakens the causal association between the trauma and the symptoms (Brewin et al., 2009; Zoellner, Bedard-Gilligan, Jun, Marks, & Garcia, 2013), because many of these symptoms are commonly seen in individuals seeking outpatient treatment for depression who are not necessarily trauma-exposed (Bodkin, Pope, Detke, & Hudson, 2007; Spitzer et al., 2007). In one study that utilized *DSM-IV-TR* (American Psychiatric Association, 2000) criteria B-F, the rate of PTSD diagnosis was equivalent in trauma-exposed versus non-trauma-exposed individuals (Bodkin et al., 2007). This equivalency indicates that there are a variety of symptoms included in the PTSD diagnosis that might not be specific to PTSD, but are rather general symptoms of distress.

According to critics, these general symptoms of distress, or “dysphoria” symptoms, represent the nonspecific symptoms that lead to comorbidity (Elhai, Biehn, et al., 2011; Liu et al., 2014; Simms, Watson, & Doebbelling, 2002). To clarify, the dysphoria symptoms include traumatic amnesia, negative beliefs, distorted blame, persistent negative emotional state, diminished interest in activities, detachment from others, inability to experience positive emotions, irritability or anger, reckless or self-destructive behavior, sleep difficulties, and difficulty concentrating. These symptoms are extremely multi-faceted such that they include cognitive (e.g., distorted blame), emotional (e.g., inability to experience positive emotions, irritability or anger), and miscellaneous symptoms (e.g., traumatic amnesia). Additionally, many symptoms overlap with mood or anxiety disorders. In fact, many of these symptoms are strongly associated with depression and anxiety (Simms et al., 2002; Witte, Domino, & Weathers, 2015) more so than other symptoms (Simms et al., 2002). Due to the strong association between dysphoria symptoms and anxiety and depression, some researchers called for the removal of the "numbing" symptoms from the *DSM-IV-TR* avoidance/numbing cluster (Criterion C) instead of moving them to NACM cluster in *DSM-5* (Brewin et al., 2009; Spitzer et al., 2007). Interestingly, in one study, removing symptoms that overlapped with mood or anxiety disorders did not substantially affect prevalence, "psychiatric comorbidity, functional impairment, or structural validity of PTSD" (Grubaugh, Long, Elhai, Frueh, & Magruder, 2010; p. 909), indicating that these symptoms might not be essential.

However, other researchers consider this cluster to be central to the PTSD diagnosis because many individuals who experience a traumatic event experience these symptoms specifically in relation to the traumatic event, along with other symptoms that are more distinct to PTSD (Friedman, Resick, Bryant, & Brewin, 2011). Although theory and empirical results remain at odds when considering dysphoria symptoms, if studies continue to show that removing these symptoms does not substantially impact prevalence, comorbidity, functional impairment, or the factor structure of PTSD, using a less complex model might be worthwhile to shorten diagnostic criteria (Grubaugh et al., 2010).

The narrow approach addresses many of the critiques of the broad approach. This approach asserts that only the core symptoms of PTSD should be included in the diagnostic criteria to distinguish PTSD from disorders that are often comorbid (Stein et al., 2014). Supporters of the narrow approach claim that by excluding non-essential symptoms, symptom profile heterogeneity will decrease, leading to fewer cases of disjoint pairs. Supporters also claim that by excluding overlapping symptoms, comorbidity will decrease. Therefore, the likelihood that symptoms are associated with trauma exposure will increase (Brewin, 2013; Maercker et al., 2013).

The narrow approach is exemplified by the *ICD-11* PTSD criteria. *ICD-11* addresses the pitfalls of *DSM-5*'s broad approach by deleting symptoms that overlap with other disorders so that there are only three clusters, including reexperiencing, avoidance, and sense of threat, with two symptoms in each category: flashbacks and nightmares; internal and external avoidance; hypervigilance and exaggerated startle response (Maercker et al., 2013). According to Reed (2010), the goal of *ICD-11* was to enhance clinical utility by logically organizing disorders based on treatment implications and reducing the number of symptoms so that clinicians would be better able to make diagnoses in the field (Reed, 2010).

Unlike the *DSM-5* workgroup, the "*ICD-11* workgroup was under no obligation to use the *DSM-IV-TR* or even the ICD-10 as a starting point" for the *ICD-11* PTSD symptom criteria, "but was charged to use their clinical and research knowledge to optimize the diagnosis in the service of clinical utility" (Brewin, 2013; p. 557). Therefore, the *ICD-11* workgroup drew from existing theoretical, clinical, and empirical sources to determine the core symptoms of PTSD. It appears as if the *ICD-11* workgroup chose to include flashbacks and nightmares based on a body of literature summarized by Brewin and colleagues (2009) that indicate that these symptoms are markers of PTSD, as they readily distinguish PTSD from non-pathological reactions to trauma exposure (Brewin, 2007) and PTSD from depression (Reynolds & Brewin, 1998) and other anxiety disorders (Sheikh, Woodward, & Leskin, 2003). Nightmares are also highly common amongst individuals with PTSD (Harvey, Jones, & Schmidt, 2003), and treating nightmares using image rehearsal therapy not only decreases nightmare intensity and frequency but also reduces overall PTSD severity (Lamarche & De Koninck, 2007). Additionally, when compared to other anxiety disorders such as panic disorder, nightmares appear to be specific to PTSD (Sheikh et al., 2003). Flashbacks are even more sensitive and specific indicators of trauma than nightmares (Duke, Allen, Rozee, & Bommaritto, 2008), and certain qualities of flashbacks are predictive of long-term PTSD severity (Michael, Ehlers, Halligan, & Clark, 2005). Further, flashbacks represent a clinically important and central feature of PTSD (Brewin et al., 2009).

Support for deleting the other intrusion symptoms and the dysphoria symptoms comes from critiques that PTSD should only include symptoms unique to the PTSD syndrome and not symptoms included in other disorders such as depression and anxiety (see Brewin et al., 2009 and Spitzer et al., 2007 for a review of these critques). Empirical support for the removal of the dysphoria symptoms comes from structural equation modeling studies that indicate that these symptoms are more predictive of anxiety and depression than the other symptoms of PTSD (Gootzeit & Markon, 2011); more highly correlated with depression and anxiety (Elklit, Hyland, & Shevlin, 2014; Simms et al., 2002); and weakly correlated with the other symptom clusters (Elklit et al., 2014). These results indicate that the dysphoria symptoms are not specific to PTSD and span PTSD, anxiety, and depression (Gootzeit & Markon, 2011) whereas reexperiencing, for example, is more specific to PTSD (Elklit et al., 2014; Gootzeit & Markon, 2011; Simms et al., 2002).

Research on the factor structure of the proposed *ICD-11* symptoms indicates good (Tay, Rees, Chen, Kareth, & Silove, 2015) to excellent fit and invariance across gender (Hansen, Hyland, Armour, Shevlin, & Elklit, 2015). Additionally, latent class analyses provide support that PTSD represents a separate class compared to complex PTSD and borderline personality disorder (Cloitre, Garvert, Brewin, Bryant, & Maercker, 2013; Cloitre, Garvert, Weiss, Carlson, & Bryant, 2014; Elklit et al., 2014).

Despite the promising factor analytic and latent class analytic results, some findings indicate that the *ICD-11* symptoms are not performing as expected. First, prevalence in *ICD-11* should be lower than in versions of the *DSM* due to the more stringent diagnostic criteria. However, findings are mixed, and research indicates that *ICD-11* produces similar (Hansen et al., 2015; Stein et al., 2014) or lower (Hansen et al., 2015; Hyland et al., 2016; O'Donnell et al., 2014; Stammel, Abbing, Heeke, & Knaevelsrud, 2015) diagnostic rates than *DSM-IV-TR* or *DSM-5* (Hansen et al., 2015). Second, the *ICD-11* diagnoses should be completely nested within *DSM-IV-TR* or *DSM-5* diagnoses. In other words, all individuals who met PTSD criteria in *DSM-IV-TR or DSM-5* should meet criteria for PTSD in *ICD-11*. However, there are some cases where individuals met criteria in the *DSM-IV-TR* or *DSM-5* but not in *ICD-11* (Green et al., 2017; Morina et al., 2015; Stammel et al., 2015; Stein et al., 2014).

Second, the association between PTSD severity and depression, anxiety, and aggression should be lower in *ICD-11* than in the *DSM-5* PTSD criteria because depression, anxiety, and aggression were explicitly deleted on *ICD-11* to reduce comorbidity. However, there is some evidence that the *ICD-11* and *DSM-5* total PTSD severity have comparable levels of association with depression, anxiety, and aggression (Hansen et al., 2015). Third, there should be lower levels of comorbidity in *ICD-11* due to the deletion of the non-specific symptoms. However, research is mixed, and studies have shown lower levels of comorbidity with major depressive disorder but not anxiety disorders (Stammel et al., 2015) or no differences in comorbidity for depression, generalized anxiety disorder, panic disorder, or alcohol use disorder (Green et al., 2017). Fourth, there should be lower levels of disjoint cases in *ICD-11* versus *DSM*. Unfortunately, no known research has examined the level of disjoint cases in *ICD-11*. Sixth, although the six *ICD-11* symptoms were apparently selected due to the ease at which clinicians can recall and interpret them, incorrect decisions about the flashback and nightmares symptoms were made more frequently than any of the other symptoms in a vignette study (Keeley et al., 2016).

Seventh, although the *ICD-11* workgroup has some evidence for including flashbacks and nightmares and removing the dysphoria symptoms, justification for including the avoidance symptoms and the sense of threat symptoms over other symptoms is lacking. There is also little evidence for not including the other three intrusion and four alterations in arousal and reactivity (AAR) symptoms. These conflicting results and the lack of evidence for including or excluding well-established PTSD symptoms are troubling, as many researchers have stated that strong empirical evidence is necessary to make determinations about the core symptoms of PTSD (Kilpatrick et al., 2013; Spitzer et al., 2007; Young et al., 2014).

Last, there is some evidence that the selected symptoms in *ICD-11* are not necessarily the most representative of PTSD. For example, there is mixed evidence about the association between trauma exposure and PTSD symptom clusters. Whereas reexperiencing was most strongly associated with trauma exposure in one study (Simms et al., 2002), dysphoria symptoms were most strongly associated with trauma exposure in another study, indicating that these symptoms are clinically important (Gootzeit & Markon, 2011). Additionally, in a study that examined the quality of specificity of the *DSM-5* PTSD symptoms for diagnostic purposes, the only *ICD-11* symptom with good specificity was flashbacks (Green et al., 2017). Further, network analyses should indicate that the *ICD-11* symptoms are the most central to the PTSD network. However, although some studies do indicate that *ICD-11* symptoms are central to the PTSD network (Armour et al., 2017; Bryant et al., 2017; McNally et al., 2015; Mitchell et al., 2017), studies also indicate that symptoms not included in the *ICD-11* criteria are central, such as

NACM symptoms (Armour et al., 2017; Mitchell et al., 2017), intrusions (Bryant et al., 2017; McNally et al., 2015), and physiological cue reactivity (Bryant et al., 2017).

Despite these limitations, the supporters of the narrow approach raise many important critiques of the broad approach. It is possible that the limitations listed above result from the *ICD-11* workgroup's symptom selection rather than a flaw in the narrow conceptualization. To better empirically address the question of which symptoms represent the core features of PTSD, item response theory (IRT) is a promising avenue that can provide information about each symptom's relative contribution to the PTSD construct. To provide a more concrete rationale for using IRT to examine the core symptoms of PTSD, I will first provide an overview of IRT, including IRT's similarities and differences with classical test theory (CTT), the advantages of using IRT over CTT, and IRT's statistical parameters and models. I will then describe the PTSD literature that utilized IRT and its limitations. Finally, I will describe the aims of the current study and explain how the current study will address limitations of previous studies.

IRT Overview

Item response theory (IRT), a statistical technique that examines the precision of items as a function of a latent construct, first appeared in the 1940's as a method for test development and refinement, scale administration, and examining individual differences (Reise & Waller, 2009). However, it was not until Lord & Novick's (1968) book that IRT became widely known (Embretson, 1996; Lord, Novick, & Birnbaum, 1968; Reise & Waller, 2009). Since then, IRT has been used by educators to develop academic tests, by state governments for licensing exams, and, more recently, by psychologists for clinical assessment (Reise & Waller, 2009) in areas such as depression (Olino et al., 2012), attention deficit hyperactivity disorder (Li, Reise, Chronis-Tuscano, Mikami, & Lee, 2016), quality of life (Sijtsma, Emons, Bouwmeester, Nyklíček, & Roorda, 2008), and PTSD (King, King, Fairbank, Schlenger, & et al, 1993).

IRT versus Classical Test Theory

Although IRT is a respected way to psychometrically evaluate measures of psychopathology, classical test theory (CTT) is the dominant framework used for test development in the field of psychology (Embretson, 1996). CTT was pioneered by Spearman (1907) in the early 1900's and rests on the assumption that observed score plus error equals true score (Embretson, 1996; Spearman, 1907). There are a variety of similarities and differences between CTT and IRT that impact the decision-making process about which framework to use (Zickar & Broadfoot, 2009). First, CTT and IRT both have estimates of an individual's construct level. CTT estimates true score, which is the observed score on a test minus the error score where the error score and true score are unknown values (Zickar & Broadfoot, 2009). IRT estimates ability level, which is a latent measure of the construct of interest. Although, the total score in CTT and ability level in IRT are often highly correlated, this correlation decreases when there are fewer items on the scale or at the upper or lower end of the ability range (Reise & Waller, 2009; Zickar & Broadfoot, 2009) indicating that these scores are not necessarily interchangeable or measuring in the same way.

Second, whereas IRT examines the data at the item level, CTT examines the data at the test level (Hambleton & Jones, 1993). In the case of construct level estimation, ability level is linked to item difficulty in IRT (Hambleton & Jones, 1993; Zickar & Broadfoot, 2010), but true score is not linked to the items in CTT. For example, in IRT, if ability level is known, it is possible to determine the likelihood of each response option to any particular item on the test, indicating that ability level provides valuable information on its own. However, for a score to be meaningful in CTT, it must be compared to a norm-referenced group, in the form of a percentile rank, for example (Embretson, 1996). Of note, it is also possible to have norm-referenced IRT scores by linearly transforming them to standard scores (Embretson, 1996). Further, both CTT and IRT have methods for estimating standard error of measurement, but CTT provides an estimate at the test level whereas IRT provides an estimate at either the item or test level. In CTT, the standard error of measurement follows a normal distribution and is applicable to all scores. However, standard error varies across respondents in IRT meaning that the precision of items varies depending on the respondent's ability level (Embretson, 1996; Reeve, 2003).

Third, dependence of the test statistics varies for IRT versus CTT. True score in CTT is test and item dependent (Hambleton & Jones, 1993; King et al., 1993), indicating that true score will change depending on the test or the items administered (Reise & Waller, 2009). However, in some models of IRT, ability level is test and item independent, indicating that ability level will not vary across tests as long as the tests measure the same construct (Embretson, 1996; Zickar & Broadfoot, 2010). Dependence also comes into play when considering the sample. Test statistics (e.g., standard error of measurement, difficulty) are sample dependent in CTT, whereas these test statistics are sample independent within a linear transformation in some IRT models (Hambleton & Jones, 1993; Reeve, 2003; Zickar & Broadfoot, 2010). Therefore, these test statistics will change every time the test is administered to a new sample in CTT, but IRT item properties remain relatively constant in some types of IRT models. In sum, the representativeness of the sample does not always impact item properties in IRT, but it does in CTT (Embretson, 1996).

Last, IRT ability level is on an interval scale whereas CTT total score is on an ordinal scale (Embretson, 1996; Reeve, 2003; Reise et al., 2005). Therefore, unlike IRT, in CTT, "equal changes on the latent trait produce unequal changes on the raw-score scale, depending on where on the latent-trait scale the changes occur. Thus, two individuals may change the same amount on the latent trait measured by an instrument, but the changes in their raw scores will not reflect this" (Reise et al., 2005; p. 99). Therefore, using total scores to examine change over time in clinical constructs is not ideal because the true change on the trait level might not be reflected in total score (Reise et al., 2005). However, Zickar and Broadfoot (2009) note that using IRT for scoring purposes only is not recommended and that CTT should be used instead.

IRT Assumptions and Components

The most commonly used IRT models have a few main assumptions that rest on the existence of a latent construct, including local independence, monotonicity, and unidimensionality (Reise & Haviland, 2005). The latent construct, or ability level, can represent a disorder or personality trait, for example. The latent construct is assumed to be responsible for the variance in individuals' responses across items (Reise & Waller, 2009). Therefore, when the latent construct's variance is removed, test items should be uncorrelated (i.e., local independence; Zickar & Broadfoot, 2010). Additionally as the level of the latent construct increases, so will the likelihood of endorsing certain items in certain ways (i.e., monotonicity; King, King, Fairbank, Schlenger, & et al, 1993; Reise & Haviland, 2005). In most cases of IRT, the latent construct is assumed to be unidimensional (Reise & Haviland, 2005). However, it is possible, yet uncommon in psychopathology measurement, to use IRT on multidimensional constructs (Raykov, 2016). In fact, Zickar and Broadfoot (2009) claim that designing unidimensional tests is "futile" in psychological research, and they recommend conceptualizing dimensionality as continuous rather than binary (i.e., unidimensional versus multidimensional). In fact, these authors cite Monte Carlo simulations that indicate that low levels of dimensionality can return accurate results (e.g., Reckase, 1979). Therefore, in some cases, using a unidimensional IRT model on a construct that has some dimensionality is permissible.

In addition to the main assumptions of IRT, there are three important measurement components of IRT: item response functions (IRFs), information functions, and scale or test information functions (TIFs). The first, and most basic, component of IRT is the IRF (Reise, Ainsworth, & Haviland, 2005), also known as the item characteristic curve (ICC; Reeve, 2003), or the category response curve (CRC; Reeve, 2003). Although both the ICC and CRC are IRFs, the ICC corresponds to a measure with dichotomous response options, and the CRC responds to a measure with more than two response options (i.e., polytomous; Reeve, 2003; Reise et al., 2005). Therefore, an ICC will have two curves per graph, and the CRC will have more than two curves per graph. The IRF denotes the non-linear association between an individual's level of a latent construct and their likelihood of responding in a particular way to an item on a test that measures that latent construct (Reeve, 2003; Reise et al., 2005). The *x*-axis is a measure of the latent construct (θ), and the *y*-axis is a measure of the probability of a response from each of the response categories (e.g., yes/no for the ICC; never, sometimes, always for the CRC; Reise et al., 2005). Because IRT generates an IRF for each item, each item on a measure can be examined independently (Reise & Haviland, 2005).

Additionally, each IRF tells about the difficulty of an item, or the level of the latent construct where an individual has a 50% chance of endorsing a particular response (Reise & Haviland, 2005). Item difficulty can be discerned by examining the *x*-axis location of the inflection point on the curve (Reise et al., 2005). Visually, when the inflection point is further to the right on the graph, the item has a higher level of difficulty. Considering psychological constructs, the more difficult an item the individual endorses, the more of the latent construct the individual possesses. The IRF also provides the item's discrimination, or the item's ability to distinguish between individuals who have different levels of the latent construct. Item discrimination involves examining the slope of the line at the inflection point. The steeper the slope of the line at the inflection point, the more discriminating an item (Reise et al., 2005).

