

St. John's University

**St. John's Scholar**

---

Theses and Dissertations

---

2021

**MULTILEVEL ITEM RESPONSE THEORY ANALYSIS OF THE ANGER  
REGULATION AND EXPRESSION SCALE: A CROSS-CULTURAL  
STUDY OF ANGER IN YOUTH**

Tovah Weinrib

Follow this and additional works at: [https://scholar.stjohns.edu/theses\\_dissertations](https://scholar.stjohns.edu/theses_dissertations)

---

MULTILEVEL ITEM RESPONSE THEORY ANALYSIS OF THE ANGER  
REGULATION AND EXPRESSION SCALE: A CROSS-CULTURAL STUDY OF  
ANGER IN YOUTH

A dissertation submitted in partial fulfillment  
of the requirements for the degree of

DOCTOR OF PHILOSOPHY

to the faculty of the

DEPARTMENT OF PSYCHOLOGY

of

ST. JOHN'S COLLEGE OF LIBERAL ARTS AND SCIENCES

at

ST. JOHN'S UNIVERSITY

New York

by

TOVAH WEINRIB

Date Submitted 8/6/2021

Date Approved 8/6/2021

---

Tovah Weinrib

---

Mark Terjesen, PhD

**© Copyright by Tovah Weinrib 2021**

**All Rights Reserved**

## **ABSTRACT**

### **MULTILEVEL ITEM RESPONSE THEORY ANALYSIS OF THE ANGER REGULATION AND EXPRESSION SCALE: A CROSS-CULTURAL STUDY OF ANGER IN YOUTH**

Tovah Weinrib

Anger research among youth is increasingly recognized as an important area to inform guided intervention for children and adolescents with clinical levels of anger. However, most psychometric assessments, including measures of anger, are developed in the United States and are based in Western culture, potentially limiting their efficacy when applied internationally. The Anger Regulation and Expression Scale Short Form (ARES-S) is a comprehensive self-report assessment of the expression and regulation of anger in youth which has been used in research across numerous countries. Using a Multilevel Item Response Theory (MLIRT) approach, we analyzed country specific modulation of the expression of anger by examining the differences in response patterns to the ARES-S between youth from Australia, Greece, Italy, Serbia, South Africa, the United States, and Vietnam. Interestingly, item discrimination between countries varied greatly, indicating differences in anger expression and regulation across these countries. Further, we applied country-specific MLIRT scoring which improved the clinical predictive power of the ARES-S internationally while maintaining a similar predictive power in the United States.

## ACKNOWLEDGEMENTS

I would like to gratefully and sincerely thank you, Dr. Mark Terjesen, for your guidance, expertise, valuable discussions, and support during my graduate studies at St. John's University. You truly embody "unconditional other acceptance" and allow your students to thrive by letting them learn from their mistakes. I can only hope to pay it forward, and become a fraction of a mentor to others as you have been to me.

To my other committee members, it is hard to know where to start. Ray, your emphasis on helping your students learn (even after surgery from their bed!) is unparalleled. You are a remarkable "evidence-based clinician" and your research directly helps your students become better therapists to their patients. Wilson, your supervision has shaped me into the clinician I am today. You have taught me not to become a "clone" of any mentor, but to find, and use, my own voice, something that will carry me throughout my career.

I would also like to thank the faculty members of the Psychology Department at St. John's University, my clinical supervisors at the Zucker Hillside Hospital, the New York State Psychiatric Institute, and the St John's Center for Psychological Services for providing me with the foundation for becoming a Clinical Psychologist. A special thanks to Dr. Meredith Owens, Dr. Stephanie Solow and Dr. Paula Yanes-Lukin, mentors who have given me years of unwavering support and education. Without you, many of my academic and professional accomplishments would not have been possible.

A special thank you is dedicated to my family. Your support, encouragement, patience, and unwavering love were undeniably the reason my graduate school studies are now complete. Evan, I could not ask for a more kindhearted and patient husband.

Spouses getting PhDs simultaneously is no easy feat! Working alongside you every step of the way, celebrating each other's "wins" and commiserating each other's "losses" has encouraged and motivated me more than you will ever know. Thank you for being my true partner.

Lastly, I would like to thank my Dad, who was my first, and very best teacher. There is no amount of gratitude that feels sufficient. You were the first one who built the foundation for me to become a "scientist-practitioner," by showing me to use my intellect to think critically, to constantly question my observations, and to respond to others with curiosity, kindness, and empathy. My years of graduate training do not compare to what you taught me as my Dad. I am earning this degree in honor of you.

## TABLE OF CONTENTS

<b>ACKNOWLEDGEMENTS .....</b>	<b>ii</b>
<b>LIST OF TABLES .....</b>	<b>vi</b>
<b>LIST OF FIGURES .....</b>	<b>vii</b>
<b>CHAPTER I: INTRODUCTION .....</b>	<b>1</b>
DEFINING ANGER.....	3
EXPRESSION OF ANGER CROSS-CULTURALLY .....	5
MEASURING ANGER.....	7
ITEM RESPONSE THEORY .....	9
MULTILEVEL IRT .....	10
THE PRESENT STUDY .....	11
RESEARCH QUESTIONS .....	12
HYPOTHESES.....	12
<b>CHAPTER II: METHODS .....</b>	<b>13</b>
PARTICIPANTS.....	13
MEASURES .....	13
<i>Anger Regulation and Expression Scales Short Form (ARES-S)</i> .....	13
<i>Conners Comprehensive Behavior Rating Scales (CBRS)</i> .....	14
DATA ANALYSIS .....	15
<i>Software</i> .....	15
<i>IRT Models</i> .....	15
<b>CHAPTER III: RESULTS.....</b>	<b>17</b>
RELIABILITY .....	17
IRT ANALYSIS OF ARES-S.....	18

UNITED STATES MULTIVARIATE LOGISTIC REGRESSION.....	20
CROSS COUNTRY LINEAR REGRESSION ON CBRS.....	22
ANGER EXPRESSION ACROSS COUNTRIES.....	23
SUMMARY.....	24
<b>CHAPTER IV: DISCUSSION.....</b>	<b>27</b>
CROSS COUNTRY ITEM DISCRIMINATION.....	28
MLIRT SCORING.....	30
COUNTRY SPECIFIC ANGER EXPRESSION.....	31
LIMITATIONS AND FUTURE DIRECTIONS.....	32
CLINICAL IMPLICATIONS.....	33
<b>APPENDIX A. MPLUS SYNTAX.....</b>	<b>35</b>
<b>APPENDIX B. MODEL FIT INFORMATION.....</b>	<b>38</b>
<b>APPENDIX C. ARES-S ITEMS.....</b>	<b>39</b>
<b>APPENDIX D. ITEM CHARACTERISTIC PARAMETERS.....</b>	<b>40</b>
<b>REFERENCES.....</b>	<b>41</b>



**LIST OF TABLES**

<b>Table 1.</b> Participant Demographics.....	13
<b>Table 2.</b> Reliability parameters by country.....	17
<b>Table 3.</b> Item discrimination parameters by country.....	19
<b>Table 4.</b> Demographic parameters of the United States ARES-S sample.....	21
<b>Table 5.</b> United States sample diagnosis multivariate logistic regression parameter estimates.....	22
<b>Table 6.</b> Multivariate logistic regression pseudo R-squared values.....	22
<b>Table 7.</b> Linear Regression Significant Predictors and Model R-values.....	23
<b>Table 8.</b> Mean differences between countries for ARES-S subscales.....	26

**LIST OF FIGURES**

<b>Figure 1.</b> Total information curves for Externalizing Anger subscale.....	19
<b>Figure 2.</b> Total information curves for Internalizing Anger subscale.....	20
<b>Figure 3.</b> Total information curves for Extent of Anger subscale. ....	20
<b>Figure 4.</b> ARES-S subscale scores by CTT and MLIRT across countries .....	25

## Chapter I: Introduction

Empirically-based psychological assessment is essential in understanding the social- emotional functioning of children and adolescents. Proper assessment allows practitioners to implement guided and effective interventions and monitor treatment progress, particularly in clinical populations (Petras et al., 2013). While there are a myriad of broad and narrow-based measures of social-emotional functioning among youth (Halle & Darling-Churchill, 2016), research on anger expression and regulation in child and adolescent populations has been overshadowed when compared to the study of other emotions. Although anger serves many useful functions when expressed at non-clinical levels (Aarts et al., 2010; Sell, 2005), dysfunctional anger is related to several adverse outcomes such as peer rejection (Hart & Ostrov, 2013), externalizing behavior (Quinn et al., 2014), health problems (Igna et al., 2009), emotional distress (Carlozzi et al., 2010), and interpersonal difficulties (Muris et al., 2004).