The second important component of IRT is the information function, which is a transformation of an item's IRF. This function reflects the precision of a measure across varying levels of the latent construct (Reeve, 2003). Specifically, it provides information about an item's contribution to the ability score (Hambleton & Jones, 1993) or at which range of the construct an item provides the most distinguishing information (Reeve, 2003). This function of IRT can inform test development and refinement because, an item provides more or less information about the latent construct across different ability level ranges (Reise et al., 2005). As such, some items will be more useful to have on a measure than others, particularly for screening tools. Similar to the IRF, the *x*-axis for the information function represents the latent construct (θ). However, the *y*-axis of the information function represents the information magnitude. Higher peaks on the information function indicate more item discrimination, and thus more relative information about the latent construct. Further, multiple items' information functions can be placed on one graph to compare the relative contribution of each item (Reeve, 2003). Considering the information function and IRF together, items with low difficulty generally discriminate individuals with low levels of the latent construct whereas items with high difficulty generally discriminate individuals with high levels of the latent construct (Reise et al., 2005).

The third important component of IRT is the TIF, which is an additive graph of the item information functions across items (Reeve, 2003; Reise et al., 2005). This function depicts the measure's functioning across levels of the latent construct (Reise et al., 2005) as well as where the test provides the most precise estimates of ability level (Hambleton & Jones, 1993). Specifically, the TIF provides information about the information magnitude (Zickar & Broadfoot, 2009), reliability, and standard error for specific levels of a construct across the entire scale (Reeve, 2003). It is possible to calculate the standard error of measurement across varying levels of the construct using the TIF by calculating the inverse of the square-root of the information function at specific levels of the latent construct (Reeve, 2003). Standard errors will be smallest where the test provides the most information about level of the latent construct (Hambleton & Jones, 1993). The TIF can also be used to make predictions about test performance based on a particular level of the latent construct (Reise et al., 2005). Like the IRF and the information function, the *x*-axis for the TIF represents the latent trait (θ). The *y*-axis of the TIF represents the magnitude of information provided by the test with the associated reliability value (Reeve, 2003).

Another defining aspect of IRT is item-parameter invariance and the ability to determine if items function differently across groups (i.e., differential item functioning [DIF]). Item-parameter invariance makes it possible to combine tests that measure the same latent construct onto a common scale (Reise & Waller, 2009) and to estimate an individual's ability level based on responses to items with known IRFs, even if the items are selected from a variety of measures (Reise et al., 2005). When an item displays DIF, the item's IRF differs across specified groups (Reise & Waller, 2009). Specifically, individuals across groups have different probabilities of responding to a particular response option, even though they have the same level of the latent construct (He, Glas, & Veldkamp, 2014). When evaluating scales, it is important to examine the potential for DIF across groups based on demographic, diagnostic, or linguistic variables because including items that display DIF might make comparisons across these groups less meaningful or inaccurate (Reise & Waller, 2009).

Model Selection in IRT

Researchers have implemented a variety of different types of models when using IRT. The first set of models are unidimensional and for binary responses. These models build on each other such that they measure increasing numbers of parameters. The simplest type of model is the one-parameter logistic (1PL) model or the Rasch model, which measures the item difficulty parameter (Embretson & Reise, 2000) and sets the discrimination parameters to equality (Raykov, 2016). Because of the restrictive nature of this model, researchers often exclude items or individual responses that do not conform to the model (e.g., Betemps & Baker, 2004). This technique has caused debate in the field. Whereas some researchers claim that the Rasch model is superior to other models because of its simplicity, Harvey (2016) claims that Rasch modeling does not fully account for the association between the items and the latent construct. Therefore, he states that it is an empirical rather than theoretical question to use the Rasch model and recommends testing a variety of models for fit first rather than defaulting to the Rasch model. Importantly, he notes that when the Rasch model appropriately fits the data it should be used (Harvey, 2016).

The two-parameter logistic (2PL) model and the three-parameter logistic (3PL) model are also unidimensional binary models. These models both estimate item difficulty and discrimination. However, the 3PL model also estimates pseudo-guessing (Embretson & Reise, 2000), which is an estimate of the probability that an individual with low levels of the construct will endorse a difficult item (Zickar & Broadfoot, 2009). However, it appears that this parameter is rarely used when examining clinical constructs because it is not relevant in most cases (Raykov, 2016).

There are also unidimensional models for polytomous response options, including the partial credit model (Masters, 1982), rating scale model (Andrich, 1978), graded response model (Samejima, 1969) and nominal response model (Bock, 1972). The partial credit model and rating scale models are extensions of the 1PL model and are considered Rasch models because they are built on successive dichotomizations of adjacent categories where the Rasch model is applied to each dichotomization (Raykov, 2016). They are used for partially ordered categorical variables in which partial credit can be received and measure ability level and item difficulty. The rating scale model is used when all items share the same rating scale structure and the partial credit model is used when items each have a unique rating scale structure. Therefore, the rating scale model is a special, more parsimonious, case of the partial credit model (Embretson & Reise, 2000).

The graded response model is used for ordered categorical variables (e.g., Likert scales), and measures ability level, item difficulty, and item discrimination (Embretson & Reise, 2000). Because this model is an extension of the 2PL model, it splits the categorical options into a series of binary items and applies the 2PL model by comparing response option 1 to response options 2, 3, and 4; response option 1 and 2 to response options 3 and 4; and response option 1, 2, and 3 to response option 4, for example. This model is the most commonly used in clinical assessment (Templin, 2015). The nominal response model is used for unordered categorical models (e.g., multiple choice) and measures ability level, item difficulty, and item discrimination (Embretson & Reise, 2000). Similar to the graded response model, the nominal response model also splits the categorical options into a series of binary items and applies the 2PL model by comparing response options. However, the nominal response model treats each response option as a separate item (Templin, 2015).

For measurement questions concerning psychopathology, unidimensional models are not always ideal because many pathological constructs are multidimensional (Reise & Waller, 2009). Unfortunately, using a unidimensional model on multidimensional data adds systematic bias (Tyek Han, 2013). Researchers have dealt with this problem by deleting items or individual responses that do not fit the model (e.g., Betemps & Baker, 2004), not seriously considering model fit (Harvey, 2016), or conducting unidimensional analyses separately on each factor (e.g., Miller et al., 2013; Zickar & Broadfoot, 2010). However, researchers have developed multidimensional IRT models to address the fact that not all constructs of interest are unidimensional. Although multidimensional models are rarely used in clinical measurement due to their complexity (Raykov, 2016), they have become more widespread in recent years as more information has become available about them (Reckase, 2009). These models can be used for dichotomous or polytomous variables, can estimate the same parameters as unidimensional models (i.e., item difficulty, item discrimination, and pseudo-guessing), and are either compensatory or non-compensatory (Embretson & Reise, 2000). Compensatory models allow for higher levels of ability in one area to compensate for lower levels of ability in another area, whereas non-compensatory models consider each area's ability level separately (Embretson & Reise, 2000). For example, the multidimensional graded response model (Muraki & Carlson, 1995) is an extension of Samejima's (1969) graded response model for unidimensional data and largely operates under the same principles within each dimension of the construct.

Models are chosen based on dimensionality of the data with guidance from exploratory and CFA, for example (Embretson & Reise, 2000; Raykov, 2016). Models are also chosen based on the most appropriate fit to the data (Embretson & Reise, 2000; Harvey, 2016). There are a variety of ways to choose which model fits best, including comparing the plots from different models, examining residual correlations, employing statistical fit tests (e.g., χ^2-based tests), or examining goodness of fit statistics (e.g., Akaike's information criterion; Embretson & Reise, 2000). Despite these solutions, model fit tends to be a problem in IRT for clinical measurement, both deciding which model provides the best fit (Hambleton & Jones, 1993; Reise & Waller, 2009) and determining the dimensionality of the construct being measured (Hambleton & Jones, 1993).

Although model selection is an important process in IRT research, some research indicates that the consequences of choosing one model over another are not necessarily severe (e.g., Cook, Dodd, & Fitzpatrick, 1999; Maydeu-Olivares, Drasgow, & Mead, 1994). For example, there is some evidence that some IRT models are robust to the unidimensionality assumption (Raykov, 2016). Further, although multidimensional models will likely produce a model that fits the data well, using a unidimensional model might still be acceptable (He et al., 2014). For example, when considering model choice, θ estimates for the partial credit model, generalized partial credit model, and graded response model correlate highly (Cook et al., 1999), indicating that these models produce comparable θ estimates. Additionally, model fit was comparable when comparing fit statistics for the partial credit model and the graded response model (Maydeu-Olivares et al., 1994).

IRT and Posttraumatic Stress Disorder

Posttraumatic stress disorder (PTSD) researchers began to use IRT 25 years ago when examining PTSD assessments to refine them, create short forms, or examine invariance across gender or test forms in different languages. However, only a limited number of studies have utilized IRT, despite many recommendations to use this technique for psychological and health outcome research (e.g., Embretson, 1996; Reeve, 2003), and PTSD specifically (Weathers, Keane, King, & King, 1997). Considering the existing studies on IRT and PTSD, it is difficult to draw firm conclusions about item properties due to the change from *DSM-IV-TR* to *DSM-5* criteria and the use of varying measures, samples, and IRT models. Additionally, whereas some researchers reported all of the item parameters in a table, other researchers chose to highlight key findings using figures or limited in-text descriptions of item properties. Therefore, the following literature review will attempt to integrate findings across IRT studies that have reported item parameters in a systematic way. Of note, some studies reported item discrimination but not item difficulty parameters. Studies that used *DSM-IV-TR* criteria are presented first.

The first use of IRT when examining a PTSD assessment was a study by King et al. (1993), which examined the Mississippi Scale for Combat-Related Stress Disorder (Keane, Caddell, & Taylor, 1988) among a sample of veterans from the National Vietnam Veterans Readjustment Study using the unidimensional graded response model (Samejima, 1969). As the Mississippi Scale is not a *DSM*-correspondent measure, it includes items that are not part of the *DSM-IV-TR* or *DSM-5* criteria. Although some of these non-*DSM* items emerged among the highest or lowest in discrimination, they will not be described here to keep the context specific to *ICD-11* and the current study, which will only use *DSM*-correspondent measures. Regarding item discrimination, nightmares (i.e., that wake me up; trauma-congruent dreams), cued distress, detachment from others, sleep difficulties (i.e., need alcohol/drugs to sleep; fear of going to sleep), and external avoidance emerged as the most discriminating items. The least discriminating items were distorted blame, sleep difficulties (i.e., I fall asleep), and traumatic amnesia. Of note, the least discriminating items were all reverse-scored items and worded negatively. The negative wording of these items might explain why they have lower levels of discrimination (King et al., 1993), particularly because other questions about sleep difficulties emerged as highly discriminating. In fact, this finding about negative wording is not unique to the King and colleagues (1993) study, as two other studies have indicated that reverse-scored items were causing problems in reliability on the Mississippi Scale – Revised and the Mississippi Scale for Combat-Related PTSD (Betemps & Baker, 2004; Conrad et al., 2004). Additionally, Betemps and Baker (2004) found that problems in precision due to wording on the Mississippi Scale – Revised extended beyond reverse-scored items, specifically items that had the word "military" in them did not perform as expected. Therefore, results from the King and colleagues (1993) study should be interpreted with these caveats in mind.

The King and colleagues (1993) study also examined discrimination parameters across the PTSD spectrum. For example, irritability or anger, sleep difficulties, and strong startle response were not very discriminating at higher levels of PTSD severity (King et al., 1993). These results indicate that these symptoms might not represent core features of PTSD. However, cued distress, detachment from others, and external avoidance had higher levels of discrimination across the PTSD dimension. Therefore, these items are "more likely to encompass critical diagnostic cut points, both for full-blown PTSD and possibly for partial PTSD" (King et al., 1993; pg. 464).

The next known use of IRT in PTSD assessment was a study by Orlando and colleagues (2002), which examined DIF across the English and Spanish versions of the Posttraumatic Stress Disorder Checklist – Civilian (PCL-C; Weathers, Litz, Huska, & Keane, 1993) using the unidimensional graded response model (Samejima, 1969) among survivors of community violence (Orlando & Marshall, 2002). Considering the English version of the PCL-C, the most discriminating items were difficulty concentrating, cued physical reactions, intrusive memories, exaggerated startle response, cued distress, flashbacks, external avoidance, and internal avoidance. The least discriminating items were traumatic amnesia, sleep difficulties, and diminished interest (Orlando & Marshall, 2002). In 2003, Betemps and colleagues examined the Clinician Administered PTSD Scale (CAPS; Blake et al., 1995) using the Rasch model among a sample of veterans (Betemps, Smith, Baker, & Rounds-Kugler, 2003). Although the authors reported very few findings about item parameters, they reported that items similar to features of depression (e.g., loss of interest) emerged among the least difficult items (Betemps et al., 2003).

In 2009, Palm and colleagues examined the item properties of the National Comorbidity Study-Replication (NCS-R; Kessler et al., 2004) survey among a nationally representative U.S. sample using the 2PL model (Palm et al., 2009). Considering item difficulty, lack of future orientation, traumatic amnesia, irritability or anger, cued physical reactions, flashbacks, exaggerated startle response, diminished interest, and numbing emerged as the most difficult items. Sleep difficulties, intrusive memories, and internal avoidance emerged as the least difficult items. Considering discrimination, intrusive memories, cued distress, cued physical reactions, exaggerated startle response, sleep difficulties, flashbacks, numbing, and nightmares emerged as the most discriminating items. Traumatic amnesia, internal avoidance, and external avoidance emerged as the least discriminating items (Palm et al., 2009).

Next, Fissette and colleagues (2014) examined the Posttraumatic Stress Disorder Checklist – Military (PCL-M; Weathers, Litz, Huska, & Keane, 1994) using the unidimensional graded response model (Samejima, 1969) in a sample of active-duty service members (Fissette et al., 2014). Traumatic amnesia, lack of future orientation, flashbacks, external avoidance, diminished interest, cued physical reactions, numbing, and intrusive memories emerged as the most difficult items. Sleep difficulties, hypervigilance, and detachment from others emerged as the least difficult items. Considering item discrimination, cued distress, flashbacks, external avoidance, nightmares, cued physical reactions, internal avoidance, detachment from others, and traumatic amnesia emerged as the most discriminating items. Exaggerated startle response, hypervigilance, and numbing emerged as the least discriminating items (Fissette et al., 2014). As for item discrimination across the PTSD spectrum, cued distress, nightmares, and cued physical reactions were most discriminating at high levels of PTSD severity and less discriminating at low levels (Fissette et al., 2014). Therefore, these symptoms might represent core features of PTSD, as they provide the most information about the PTSD construct at a level that is diagnostically important. On the other hand, exaggerated startle response, numbing, and difficulty concentrating were not very discriminating at high levels of PTSD severity (Fissette et al., 2014), indicating these symptoms might not represent core features of PTSD.

Moving to *DSM-5* criteria, there is only one study that has examined *DSM-5* PTSD symptoms. Miller and colleagues (2013) examined the National Stressful Events Survey among a nationally representative U.S. sample (Kilpatrick, Resnick, Baber, Guille, & Gros, 2011). Whereas the studies described thus far used unidimensional models with PTSD as θ, this study implemented a unidimensional model within each cluster of PTSD, meaning that there is a θ for each cluster. Of note, the avoidance cluster did not converge because it had only two indicators (Miller et al., 2013). Traumatic amnesia emerged as the most difficult symptom in the NACM symptom cluster and irritability or anger and reckless or self-destructive behavior emerged as the most difficult symptoms in the AAR cluster. Sleep difficulties emerged as the least difficult symptom in the AAR cluster. However, all items had similar levels of difficulty in the intrusions cluster. Additionally, traumatic amnesia, reckless or self-destructive behavior, and irritable or aggressive behavior all emerged as poorly discriminating items (Miller et al., 2013).

A few patterns emerged from the existing IRT and PTSD studies. Overall there is some support for the items chosen by *ICD-11* with the exception of hypervigilance. This item never emerged as highly discriminating, and in one study, it emerged among the least discriminating items (Fissette et al., 2014). Interestingly, the intrusions symptoms not included in *ICD-11* (i.e., intrusive memories, cued distress, cued physical reactions) emerged as highly discriminating in at least two studies (Fissette et al., 2014; D. W. King et al., 1993; Orlando & Marshall, 2002; Palm et al., 2009). Additionally, the AAR symptoms not included in *ICD-11* (i.e., sleep difficulties; concentration difficulties) emerged as highly discriminating in at least one study (King et al., 1993; Orlando & Marshall, 2002; Palm et al., 2009). However, there is not much evidence that the NACM cluster contains highly discriminating items, as only a few symptoms emerged as highly discriminating with no clear pattern. Further, traumatic amnesia emerged as poorly discriminating across four studies (King et al., 1993; Miller et al., 2013; Orlando & Marshall, 2002; Palm et al., 2009). In sum, no firm conclusion can be made about which symptoms represent the core symptoms of PTSD due to conflicting findings across studies and lack of replication using *DSM-5* criteria.

DIF across Gender

Gender differences in PTSD are a well-established phenomenon in the field such that women are consistently more likely than men to have more severe total PTSD scores (Bovin et al., 2015: Carmassi et al., 2014; Vazquez et al., 2006) and to meet criteria for PTSD (Carmassi et al., 2014; Dell'Osso et al., 2010; Holbrook et al., 2002; Luxton et al., 2010; Olff et al., 2007; Zlotnick, 2001), even after experiencing the same index trauma (e.g., Dell'Osso et al., 2010;

Fullerton et al., 2001). Researchers posit a variety of reasons for this phenomenon. First, women are exposed to traumas that are more likely to lead to PTSD such as rape and sexual assault (McLean et al., 2009; Olff et al., 2007; Zlotnick, 2001). Second, gender differences in cognitive appraisal of the traumatic event and coping strategies after the traumatic event might put women at greater risk for PTSD as they are more likely to view the trauma as more threatening and use rumination as a coping strategy (See McLean et al., 2009; Olff et al., 2007 for a review). Third, acute reactions to trauma might also play a role in differential rates of PTSD, as women are more likely to experience peritraumatic dissociation, which is strongly associated with subsequent PTSD (Fullerton et al., 2001; Irish et al., 2011; Olff et al., 2007; Ozer et al., 2003). Last, biological factors such as gender differences in HPA axis functioning, oxytocin (Olff et al., 2007), and orbitofrontal cortex and amygdala activation (McClure et al., 2004) might also serve as protective factors against PTSD for men.

Although much of the research on gender differences in PTSD is at the syndrome level, there is also evidence that gender differences vary by symptom cluster and individual symptoms. Studies using *DSM-IV* criteria have varying results and indicate that women endorse significantly greater reexperiencing (Vazquez et al., 2006; Zlotnick, 2001), avoidance/numbing (Fullerton et al., 2001), and hyperarousal (Fullerton et al., 2001; Vazquez et al., 2006) symptoms than men. As for individual symptoms, women endorse higher rates of internal and external avoidance (Fullerton et al., 2001; Peters et al., 2006), anhedonia, sense of foreshortened future, sleeping difficulties (Fullerton et al., 2001; Peters et al., 2006), concentration difficulties (Fullerton et al., 2001), and startle response (Fullerton et al., 2001; Peters et al., 2006).

Studies that examined gender differences using *DSM-5* criteria also vary and indicate that women endorse significantly greater reexperiencing, avoidance, NACM (Bovin et al., 2015;

Carmassi et al., 2014), and AAR (Carmassi et al., 2014) symptoms than men. As for individual symptoms, women endorse higher rates of intrusive memories, nightmares, flashbacks, internal avoidance, external avoidance, strong negative feelings, decreased interest in activities, hypervigilance, startle response, and difficulty sleeping (Carmassi et al., 2014).

Due to the strong literature on gender differences in PTSD symptoms, researchers have begun to examine this phenomenon using DIF across gender (e.g., He et al., 2014; King et al., 2013; Palm et al., 2009) for a more nuanced understanding of gender differences on individual PTSD symptoms. However, there is no consensus on which items display DIF, likely due to varying index traumas and the switch from *DSM-IV-TR* to *DSM-5*. First, whereas two studies indicated that persistent negative emotional state displays DIF, with women having a higher likelihood of endorsement (He et al., 2014; Palm et al., 2009), one study indicated the opposite (i.e., men have a higher likelihood of endorsement [King et al., 2013]). Two studies also indicate that nightmares displays DIF, with men having a higher likelihood of endorsement (King et al., 2013; Palm et al., 2009). Similarly, two studies indicate that intrusive memories displays DIF, with men having a higher likelihood of endorsement (He et al., 2014; Palm et al., 2009). Further, traumatic amnesia (He et al., 2014), cued distress, concentration difficulties (King et al., 2013), and exaggerated startle response (Palm et al., 2009) have all been indicated to have DIF with women having a higher likelihood of endorsement. Additionally, hypervigilance (King et al., 2013) and irritability or anger (Palm et al., 2009) have been indicated to have DIF with men having a higher likelihood of endorsement. Finally, two studies indicate that the lack of plan for future item displays DIF, with men having a higher likelihood of endorsement (He et al., 2014; Palm et al., 2009). However, this item is no longer relevant as it is not in the *DSM-5* criteria. In sum, it is difficult to draw firm conclusions about DIF across gender from the existing research. Therefore, more research is needed, particularly with *DSM-5*.

Limitations of Existing PTSD and IRT Literature

The first limitation of existing IRT research on measures of PTSD is that all of these studies used unidimensional IRT models. Using a unidimensional model is in direct contrast with the body of PTSD research that indicates PTSD is multidimensional with a four- (King, Leskin, King, & Weathers, 1998), five- (Elhai, Biehn, et al., 2011), six- (Liu et al., 2014; Tsai et al., 2015), or seven- (Armour et al., 2015) factor structure. Additionally, according to Reise and colleauges (2007), the more heterogeneous the items on a measure, the more reason to use a multidimensional model. As discussed previously, the items on PTSD measures are quite heterogeneous, which provides more support not to use a unidimensional model. Despite theory and the confirmatory factor analysis (CFA) literature that indicates a multidimensional structure, most researchers have concluded that PTSD can be considered unidimensional for the purposes of IRT analyses through Rasch-based principal components analysis and CFA (e.g., E. Betemps & Baker, 2004; Elhai, de Francisco Carvalho, et al., 2011; Fissette et al., 2014; D. W. King et al., 1993; Palm et al., 2009). Therefore, it remains unclear if the conclusions drawn from these methods for assessing unidimensionality are acceptable (Miles, Marshall, & Schell, 2008; Miller et al., 2013), and it is possible that results of these studies do not accurately depict the properties of individual items on the measures (Miles et al., 2008). Some researchers have handled this potential problem of unidimensionality by examining each PTSD symptom cluster independently using IRT (Miller et al., 2013), which is a recommended technique for high levels of multidimensionality (Zickar & Broadfoot, 2010). However, no researchers have employed multidimensional models, which are also recommended for constructs with high levels of multidimensionality (Zickar & Broadfoot, 2010).

Opposing evidence for multidimensionality of the PTSD construct has emerged from CFA studies that found PTSD's factors to be highly intercorrelated (e.g., Forbes et al., 2015). As stated previously, dimensionality can be conceptualized as continuous (Zickar & Broadfoot, 2009) and unidimensional IRT models stand up relatively well to low levels of multidimensionality (e.g., Reckase, 1979). Therefore, there is some evidence that PTSD has the level of dimensionality necessary to not violate the assumption of unidimensionality. Despite this possibility, researchers should attempt to employ the multidimensional model to PTSD to be consistent with suggestions from researchers (Embretson & Reise, 2000; Raykov, 2016; Zickar & Broadfoot, 2010).

The second limitation of these studies concerns model fit. Some studies selected the IRT model based on variable characteristics alone (Betemps & Baker, 2004; Betemps et al., 2003; Choi et al., 2006; Conrad et al., 2004; M. W. King et al., 2013; Orlando & Marshall, 2002). However, researchers suggest fitting the data to the model (Embretson & Reise, 2000; Harvey, 2016), rather than picking the most widely used model (e.g., Rasch; Harvey, 2016). For example, using the Rasch model often leads to the application of an overly-restrictive model. In turn, researchers choose to delete items that do not fit the model rather than choosing a less restrictive model that accounts for all of the items (e.g., Betemps & Baker, 2004). However, test validity should be considered before model fit, and it is an empirical rather than theoretical question to use the Rasch model (Harvey, 2016; Raykov, 2016), and not all studies have used this contraindicated approach. As suggested by Harvey (2016), some studies have chosen a model according to these standards by comparing model fits or choosing model fit based on suggested fit statistics (He et al., 2014; Palm et al., 2009), which likely resulted in results that more accurately reflect item characteristics (Tyek Han, 2013).

The third limitation of the existing IRT research on measures of PTSD involves the types of measures examined. Studies used measures with a dichotomous rating scale (e.g., NCS-R; He et al., 2014; Palm et al., 2009), which is not common for assessing PTSD severity. In fact, measures of psychopathology that utilize dichotomous rating scales have lower levels of test-retest reliability, internal consistency, and convergent validity (Preston & Colman, 2000). Additionally, research thus far has examined measures that were not well-validated (e.g., National Stressful Events Survey; Miller et al., 2013). In sum, conclusions drawn from these studies might not be generalizable to more widely-used or well-validated measures of PTSD. Further, as only one study has examined *DSM-5* PTSD criteria using IRT (Miller et al., 2013), IRT research is needed that examines PTSD assessments that have been updated to reflect *DSM-5* criteria. As many of the studies described above used outdated measures, it is difficult to draw definitive conclusions on how items are functioning across the PTSD severity spectrum, particularly because three new items have been added to the PTSD criteria and many items have been reworded.

Fourth, the majority of IRT and PTSD studies have examined item parameters in a combined sample of men and women. As mentioned previously, there is a strong literature that suggests different symptom profiles for men and women and DIF across men and women (He et al., 2014; King et al., 2013; Palm et al., 2009). As such, it is possible that combining these populations in IRT and PTSD studies leads researchers to draw conclusions about item difficulty and discrimination that do not necessarily apply to men individually or women individually.

Finally, only two known studies examined item discrimination across the PTSD spectrum (Fissette et al., 2014; King et al., 1998). In theory, symptoms that are highly discriminating at moderate to high levels of a disorder represent key diagnostic features because they contribute relatively more information to the disorder (Raykov, 2016). Without more research on how PTSD symptoms perform at moderate to high levels of PTSD rather than their general performance, it will not be possible to draw firm conclusions about the core PTSD symptoms.

The overarching aim of the current study was to conduct an IRT analysis using a widely used measure of PTSD to identify the optimally performing items with an eye towards a smaller criterion set, much like the narrow approach of *ICD-11*. The first aim was to choose an IRT model that accurately reflects the dimensionality of the PTSD construct. This aim was accomplished through comparing the one-factor and four-factor *DSM-5* models of PTSD (see Table 1 for item loadings for the four-factor *DSM-5* model). Although previous research has shown that the six- and seven-factor models also provide good fit, testing these models is outside of the scope of the current paper for a few reasons. First, the four-factor *DSM-5* model has repeatedly provided a good fit (see Armour, Műllerová, & Elhai, 2016 for a review), and according to some researchers, choosing a parsimonious model is important when deciding which factor structure best represents PTSD (Armour, 2015; Brown, 2015). Second, the factors of the six- and seven-factor models are highly intercorrelated (Armour et al., 2015; Liu et al., 2014), indicating the factors might not represent distinct constructs.

Third, little research on the construct validity of the further divisions of the PTSD factors has emerged. Therefore, there is very little empirical evidence that these new factors represent meaningful distinctions (Armour et al., 2016), and it appears that these more complex models are post hoc attempts to label empirically, rather than theoretically, derived factors. Fourth, as this is the first IRT study that attempted to test a multidimensional model, it is important to note that the more complex the multidimensional model, the less likely it will converge (Zickar & Broadfoot, 2010). Considering the symptoms of PTSD, there is already one factor with only two symptoms (i.e., avoidance). If the PTSD factors are further broken apart, there will be even more factors with only two symptoms, making it even more unlikely for the model to converge. Further, Brown (2015) cautions the use of models with only two items as an indicator for a factor (Brown, 2015).

The second aim was to examine DIF across men and women to determine if items display DIF. If any items display DIF, the following analyses were conducted independently for men and women to get the most accurate difficulty and discrimination parameters (Walker, 2011). The third aim was to examine the CRCs and item difficulty parameters for descriptive information about the relative difficulty of items on the PTSD measures. This information informed conclusions about which items are likely to be endorsed by individuals across the PTSD severity spectrum. The fourth aim was to examine the information functions and discrimination parameters to determine which items provide the most information about the PTSD construct at varying levels of PTSD. Together, these analyses served to determine which items represent core symptoms of PTSD. Considering the narrow approach used by *ICD-11*, the goal was to choose items that are highly discriminating at moderate to high levels of PTSD where a PTSD diagnosis will be most relevant. Specifically, the goal was to choose items that peak at high levels of θ, approximately 1.5 to 3 standard deviations above the mean (Raykov, 2016).

Hypotheses for the current study were largely exploratory due to the limited research on *DSM-5* correspondent measures of PTSD. However, some conclusions based on theory and past research were drawn. The first hypothesis was that a multidimensional model will fit better than a unidimensional model. This hypothesis is based on the extensive CFA literature that indicates that PTSD is a multidimensional construct and past research that has shown the four-factor model has a good fit (Armour et al., 2016). The second hypothesis was that there will be items that display DIF across men and women. Based on theory and past research, persistent negative emotional state will likely display DIF, with women having a higher likelihood of endorsement (He et al., 2014; Palm et al., 2009). Intrusive memories and nightmares will also likely display DIF, with men having a higher likelihood of endorsement (He et al., 2014; King et al., 2013; Palm et al., 2009).

The third hypothesis was that there will be well-performing items that should be retained for the narrow approach. Examples of well-performing items include flashbacks, nightmares, and cued distress. Flashbacks and nightmares, in particular, have strong theoretical support for representing core symptoms of PTSD (Brewin et al., 2009), and past IRT studies have indicated these items to be among the most discriminating (Fissette et al., 2014; King et al., 1998; Palm et al., 2009).

The fourth hypothesis was that there will be items that do not perform well. The first category of items that are expected to perform poorly will have low prevalence but might be pathognomonic. For example, the traumatic amnesia and reckless or self-destructive behavior symptoms were among the least discriminating items in previous IRT research (King et al., 1998; Miller et al., 2013; Orlando & Marshall, 2002; Palm et al., 2009) load poorly on their designated factor (Forbes et al., 2015; Liu et al., 2014), and have low diagnostic utility (Green et al., 2017), indicating that they might not be characteristic of PTSD. Additionally, the association between reckless or self-destructive behavior and PTSD is weak (Thomsen, Stander, McWhorter, Rabenhorst, & Milner, 2011) and is better accounted for by other factors, such as co-occurring psychopathology (Panagioti, Gooding, Taylor, & Tarrier, 2013; Zlotnick et al., 1997; Zoellner et al., 2013). It is important to note these empirical findings might be due to the fact that traumatic amnesia and reckless or self-destructive behavior have very low prevalence rates compared to other *DSM-5* symptoms of PTSD (Miller et al., 2013). Therefore, it could be that restriction of range is causing it to seem like these symptoms are not core PTSD symptoms. Although the results of the current study might also be impacted by restriction of range, it is important to examine these symptoms because, theoretically and clinically, they are important (Friedman et al., 2011; Weathers, Marx, Friedman, & Schnurr, 2014).

The second category of items that were expected to perform poorly will have high prevalence but low levels of discrimination. For example, symptoms such as persistent negative emotional state, diminished interest, and inability to experience positive emotions are highly prevalent (Miller et al., 2013) and emerged among the least difficult (Betemps et al., 2003) and least discriminating (Fissette et al., 2014; Orlando & Marshall, 2002) items in past IRT research. Additionally, theory and empirical research indicate that these symptoms might represent general symptoms of distress, rather than PTSD-specific symptoms (Brewin et al., 2009; Elklit et al., 2014; Gootzeit & Markon, 2011; Simms et al., 2002; Spitzer et al., 2007).

Method

Participants and Procedure

Undergraduate students 18 and older enrolled in a psychology course at a large, public, southeastern university were invited to complete an online survey related to *a very stressful life event*. Participants who consented completed an online self-report battery and were compensated with extra credit. All procedures were approved by the university institutional review board. Considering the models and parameter estimates of interest measuring (see below for a full description), a sample size of at least 500 is recommended (Embretson & Reise, 2000; Jiang, Wang, & Weiss, 2016; Reeve & Fayers, 2005).

Trauma exposure was coded by reviewing both participant responses on the Life Events Checklist-5 (LEC-5; Weathers et al., 2013) and written narratives of their index event. First, index events were coded as meeting *DSM-5* Criterion A if they endorsed their worst event on the LEC-5 as having either *happened to me directly* or *witnessed it* and endorsed that *my life was in danger*, *someone else's life was in danger*, or the event *involved sexual violence*. Index events were also coded as *DSM-5* Criterion A if they endorsed having *learned about it happening to a close family member or close friend* and the event involved *accident or violence* or *sexual violence*. If participants did not endorse any of these response patterns, index events were coded as not meeting *DSM-5* Criterion A. Second, two graduate students verified the *DSM-5* Criterion A status by reviewing each narrative. The raters independently read each narrative, and either agreed or disagreed with the code based on the LEC-5 responses following *DSM-5* Criterion A guidelines. When participants did not provide a narrative, responses were not used in the analyses.

Third, raters independently reviewed narratives for type of event, and categories included transportation accident, sexual assault, suicide, serious accident at work, home, or during a recreational activity, physical assault, cancer, divorce, death of a grandparent, life-threatening chronic illness or serious but nonlife-threatening injury (e.g., broken bone), heart problems, and natural disaster. Disagreements between the raters for syntax agreement, confidence ratings, and event type were resolved through discussion with the raters and an expert in trauma exposure and PTSD assessment.

Of the 2,233 individuals who completed the survey, 1,213 endorsed experiencing an event that met *DSM-5* Criterion A. Common event types for individuals included in the analyses included transportation accident (35.4%, *n* = 430), sexual assault (20.9%, *n* = 253), serious accident at work, home, or during a recreational activity (11.2%, *n* = 136), natural disaster (8.1%, *n* = 98), physical assault (5.0%, *n* = 61), and suicide (4.5%, *n* = 54).

The final sample consisted of 1,213 individuals with an average age of 19.30 (*SD* = 1.40; range = 18.0 – 30.0). The majority of participants identified as female (*n* = 927; 76.6%). Racially, the majority of participants identified as White (*n* = 1053; 87.0%) with the remaining participants identifying as African American/Black (*n* = 93; 7.7%), Asian (*n* = 31; 2.6%), Other/multi-racial (*n* = 29; 2.4%), American Indian/Alaskan Native (*n* = 4; 0.3%), and Native Hawaiian/Pacific Islander (*n* = 1; 0.1%). Ethnically, a minority of participants identified as Latino/Hispanic (*n* = 51; 4.3%).

Measures

A demographics questionnaire was used to assess sex, age, and race information. As described above, the LEC-5 (Weathers, et al., 2013) was used in conjunction with provided narratives to determine if the described index event satisfied *DSM-5* Criterion A. The LEC-5 is a self-report measure that consists of 17 categories of traumatic stressors (e.g., natural disaster, fire or explosion, transportation accident, serious accident, physical assault, sexual assault, combat exposure, life-threatening illness or injury). Respondents indicate the degree to which they have experienced each category of traumatic stressor: *happened to me, witnessed it, learned about it, part of my job, not sure,* or *does not apply*. Additionally, respondents identify their worst event. The LEC-5 also includes 14 items where respondents describe the worst event in detail (e.g., resulting injuries, age at time of event) to help clarify Criterion A status. Previous versions of the

LEC have been shown to be reliable and valid in a variety of samples (Gray, Litz, Hsu, & Lombardo, 2004).

PTSD symptoms were assessed using the PCL-5 (Weathers, Litz, Keane, Palmieri, Marx, & Schnurr, 2013), a 20-item self-report measure of *DSM-5* PTSD symptoms. The PCL-5 instructs respondents to rate how much they have been bothered by PTSD symptoms (e.g., *repeated, disturbing memories, thoughts, or images of the stressful experience*) in the past month, using a 5-point scale ranging from *not at all* to *extremely*. Higher PCL-5 scores indicate greater PTSD symptom severity, and possible scores range from 0 to 80. PCL-5 scores have strong reliability and validity (Blevins, Weathers, Davis, Witte, & Domino, 2015). In the current sample, $\alpha = 0.95$ for the full scale, 0.91 for the reexperiencing subscale, 0.88 for the avoidance subscale, 0.90 for the NACM subscale, and 0.88 for the AAR subscale.

Data Analytic Plan

The following analyses were conducted using Mplus version 8 (Muthén & Muthén, 1998-2017), unless otherwise indicated, and individual items were used as indicators for all latent variable models. Per Raykov's (2016) suggestion, items were treated as ordinal, rather than continuous, because of the small number of response options for each measure and due to the non-normally distributed data (Flora & Curran, 2004; Wirth & Edwards, 2007). Accordingly, parameters were estimated with mean- and variance-adjusted weighted least squares (WLSMV), which provides a robust χ^2. Missing data were handled using pairwise deletion. While pairwise deletion has limitations (Brown, 2015), in the current study it was the ideal method due to the small portion of the data that was missing, the absence of techniques for pooling most fit indices across estimates in multiple imputation (Enders, 2010), and inability to use full information maximum likelihood (FIML) procedures with the WLSMV estimator. Additionally, model fit for each model was assessed using multiple indices: χ^2, comparative fit index (CFI), Tucker-Lewis index (TLI), and root mean square error of approximation (RMSEA). Proposed fit statistics cutoffs outlined by Hu and Bentler (1999) and Kline (2005) will be used (RMSEA ≤ .06, CFI and TLI ≥ .95). Overall fit of each model was interpreted by taking all fit statistics into account (Brown, 2015).

Like previous researchers (e.g., Miller et al., 2013; Palm et al., 2009), and as outlined by Reise and colleagues (2007), measurement models of PTSD were compared using a series of nested models using robust χ^2 difference testing to determine the best-fitting model. χ^2 difference testing was conducting using the DIFFTEST procedure in Mplus, which generates scaled χ^2 difference tests that can be used to compare nested models, because regular χ^2 difference testing cannot be conducting when using WLSMV (Brown, 2015). The following models were compared:

1. Unidimensional graded response model with all 20 items loading on to a single general PTSD factor for the PCL-5. This model has been used multiple times in the IRT and PTSD literature (e.g., King et al., 2013; Orlando & Marshall, 2002).
2. Multidimensional graded response model that corresponds to the theoretically and empirically supported *DSM-5* four-factor model of PTSD for the PCL-5. No PTSD study that used IRT has used this type of model. However, researchers have called for the use of this model in multidimensional psychological constructs.