Anger research among youth is growing especially in terms of measure development and validation (Kerr & Schneider, 2008; Potegal et al., 2010) reinforcing that proper assessment to inform guided intervention for children and adolescents with clinical levels of anger is critical for their well-being. However, most psychometric assessments, including measures of anger, are developed in the United States and are based in Western culture (Mooney, 2014; Moscoso & Spielberger, 2011). This is problematic as the majority of humans live in non-Western cultures, meaning that psychological studies are conducted on the minority of the human population (Her & Joo, 2018). Few anger measures have been validated for use outside the United States; thus,

international clinicians have limited options for empirically-supported anger assessments (Moscoso & Spielberger, 2011).

Given the lack of empirically-based anger assessments for cross-cultural use, when seeking to administer psychological measures internationally, the primary approach thus far has been to back-translate English measures into different languages while incorporating cultural consideration (Alaka Mani et al., 2018; Cheung, 2012; Moscoso & Spielberger, 2011). This allows for the empirical knowledge base founded in English-language and Western cultures to be widely applied around the world rather than taking on the laborious task of creating distinct measures for each culture (Byrne, 2016; Cheung, 2012). Despite the utility of this method, it assumes that the construct(s) being measured are equivalent across cultures, which, if untested, can unfortunately call into question the validity of the conclusions (Byrne, 2016).

The purpose of this study is to analyze the psychometric properties of the Anger Regulation and Expression Scale Short Form (ARES-S; DiGiuseppe & Tafrate, 2011) internationally, a scale developed and normed in the United States, and determine its utility as a measure of anger across distinct cultures. Using a multi-level and multi-dimensional Item Response Theory (MLIRT) approach, each item of the ARES-S will be analyzed for the information it provides to better understand how the different dimensions of anger are reported across cultures (de Jong & Steenkamp, 2010). ARES-S items can then be weighted in a country-specific manner so that they can be reliably compared to determine whether there are differences in anger expression and regulation in children and adolescents across countries (Stucky, 2009; van Lier et al., 2018). This method will allow for measurement equivalence, decreasing the bias inherently observed

when administering a psychometric measure normed in one country to individuals in another (Kankaras & Moors, 2010).

### **Defining Anger**

Anger is considered to be a basic emotion which has evolved as a reaction to distress, physical discomfort, pain, or frustration (Ekman, 2005; Panksepp & Biven, 2012). Despite its universality and commonplace in clinical problems, anger is complex and hard to define (DiGiuseppe & Tafrate, 2007). As such, inconsistencies can be found throughout the anger literature, specifically regarding anger origin and expression (Jasinski et al., 2016). However, the lack of empirical research into anger is surprising due to its many known negative consequences that impact individuals across the lifespan (Barrett et al., 2013; Cassiello-Robbins & Barlow, 2016; C. I. Eckhardt et al., 2008; Quinn et al., 2014).

Despite inconsistent views and theories of anger, such as the characteristics of state and trait anger, the distinction between anger and aggression, individual differences in the experience of anger, and typologies of dysfunctional anger, certain constants can be found across the literature of clinical anger (Ahmed et al., 2012; Cassiello-Robbins & Barlow, 2016; Kerr & Schneider, 2008; Kidwell et al., 2016; Moscoso & Spielberger, 2011). Researchers agree that, like sadness and fear, anger can range in intensity from irritation to fury and in duration from episodic to general temperament (Kerr & Schneider, 2008). Further, the expression of anger is often categorized into internalizing anger and externalizing anger (Steele et al., 2007). Internalizing anger, which often involves an individual's attempt to suppress or deny their anger, often overlaps with feelings of bitterness and rejection and can stem from an individual's difficulty in acting

assertively. Externalizing anger is categorized by forms of aggression and other, non-aggressive physical gestures (DiGiuseppe & Tafrate, 2007). Importantly, internalizing and externalizing anger, constructs originally studied in adult anger, have also been validated in childhood and adolescent anger (Kerr & Schneider, 2008; Kidwell et al., 2016).

Anger has been linked as a key component to a number of diagnoses among youth. It is present among the diagnostic criteria of oppositional defiant disorder (ODD), a disorder represented by negativistic, hostile, and defiant behavior, intermittent explosive disorder (IED), a disorder characterized by recurrent angry aggressive episodes that are disproportionate to a stimulus or provocation, and the newly classified disruptive mood dysregulation disorder (DMDD), a disorder only diagnosed in children and adolescents with a chronic component of irritability or anger (American Psychiatric Association, 2013; Copeland et al., 2013; Fernandez & Johnson, 2016). Although anger is considered in the above diagnoses, DiGiuseppe and Tafrate argue that the focus is typically on the violent behavior that anger incites, rather than the emotion of anger itself (2007).

Regardless of whether criteria for a specific diagnosis is met, clinical levels of anger, if left untreated in childhood or adolescence, can lead to additional diagnoses like conduct disorder, which is known to have more problematic behavioral consequences (Rowe et al., 2010). There are also well-established associations between anger and negative mood in children and adolescents. Anger's manifestation in many clinical disorders highlights the importance of accurate and valid assessment, particularly in youth, where early intervention can be critical (Cassidello-Robbins & Barlow, 2016;

Kidwell et al., 2016).

### **Expression of Anger Cross-Culturally**

Cross-cultural studies are particularly relevant in the field of psychology where contemporary psychological theories are typically founded and subsequently tested in Western cultures (Cheon et al., 2020). Thus, for psychological theories to be considered universal and for experiments to have high external validity, they must be studied in a cross-cultural context where the people of various cultures are considered (Moscoso & Spielberger, 2011). While there are many definitions of culture, it is typically agreed upon that culture consists of shared elements of perception, belief, evaluation, and communication among those who share a language, historic period, or geographic location (Matsumoto et al., 2010).

Regardless of culture, there is a body of literature describing the universality of certain emotions such as anger, joy, and sadness, emotions believed to be intrinsic (or basic) to all humans. Basic emotions are associated with unique physiological responses in both the autonomic and central nervous systems (Davidson et al., 1990; Levenson, 2006). These physiological responses as they relate to anger have been identified in many distinct and varied cultural groups (Mauss et al., 2005). For example, the experience of anger universally and across the lifespan produces increased heart rate, blood pressure, and blood flow to the arms and hands, a phenomenon that can sometimes result in health problems if the anger becomes dysfunctional, and particularly worrisome when it impacts youth (Igna et al., 2009; Levenson, 2006).

Importantly, however, more recent literature allows for cultural context to modulate the expression and experience of emotion, even for those considered “basic” or

universal. For instance, expressions of anger may be appropriate within an individual's family in one culture but not in another (Matsumoto et al., 2010). Matsumoto et al. found that in individualistic cultures there was a greater endorsement of angry expressions toward in-groups, such as family units, as compared to out-groups (2010). Conversely, they found in collectivistic cultures that there was a greater endorsement of angry expressions toward out-groups than in-groups. This discrepancy represents the difference in the meanings of social relationships from one culture to the next and produces different guidelines for the regulation of expressive behavior.

Anger has been studied cross-culturally in many contexts and has been shown to be modified in its expression, response, and regulation based on the culture of the individual, both in naturalistic and in laboratory settings (Matsumoto et al., 2010). For instance, a seminal ethnography detailing the experience of anger amongst the Utku Inuits found them to rarely express anger or interpret others as feeling angry, a phenomenon attributed to the importance of the group working collaboratively (Briggs, 1975). In contrast, a similar ethnography of the Kaluli people of Papua New Guinea demonstrated that displays of anger are common and serve a purpose in what has been described as a reciprocity-based society, where displays of anger are meant to be forceful pleas for support (Schieffelin, 1983). Additionally, individuals with high social status in Japan express more anger due to their "anger privilege" when Western research has shown that individuals with lower social status have been reported to express more anger (Park et al., 2013).

Researchers therefore argue that evidence collected on anger in Western societies does not necessarily hold true across other cultures (Byrne, 2016; Cheon et al., 2020;



Kankaras & Moors, 2010). Taken together, studying anger in a cross-cultural manner is an important step in generalizing the results found in Western anger research. As children internalize messages embedded in their culture in regards to how to regulate and express emotion (Holodynski, 2013), studying anger cross-culturally can provide unique insight as to the cross-cultural differences in how children and adolescents express and regulate anger.

### **Measuring Anger**

Anger is a difficult construct to measure due to its complexity; the inconsistencies in theories of anger become pronounced when trying to accurately measure its expression (Ahmed et al., 2012; Fernandez et al., 2015). Early attempts to measure anger primarily focused on physiological manifestations (Ax, 1953). Research has since moved towards subjective measures of the intensity of the experience of anger as well as anger specific thoughts (C. Eckhardt et al., 2004; Kerr & Schneider, 2008). However, few assessments focus on measuring clinical or dysfunctional anger.