After choosing the best fitting model, now referred to as the baseline model, DIF analyses were conducted across men and women. Specifically, a multigroup confirmatory factor analysis (MGCFA) was conducted following established procedures (Brown, 2015; Dimitrov, 2010; Gregorich, 2006; Millsap & Olivera-Aguilar, 2012) to determine if observed scores on the PCL-

5 could be compared across gender. Once a consistent measurement model was established across groups, equality of factor loadings and the intercepts across groups was tested. First, the MGCFA was run with no cross-group constraints on intercepts and loadings and latent means were fixed at zero across groups to serve as the baseline model for the χ^2 difference test (i.e., baseline model). In the second step, an MGCFA was run with latent means fixed at zero across groups and cross-group equality constraints on the factor loadings while the intercepts were allowed to vary freely across groups (i.e., Model 2) and compared to the baseline model using a χ^2 difference test. In the third step of the MGCFA, cross-group equality constraints were placed on the intercepts, and factor means were fixed at zero for the Criterion A group (i.e., Model 3) and allowed to vary for the non-Criterion A group. This model was compared to Model 2 using a χ^2 difference test.

After establishing that observed scores could be compared across gender, DIF analyses were conducted using SPSS Version 23 (IBM, 2015) as outlined by Raykov (2016) and the Mantel-Haenszel (M-H) test across gender. The M-H test is a χ^2-based test that examines conditional independence. The null hypothesis states that there is no partial association between the two variables, the conditional odds ratio is equal to one, or response is not dependent on group. The test produces a *p* value for the significance of the χ^2 value, as well as the conditional odds ratio and corresponding 95% confidence intervals. An insignificant *p* value, an odds ratio close to one, or a confidence interval that includes one indicates that the null hypothesis should be retained. To run the M-H test, each PCL-5 response option (i.e., 0, 1, 2, 3, 4) was dummy coded to determine the exact source of DIF. Each PCL-5 response option within each item served as the dependent variable, whereas gender served as the independent variable. One hundred M-H tests were conducted, one for each of the five response options for each of the 20

PCL-5 items. Finally, due to the large number of tests, the Benjamini and Hochberg (1995) procedure to control for the false discovery rate was implemented, which downwardly adjusts the *p* value necessary for significance based on the number of tests conducted.

Next, item and test characteristics were examined using parameter estimates and graphs generated by Mplus version 8 (Muthén & Muthén, 1998-2017). Specifically, Mplus produced item difficulty parameters for each threshold and an item discrimination parameter for each item. Additionally, Mplus produced CRCs and IIFs for all items within each symptom cluster.

Results

Covariance coverage ranged from 0.985 to 0.992. For descriptive statistics on the PCL-5, see Table 2. 6.3% (*n* = 77) of participants score patterns were in line with a PTSD diagnosis.

Best-fitting Model

Fit statistics for the unidimensional graded response model were poor (χ^2 = 1420.15, *df* = 170, *p* < .001; CFI = 0.96; TLI = 0.96; RMSEA = 0.10 [0.09; 0.10]). Fit statistics for the multidimensional four-factor graded response model were acceptable (χ^2 = 744.91, *df* = 164, *p* < .001; CFI = 0.98; TLI = 0.98; RMSEA = 0.07 [0.06; 0.07]). χ^2 difference testing of the one-factor model nested within the four-factor model indicated that the one-factor model significantly worsened the fit (χ^2 = 270.97, *df* = 6, *p* < .001). Therefore, the four-factor model was retained for the remaining analyses. See Table 4 for difficulty parameters for each threshold and for discrimination parameters for each item.

DIF

The baseline model was run with no cross-group constraints on intercepts and loadings and latent means were fixed at zero across groups. As with the separate CFAs in each group, this model resulted in mediocre fit (χ^2 = 928.694, *df* = 328, *p* < .001; CFI = 0.88; TLI = 0.86;

RMSEA = 0.07 [0.06; 0.07]). Next, Model 2 was run, which was identical to the baseline model other than constraining the factor loadings to equality across groups. This model also resulted in mediocre fit (χ^2 = 938.028, *df* = 344, *p* < .001; CFI = 0.88; TLI = 0.87; RMSEA = 0.07 [0.06; 0.07]); however, constraining the factor loadings to equality did not significantly worsen model fit (χ^2 = 17.496, *df* = 16, *p* = .354), indicating factor loading equivalence across groups. Next, Model 3 was run with cross-group equality constraints placed on the intercepts, but allowing the latent means to freely vary for the female group. As with previous models, this model provided mediocre fit (χ^2 = 972.863, *df* = 360, *p* < .001; CFI = 0.88; TLI = 0.87; RMSEA = 0.07 [0.06; 0.07]); however, constraining the intercepts did not significantly worsen model fit (χ^2 = 24.540, *df* = 16, *p* = 0.078), indicating item intercepts were equivalent across groups. Collectively, these results indicate that the factor structure of PTSD is equivalent across gender and that observed scores on the PCL-5 can be compared across gender.

DIF analyses indicated that six response options displayed DIF, including response option 0 of the following items: nightmares, cued distress, internal avoidance, external avoidance; response option 2 of inability to experience positive emotions; and response option 4 of negative beliefs. See Table 3 for associated M-H results. However, after conducting the Benjamini and Hochberg (1995) test to control for the false discovery rate, no M-H tests remained statistically significant, which provided justification to report item and test parameters in the combined sample.

Item Difficulty

Within the reexperiencing factor, cued distress, intrusive memories, cued physical reactions, nightmares, and flashbacks emerged as the least to the most difficult, respectively. Within the avoidance factor, internal avoidance and external avoidance emerged as the least to the most difficult, respectively. Within the NACM factor, item difficulty varied depending on the level of theta. At levels of theta below the mean, traumatic amnesia, distorted blame, persistent negative emotional state, negative beliefs, detachment from others, inability to experience positive emotions, and diminished interest emerged as the least to the most difficult, respectively. At levels of theta above the mean, persistent negative emotional state, detachment from others inability to experience positive emotions, diminished interest, negative beliefs, blame, and traumatic amnesia emerged as the least to the most difficult, respectively. Within the AAR factor, item difficulty varied depending on the level of theta. At levels of theta below the mean, hypervigilance, difficulty sleeping, exaggerated startle response, difficulty concentrating, irritability or anger, and reckless or self-destructive behavior emerged as the least to the most difficult, respectively. At levels of theta above the mean, difficulty concentrating, difficulty sleeping, hypervigilance, exaggerated startle response, irritability or anger, and reckless or self-destructive behavior emerged as the least to the most difficult, respectively. See Figures 1 – 4 for CRCs within the reexperiencing, avoidance, NACM, and AAR factors.

Item Discrimination

Within the reexperiencing factor, nightmares, intrusive memories, cued distress, flashbacks, and cued physical reaction emerged as the least to most discriminating at approximately 1.5 to 3 standard deviations above the mean, respectively. Of note, although flashbacks is comparatively less discriminating than cued physical reactions at 1.5 to 3 standard deviations above the mean, this symptom peaks at slightly higher levels of theta, indicating it is more discriminating among individuals with more severe presentations. Within the avoidance factor, external avoidance and internal avoidance emerged as the least to most discriminating at approximately 1.5 to 3 standard deviations above the mean, respectively.

Within the NACM factor, traumatic amnesia, distorted blame, negative beliefs, persistent negative emotional state, diminished interest, detachment from others, and inability to experience positive emotions emerged as the least to most discriminating at approximately 1.5 to 3 standard deviations above the mean, respectively. Of note, although traumatic amnesia is the least discriminating at 1.5 to 3 standard deviations above the mean, it is the most discriminating at 3 to 6 standard deviations above the mean, indicating this symptom is performing well among individuals with the highest levels of PTSD. Within the AAR factor, reckless or self-destructive behavior, hypervigilance, exaggerated startle response, irritability or anger, sleep difficulties, and concentration difficulties emerged as the least to most discriminating at approximately 1.5 to 3 standard deviations above the mean, respectively. Of note, although exaggerated startle response, irritability or anger, and reckless or self-destructive behavior are among the least discriminating 1.5 to 3 standard deviations above the mean, they are the most discriminating approximately 3.5 to 5 standard deviations above the mean, indicating these symptoms are performing better among individuals with the highest levels of PTSD. See Figures 5 – 8 for IIFs within the reexperiencing, avoidance, NACM, and AAR factors.

Discussion

This is the first known study to examine DIF across gender and to calculate item difficulty and discrimination parameters using a multidimensional graded response model on a measure of PTSD. As predicted, findings indicated that the four-factor *DSM-5* model fit the data better than the unidimensional model, which provided justification to conduct the rest of the IRT analyses using the four-factor model. These results are in line with theory and empirical studies that indicate PTSD is a multidimensional construct, and that the four-factor model adequately fits the data.

Contrary to hypotheses, no items displayed DIF across gender after conducting the Benjamini and Hochberg (1995) procedure to control for the false discovery rate. These results are contrary to previous studies that indicate DIF for symptoms such as persistent negative emotional state, intrusive memories, and nightmares (He et al., 2014; King et al., 2013; Palm et al., 2009). It is likely that results differed in the current study due to the use of a non-clinical sample. Because of the absence of DIF across gender, the remaining analyses were conducted in a combined sample of men and women.

Consistent with hypotheses about well-performing items, cued physical reactions, cued distress, and flashbacks emerged as most discriminating at 1.5 to 3 standard deviations above the mean. Results suggest that these items represent core features of the reexperiencing factor. Inconsistent with hypotheses about well-performing items, nightmares emerged as the least discriminating symptom at 1.5 to 3 standard deviations above the mean. This result was unexpected based on past theoretical and empirical research that nightmares represents a core feature of PTSD (Brewin et al., 2009; Fissette et al., 2014; King et al., 1998; Palm et al., 2009).

Consistent with hypotheses about poorly performing items, traumatic amnesia and reckless or self-destructive behavior emerged as the least discriminating items at 1.5 to 3 standard deviations above the mean. However, these symptoms were among the most discriminating at higher levels of PTSD, indicating that, in the current sample, this symptom might represent a core feature of PTSD only among individuals with the most severe symptom presentation. Inconsistent with hypotheses about poorly performing items, inability to experience positive emotions emerged as most discriminating at 1.5 to 3 standard deviations above the mean. This result was unexpected based on theory and empirical research that indicates this symptom represents a general symptom of distress (Brewin et al., 2009; Elklit et al., 2014;

Gootzeit & Markon, 2011; Simms et al., 2002; Spitzer et al., 2007) and past IRT research that indicates this symptom is among the least discriminating (Fissette et al., 2014; Orlando & Marshall, 2002).

Although the *ICD-11* workgroup's aim was to select the "core" symptoms of PTSD, according to the results of the current study, only some of the selected symptoms in the *ICD-11* criteria actually represent the core symptoms, as defined by the most discriminating at 1.5 to 3 standard deviations above the mean. Specifically, flashbacks and the avoidance symptoms were the only two symptoms that emerged as the most discriminating that are included in the *ICD-11* criteria. These results are partially in line with a specificity analysis conducted by Green and colleagues (2017), which indicated that flashbacks was the only *ICD-11* symptom with good specificity. Based on the results of the current study, a reduced criterion set for the narrow approach should include flashbacks, cued physical reactions, internal avoidance, external avoidance, detachment from others, inability to experience positive emotions, sleep difficulties, and concentration difficulties. Of note, due to the fact that there were only two symptoms in the avoidance factor and both had high levels of discrimination, it is assumed that these should be included in a reduced criterion set.

These results also provide valuable information for the broad approach. Researchers have recommended weighting PTSD symptoms differently based on their relative importance to the PTSD diagnosis (Betemps, Smith, Baker, & Rounds-Kugler, 2003; Bliese et al., 2008; Fissette et al., 2014; King, 2013). Considering this approach, if all 20 symptoms in *DSM-5* are being considered, flashbacks, cued physical reactions, internal avoidance, external avoidance, detachment from others, inability to experience positive emotions, sleep difficulties, and concentration difficulties should be given more weight in the PTSD diagnosis by either making them a requirement or having them worth double the points toward the total score compared to the other symptoms due to their higher levels of discrimination.

Limitations

The first limitation of the current study involves the generalizability of the findings. Although the main goal of the current study was to choose a single criterion set with the intention for it to optimally generalize across a wide variety of trauma populations and other demographics, it is important to note that our selected criterion set might not be optimal for subgroups within our sample or groups outside of our sample. For example, research indicates that factor structures vary based on trauma type (Chung & Breslau, 2008; Frankfurt et al., 2016). As participants experienced a variety of types of traumas, there is a possibility for DIF across these groups. To mitigate this limitation, DIF analysis on gender was conducted, which is likely a major subgroup of the current sample. Further, the current sample is largely the same age and race, so it is possible that the same criterion set might apply to most of the individuals. Future research should conduct DIF analyses in a variety of subgroups to ensure the most accurate criterion set.

The second limitation is the use of a non-clinical sample due to restriction of range. However, presumptive PTSD was 6.3%, all individuals met *DSM-5* Criterion A. This could have impacted results, particularly for the examination of the most discriminating items in the AAR cluster. Specifically, the most discriminating symptoms found in the current study are often present in individuals without PTSD (i.e., difficulty concentrating and sleeping), particularly college students, due to the stressful transition period and increase in academic and social demands (Hershner & Chervin, 2014). Despite this limitation, some symptoms that emerged as highly discriminating do not appear to be specific to college students such as inability to experience positive emotion, detachment from others, and flashbacks. Further, some of our findings are comparable to IRT studies that used clinical samples. For example, similar symptoms were found to be highly discriminating (e.g., cued physical reactions [Orlando & Marshall, 2002, Palm et al., 2009], detachment from others [Fissette et al., 2014; King et al., 1998], sleep difficulties [King et al., 1998; Palm et al., 2009], concentration difficulties [Orlando & Marshall, 2002]) or poorly discriminating (e.g., traumatic amnesia [King et al., 1998, Orlando & Marshall, 2002; Miller et al., 2013; Palm et al., 2009], distorted blame [King et al., 1998], reckless or self-destructive behavior [Miller et al., 2013], hypervigilance [Fissette et al., 2014]). Future research should conduct a similar study using a clinical sample to examine if the most discriminating symptoms in a clinical sample align with the *ICD-11* criteria.

The third limitation was using a self-report measure to measure PTSD symptoms and to subsequently draw conclusions about the core symptoms of PTSD. Due to the nature of self-report measures, it is possible that participants over- or under-reported symptoms. Therefore, future research on the core symptoms of PTSD should employ a clinician-administered measure to ensure participants are correctly reporting symptoms.

Fourth, *ICD-11* conceptualizes PTSD with three factors whereas the current study models the data using four factors. This difference in modeling might make direct comparisons between the results of the current study and *ICD-11*'s symptom selection difficult to make. Additionally, it appears that the *ICD-11* approach was to choose the core symptoms of PTSD, whereas the current study's approach was to choose the core symptoms within each symptom cluster. In fact, Brewin (2017) raises this critique of network analytic studies that examined all 20 *DSM-5* PTSD symptoms. To mitigate this limitation, the difficulty and discrimination parameters were conducted in the combined sample using a unidimensional model, despite poor fit. These parameters are available in Supplemental Table 1. Results indicate that similar symptoms emerge as most discriminating at 1.5 to 3 standard deviations above the mean when using a unidimensional model, and NACM and avoidance symptoms are among the most discriminating items. Specifically, traumatic amnesia, reckless or self-destructive behavior, hypervigilance, irritability or anger, distorted blame, exaggerated startle response, nightmares, intrusive memories, difficulty sleeping, flashbacks, negative beliefs, cued distress, cued physical reactions, difficulty concentrating, persistent negative emotional state, external avoidance, internal avoidance, diminished interest, detachment from others, and inability to experience positive emotions emerged as the least to most discriminating at 1.5 to 3 standard deviations above mean, respectively. See Supplemental Figures 1 and 2 for the unidimensional model CRC and IIF.

These results are in line with network analytic studies that indicated that NACM symptoms are central to the PTSD network (Armour et al., 2017; Mitchell et al., 2017). Therefore, when selecting a reduced criterion set for the narrow approach using the one-factor model, negative feelings, external avoidance, internal avoidance, diminished interest, detachment from others, and inability to experience positive emotions would be recommended. In this case, only the avoidance symptoms are in line with *ICD-11* criteria. When considering a reduced criterion set without NACM symptoms as Brewin (2017) recommended, flashbacks, cued distress, cued physical reaction, difficulty concentrating, external avoidance, and internal avoidance would be recommended. In this case, flashback and the avoidance symptoms are the only ones in line with *ICD-11* criteria. Therefore, even when excluding NACM symptoms, there is a lack of evidence for some of the proposed *ICD-11* symptoms (i.e., nightmares, exaggerated startle response, hypervigilance).

The narrow approach includes only the core symptoms of PTSD and offers the possibility to reduce comorbidity while increasing the likelihood that symptoms are associated with trauma. Additionally, the narrow approach likely reduces heterogeneity in the PTSD diagnosis, which will be helpful for exploring the genetic markers of PTSD. Although the narrow approach has many merits, it appears that *ICD-11*'s attempt at this approach is not supported by sufficient evidence. When considering the results of previous specificity analyses, network analyses, research on comorbidity, prevalence studies, and the current study, there is limited support for the *ICD-11* criteria for PTSD.

References

American Psychiatric Association. (2000). *Diagnostic and Statistical Manual of Mental Disorders, Fourth Edition, Text Revision (DSM-IV-TR)* (4th ed., Vol. 1). Arlington, VA: American Psychiatric Association.

American Psychiatric Association. (2013). *Diagnostic and Statistical Manual of Mental Disorders* (Fifth Edition). American Psychiatric Association.

Andrich, D. (1978). A rating formulation for ordered response categories. *Psychometrika*, *43*, 561–573. doi:10.1007/BF02293814

Armour, C. (2015). The underlying dimensionality of PTSD in the diagnostic and statistical manual of mental disorders: Where are we going? *European Journal of Psychotraumatology*, *6*, 1–8. doi:10.3402/ejpt.v6.28074

Armour, C., Fried, E. I., Deserno, M. K., Tsai, J., & Pietrzak, R. H. (2017). A network analysis of *DSM-5* posttraumatic stress disorder symptoms and correlates in US military veterans. *Journal of Anxiety Disorders, 45*, 49-59. doi:10.1016/j.janxdis.2016.11.008

Armour, C., Műllerová, J., & Elhai, J. D. (2016). A systematic literature review of PTSD's latent structure in the Diagnostic and Statistical Manual of Mental Disorders: *DSM-IV* to *DSM-5*. *Clinical Psychology Review*, *44*, 60–74. doi:10.1016/j.cpr.2015.12.003

Armour, C., Tsai, J., Durham, T. A., Charak, R., Biehn, T. L., Elhai, J. D., & Pietrzak, R. H. (2015). Dimensional structure of *DSM-5* posttraumatic stress symptoms: Support for a hybrid anhedonia and externalizing behaviors model. *Journal of Psychiatric Research*, *61*, 106–113. doi:10.1016/j.jpsychires.2014.10.012

Benjamini, Y., & Hochberg, Y. (1995). Controlling the false discovery rate: A practical and powerful approach to multiple testing. *Journal of Royal Statistical Society*, *57*, 289–300.

Betemps, E., & Baker, D. G. (2004). Evaluation of the Mississippi PTSD Scale—Revised using Rasch measurement. *Mental Health Services Research*, *6*, 117–125. doi:10.1023/B:MHSR.0000024355.38667.e5

Betemps, E. J., Smith, R. M., Baker, D. G., & Rounds-Kugler, B. A. (2003). Measurement precision of the Clinician Administered PTSD Scale (CAPS): A Rasch model analysis. *Journal of Applied Measurement*, *4*, 59–69.

Blevins, C. A., Weathers, F. W., Davis, M. T., Witte, T. K., & Domino, J. L. (2015). The Posttraumatic Stress Disorder Checklist for *DSM-5* (PCL-5): Development and initial psychometric evaluation. *Journal of Traumatic Stress*, *28*, 489–498. doi:10.1002/jts.22059

Bliese, P. D., Wright, K. M., Adler, A. B., Cabrera, O., Castro, C. A., & Hoge, C. W. (2008). Validating the Primary Care Posttraumatic Stress Disorder Screen and the Posttraumatic Stress Disorder Checklist with soldiers returning from combat. *Journal of Consulting and Clinical Psychology*, *76*, 272–281. http://doi.org/10.1037/0022-006X.76.2.272

Bock, D. R. (1972). Estimating item parameters and latent ability when responses are scored in two or more nominal categories. *Psychometrika*, *37*, 29–51. doi:10.1007/BF02291411

Bodkin, J. A., Pope, H. G., Detke, M. J., & Hudson, J. I. (2007). Is PTSD caused by traumatic stress? *Journal of Anxiety Disorders*, *21*, 176–182. doi:10.1016/j.janxdis.2006.09.004

Brewin, C. R. (2007). Autobiographical memory for trauma: Update on four controversies. *Memory*, *15*, 227–248. doi:10.1080/09658210701256423

Brewin, C. R. (2013). “I wouldn’t start from here” - An alternative perspective on PTSD from the *ICD-11* : Comment on Friedman (2013): Perspective on PTSD from *ICD-11*. *Journal of Traumatic Stress*, *26*, 557–559. doi:10.1002/jts.21843

Brewin, C. R., Lanius, R. A., Novac, A., Schnyder, U., & Galea, S. (2009). Reformulating PTSD for *DSM-V* : Life after Criterion A. *Journal of Traumatic Stress*, *22*(5), 366–373. doi:10.1002/jts.20443

Brown, T. A. (2015). *Confirmatory factor analysis for applied research, Second Edition*. New York; London: The Guilford Press.

Bryant, R. A., Creamer, M., O'Donnell, M., Forbes, D., McFarlane, A. C., Silove, D., & Hadzi-Pavlovic, D. (2017). Acute and chronic posttraumatic stress symptoms in the emergence of posttraumatic stress disorder: A network analysis. *JAMA Psychiatry,74*, 135-142. doi:10.1001/jamapsychiatry.2016.3470

Choi, Y., Mericle, A., & Harachi, T. W. (2006). Using Rasch analysis to test the cross-cultural item equivalence of the Harvard Trauma Questionnaire and the Hopkins Symptom Checklist across Vietnamese and Cambodian immigrant mothers. *Journal of Applied Measurement*, *7*, 16.

Chung, H., & Breslau, N. (2008). The latent structure of posttraumatic stress disorder: Tests of invariance by gender and trauma type. *Psychological Medicine*, *38*, 563-573. doi:10.1017/s0033291707002589

Cloitre, M., Garvert, D. W., Brewin, C. R., Bryant, R. A., & Maercker, A. (2013). Evidence for proposed *ICD-11* PTSD and complex PTSD: A latent profile analysis. *European Journal of Psychotraumatology*, *4*, 20706. doi:10.3402/ejpt.v4i0.20706

Cloitre, M., Garvert, D. W., Weiss, B., Carlson, E. B., & Bryant, R. A. (2014). Distinguishing PTSD, complex PTSD, and borderline personality disorder: A latent class analysis. *European Journal of Psychotraumatology*, *5*, 25097. doi:10.3402/ejpt.v5.25097

Conrad, K. J., Wright, B. D., McKnight, P., McFall, M., Fontana, A., & Rosenheck, R. (2004). Comparing traditional and Rasch analyses of the Mississippi PTSD Scale: Revealing limitations of reverse-scored items. *Journal of Applied Measurement*, *5*, 15–30.

Cook, K. F., Dodd, B. G., & Fitzpatrick, S. J. (1999). A comparison of three polytomous item response theory models in the context of testlet scoring. *Journal of Outcome Measurement*, *3*(1), 1–20.

Cuthbert, B. N. (2014). The RDoC framework: Facilitating transition from ICD/DSM to dimensional approaches that integrate neuroscience and psychopathology. *World Psychiatry*, *13*, 28–35.

Duke, L. A., Allen, D. N., Rozee, P. D., & Bommaritto, M. (2008). The sensitivity and specificity of flashbacks and nightmares to trauma. *Journal of Anxiety Disorders*, *22*, 319–327. doi:10.1016/j.janxdis.2007.03.002

Elhai, J. D., Biehn, T. L., Armour, C., Klopper, J. J., Frueh, B. C., & Palmieri, P. A. (2011). Evidence for a unique PTSD construct represented by PTSD's D1–D3 symptoms. *Journal of Anxiety Disorders*, *25*, 340–345. doi:10.1016/j.janxdis.2010.10.007

Elhai, J. D., de Francisco Carvalho, L., Miguel, F. K., Palmieri, P. A., Primi, R., & Christopher Frueh, B. (2011). Testing whether posttraumatic stress disorder and major depressive disorder are similar or unique constructs. *Journal of Anxiety Disorders*, *25*, 404–410. doi:10.1016/j.janxdis.2010.11.003

Elklit, A., Hyland, P., & Shevlin, M. (2014). Evidence of symptom profiles consistent with posttraumatic stress disorder and complex posttraumatic stress disorder in different trauma samples. *European Journal of Psychotraumatology*, *5*. doi:10.3402/ejpt.v5.24221

Embretson, S. E. (1996). The new rules of measurement. *Psychological Assessment*, *8*, 341–349. doi:10.1037/1040-3590.8.4.341

Embretson, S. E., & Reise, S. P. (2000). *Item response theory for psychologists*. Mahwah, N.J: L. Erlbaum Associates.

Fissette, C. L., Snyder, D. K., Balderrama-Durbin, C., Balsis, S., Cigrang, J., Talcott, G. W., … Smith Slep, A. M. (2014). Assessing posttraumatic stress in military service members: Improving efficiency and accuracy. *Psychological Assessment*, *26*, 1–7. doi:10.1037/a0034315

Forbes, D., Lockwood, E., Elhai, J. D., Creamer, M., Bryant, R., McFarlane, A., … O'Donnell, M. (2015). An evaluation of the *DSM-5* factor structure for posttraumatic stress disorder in survivors of traumatic injury. *Journal of Anxiety Disorders*, *29*, 43–51. doi:10.1016/j.janxdis.2014.11.004

Frankfurt, S. B., Armour, C., Contractor, A. A., & Elahi, J. D. (2016). Do gender and directness of trauma exposure moderate PTSD's latent factor structure? *Psychiatry Research*, *245*, 365-370. doi:10.1016/j.psychres.2016.08.049

Friedman, M. J. (2013). Finalizing PTSD in *DSM-5*: Getting here from there and where to go next. *Journal of Traumatic Stress*, *26*, 548–556. doi:10.1002/jts.21840

Friedman, M. J., Resick, P. A., Bryant, R. A., & Brewin, C. R. (2011). Considering PTSD for *DSM-5*. *Depression and Anxiety*, *28*, 750–769. doi:10.1002/da.20767

Galatzer-Levy, I. R., & Bryant, R. A. (2013). 636,120 ways to have posttraumatic stress disorder. *Perspectives on Psychological Science*, *8*, 651–662. doi:10.1177/1745691613504115

Gootzeit, J., & Markon, K. (2011). Factors of PTSD: Differential specificity and external correlates. *Clinical Psychology Review*, *31*(6), 993–1003. doi:10.1016/j.cpr.2011.06.005

Green, J., Annunziata, A., Kleiman, S., Bovin, M., Harwell, A., Fox, A., Black, S., Schnurr, P., Holowka, D., Rosen, R., Keane, T., & Marx, B. (2017). Examining the diagnostic utility of the *DSM-5* PTSD symptoms among male and female returning veterans. *Depression and Anxiety, 34*, 752-760. doi:10.1002/da.22667

Grubaugh, A. L., Long, M. E., Elhai, J. D., Frueh, B. C., & Magruder, K. M. (2010). An examination of the construct validity of posttraumatic stress disorder with veterans using a revised criterion set. *Behaviour Research and Therapy*, *48*, 909–914. doi:10.1016/j.brat.2010.05.019

Hambleton, R. K., & Jones, R. W. (1993). An NCME instructional module on. *Educational Measurement: Issues and Practice*, *12*, 38–47.

Hansen, M., Hyland, P., Armour, C., Shevlin, M., & Elklit, A. (2015). Less is more? Assessing the validity of the *ICD-11* model of PTSD across multiple trauma samples. *European Journal of Psychotraumatology*, *6*. doi:10.3402/ejpt.v6.28766

Harvey, A. G., Jones, C., & Schmidt, D. A. (2003). Sleep and posttraumatic stress disorder: A review. *Clinical Psychology Review*, *23*, 377–407.

Harvey, R. J. (2016). Improving measurement via item response theory: Great idea, but hold the Rasch. *The Counseling Psychologist*, *44,* 195–204. doi:10.1177/0011000015615427

He, Q., Glas, C. A. W., & Veldkamp, B. P. (2014). Assessing impact of differential symptom functioning on post-traumatic stress disorder (PTSD) diagnosis: Assessing impact of DIF on PTSD diagnosis. *International Journal of Methods in Psychiatric Research*, *23*, 131–141. doi:10.1002/mpr.1417

Hershner, S. D., & Chervin, R. D. (2014). Causes and consequences of sleepiness among college students. *Nature and Science of Sleep*, *6*, 73–84. doi:10.2147/NSS.S62907

Hickling, E. J., Barnett, S. D., & Gibbons, S. (2013). The many presentations of posttraumatic stress disorder: An empirical examination of theoretical possibilities. *SAGE Open*, *3*. doi:10.1177/2158244013480151

Hyland, P., Shevlin, M., McNally, S., Murphy, J., Hansen, M., & Elklit, A. (2016). Exploring differences between the *ICD-11* and *DSM-5* models of PTSD: Does it matter which model is used? *Journal of Anxiety Disorders*, *37*, 48–53. doi:10.1016/j.janxdis.2015.11.002

Jiang, S., Wang, C., & Weiss, D. J. (2016). Sample size requirements for estimation of item parameters in the multidimensional graded response model. *Frontiers in Psychology*, *7*. doi:10.3389/fpsyg.2016.00109

Keane, T. M., Caddell, J. M., & Taylor, K. L. (1988). Mississippi Scale for Combat-Related Posttraumatic Stress Disorder: Three studies in reliability and validity. *Journal of Consulting and Clinical Psychology*, *56*, 85–90. doi:10.1037//0022-006X.56.1.85

Keeley, J. W., Reed, G. M., Roberts, M. C., Evans, S. C., Robles, R., Matsumoto, C., … Maercker, A. (2016). Disorders specifically associated with stress: A case-controlled field study for *ICD-11* mental and behavioural disorders. *International Journal of Clinical and Health Psychology*, *16*, 109–127. doi:10.1016/j.ijchp.2015.09.002

Kessler, R. C., Berglund, P., Chiu, W. T., Demler, O., Heeringa, S., Hiripi, E., … Zheng, H. (2004). The US National Comorbidity Survey Replication (NCS-R): Design and field procedures. *International Journal of Methods in Psychiatric Research*, *13*, 69–92. doi:10.1002/mpr.167

Kilpatrick, D. G., Resnick, H. S., Baber, B., Guille, C., & Gros, K. (2011). The National Stressful Events Web Survey (NSES-W). Charleston, SC: Medical University of South Carolina.

Kilpatrick, D. G., Resnick, H. S., Milanak, M. E., Miller, M. W., Keyes, K. M., & Friedman, M. J. (2013). National estimates of exposure to traumatic events and PTSD prevalence using *DSM-IV* and *DSM-5* Criteria: *DSM-5* PTSD Prevalence. *Journal of Traumatic Stress*, *26*, 537–547. doi:10.1002/jts.21848

King, D. W., King, L. A., Fairbank, J. A., Schlenger, W. E., & et al. (1993). Enhancing the precision of the Mississippi Scale for Combat-Related Posttraumatic Stress Disorder: An application of item response theory. *Psychological Assessment*, *5*, 457–471. doi:10.1037/1040-3590.5.4.457

King, D. W., Leskin, G. A., King, L. A., & Weathers, F. W. (1998). Confirmatory factor analysis of the clinician-administered PTSD Scale: Evidence for the dimensionality of posttraumatic stress disorder. *Psychological Assessment*, *10*, 90–96. doi:doi:10.1037//1040-3590.10.2.90

King, M. W., Street, A. E., Gradus, J. L., Vogt, D. S., & Resick, P. A. (2013). Gender differences in posttraumatic stress symptoms among OEF/OIF veterans: An item response theory analysis. *Journal of Traumatic Stress*, *26*, 175–183. doi:10.1002/jts.21802

Lamarche, L. J., & De Koninck, J. (2007). Sleep disturbance in adults with posttraumatic stress disorder: A review. *The Journal of Clinical Psychiatry*, *68*, 1257–1270.

Li, J. J., Reise, S. P., Chronis-Tuscano, A., Mikami, A. Y., & Lee, S. S. (2016). Item response theory analysis of ADHD symptoms in children with and without ADHD. *Assessment*, *23*, 655–671. doi:10.1177/1073191115591595

Liu, P., Wang, L., Cao, C., Wang, R., Zhang, J., Zhang, B., … Elhai, J. D. (2014). The underlying dimensions of *DSM-5* posttraumatic stress disorder symptoms in an epidemiological sample of Chinese earthquake survivors. *Journal of Anxiety Disorders*, *28*, 345–351. doi:10.1016/j.janxdis.2014.03.008

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

Maercker, A., Brewin, C. R., Bryant, R. A., Cloitre, M., Reed, G. M., van Ommeren, M., … others. (2013). Proposals for mental disorders specifically associated with stress in the International Classification of Diseases-11. *The Lancet*, *381*, 1683–1685.

Masters, G. N. (1982). A Rasch model for partial credit scoring. *Psychometrika*, *47*, 149–174. doi:10.1007/BF02296272

Maydeu-Olivares, A., Drasgow, F., & Mead, A. D. (1994). Distinguishing among paranletric item response models for polychotomous ordered data. *Applied Psychological Measurement*, *18*, 245–256. doi:10.1177/014662169401800305

McNally, R. J., Robinaugh, D. J., Wu, G. W. Y., Wang, L., Deserno, M. K., & Borsboom, D. (2015). Mental disorders as causal systems: A network approach to posttraumatic stress disorder. *Clinical Psychological Science, 3*, 836-849. doi:10.1177/2167702614553230

Michael, T., Ehlers, A., Halligan, S. L., & Clark, D. M. (2005). Unwanted memories of assault: What intrusion characteristics are associated with PTSD? *Behaviour Research and Therapy*, *43*, 613–628. doi:10.1016/j.brat.2004.04.006

Miles, J. N. V., Marshall, G. N., & Schell, T. L. (2008). Spanish and English versions of the PTSD Checklist-Civilian version (PCL-C): Testing for differential item functioning. *Journal of Traumatic Stress*, *21*, 369–376. doi:10.1002/jts.20349

Miller, M. W., Wolf, E. J., Kilpatrick, D., Resnick, H., Marx, B. P., Holowka, D. W., … Friedman, M. J. (2013). The prevalence and latent structure of proposed *DSM-5* posttraumatic stress disorder symptoms in U.S. national and veteran samples. *Psychological Trauma: Theory, Research, Practice, and Policy*, *5*, 501–512. doi:10.1037/a0029730

Mitchell, K. S., Wolf, E. J., Bovin, M. J., Lee, L. O., Green, J. D., Rosen, R. C., . . . Marx, B. P. (2017). Network models of *DSM-5* posttraumatic stress disorder: Implications for *ICD-11*. *Journal of Abnormal Psychology, 126*, 355-366. doi:10.1037/abn0000252

Muraki, E., & Carlson, J. E. (1995). Full-information factor analysis for polytomous item responses. *Applied Psychological Measurement*, *19*, 73–90. doi:10.1177/014662169501900109

Muthén, L.K. and Muthén, B.O. (1998-2017). Mplus User's Guide. Eighth Edition. Los Angeles, CA: Muthén & Muthén

Nandi, A., Beard, J. R., & Galea, S. (2009). Epidemiologic heterogeneity of common mood and anxiety disorders over the lifecourse in the general population: a systematic review. *BMC Psychiatry*, *9*. doi:10.1186/1471-244X-9-31

O'Donnell, M. L., Alkemade, N., Nickerson, A., Creamer, M., McFarlane, A. C., Silove, D., … Forbes, D. (2014). Impact of the diagnostic changes to post-traumatic stress disorder for *DSM-5* and the proposed changes to *ICD-11*. *The British Journal of Psychiatry*, *205*, 230–235. doi:10.1192/bjp.bp.113.135285

Olbert, C. M., Gala, G. J., & Tupler, L. A. (2014). Quantifying heterogeneity attributable to polythetic diagnostic criteria: Theoretical framework and empirical application. *Journal of Abnormal Psychology*, *123*, 452–462. doi:10.1037/a0036068

Olino, T. M., Yu, L., Klein, D. N., Rohde, P., Seeley, J. R., Pilkonis, P. A., & Lewinsohn, P. M. (2012). Measuring depression using item response theory: An examination of three measures of depressive symptomatology. *International Journal of Methods in Psychiatric Research*, *21*, 76–85. doi:10.1002/mpr.1348

Orlando, M., & Marshall, G. N. (2002). Differential item functioning in a Spanish translation of the PTSD Checklist: Detection and evaluation of impact. *Psychological Assessment*, *14*, 50–59. doi:10.1037//1040-3590.14.1.50

Palm, K. M., Strong, D. R., & MacPherson, L. (2009). Evaluating symptom expression as a function of a posttraumatic stress disorder severity. *Journal of Anxiety Disorders*, *23*, 27–37. doi:10.1016/j.janxdis.2008.03.012

Panagioti, M., Gooding, P., Taylor, P. J., & Tarrier, N. (2013). A model of suicidal behavior in posttraumatic stress disorder (PTSD): The mediating role of defeat and entrapment. *Psychiatry Research*, *209*, 55–59. doi:10.1016/j.psychres.2013.02.018

Preston, C. C., & Colman, A. M. (2000). Optimal number of response categories in rating scales: Reliability, validity, discriminating power, and respondent preferences. *Acta Psychologica*, *104*, 1–15. doi:10.1016/S0001-6918(99)00050-5

Raykov, T. (2016). *Item Response Theory. Course Book.* Philadelphia, PA: Statistical Horizons.

Reckase, M. (2009). *Multidimensional item response theory*. Dordrecht ; New York: Springer.

Reckase, M. D. (1979). Unifactor latent trait models applied to multifactor tests: Results and implications. *Journal of Educational Statistics*, *4*, 207. doi:10.2307/1164671

Reed, G. M. (2010). Toward *ICD-11*: Improving the clinical utility of WHO’s International Classification of mental disorders. *Professional Psychology: Research and Practice*, *41*, 457–464. doi:10.1037/a0021701

Reeve, B. B. (2003). Item response theory modeling in health outcomes measurement. *Expert Review of Pharmacoeconomics & Outcomes Research*, *3*, 131–145. doi:doi:10.1586/14737167.3.2.131

Reeve, B. B., & Fayers, P. (2005). Applying item response theory modeling for evaluating questionnaire item and scale properties. In P. Fayers & R. D. Hays (Eds.), *Assessing quality of life in clinical trials: Methods of practice* (2nd ed., pp. 55–73). New York: Oxford University Press.