One of the most widely used measures of adult anger, the State-Trait Anger Expression Inventory (STAXI), was both validated in an adolescent population (Spielberger, 1999) and modified for use in a child and adolescent population (del Barrio et al., 2004), but suffers from fairly low test/retest validity. A newer scale of anger, the Patient Reported Outcomes Measurement Information System Pediatric Anger Scale (PROMIS; Irwin et al., 2012) was developed using modern psychometric techniques including IRT (Irwin et al., 2012). The PROMIS Pediatric Anger Scale encountered the issue of multidimensionality, or the intersection of many correlated factors. Taken together, the issues encountered with these measures indicate ambiguity in the constructs

of anger being measured (Kerr & Schneider, 2008).

This concern was addressed by the ARES, a measure developed by DiGiuseppe and Tafrate to assess and measure pathologic or disturbed anger in youth (2011). The subscales of the ARES were constructed based on several conceptual considerations including consistency with current emotion research (DiGiuseppe & Tafrate, 2011). They also assess clinically relevant patterns of behavior, physiological arousal, and thinking, which may be useful as targets for clinical intervention (Cavlazoglu et al., 2013). The final version of the ARES was organized into a higher order structure in which the Total Score was broken down into the conceptual areas of Internalizing Anger, Externalizing Anger, and Extent of Anger, a measure of anger intensity and duration, a three factor model (DiGiuseppe & Tafrate, 2011). A second, short form version of the ARES (ARES-S), was later created to allow for quicker and more widespread assessment.

Importantly, the study of anger across nations and cultures relies on the validation of the scales that measure anger, and the generalizability of their results (Moscoso & Spielberger, 2011). International validation of the ARES-S have been performed and found to have good internal reliability (Harris, 2014; Pzena, 2018; Smajlaj, 2014). Factor analysis of the ARES-S in various nations, however, supported different factor solutions in each country. For instance, Exploratory Factor Analysis (EFA) of an Israeli sample yielded a 1-factor solution (Harris, 2014), whereas a Vietnamese sample produced a 3-factor model (Pzena, 2018). Of note, one limitation of these cross-national validity and factor analyses is the significantly smaller sample sizes collected in each country as compared to the American sample sets, making statistical comparison and modeling difficult.

This difference in factor loadings makes cross cultural comparison of scores difficult and is in part due to the ARES-S being developed using Classical Test Theory (CTT) which requires normative samples to compare across distinct groups.

### **Item Response Theory**

IRT offers a useful framework for understanding the relationship between a latent trait and observed scores on a measure. IRT is often discussed in the context of the limitations of its predecessor classical test theory (CTT) and differs from CTT in several important ways (Muris et al., 2004). CTT posits that the pattern of responses on a test are the result of a “true score” plus some error associated with the test. The true score represents the ideal responses of an individual given a perfectly crafted measure without any error introduced by factors such as the environment or the current mental state of the individual. This, however, is functionally impossible, as no measure is completely devoid of error. IRT, in contrast, is a model that posits that there is an underlying trait whose level accounts for an individual’s choices on a measure (Cai et al., 2016). While CTT interpretations are completely based on the responses to the measure, IRT assumes that the trait of the individual is independent of the content of the measure and can be modeled regardless of the specific item. Further, where CTT is concerned with the score of the whole measure, IRT provides information about each of the items which comprise the measure including the difficulty of an item and the probability of each response based on the amount of latent trait (Cai et al., 2016).

The underlying trait being assessed by an item on a measure is denoted by the Greek letter theta ( $\theta$ ). Theta is then related to the probability of endorsement of a certain response on a test item resulting in what is termed the item characteristic curve (ICC;

(Andrich, 2010). The ICC is a typically non-linear regression line denoting the probability of endorsing a response of a test item based on a theoretical level of the latent trait that the test item aims to measure. Each item response has its own ICC and together they describe the probability of each response at any given level of theta. ICCs are rarely uniform across all levels of a latent trait. This is important as it implies that an item's ability to discriminate varies between levels of the latent trait and has better discrimination within certain regions of the latent trait spectrum.

### **Multilevel IRT**

Measurement equivalence is a key component to cross cultural research (Byrne, 2016; Kankaras & Moors, 2010) and a multilevel IRT (MLIRT) approach provides an estimate of a latent trait that is directly comparable between populations without the need of a normative sample. This is particularly appealing for cross-cultural research which, using CTT, would require a normative sample in each country which can quickly become prohibitively expensive and labor intensive. A MLIRT approach considers all participants as one group in the first level, and then separates by subgroups in the second level allowing for item thresholds to vary across respondents from different groups (Johnson, 2003; King et al., 2004; Wolfe & Firth, 2002). Multilevel models have made much progress since IRT was first developed. While originally it was thought that modeling item parameters should be done in one step and then applied to a combined IRT analysis, these models assume that items are invariant across groups, an assumption that is increasingly less likely as the number of distinct groups increases (De Jong et al., 2007). Another approach is to model the item parameters in the model development stage (De Jong et al., 2007). A major advantage to this approach is that by considering all

participants as one group in the first level, the power of each group is combined and therefore less restrictive assumptions on measurement invariance are necessary.

Measurement invariance is the assumption that the measure behaves the same in every group. This then allows for the possibility of letting the various parameters in the model fluctuate across countries while still measuring the same latent trait (De Jong et al., 2007).

### **The Present Study**

The present study seeks to expand the international utility of the ARES-S as a measure of anger among youth across several distinct cultures. Anger, although difficult to define and challenging to study across the lifespan, is particularly important to measure and treat in youth due to the abundant adverse outcomes with which it is associated such as peer rejection (Hart & Ostrov, 2013), externalizing behavior (Quinn et al., 2014), and health problems (Igna et al., 2009). The ARES-S, a measure developed in the United States to assess anger expression and regulation among children and adolescents, has sound psychometric properties, as demonstrated by several studies (DiGiuseppe & Tafrate, 2011; Neuhaus, 2013; Smajlaj, 2014). Although the ARES-S has been used in research in several international populations (Harris, 2014; Pzena, 2018), it is unclear whether its psychometric properties will translate to other cultures. Using modern psychometric techniques, we can analyze this international archival data to assess the culture-specific changes in item information and discrimination and weight them accordingly to improve measurement equivalence (de Jong & Steenkamp, 2010; Kankaras & Moors, 2010). This will allow for the accurate cross-cultural comparison of anger expression and regulation in children and adolescents, as measured by the ARES-S.

### **Research Questions**

- 1) Does the ARES-S demonstrate adequate internal reliability internationally?
- 2) Do different ARES-S items provide different amounts of information and discrimination  
depending on the country in which it was administered?
- 3) Can we use the United States normative sample to validate the efficacy of MLIRT scoring as compared to previously normed CTT scoring?
- 4) Does IRT scoring compared to CTT scoring improve the ARES's ability to make clinical predictions across countries?
- 5) Is anger expressed or regulated differently across countries as measured by the ARES-S with IRT scoring?

### **Hypotheses**

- 1) Using MLIRT, we can improve the ARES-S's ability to measure clinically relevant anger across countries and make statistically sound comparisons across countries, allowing for true understanding of the cross-cultural expression of anger.
- 2) MLIRT ARES-S anger subscale-scores will differ across countries due to the cultural modulation of anger expression and regulation.

## Chapter II: Methods

### Participants

Archival data used for the current study was collected as part of the standardization process of the ARES in the United States as well as from studies validating the ARES-S across countries including Australia, Greece, Italy, Serbia, South Africa, and Vietnam. This dataset represents 2,074 completed ARES-S forms collected between 2006 and 2019 and has an approximately even distribution of both males and females (49% male, 46% female, 6% unknown), between the ages of 7 and 20 (Table 1). Additionally, participants from Australia, Greece, South Africa, Serbia, and Vietnam also completed the Conner's Comprehensive Behavior Rating Scales (CBRS) which was used to validate the efficacy of IRT scoring in predicting clinical parameters.

**Table 1**

*Participant Demographics*

	Australia	Greece	Italy	United States	Serbia	South Africa	Vietnam
Mean Age (range)	15	13	9	14	14	14	14
% Female	61	50	47	45	38	57	71

### Measures

***Anger Regulation and Expression Scales Short Form (ARES-S)***

Anger was measured using the ARES-S which is a 17-item self-report questionnaire adapted from the ARES designed to assess anger in terms of thoughts, emotions, and behaviors. The ARES-S is subdivided into multiple subscales and clusters including Internalizing Anger, Externalizing Anger, and Extent of Anger.