Reise, S. P., Ainsworth, A. T., & Haviland, M. G. (2005). Item response theory fundamentals, applications, and promise in psychological research. *Current Directions in Psychological Science*, *14*, 95–101. doi:doi:10.1111/j.0963-7214.2005.00342.x

Reise, S. P., & Haviland, M. G. (2005). Item response theory and the measurement of clinical change. *Journal of Personality Assessment*, *84*, 228–238.

Reise, S. P., & Waller, N. G. (2009). Item response theory and clinical measurement. *Annual Review of Clinical Psychology*, *5*, 27–48. doi:10.1146/annurev.clinpsy.032408.153553

Resick, P. A., & Miller, M. W. (2009). Posttraumatic stress disorder: Anxiety or traumatic stress disorder? *Journal of Traumatic Stress*, *22*, 384–390. doi:10.1002/jts.20437

Reynolds, M., & Brewin, C. R. (1998). Intrusive cognitions, coping strategies and emotional responses in depression, post-traumatic stress disorder and a non-clinical population. *Behaviour Research and Therapy*, *36*, 135–147. doi:10.1016/S0005-7967(98)00013-8

Samejima, F. (1969). Estimation of latent ability using a response pattern of graded scores. *Psychometrika*, *34*, 1–97. doi:10.1007/BF03372160

Sheikh, J. I., Woodward, S. H., & Leskin, G. A. (2003). Sleep in post-traumatic stress disorder and panic: Convergence and divergence. *Depression and Anxiety*, *18*, 187–197. doi:10.1002/da.10066

Sijtsma, K., Emons, W. H. M., Bouwmeester, S., Nyklíček, I., & Roorda, L. D. (2008). Nonparametric IRT analysis of Quality-of-Life Scales and its application to the World Health Organization Quality-of-Life Scale (WHOQOL-Bref). *Quality of Life Research*, *17*, 275–290. doi:10.1007/s11136-007-9281-6

Simms, L. J., Watson, D., & Doebbelling, B. N. (2002). Confirmatory factor analyses of posttraumatic stress symptoms in deployed and nondeployed veterans of the Gulf War. *Journal of Abnormal Psychology*, *111*, 637–647. doi:10.1037//0021-843X.111.4.637

Spearman, C. (1907). Demonstration of formulae for true measurement of correlation. *The American Journal of Psychology*, *18*, 161–169. doi:10.2307/1412408

Spitzer, R. L., First, M. B., & Wakefield, J. C. (2007). Saving PTSD from itself in *DSM-V*. *Journal of Anxiety Disorders*, *21*, 233–241. doi:10.1016/j.janxdis.2006.09.006

Stammel, N., Abbing, E. M., Heeke, C., & Knaevelsrud, C. (2015). Applicability of the *ICD-11* proposal for PTSD: A comparison of prevalence and comorbidity rates with the *DSM-IV* PTSD classification in two post-conflict samples. *European Journal of Psychotraumatology*, *6*. doi:10.3402/ejpt.v6.27070

Stein, D. J., McLaughlin, K. A., Koenen, K. C., Atwoli, L., Friedman, M. J., Hill, E. D., … Kessler, R. C. (2014). *DSM-5* and *ICD-11* definition of posttraumatic stress disorder: Investigating "narrow" and "broad" approaches. *Depression and Anxiety*, *31*, 494–505. doi:10.1002/da.22279

Tay, A. K., Rees, S., Chen, J., Kareth, M., & Silove, D. (2015). The structure of post-traumatic stress disorder and complex post-traumatic stress disorder amongst West Papuan refugees. *BMC Psychiatry*, *15*. doi:10.1186/s12888-015-0480-3

Templin, J. (2015, June). *IRT models for polytomous response data*. Presented at the Interuniversity Consortium for Political and Social Research. Retrieved from http://jtemplin.coe.uga.edu/files/irt/irt11icpsr/irt11icpsr_lecture04.pdf

Thomsen, C. J., Stander, V. A., McWhorter, S. K., Rabenhorst, M. M., & Milner, J. S. (2011). Effects of combat deployment on risky and self-destructive behavior among active duty military personnel. *Journal of Psychiatric Research*, *45*, 1321–1331. doi:10.1016/j.jpsychires.2011.04.003

Tsai, J., Harpaz-Rotem, I., Armour, C., Southwick, S. M., Krystal, J., & Pietrzak, R. H. (2015). Dimensional structure of *DSM-5* posttraumatic stress disorder symptoms: Results from the National Health and Resilience in Veterans study. *Journal of Clinical Psychiatry*, *76*, 546–553.

Tyek Han, C. (2013). Item response models used within WinGen. University of Massachusetts Amherst: Research and Evaluation Program Methods. Retrieved from https://www.umass.edu/remp/software/simcata/wingen/modelsF.html

Walker, C. (2011). What's the DIF? Why differential item functioning analyses are an important part of instrument development and validation. *Journal of Psychoeducational Assessment, 29*, 364-376. doi:10.1177/0734282911406666

Weathers, F. W., Keane, T. M., King, L. A., & King, D. W. (1997). Psychometric theory in the development of posttraumatic stress disorder assessment tools. In J. P. Wilson & T. M.

Keane (Eds.), Assessing psychological trauma and PTSD (pp. 98-135). New York: Guilford Press.

Weathers, F., Litz, B., Herman, D., Huska, J., & Keane, T. (October 1993). *The PTSD Checklist (PCL): Reliability, validity, and diagnostic utility*. Paper presented at the Annual Convention of the International Society for Traumatic Stress Studies, San Antonio, TX.

Weathers, F. W., Marx, B. P., Friedman, M. J., & Schnurr, P. P. (2014). Posttraumatic stress disorder in *DSM-5*: New criteria, new measures, and implications for assessment. *Psychological Injury and Law*, *7*, 93–107. doi:10.1007/s12207-014-9191-1

Witte, T. K., Domino, J. L., & Weathers, F. W. (2015). Item order effects in the evaluation of posttraumatic stress disorder symptom structure. *Psychological Assessment*, *27*, 852–864. doi:10.1037/pas0000089

Young, G., Lareau, C., & Pierre, B. (2014). One quintillion ways to have PTSD comorbidity: Recommendations for the disordered *DSM-5*. *Psychological Injury and Law*, *7*, 61–74. doi:10.1007/s12207-014-9186-y

Zickar, M. J., & Broadfoot, A. A. (2010). The partial revival of a dead horse? Comparing classical test theory and item response theory. In C. E. Lance & R. J. Vandenberg (Eds.), *Statistical and methodological myths and urban legends: Doctrine, verity and fable in organizational and social sciences* (pp. 37-59). New York: Routledge.

Zlotnick, C., Shea, T. M., Rosen, K., Simpson, E., Mulrenin, K., Begin, A., & Pearlstein, T. (1997). An affect-management group for women with posttraumatic stress disorder and histories of childhood sexual abuse. *Journal of Traumatic Stress*, *10*, 425–436. doi:10.1002/jts.2490100308

Zoellner, L. A., Bedard-Gilligan, M. A., Jun, J. J., Marks, L. H., & Garcia, N. M. (2013). The evolving construct of posttraumatic stress disorder (PTSD): *DSM-5* criteria changes and legal implications. *Psychological Injury and Law*, *6*, 277–289. doi:10.1007/s12207-013-9175-6

Table 1
Symptom Mappings of the DSM-5 PTSD Symptoms

DSM-5 Symptoms	*DSM-5*	One-Factor	*ICD-11*
B1: intrusive memories	R	PTSD	-
B2: nightmares	R	PTSD	R
B3: flashbacks	R	PTSD	R
B4: cued distress	R	PTSD	-
B5: cued physical reactions	R	PTSD	-
C1: internal avoidance	AV	PTSD	AV
C2: external avoidance	AV	PTSD	AV
D1: traumatic amnesia	NACM	PTSD	-
D2: negative beliefs	NACM	PTSD	-
D3: distorted blame	NACM	PTSD	-
D4: persistent negative emotional state	NACM	PTSD	-
D5: diminished interest in activities	NACM	PTSD	-
D6: feelings of detachment from others	NACM	PTSD	-
D7: inability to experience positive emotions	NACM	PTSD	-
E1: irritability or anger	AAR	PTSD	-
E2: reckless or self-destructive behavior	AAR	PTSD	-
E3: hypervigilance	AAR	PTSD	ST
E4: exaggerated startle response	AAR	PTSD	ST
E5: difficulty concentrating	AAR	PTSD	-
E6: sleeping difficulties	AAR	PTSD	-

Note. *DSM-5* = Diagnostic and Statistical Manual of Mental Disorders, 5th Edition (APA, 2013); PTSD = posttraumatic stress disorder; R = reexperiencing; AV = avoidance; NACM = negative alterations in cognitions and mood; AAR = alterations in arousal and reactivity; ST = sense of threat.

Table 2

Descriptive Statistics of the PCL-5

Scale	*M* (*SD*)	Median	Range	Skewness	Kurtosis
Total Score	8.94 (13.59)	2.00	0 – 77	2.02	3.93
Rex	2.23 (3.82)	0.00	0 – 20	2.21	4.70
Avoid	1.24 (2.03)	0.00	0 – 8	1.74	2.22
NACM	3.32 (5.55)	0.00	0 – 28	2.00	3.37
AAR	2.80 (4.60)	0.00	0 – 24	1.95	3.37

Note. PCL-5 = Posttraumatic Stress Disorder Checklist – 5; Rex = reexperiencing; Avoid = avoidance, NACM = negative alterations in cognition and mood; AAR = alterations in arousal and reactivity.

Table 3

Mantel-Haenszel Test Results of the PCL-5 across Gender

				95% Confidence Interval	
Item	χ^2	*p*	Odds Ratio	Lower Bound	Upper Bound
1.1	3.447	0.063	0.693	0.470	1.022
1.2	0.025	0.874	0.939	0.600	1.470
1.3	0.852	0.356	1.501	0.721	3.125
1.4	3.335	0.068	3.23	0.979	10.65
1.5	2.439	0.118	-	-	-
2.1	5.387	0.020*	0.543	0.33	0.893
2.2	0.585	0.444	1.336	0.716	0.362
2.3	0.852	0.356	1.730	0.662	4.521
2.4	2.007	0.157	3.224	0.752	13.816
2.5	0.770	0.380	3.322	0.429	25.736
3.1	1.524	0.217	0.707	0.427	1.170
3.2	0.66	0.417	1.392	0.711	2.725
3.3	0.039	0.843	1.008	0.430	2.366
3.4	0.777	0.378	2.344	0.536	10.251
3.5	0.000	0.996	1.635	0.195	13.673
4.1	4.410	0.036*	0.661	0.456	0.959
4.2	0.276	0.599	1.168	0.727	1.879
4.3	0.605	0.437	1.304	0.738	2.306
4.4	1.858	0.173	2.094	0.810	5.415
4.5	0.554	0.457	1.854	0.544	6.318
5.1	3.080	0.079	0.659	0.424	1.025
5.2	1.136	0.286	1.445	0.792	2.633
5.3	0.016	0.898	0.983	0.493	1.961
5.4	2.188	0.139	3.340	0.781	14.282
5.5	0.196	0.658	1.776	0.397	7.949
6.1	6.841	0.009*	0.592	0.404	0.867
6.2	3.112	0.078	1.644	0.976	2.770
6.3	2.175	0.140	1.763	0.883	3.520
6.4	0.004	0.947	0.961	0.481	1.921
6.5	0.368	0.544	1.585	0.540	4.649
7.1	5.738	0.017*	0.592	0.391	0.897
7.2	1.621	0.203	1.496	0.850	2.631
7.3	1.600	0.206	1.807	0.799	4.086
7.4	0.326	0.568	1.381	0.602	3.169
7.5	0.175	0.676	1.442	0.488	4.262

8.1	2.258	0.133	0.716	0.475	1.079
8.2	0.129	0.720	1.143	0.670	1.949
8.3	1.378	0.241	1.637	0.790	3.396
8.4	1.093	0.296	1.962	0.678	5.674
8.5	0.03	0.862	0.749	0.235	2.382
9.1	2.886	0.089	0.671	0.433	1.037
9.2	3.052	0.081	1.988	0.966	4.090
9.3	0.00	0.988	1.069	0.503	2.272
9.4	1.312	0.252	0.606	0.292	1.259
9.5	4.754	0.029*	7.993	1.079	59.188
10.1	1.264	0.261	0.776	0.516	1.166
10.2	1.76	0.185	1.582	0.853	2.934
10.3	0.005	0.942	0.914	0.455	1.833
10.4	0.036	0.850	0.867	0.417	1.803
10.5	1.019	0.313	2.138	0.634	7.209
11.1	3.812	0.051	0.659	0.441	0.984
11.2	2.245	0.134	1.612	0.904	2.876
11.3	0.001	0.980	1.071	0.540	2.127
11.4	0.075	0.785	1.205	0.548	2.650
11.5	0.915	0.339	1.88	0.648	5.451
12.1	0.002	0.967	0.983	0.626	1.545
12.2	1.336	0.248	0.665	0.362	1.220
12.3	0.193	0.660	1.345	0.549	3.296
12.4	0.650	0.420	1.749	0.600	5.097
12.5	0.001	0.975	1.195	0.337	4.244
13.1	0.000	0.989	1.019	0.681	1.522
13.2	0.218	0.640	0.847	0.496	1.446
13.3	0.067	0.797	1.244	0.505	3.064
13.4	0.028	0.867	1.195	0.484	2.954
13.5	0.010	0.920	0.957	0.428	2.140
14.1	0.407	0.524	1.170	0.774	1.768
14.2	0.082	0.775	1.146	0.622	2.110
14.3	4.690	0.030*	0.446	0.225	0.884
14.4	0.026	0.872	1.192	0.482	2.946
14.5	0.057	0.811	1.025	0.336	3.131
15.1	0.038	0.844	0.933	0.598	1.454
15.2	1.588	0.208	0.682	0.399	1.165
15.3	1.620	0.203	2.014	0.778	5.216
15.4	0.610	0.435	2.203	0.501	9.678
15.5	0.002	0.965	1.363	0.296	6.282
16.1	0.024	0.878	0.932	0.569	1.528

16.2	0.007	0.932	1.025	0.542	1.938
16.3	0.016	0.900	1.201	0.448	3.222
16.4	0.099	0.753	1.459	0.420	5.070
16.5	0.224	0.636	0.404	0.067	2.439
17.1	0.810	0.368	0.821	0.557	1.210
17.2	0.014	0.907	1.074	0.620	1.859
17.3	0.104	0.747	1.185	0.600	2.339
17.4	0.042	0.838	1.174	0.533	2.587
17.5	0.265	0.606	1.389	0.568	3.395
18.1	1.384	0.239	0.761	0.500	1.157
18.2	1.822	0.177	1.655	0.851	3.216
18.3	1.052	0.305	0.694	0.379	1.269
18.4	1.237	0.266	2.024	0.701	5.843
18.5	0.407	0.524	1.749	0.511	5.984
19.1	1.796	0.180	0.748	0.503	1.112
19.2	0.456	0.500	1.258	0.721	2.193
19.3	0.062	0.804	1.147	0.596	2.206
19.4	0.011	0.915	1.164	0.471	2.876
19.5	0.631	0.427	1.633	0.623	4.277
20.1	1.100	0.294	0.796	0.538	1.177
20.2	1.136	0.286	1.445	0.792	2.633
20.3	0.095	0.757	0.859	0.459	1.608
20.4	0.01	0.920	0.904	0.451	1.815
20.5	2.136	0.144	2.697	0.811	8.968

Note. PCL-5 = Posttraumatic Stress Disorder Checklist – 5; χ^2 = chi-square test statistic; .1 = response option 0; .2 = response option 1; .3 = response option 2; .4 = response option 3; .5 = response option 4. No males endorsed response option 4 on item 1, so no M-H odds ratio was generated.

*p < .05.

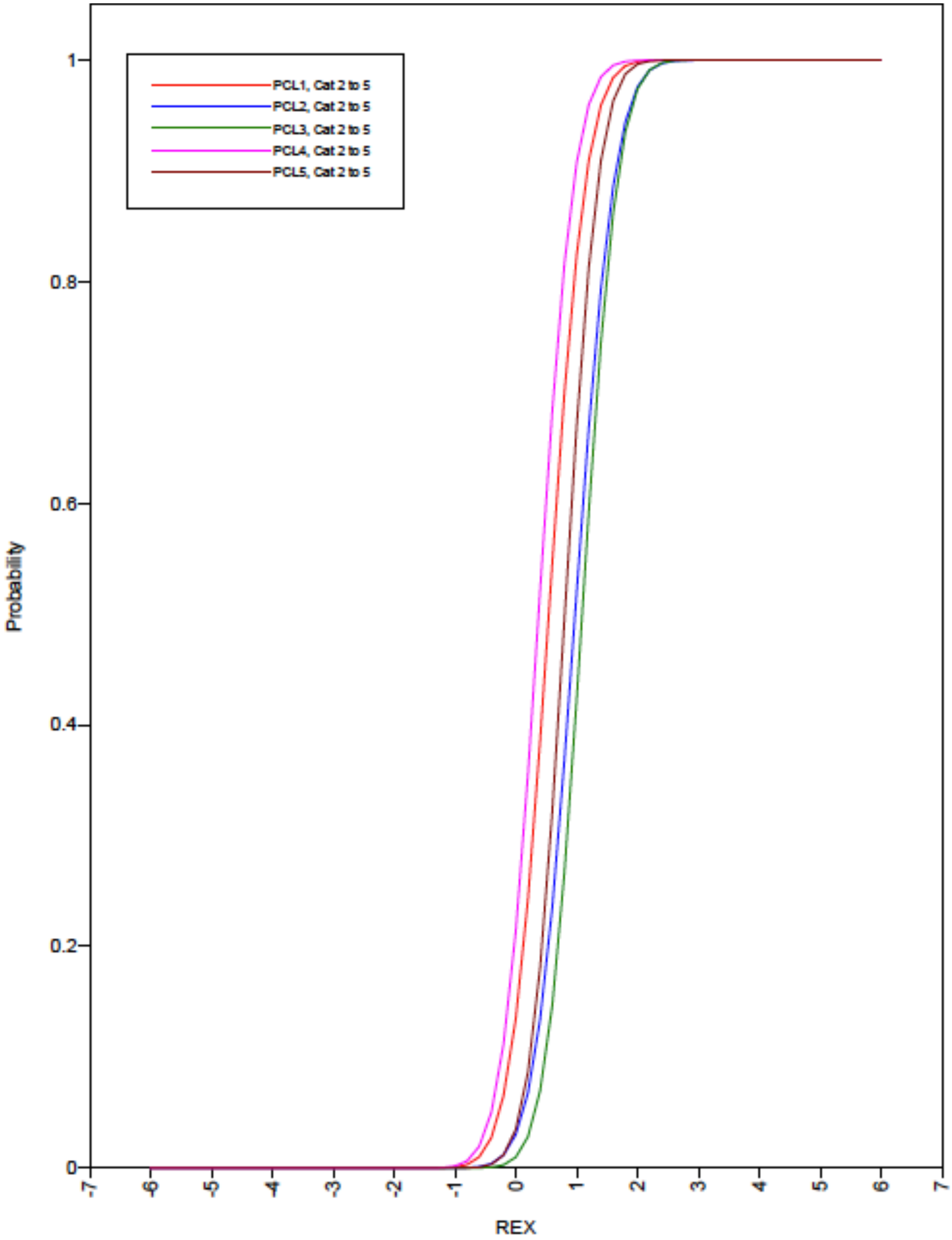
Table 4

Discrimination and Difficulty Parameters for the DSM-5 Model

Item	*a*	b_1	b_2	b_3	b_4
REX by					
PCL1	0.897	0.487	1.087	1.517	2.124
PCL2	0.888	0.865	1.308	1.654	2.125
PCL3	0.906	0.988	1.431	1.83	2.364
PCL4	0.904	0.339	0.840	1.378	1.884
PCL5	0.914	0.736	1.178	1.615	2.065
AVOID by					
PCL6	0.949	0.366	0.868	1.284	1.813
PCL7	0.935	0.569	1.020	1.362	1.847
NACM by					
PCL8	0.667	0.627	1.082	1.548	2.066
PCL9	0.884	0.725	1.076	1.378	1.780
PCL10	0.853	0.64	1.034	1.351	1.828
PCL11	0.919	0.537	0.970	1.305	1.748
PCL12	0.927	0.929	1.260	1.569	2.038
PCL13	0.942	0.713	1.123	1.353	1.679
PCL14	0.944	0.810	1.197	1.504	1.966
AAR by					
PCL15	0.868	0.878	1.330	1.765	2.157
PCL16	0.850	1.054	1.507	1.866	2.486
PCL17	0.855	0.554	0.931	1.276	1.679
PCL18	0.867	0.690	1.048	1.463	1.901
PCL19	0.921	0.547	0.959	1.359	1.691
PCL20	0.893	0.548	0.914	1.260	1.733

Note. DSM-5 = Diagnostic and Statistical Manual of Mental Disorders, 5th Edition (APA, 2013); REX = reexperiencing; AVOID = avoidance, NACM = negative alterations in cognition and mood; AAR = alterations in arousal and reactivity; PCL = Posttraumatic Stress Disorder Checklist – 5; *a* = IRT discrimination parameter; $b_{1\text{-}4}$ = IRT difficulty parameters for thresholds 1-4.