The Externalizing Anger scale was designed to measure aggression, bullying, and revenge, whereas the Internalizing Anger scale was designed to measure arousal, rejection, and bitterness. The Extent of Anger scale was designed to measure the duration of anger episodes and length of time experienced. In the normative sample, across age and gender groups, Cronbach's alpha coefficient for total scores ranged from .97 to .99. The Internalizing Anger scale alpha coefficient ranged from .77 to .96, the Externalizing Anger alpha coefficient ranged from .68 to .95, and the Extent of Anger alpha coefficient ranged from .65 to .95. Test-retest values ranged from .58 to .92.

Discriminative validity analyses indicated that the ARES-S scores accurately discriminate between clinical groups and the general population. In terms of the classification accuracy of the scores, the overall correct classification rates ranged from 86.8% to 90.8% for Conduct Disorder and from 83.0% to 85.0% for Oppositional Defiant Disorder. ARES-S scores were found to be correlated with scores from other measures of early psychopathology such as the Conners Comprehensive Behavior Rating Scales (Conners, 2010).

### ***Conners Comprehensive Behavior Rating Scales (CBRS)***

The CBRS is a comprehensive assessment tool used to screen for behavioral, emotional, social and academic concerns in children. The CBRS yields scale scores for Content Scales, DSM-IV TR Symptom Scales and Validity Scales. Additionally, they produce a Clinical Index score and Other Clinical Indicators score, such as the Violence Potential (Conners, 2010). The CBRS has strong psychometric properties with reliability, internal consistency, test-retest and interrater reliability coefficients above .70, and validity, with discriminative analyses demonstrating that the scales can distinguish children with disruptive behavior disorders from general population as well as other



clinical groups. The CBRS has previously been validated and normed in several international populations including in Vietnam and Turkey.

## **Data Analysis**

### ***Software***

The IRT analysis in this study was conducted using the Mplus software (Muthén & Muthén, 2017). SPSS (IBM corp.) was used to prepare data files and transform the dataset into the form required for Mplus import as well as to perform regression, MANOVA, and reliability analysis. Cronbach's  $\alpha$  and McDonald's  $\omega$  were calculated with R (v 4.04) using the "psych" and "psychTools" packages (Revelle, 2017).

### ***IRT Models***

Two IRT models were prepared and performed: a single-level multidimensional IRT model (SLIRT) and a multidimensional multilevel IRT model (MLIRT). All countries were included in this analysis, regardless of sample size, since the modelling accuracy for each country is independent. The SLIRT model was stratified by country of origin but did not consider country of origin as a known latent class. The MLIRT model considered country of origin as a known latent class. In SLIRT, the country of origin is used as a stratifying factor, but assumes measure invariance and that the latent trait (i.e. dimensions of anger), and the items on the measure, will behave similarly in each country. Both models used a multidimensional approach which describes anger through the validated three-factor model of Internalizing Anger, Externalizing Anger, and Extent of Anger. Both models used a maximum likelihood estimator (MLE) with robust standard errors to fit a graded response model (GRM) which is used for ordinal categorical variables, such as Likert scales. IRT-based

scores were calculated for both models and used for further comparison. Although scores were computed for all countries, data from Australia and Greece was not used in subsequent analyses; their low sample size limited the utility in certain analyses (such as regression).

### Chapter III: Results

#### Reliability

Before IRT can be performed, reliability of the assessment must be established. To address Research Question 1 (Does the ARES-S demonstrate internal reliability internationally?), reliability was calculated for the ARES-S separately for each country and across the entire sample. Cronbach's alpha was 0.9 across the whole sample which is similar to the reliability estimate calculated in the United States normative sample (0.92; DiGiuseppe & Tarfate, 2011). Individual country Cronbach's alpha values ranged from 0.81 to 0.9 (Table 2). Cronbach's alpha often underestimates the true reliability and so an additional measure of internal reliability, McDonald's omega, was calculated for each country to verify the level of reliability (Deng & Chan, 2017). Both methods yielded similar levels of reliability. However, the reliability for Greece was unable to be calculated by this method due to small sample size (Table 2). Reliability was deemed high enough to proceed to IRT analysis.

**Table 2**

*Reliability parameters by country*

	Number of individuals	Cronbach's $\alpha$	McDonald's $\omega$
Australia	24	0.84	0.87
Greece	14	0.9	Not calculated
Italy	225	0.82	0.84
Serbia	225	0.81	0.89
South Africa	112	0.81	0.83
United States	1305	0.92	0.93
Vietnam	169	0.85	0.89
Total	2074	0.9	0.9

### **IRT Analysis of ARES-S**

To determine the benefit of using MLIRT as opposed to SLIRT for measuring the validated factors of anger, Internalizing Anger (6 items), Externalizing Anger (8 items), and Extent of Anger (3 items), a head-to-head comparison of measure information was performed. Completed ARES-S responses from 2,074 individuals collected in Australia, Greece, Italy, Serbia, South Africa, United States, and Vietnam were analyzed by using both MLIRT and SLIRT. Total information curves (TICs) were compared (Figures 1-3). TICs for each factor were highly variable between countries when analyzed by MLIRT. TICs from the MLIRT analysis, including the United States, also varied greatly from the SLIRT TICs. To address Research Question 2 (Do different ARES-S items provide different amounts of information and discrimination depending on the country in which it was administered?), item discrimination parameters were compared. Item discrimination, which describes an item's ability to correctly identify levels of latent trait, varied greatly between countries, indicating that in different countries, items provided different amounts of information about the dimension of anger that the subscale was designed to measure. Certain items had very high variability across countries; for instance, Item 1 had an item discrimination of 2.39 in Greece but had a -1.79 value in Vietnam indicating it to be a good item for measuring externalizing anger in Greece, but it was not useful at all in Vietnam (Table 3). Items 1, 10, and 16 had the highest variation across countries with standard deviations of 1.28, 1.08, and 1.47, respectively. Additionally, IRT scoring was applied to generate Internalizing Anger, Externalizing Anger, and Extent of Anger scores for each participant. These scores for each ARES-S subscale provide a direct estimation of the level of the latent trait in each individual and were thus used for further analysis to

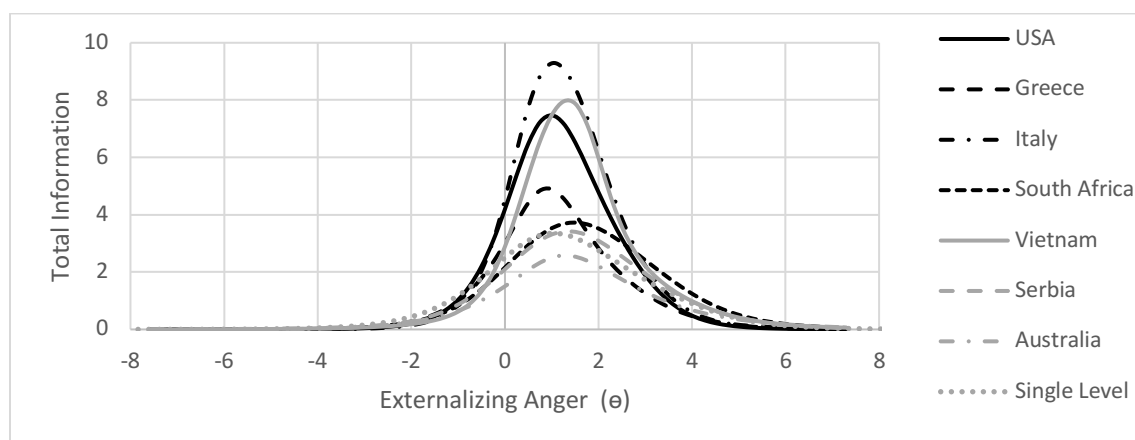
assess the utility of scoring by MLIRT rather than CTT.

**Table 3**

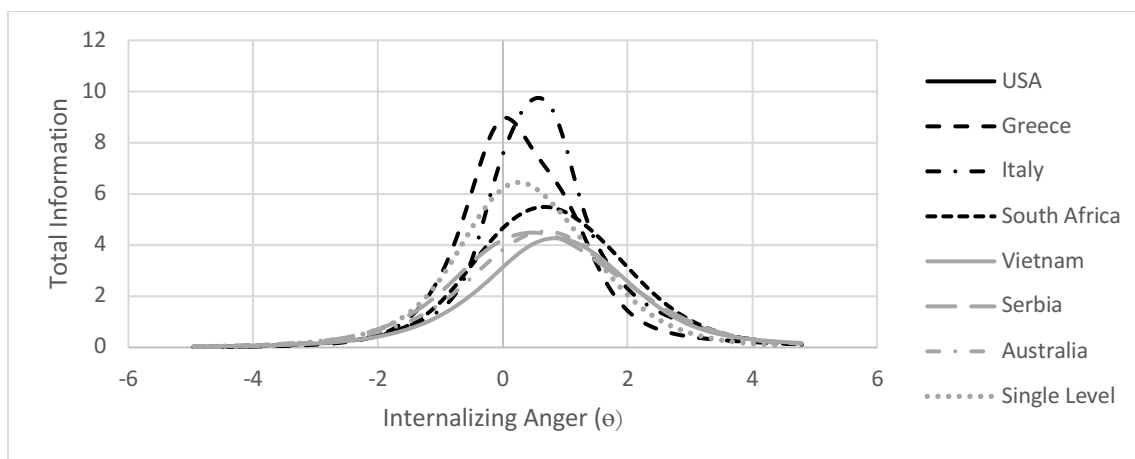
*Item discrimination parameters by country*

	Australia	Greece	Italy	Serbia	South Africa	United States	Vietnam	SLIRT
Item 1 <sup>b</sup>	0.75	2.39	1.36	1.29	0.82	1.03	-1.79	0.97
Item 2 <sup>a</sup>	0.94	1.02	1.40	1.10	0.79	1.73	3.17	1.32
Item 3 <sup>b</sup>	2.12	3.78	1.58	1.47	2.16	1.98	0.92	1.83
Item 4 <sup>b</sup>	1.25	3.13	1.26	1.67	1.44	1.83	0.34	1.68
Item 5 <sup>a</sup>	0.54	1.35	2.33	1.08	1.17	1.64	2.59	1.40
Item 6 <sup>b</sup>	1.87	1.31	3.46	1.85	1.99	2.96	2.55	2.70
Item 7 <sup>a</sup>	1.35	1.52	1.51	1.33	1.49	1.96	0.82	1.65
Item 8 <sup>a</sup>	1.19	1.44	3.17	1.56	1.61	2.25	1.46	2.05
Item 9 <sup>a</sup>	0.88	2.40	2.08	1.23	1.41	2.35	1.61	2.05
Item 10 <sup>b</sup>	1.25	2.38	4.76	2.24	2.24	2.60	1.96	2.65
Item 11 <sup>a</sup>	0.84	1.33	1.76	1.27	1.35	1.37	0.87	1.33
Item 12 <sup>a</sup>	-0.11	0.22	2.42	1.11	1.24	2.04	1.59	1.86
Item 13 <sup>b</sup>	1.90	0.93	1.38	1.22	2.03	1.42	1.05	1.41
Item 14 <sup>c</sup>	1.99	3.29	2.24	2.17	1.62	1.77	-0.03	1.56
Item 15 <sup>a</sup>	1.85	2.01	2.21	1.53	1.45	1.93	1.90	1.75
Item 16 <sup>c</sup>	2.47	2.60	2.17	0.64	1.02	5.15	2.99	2.53
Item 17 <sup>c</sup>	1.48	1.18	1.51	0.98	1.04	2.17	0.61	1.69

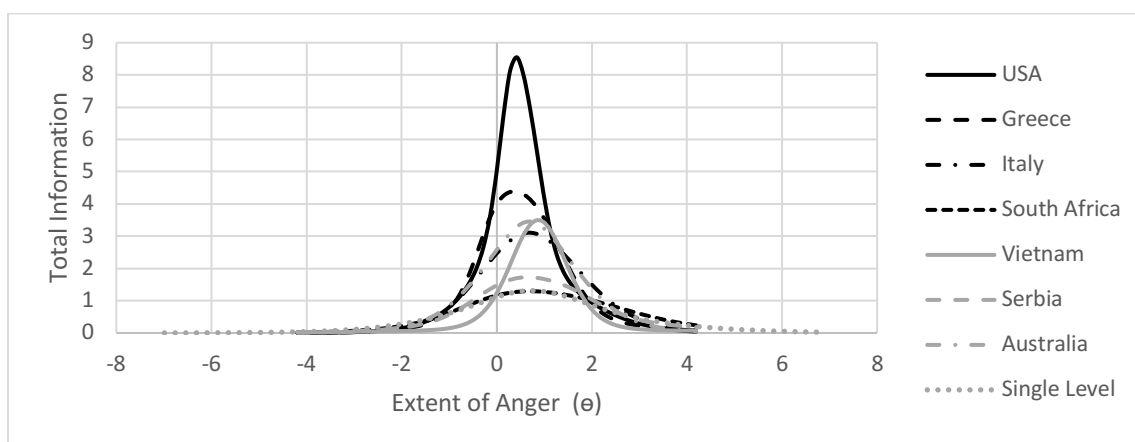
- a. Externalizing Anger subscale item  
 b. Internalizing Anger subscale item  
 c. Extent of Anger subscale item



**Figure 1.** *Total information curves for Externalizing Anger subscale*



**Figure 2.** *Total information curves for Internalizing Anger subscale.*



**Figure 3.** *Total information curves for Extent of Anger subscale.*

### United States Multivariate Logistic Regression

To validate the ability of MLIRT scored ARES-S subscales to predict known diagnoses in the normative United States sample, multivariate logistic regression analysis was performed and compared to CTT scores in predicting the same diagnoses. This analysis was completed to address Research Question 3 (Can we use the United States normative sample to validate the efficacy of MLIRT scoring as compared to previously normed theory (CTT) scoring?). The United States sample included data on known diagnoses of ODD, CD, ADHD, and GAD which were used as dependent variables in the

regression model (Table 4).

**Table 4.**

*Demographic parameters of the United States ARES-S sample*

Demographic parameter		N	Percentage
Diagnosis	No Diagnosis	783	60.0%
	ODD	100	7.7%
	CD	52	4.0%
	ADHD	61	4.7%
	GAD	22	1.7%
	Other	288	22.1%
Age	10	152	11.6%
	11	144	11.0%
	12	152	11.6%
	13	203	15.5%
	14	164	12.6%
	15	169	12.9%
	16	164	12.6%
Sex	17	158	12.1%
	Male	706	54.1%
	Female	600	45.9%
Ethnicity	Black	275	21.1%
	Hispanic	168	12.9%
	White	779	59.6%
	Asian/Multi/Native/Other	84	6.4%

Age, sex, and ethnicity were entered into the model as factors and were each significant predictors of all dependent variables. For all diagnoses, using MLIRT scoring yielded similar results to CTT scoring. B values, Wald test statistics, and *p* values for each diagnosis are reported in Table 5. Additionally, pseudo R-squared values, which are a measure of goodness of fit of the model, were similar between MLIRT and CTT, indicating that both scoring methods were equivalent at correctly predicting diagnoses (Table 6).

**Table 5.***United States sample diagnosis multivariate logistic regression parameter estimates*

Diagnosis	ARES Subscale	CTT model			MLIRT model		
		B	Wald-test	p-value	B	Wald-test	p-value
No Diagnosis	Externalizing	-.03	1.63	.20	.25	.91	.34
	Internalizing	.000	.00	.99	.35	2.33	.13
	Extent	-.13*	12.05	.00	-1.99*	10.09	.00
ODD	Externalizing	.14*	18.55	.00	1.47*	10.27	.00
	Internalizing	.04	1.17	.28	.02	.00	.96
	Extent	.04	.53	.47	-.96	.81	.37
CD	Externalizing	.11*	8.21	.00	.73	1.45	.23
	Internalizing	-.07	2.34	.13	-1.62*	9.65	.00
	Extent	.30*	17.66	.00	4.16*	8.07	.00
ADHD	Externalizing	-.03	.50	.48	-.62	1.47	.23
	Internalizing	.04	1.04	.31	.12	.07	.79
	Extent	.09	1.84	.16	1.76	2.13	.14
GAD	Externalizing	.06	1.12	.29	.49	.35	.56
	Internalizing	.21*	8.12	.00	2.06*	6.50	.01
	Extent	.02	.04	.84	-2.65	1.80	.18

**Table 6.***Multivariate logistic regression pseudo R-squared values*

	CTT model	MLIRT model
Cox and Snell	0.311	0.337
Nagelkerke	0.344	0.373
McFadden	0.158	0.175

**Cross Country Linear Regression on CBRS**

Once validated in the United States sample, to address Research Question 4 (Does IRT scoring improve the ARES-S's ability to make clinical predictions across countries?), MLIRT was compared to CTT for predicting CBRS scores that were administered to the same participants who completed the ARES-S in Australia, Greece,



Serbia, South Africa, and Vietnam. CBRS subscales for ODD, CD, ADHD, violence potential, aggressive behaviors, and social phobia were used as dependent variables in separate stepwise linear regression analysis using either MLIRT or CTT scores for the ARES-S subscales. Additionally, age and sex were added as variables to the model in the first step. R values for models using MLIRT scores were higher for each CBRS subscale, indicating greater goodness of fit and predictive power as compared to CTT scores (Table 7). R values ranged from 0.574 to 0.905 using MLIRT and from 0.258 to 0.874 using CTT. Additionally, for each CBRS subscale, all three ARES-S subscales were significant predictors when using MLIRT scores but not when using CTT scores, indicating that scoring by MLIRT expanded the ability of the ARES-S to provide information about diverse clinically relevant scores.