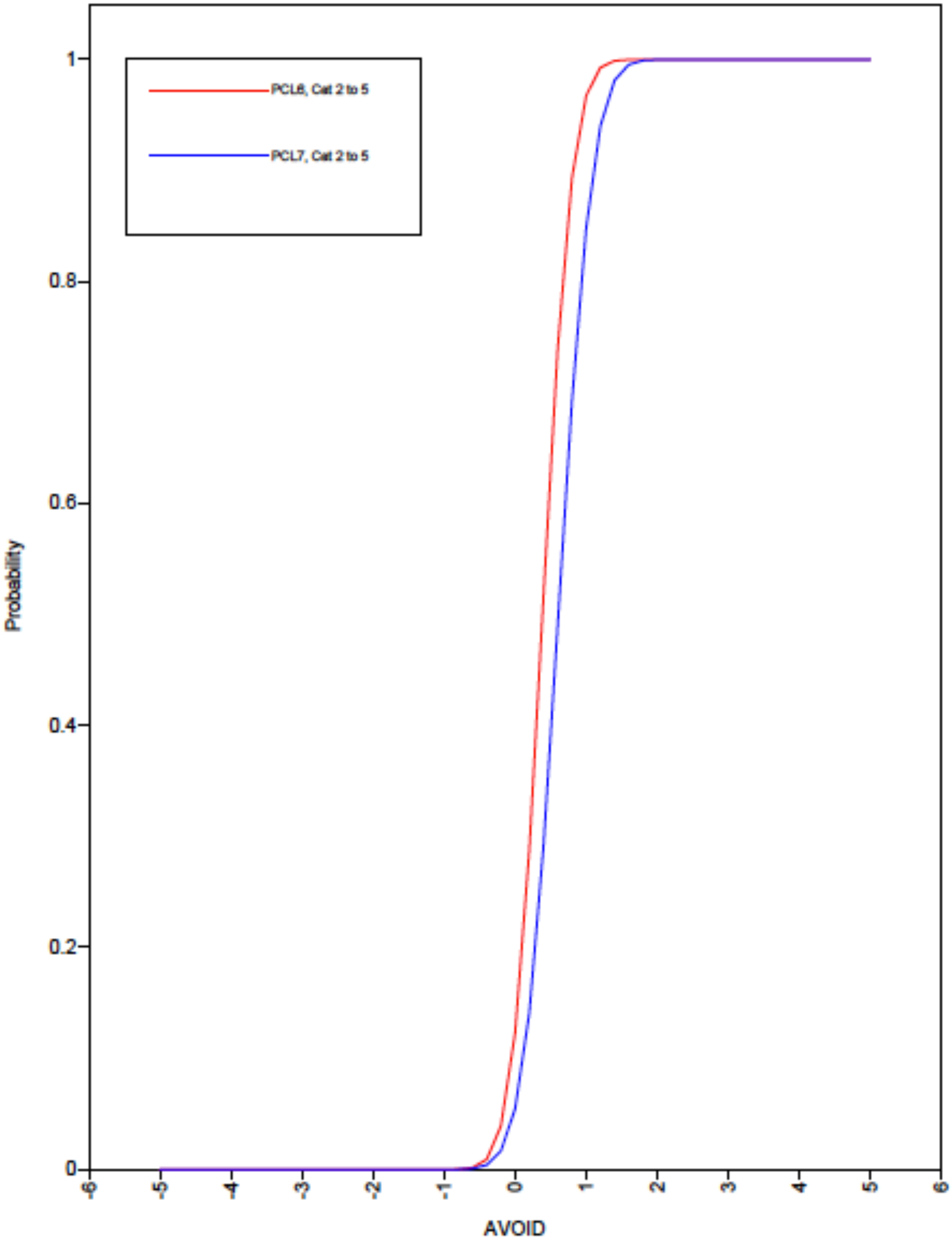
Figure 1
Reexperiencing Category Response Curve



Note. PCL = Posttraumatic Stress Disorder Checklist – 5 (Weathers et al., 2013), Cat = category, REX = reexperiencing.

Figure 2

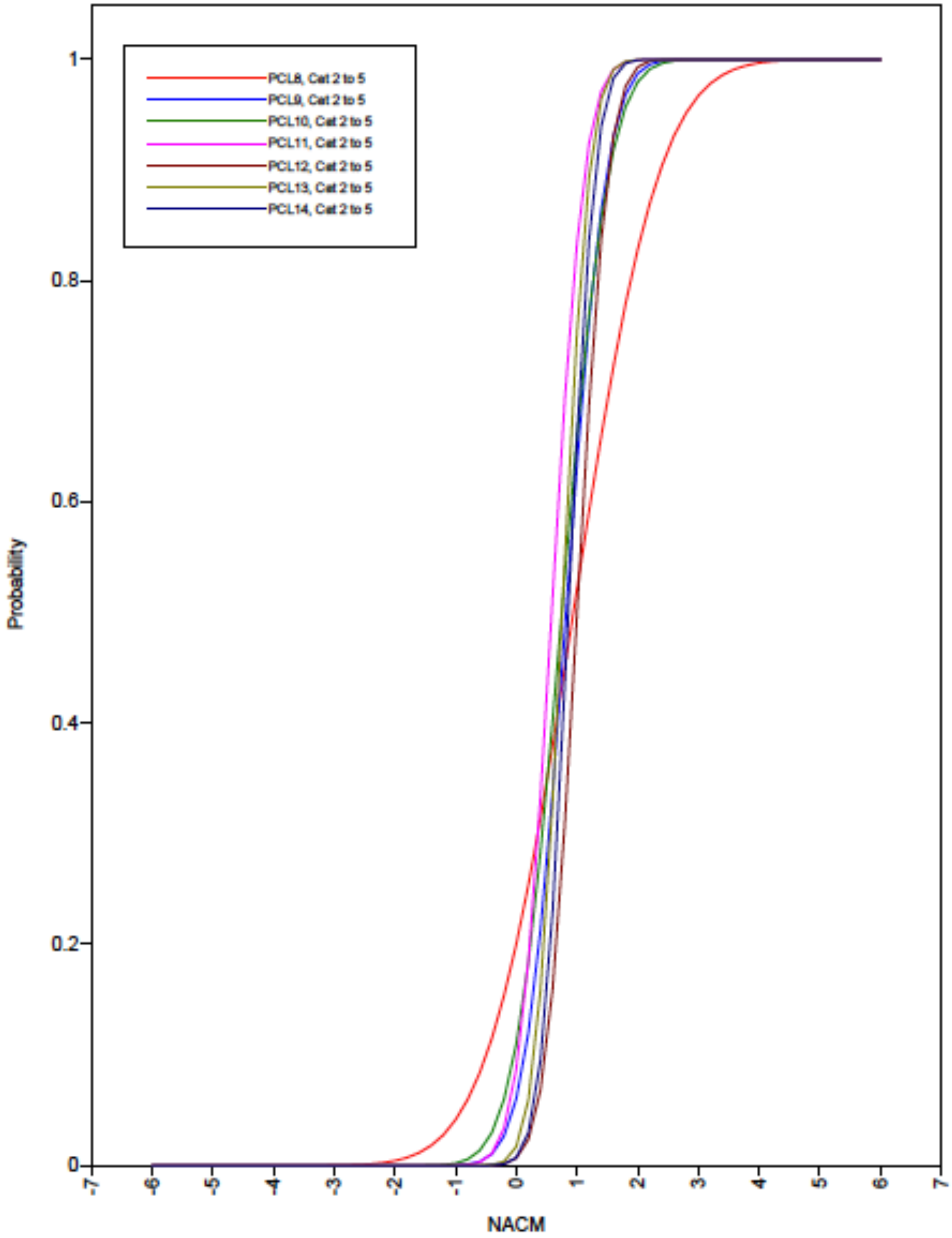
Avoidance Category Response Curve



Note. PCL = Posttraumatic Stress Disorder Checklist – 5 (Weathers et al., 2013), Cat = category, AVOID = avoidance.

Figure 3

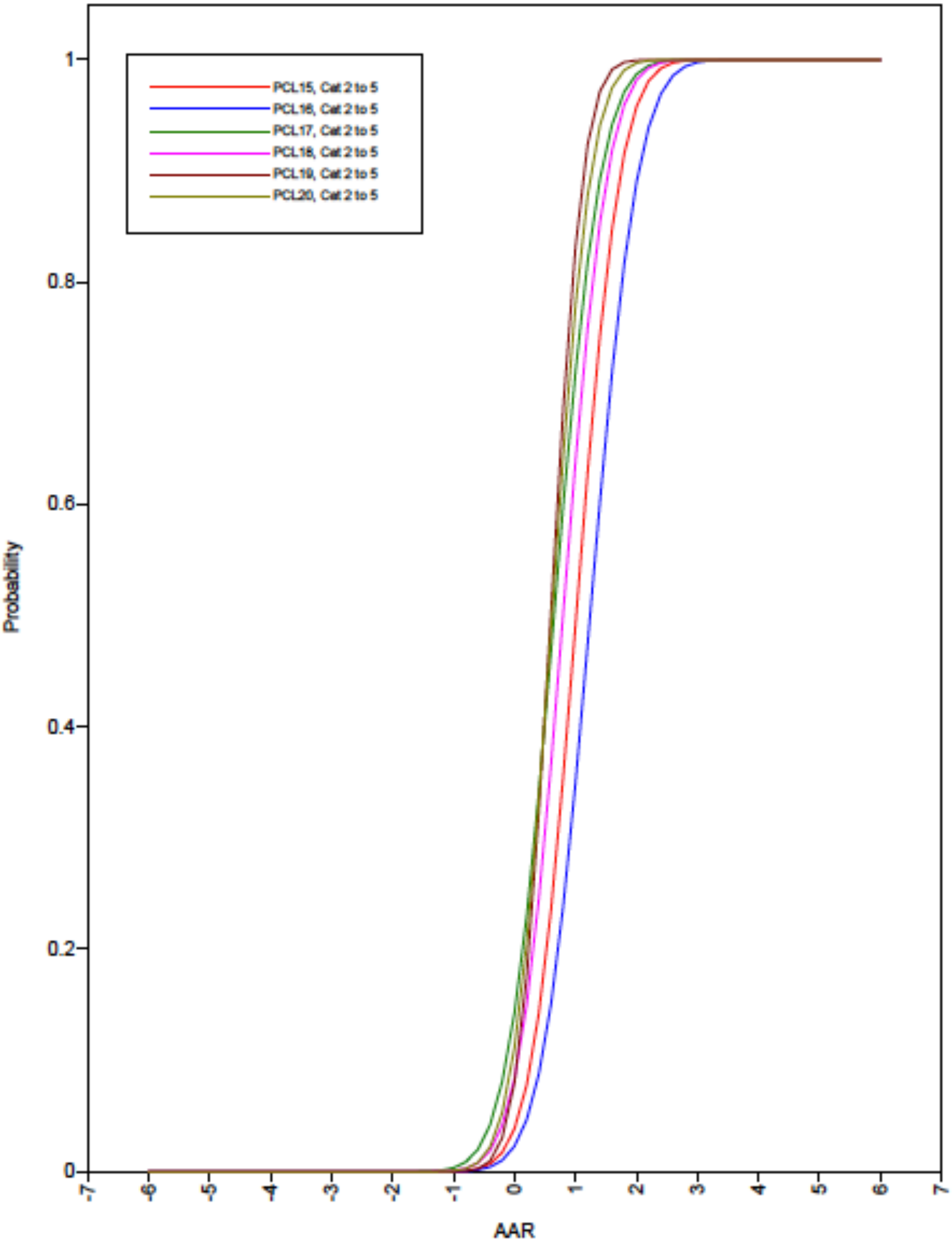
NACM Category Response Curve



Note. PCL = Posttraumatic Stress Disorder Checklist – 5 (Weathers et al., 2013), Cat = category, NACM = negative alterations in cognition and mood.

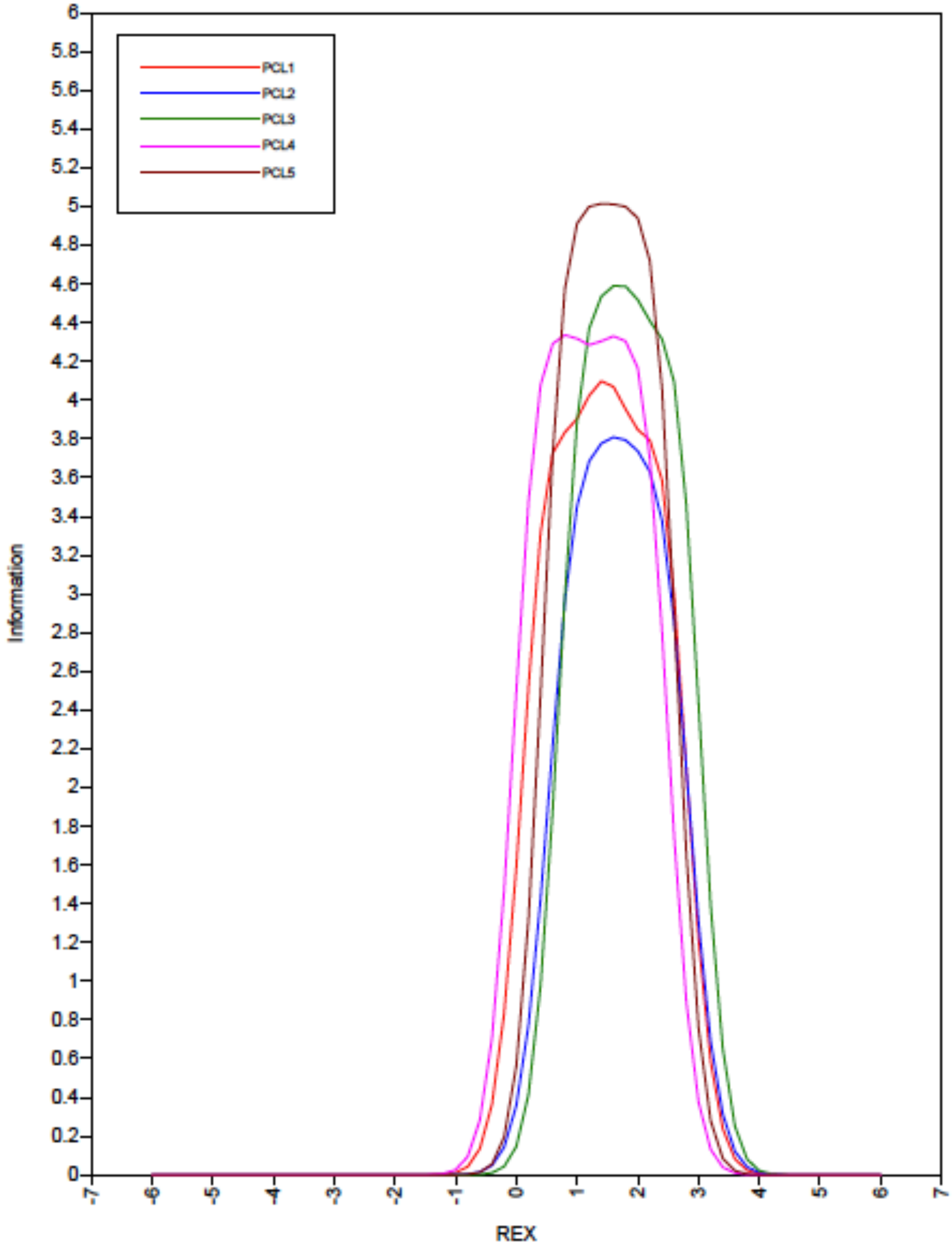
Figure 4

AAR Category Response Curve



Note. PCL = Posttraumatic Stress Disorder Checklist – 5 (Weathers et al., 2013), Cat = category, AAR = alterations in arousal and reactivity.