**Table 7**

*Linear Regression Significant Predictors<sup>a</sup> and Model R-values*

CBRS subscale	Significant ARES-S subscale Predictor		R-value	
	MLIRT	CTT	MLIRT	CTT
Oppositional Defiant Disorder	Int <sup>b</sup> , Ext <sup>c</sup> , Extent <sup>d</sup>	Int	0.898	0.793
Aggressive Behavior	Int, Ext, Extent	Int, Ext	0.899	0.794
Violence Potential	Int, Ext, Extent	Int, Ext, Extent	0.799	0.479
Conduct Disorder	Int, Ext, Extent	Ext	0.854	0.794
Social Phobia	Int, Ext, Extent	Ext, Int	0.900	0.383
ADHD	Int, Ext, Extent	Extent	0.899	0.377

a. Significant predictors were  $p < 0.05$

b. Internalizing Anger subscale

c. Externalizing Anger subscale

d. Extent of Anger subscale

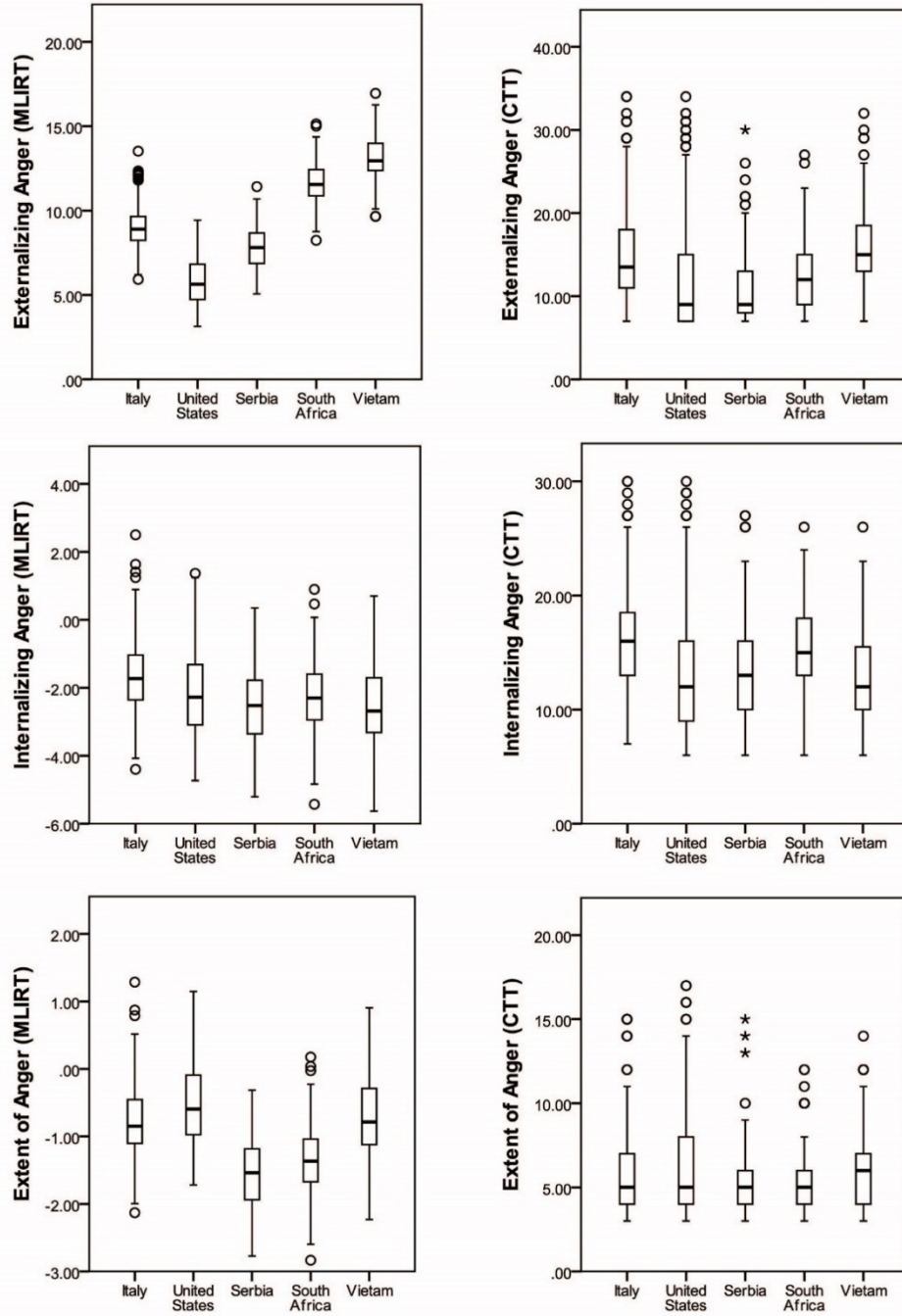
### Anger Expression across Countries

To determine if anger expression varied across countries, mean ARES-S subscale scores were compared by either MLIRT or CTT (Figure 4). To address Research

Question 5 (Is anger expressed or regulated differently across countries?), significant differences in mean anger scores between countries were determined with a two-way MANOVA and individual comparisons were determined by a Tukey's post-hoc test. MLIRT or CTT scores were dependent variables, and country, age and sex were independent variables. Differences in means were more significant across countries when using MLIRT scores as opposed to CTT scores (Table 8).

### **Summary**

In this study, MLIRT was applied to the ARES-S collected in Australia, Greece, Italy, Serbia, South Africa, the United States, and Vietnam. Item discrimination and variation were examined across countries where large amounts of variation were observed. This variation in item discrimination between countries provides an interesting perspective on differences in anger expression and regulation across countries. Further, when we applied MLIRT scoring to the ARES-S, we found that it improved the clinical predictive power of the measure in other countries while maintaining a similar predictive power in the United States, the country where it was normed. MLIRT scoring was able to account for country-specific response patterns, enabling us to directly compare levels of Internalizing Anger, Externalizing Anger, and Extent of Anger between countries. We found that there are significant differences in Internalizing Anger, Externalizing Anger, and Extent of Anger between countries, independent of age or sex.



**Figure 4.** *ARES-S subscale scores by CTT and MLIRT across countries*

**Table 8***Mean differences between countries for ARES-S subscales*

		Italy	Serbia	South Africa	United States	Vietnam
Externalizing Anger <sup>a</sup>	Italy	N/A	4.08*	2.65*	3.23*	-0.96
	Serbia	1.24*	N/A	3.80*	-2.05*	5.28*
	South Africa	-2.56*	1.42	N/A	0.57	3.6*
	United States	3.29*	0.85	5.85*	N/A	-4.20*
	Vietnam	-4.04*	-5.05*	-1.48*	-7.34*	N/A
Internalizing Anger <sup>a</sup>	Italy	N/A	2.6*	0.49	2.90*	3.55*
	Serbia	.96*	N/A	0.34	0.35*	-0.04
	South Africa	.61*	2.10*	N/A	2.41*	3.05*
	United States	.61*	-0.30	-0.01	N/A	0.64
	Vietnam	1.01*	-0.94	0.39	0.40*	N/A
Extent of Anger <sup>a</sup>	Italy	N/A	0.63	0.59	-0.17	0.15
	Serbia	0.80*	N/A	.21*	1.00*	.78*
	South Africa	0.59*	0.03	N/A	-0.77	-0.44
	United States	-0.20*	.81*	-.79*	N/A	0.32
	Vietnam	0.02	0.48	-.57*	0.22*	N/A

a. Top half of table refers to CTT, bottom half refers to MLIRT

\* difference in mean p value &lt;0.05

## Chapter IV: Discussion

The purpose of the current study was to expand the international utility of the ARES-S as a measure of anger among youth across several distinct cultures. The current study sought to use an MLIRT approach to score and compare the ARES-S across several countries. In addition, the data were analyzed to identify items which provided more or less information within each specific country. Further, the clinical predictive value of the ARES-S, when scored using an MLIRT approach, was compared to a standard CTT-based score.

Anger is considered a universally experienced, basic emotion, and is known to be key in a number of pathologies. Efforts to create clinical measures of anger have yielded useful scales for the identification of pathological anger, such as the ARES-S (DiGiuseppe & Tafrate, 2011). The experience of anger, however, varies greatly across cultures, making it difficult to study cross-culturally and even harder to generalize results from one study to another (Kankaras & Moors, 2010). Conceptually, the easy solution would be to re-norm the ARES-S for each country in which it is administered. However, aside from being logistically and monetarily cumbersome, this approach would validate the ARES-S within each country but prevent statistically sound generalization or comparison of results across countries. Modern psychometric tools, such as IRT, provide a more elegant solution to this problem by allowing the information provided by each item to create a model where the response pattern of each individual estimates their true level of anger dimensions. When used with an MLIRT approach, where all participants are considered together in the first level, and then subdivided by country in the second level, MLIRT can account for the country-specific response patterns while still estimating

dimensions of anger (De Jong et al., 2007; King et al., 2004). In this way, individual MLIRT scores are then generalizable across countries since they are measuring the same latent trait.

### **Cross Country Item Discrimination**

Our study demonstrated that the initial construction of the ARES-S created items with good psychometric properties across a variety of countries. As such, we found that no singular item would benefit from being removed from the ARES-S, as each item displayed high discrimination in at least one of the countries analyzed.

Although it is beyond the scope of this study to hypothesize why each item performed differently in each country, it is possible that cultural norms in some countries prevented endorsement of the highest levels of anger, reducing the maximum possible information an item can provide. Taking a further look into the item response patterns in Vietnam specifically, illustrates the utility of MLIRT for analyzing a psychometric measure. For example, Item 1, part of the Internalizing Anger subscale, had the lowest average item discrimination due to a highly negative item discrimination in Vietnam. A negative item discrimination parameter here indicates that those with the highest level of Internalizing Anger were endorsing the opposite of what the item was designed to measure (i.e. indicating they were unlikely to “get angry in response to hurt feelings.”). Often, a negative item discrimination parameter indicates that the item is not actually measuring the same construct as the other items on the subscale. However, since the rest of the Internalizing Anger subscale items performed well in Vietnam, we should not conclude that the Internalizing Anger latent trait is fundamentally different. In this case, we believe this response pattern is more likely a reflection of participants’

misunderstanding of the item wording, and perhaps was interpreted as *expressing* anger in response to hurt feelings which would also explain why the item aligned more with the Externalizing Anger trait.

Interestingly, when removing Vietnam's highly negative value from the average, this item had the lowest average item discrimination across all countries. This possibly suggests that amending the wording of this item may increase its discriminative power.

Unlike the other countries in this analysis, Vietnam also had an item which had a near zero item discrimination, indicating that the item was essentially answered randomly. This was Item 14 (How many things set off your anger with friends?). This may suggest that the concept itself is not well understood in Vietnam or that the item was poorly translated. This conclusion is supported since in the original factor analysis of the Vietnam sample, this item did not load significantly onto any factor (Pzena, 2018). Item 14 had high discrimination in Italy and Serbia indicating that the item itself is valid.

Interestingly, however, the total information curves for Externalizing Anger were similar between the United States, Vietnam, and Italy despite Items 2 (When I am angry, I bully others to make them do what I want) and Item 5 (If somebody gets me angry, I will do whatever it takes to get even with that person) providing much more information in Vietnam. This is accounted for by Item 8 (How often do you get so angry that you lose your self-control?), Item 9 (How often do you get so angry with people that you push or shove them around?), and Item 11 (How often do you get so angry at someone that you find ways to not play or work with that person), which were much higher in Italy and the United States compared to Vietnam. This is because the items that provided the most information in the United States and Italy were designed to measure impulsivity, physical

aggression, and passive aggression, whereas the items that provided the most information in Vietnam, were designed to measure bullying and revenge. This demonstrates how MLIRT, when applied to the ARES-S, can be used to differentiate anger expression across cultures even if the level of the latent trait of Externalizing Anger is the same. This method of cross-cultural analysis, however, is not restricted to the ARES-S and can be applied to many psychometric measures that were normed in one country yet still provide meaningful and valid clinical data for individuals in another.

### **MLIRT Scoring**

Our study also demonstrated the utility of IRT scoring for use in cross-cultural research when compared to its CTT counterpart. Since the ARES-S is a clinical measure, it was important to not only look at country specific differences, but also at clinical data at the individual level. Although CTT-based scoring has long been the mainstay of scoring psychometric measures, we intended to use IRT scoring, a superior method, that yields a trait level estimate for an individual on a continuous scale that is independent of the number or difficulty of items. IRT is particularly attractive for cross-cultural research because it allows for the comparison of individuals on a common scale despite being part of distinct groups. However, in order to use IRT scoring, we had to first demonstrate that this method was non-inferior to CTT-based scoring.

Although theoretically superior, it is important that we confirmed that MLIRT-based scoring provided an improvement in clinical predictive power when directly compared to CTT scoring. Logistic regression analysis provided a useful positive control for evaluating the utility of MLIRT scoring in the United States sample where individuals had known diagnoses. This allowed for direct comparison of MLIRT to CTT scoring in



an ideal scenario for CTT-based scoring, in the country where the ARES-S was created and normed. MLIRT scoring performed slightly better than CTT scoring in predicting clinical diagnoses. This finding was important as it enabled us to use MLIRT scores for individuals in other countries, which provided confidence in comparing the ARES-S across countries.

Comparison of MLIRT and CTT in other countries was performed by using CBRS scores as a surrogate for clinical diagnoses. In these populations, where the ARES-S was not normed, MLIRT performed significantly better in predicting CBRS diagnostic scores and provided much higher explanation of variance. Surprisingly, when using MLIRT scores, all three dimensions of anger were significant predictors of CBRS scores, which was not true for CTT scoring. This indicates that each dimension of anger, as measured by the ARES-S, is useful for predicting clinical diagnoses but that CTT scoring is not able to accurately measure the dimensions in every country. Overall, MLIRT scoring provides more accurate scoring for individual participants. This is particularly relevant when considering psychometric measures across clinical populations where an individual's score directly impacts treatment planning and progress monitoring. Our study showed that we can accurately do this, not only within a United States sample, but across cultures as well.

### **Country Specific Anger Expression**

Since MLIRT scoring allows for all individuals to be scored on the same scale, comparison between individuals from different groups is possible. We intended to explore the universality of anger, as defined by the ARES-S, using MLIRT scoring. Interestingly, there was not a single country which demonstrated the highest level of all

three dimensions of anger. Rather, Vietnam had the highest levels of Externalizing Anger, Italy demonstrated the highest levels of Internalizing Anger, and the United States had the highest levels of the Extent of Anger. This is particularly interesting because it has been shown previously that all three of these dimensions are intercorrelated and it might have been expected that there would be an “Angriest” country overall. This finding highlights the importance of culture in modulating the expression of anger.

### **Limitations and Future Directions**

Several assumptions were necessary to perform the analyses here which may require further investigation to validate. We assumed a three-factor model of anger, and while this model has been validated in the United States it does not preclude the possibility of alternative models. Nevertheless, the method proposed and performed here can be applied to any number of dimensions, provided computing power is not limited. Additionally, we assumed that anger itself is an emotion which exists across all humans regardless of country. While this has been demonstrated previously, it should be remembered that this assumption is a requirement for the analysis to be performed. However, if this assumption were not true, our analysis would likely have produced an anomalous result.

While MLIRT has been shown here to be a useful tool in the cross-cultural study of anger, there are several limitations to the work presented. Firstly, while several countries were considered here, and proof of concept was demonstrated, a larger sample of countries is needed to better define the potential of MLIRT-based scoring for the ARES-S. It is especially important to obtain data from countries from distinct areas of the world, such as Africa, East Asia, and South America, which were not represented in this

sample. Secondly, while country of origin is certainly an important factor influencing the experience and expression of anger, there are likely additional factors which may be important to nest into the MLIRT model to better understand the dimensions of anger and make research more generalizable. Exactly how to identify which factors are important to add is not immediately clear. Countries themselves can be considered arbitrary; multiple cultures exist within one country and cultures cross country lines as well. Here, we have treated each country as equivalent to culture as the countries studied are geographically distinct. A more precise approach in defining culture, such as using the six dimension model proposed by Hofstede, may yield even better results (Hofstede & Bond, 1984).

Additionally, although the utility of MLIRT scoring has been demonstrated in the samples analyzed here, it is necessary to demonstrate the external validity these results in an independent cohort with blinded diagnoses. This would validate not only the utility of MLIRT scoring, but also the regression models that were built here.