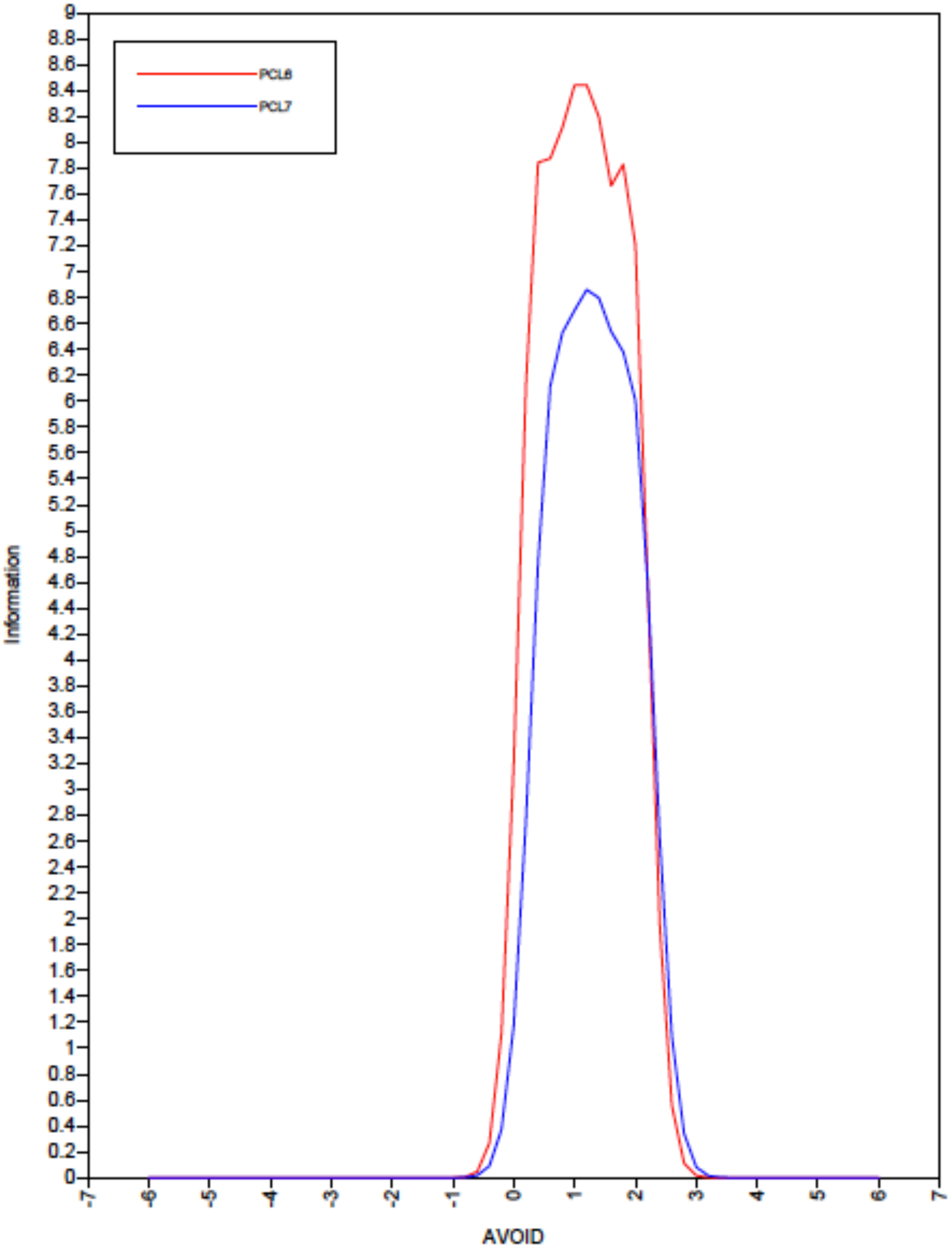
Figure 5
Reexperiencing Information Function



Note. PCL = Posttraumatic Stress Disorder Checklist – 5 (Weathers et al., 2013), REX = reexperiencing.

Figure 6

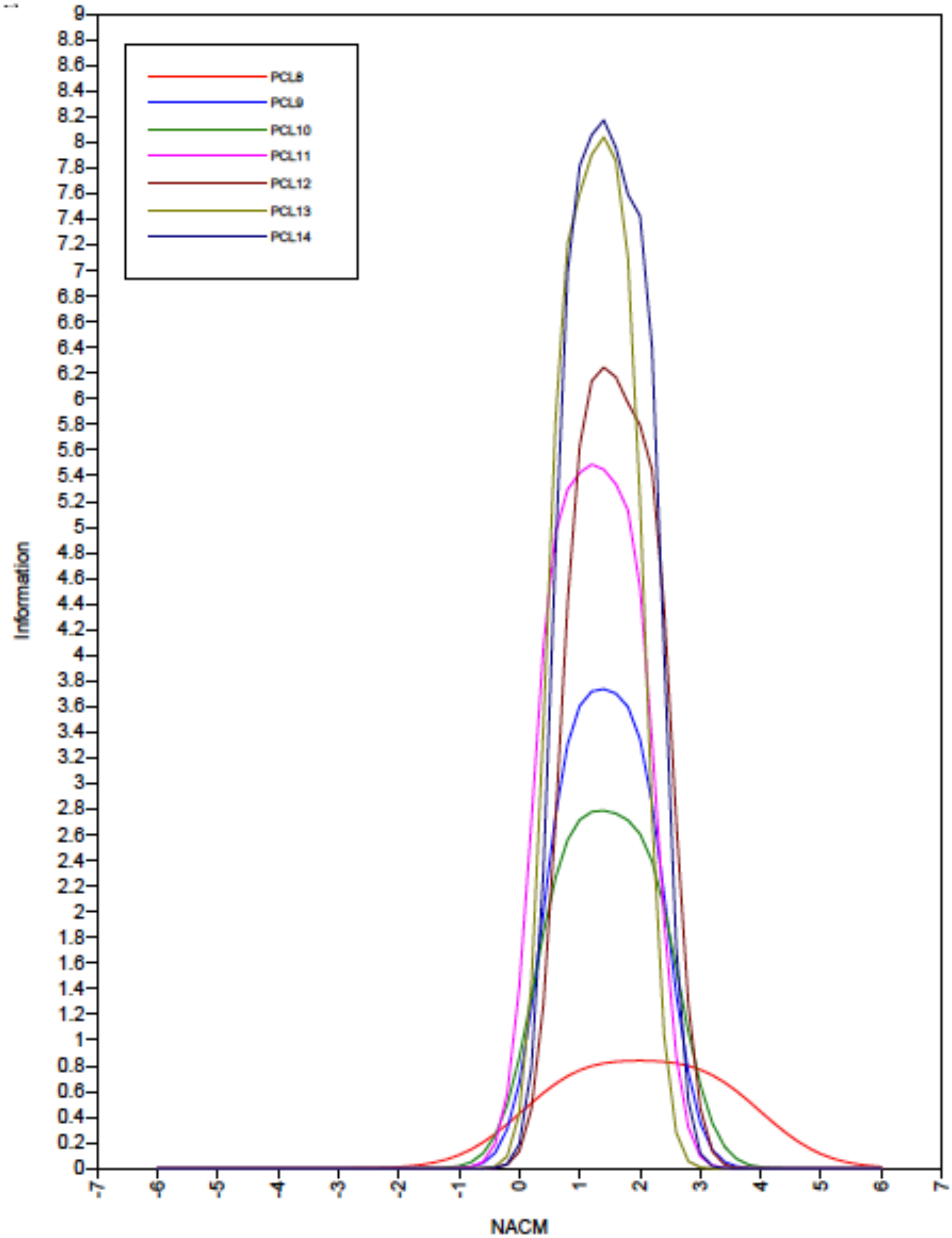
Avoidance Information Function



Note. PCL = Posttraumatic Stress Disorder Checklist – 5 (Weathers et al., 2013), AVOID = avoidance.

Figure 7

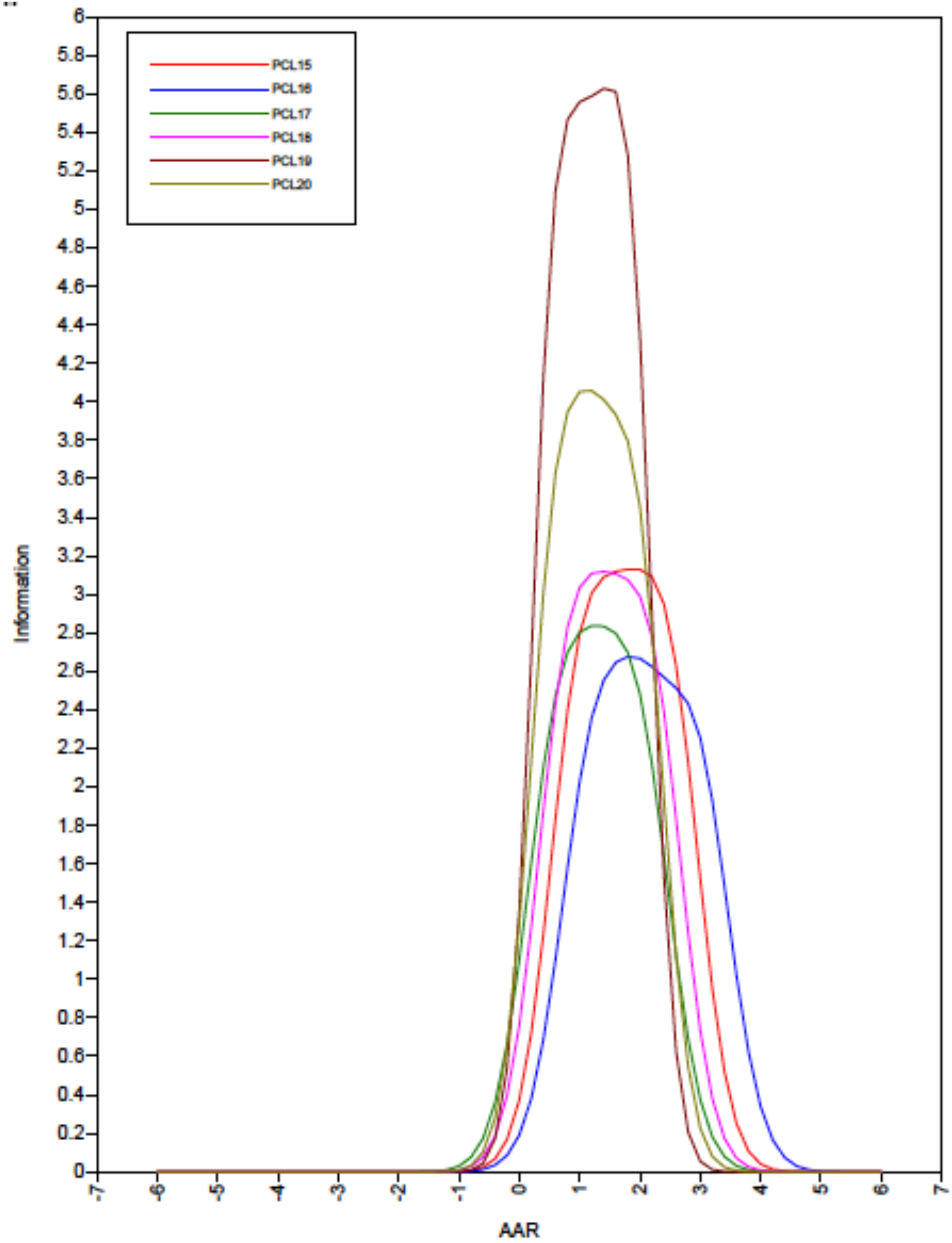
NACM Information Function



Note. PCL = Posttraumatic Stress Disorder Checklist – 5 (Weathers et al., 2013), NACM = negative alterations in cognition and mood.

Figure 8

AAR Information Function



Note. PCL = Posttraumatic Stress Disorder Checklist – 5 (Weathers et al., 2013), AAR = alterations in arousal and reactivity.

Supplemental Table 1

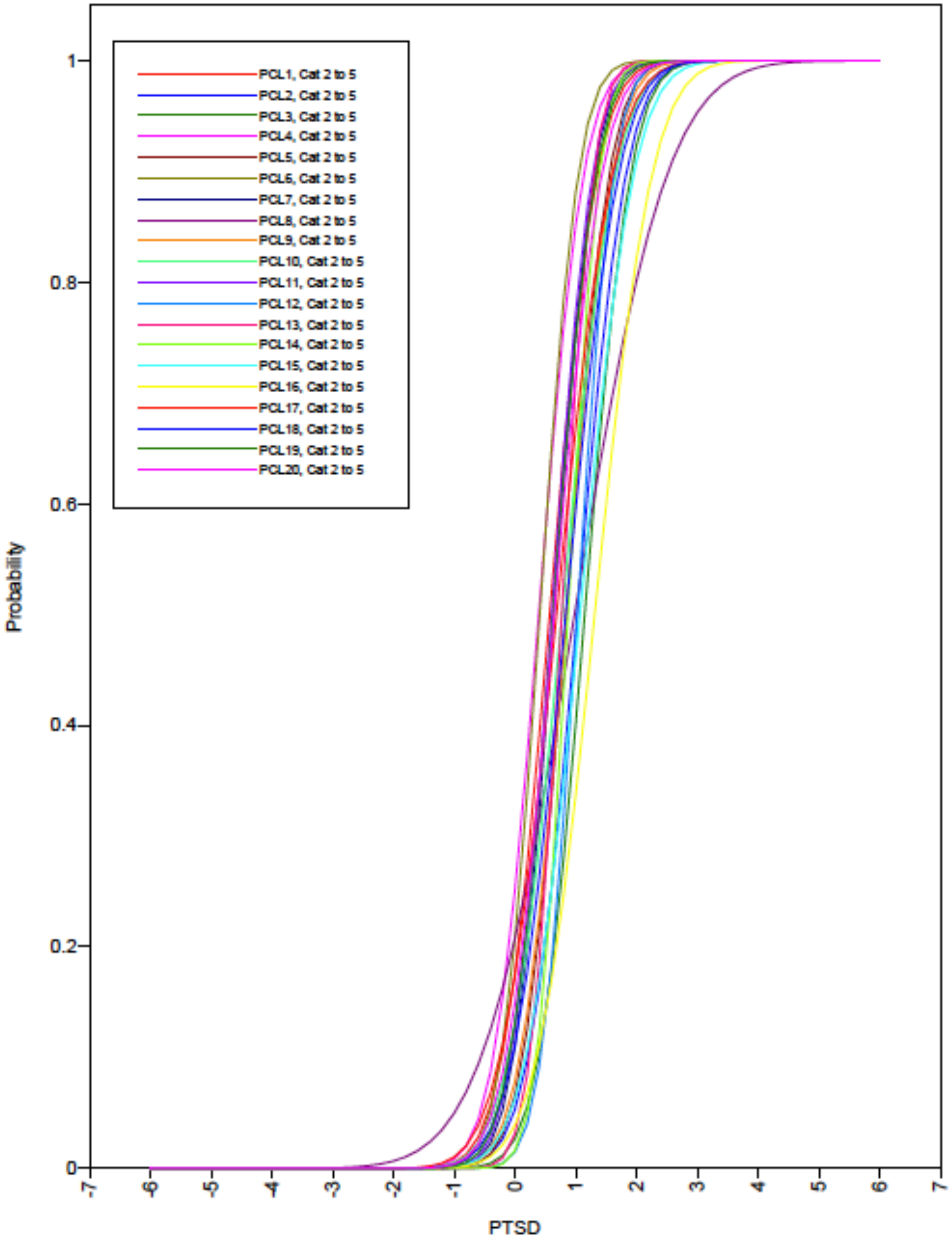
Discrimination and Difficulty Parameters for the One-factor Model

PTSD by	*a*	b_1	b_2	b_3	b_4
PCL1	0.852	0.487	1.087	1.517	2.124
PCL2	0.845	0.865	1.308	1.654	2.125
PCL3	0.860	0.988	1.431	1.830	2.364
PCL4	0.864	0.339	0.840	1.378	1.884
PCL5	0.873	0.736	1.178	1.615	2.065
PCL6	0.894	0.366	0.868	1.284	1.813
PCL7	0.887	0.569	1.020	1.362	1.847
PCL8	0.639	0.627	1.082	1.548	2.066
PCL9	0.860	0.725	1.076	1.378	1.78
PCL10	0.823	0.640	1.034	1.351	1.828
PCL11	0.885	0.537	0.970	1.305	1.748
PCL12	0.905	0.929	1.260	1.569	2.038
PCL13	0.924	0.713	1.123	1.353	1.679
PCL14	0.927	0.810	1.197	1.504	1.966
PCL15	0.821	0.878	1.330	1.765	2.157
PCL16	0.802	1.054	1.507	1.866	2.486
PCL17	0.809	0.554	0.931	1.276	1.679
PCL18	0.825	0.690	1.048	1.463	1.901
PCL19	0.874	0.547	0.959	1.359	1.691
PCL20	0.850	0.548	0.914	1.260	1.733

Note. PTSD = posttraumatic stress disorder; PCL = Posttraumatic Stress Disorder Checklist – 5; *a* = IRT discrimination parameter; $b_{1\text{-}4}$ = IRT difficulty parameters for thresholds 1-4.

Supplemental Figure 1

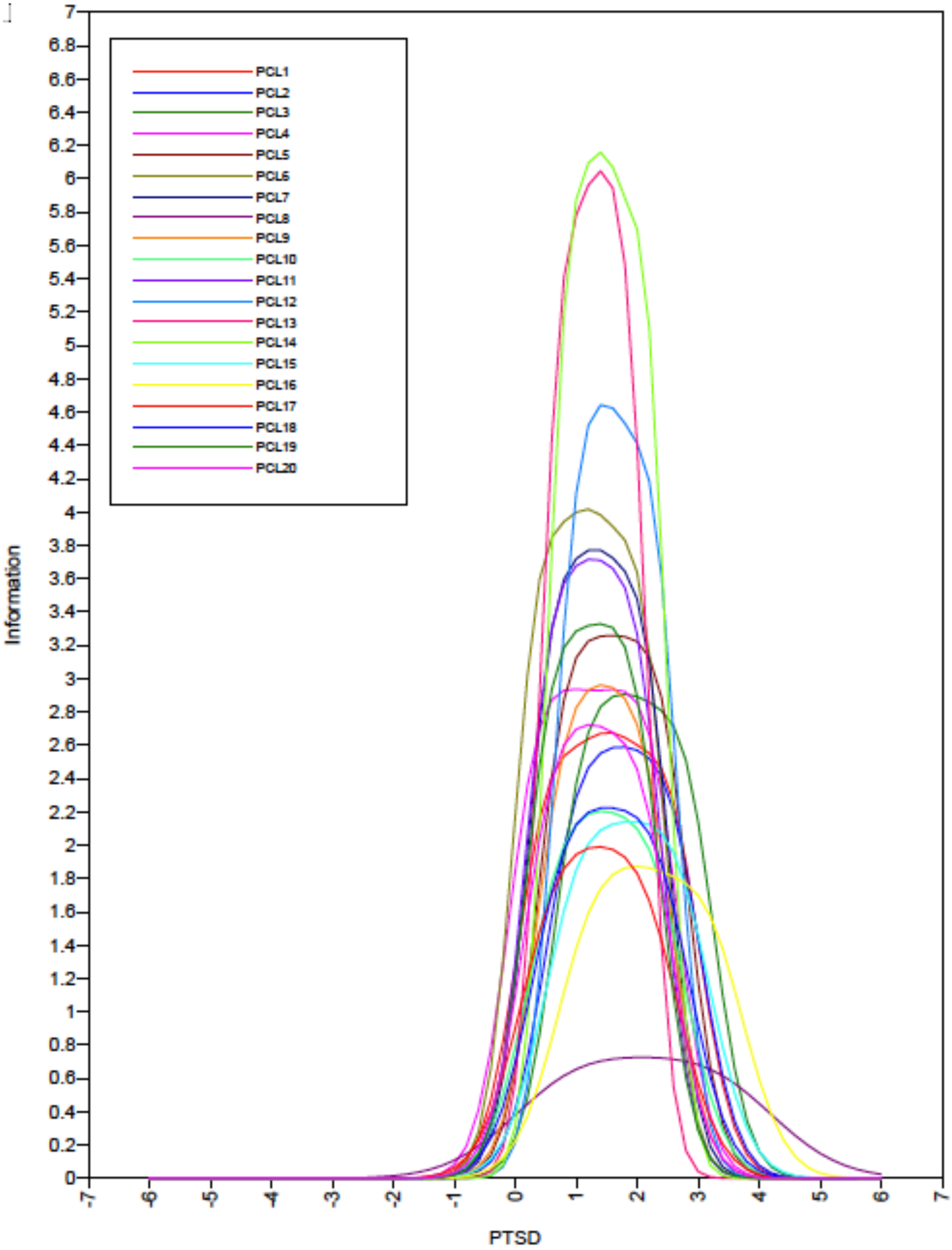
PTSD Category Response Curve



Note. PCL = Posttraumatic Stress Disorder Checklist – 5 (Weathers et al., 2013), Cat = category, PTSD = posttraumatic stress disorder.

Supplemental Figure 2

PTSD Information Function



Note. PCL = Posttraumatic Stress Disorder Checklist – 5 (Weathers et al., 2013), PTSD = posttraumatic stress disorder.