### **Clinical Implications**

Accurate measures are critical to provide quality treatment and to assess and document changes in clinical parameters in patients over time (*Handbook of Psychological Assessment*, 2019). This is especially true for anger, which although is not a clinical diagnosis itself, is involved and contributes to many pathological states (Ahmed et al., 2012). While accurate and clinically useful measurements for anger, such as the ARES-S, have been developed and used in the United States, their application in other countries and cultures remained unclear. The work presented here indicates a path by which the ARES-S, and indeed other measures, can be applied in a cross-cultural manner which takes into account the different modulating factors that a culture may introduce,

but still accurately measures the same trait. Importantly, the MLIRT model used here does not require a normative sample in each country or prohibitively large sample sizes from any single country, as long as the total sample size across countries is sufficient. Beyond simply improving the scoring of the ARES-S across countries, MLIRT provided important information about how specific items varied from country to country giving insight into the way anger is differentially expressed between cultures. This phenomenon can be used to provide evidence-based modification of manualized and modular treatments to be tailored to specific cultures by addressing those areas which are more or less important within specific countries. Additionally, better understanding the differences in anger expression and regulation between countries will help facilitate cross-cultural discussion amongst practitioners by clarifying and highlighting the salient differences in their experiences with clients.

**Appendix A. Mplus Syntax**

DATA: FILE IS ForMplus4.dat;

VARIABLE: NAMES ARE u1-u17 g;

USEVARIABLES = u1-u17;

CATEGORICAL = u1-u17;

CLASSES = c(5);

KNOWNCLASS = c(g = 5 g = 3 g = 4 g = 6 g = 7);

MISSING = ALL(9);

ANALYSIS:

TYPE= MIXTURE;

PROCESSORS = 8;

ALGORITHM = INTEGRATION;

ESTIMATOR = MLR;

INTEGRATION = MONTECARLO;

ITERATIONS = 1000;

MODEL:

%OVERALL%

v1 BY u2\* u5 u7 u8 u9 u11 u12 u15;

v2 BY u1\* u3 u4 u6 u10 u13;

v3 BY u14\* u16 u17;

[v1-v3@0];

%c#1%

v1 BY u2\* u5 u7 u8 u9 u11 u12 u15;

v2 BY u1\* u3 u4 u6 u10 u13;

v3 BY u14\* u16 u17;

%c#2%

v1 BY u2\* u5 u7 u8 u9 u11 u12 u15;

v2 BY u1\* u3 u4 u6 u10 u13;

v3 BY u14\* u16 u17;;

%c#3%

v1 BY u2\* u5 u7 u8 u9 u11 u12 u15;

v2 BY u1\* u3 u4 u6 u10 u13;

v3 BY u14\* u16 u17;

%c#4%

v1 BY u2\* u5 u7 u8 u9 u11 u12 u15;

v2 BY u1\* u3 u4 u6 u10 u13;

v3 BY u14\* u16 u17;

%c#5%

v1 BY u2\* u5 u7 u8 u9 u11 u12 u15;

v2 BY u1\* u3 u4 u6 u10 u13;

v3 BY u14\* u16 u17;

OUTPUT: TECH1 TECH2 TECH4 TECH8;

PLOT: TYPE = PLOT3;

SAVEDATA:

RESULTS = results.sav;

**Appendix B. Model Fit Information**

Number of Free Parameters: 201

Information Criteria

Akaike (AIC): 86340.886

Bayesian (BIC): 87483.805

Loglikelihood

H0 Value: -42969.443

H0 Scaling Correction Factor: 1.060



### Appendix C. ARES-S Items

1. If someone hurts my feelings, I get angry.
2. When I am angry, I bully others to make them do what I want.
3. I feel jealous that life seems to be easier for other people.
4. I think that I cannot trust other people.
5. If somebody gets me angry, I will do whatever it takes to get even with that person.
6. How often do you get so angry that you start to breathe quickly?
7. How often do you get so angry that you swear or say nasty things to people?
8. How often do you get so angry that you lose your self-control?
9. How often do you get so angry with people that you push or shove them around?
10. How often do you get so angry about something that you cannot stop thinking about it?
11. How often do you get so angry at someone that you find ways to not play or work with that person?
12. How often do you get so angry at someone that you tell stories or lies about that person?
13. How often do you get angry but hold it in and do not let anyone know?
14. How many things set off your anger with friends?
15. When I have been angry, I have secretly tried to destroy the person with whom I was angry.
16. My family would say that I have had an anger problem for...
17. When I get angry at home, my anger usually lasts for...

**Appendix D. Item Characteristic Parameters**

	b1	b2	b3	b4
Item 1	-1.90	-0.60	1.02	2.32
Item 2	1.66	3.00	4.65	5.91
Item 3	0.01	1.39	2.86	4.19
Item 4	-0.73	0.72	2.56	3.84
Item 5	0.48	1.93	3.22	4.23
Item 6	1.02	2.58	4.01	5.45
Item 7	0.49	1.94	3.35	4.74
Item 8	1.40	3.00	4.46	5.86
Item 9	1.89	3.45	4.98	6.56
Item 10	-0.47	1.64	3.13	4.37
Item 11	0.46	2.07	3.31	4.42
Item 12	2.35	3.61	4.84	6.31
Item 13	-0.46	1.11	2.22	3.12
Item 14	-0.28	1.26	3.14	4.54
Item 15	2.02	3.15	4.59	5.81
Item 16	1.63	2.47	3.34	4.02
Item 17	0.08	1.93	3.25	4.25

## References

- Aarts, H., Ruys, K. I., Veling, H., Renes, R. A., de Groot, J. H. B., van Nunen, A. M., & Geertjes, S. (2010). The art of anger: Reward context turns avoidance responses to anger-related objects into approach. *Psychological Science, 21*(10), 1406–1410. <https://doi.org/10.1177/0956797610384152>
- Ahmed, A. G., Kingston, D. A., DiGiuseppe, R., Bradford, J. M., & Seto, M. C. (2012). Developing a clinical typology of dysfunctional anger. *Journal of Affective Disorders, 136*(1–2), 139–148. <https://doi.org/10.1016/j.jad.2011.09.008>
- Alaka Mani, T. L., Sharma, M. K., Omkar, S. N., & Nagendra, H. R. (2018). Holistic assessment of anger in adolescents—Development of a rating scale. *Journal of Ayurveda and Integrative Medicine, 9*(3), 195–200. <https://doi.org/10.1016/j.jaim.2017.04.012>
- American Psychiatric Association. (2013). *Diagnostic and Statistical Manual of Mental Disorders* (Fifth Edition). American Psychiatric Association. <https://doi.org/10.1176/appi.books.9780890425596>
- Andrich, D. (2010). Rasch Models. In P. Peterson, E. Baker, & B. McGaw (Eds.), *International Encyclopedia of Education (Third Edition)* (pp. 111–122). Elsevier. <https://doi.org/10.1016/B978-0-08-044894-7.00258-X>
- Ax, A. F. (1953). The physiological differentiation between fear and anger in humans. *Psychosomatic Medicine, 15*(5), 433–442.
- Barrett, E. L., Mills, K. L., & Teesson, M. (2013). Mental health correlates of anger in the general population: Findings from the 2007 National Survey of Mental Health and Wellbeing. *Australian & New Zealand Journal of Psychiatry, 47*(5), 470–476.

<https://doi.org/10.1177/0004867413476752>

- Briggs, J. L. (1975). *Never in anger: Portrait of an Eskimo family* (11. print). Harvard Univ. Press.
- Byrne, B. M. (2016). Adaptation of assessment scales in cross-national research: Issues, guidelines, and caveats. *International Perspectives in Psychology: Research, Practice, Consultation*, 5(1), 51–65. <https://doi.org/10.1037/ipp0000042>
- Cai, L., Choi, K., Hansen, M., & Harrell, L. (2016). Item Response Theory. *Annual Review of Statistics and Its Application*, 3(1), 297–321. <https://doi.org/10.1146/annurev-statistics-041715-033702>
- Carlozzi, B. L., Winterowd, C., Steven Harrist, R., Thomason, N., Bratkovich, K., & Worth, S. (2010). Spirituality, Anger, and Stress in Early Adolescents. *Journal of Religion and Health*, 49(4), 445–459. <https://doi.org/10.1007/s10943-009-9295-1>
- Cassielo-Robbins, C., & Barlow, D. H. (2016). Anger: The Unrecognized Emotion in Emotional Disorders. *Clinical Psychology: Science and Practice*, 23(1), 66–85. <https://doi.org/10.1111/cpsp.12139>
- Cavlazoglu, B., Erdogan, N., Paine, T., & Jones, M. (2013). Test Review: Anger Regulation and Expression Scale. *Journal of Psychoeducational Assessment*, 31(1), 84–88. <https://doi.org/10.1177/0734282912447762>
- Cheon, B. K., Melani, I., & Hong, Y. (2020). How USA-Centric Is Psychology? An Archival Study of Implicit Assumptions of Generalizability of Findings to Human Nature Based on Origins of Study Samples. *Social Psychological and Personality Science*, 11(7), 928–937. <https://doi.org/10.1177/1948550620927269>
- Cheung, F. M. (2012). Mainstreaming culture in psychology. *American Psychologist*,

67(8), 721–730. <https://doi.org/10.1037/a0029876>

Conners, C. K. (2010). *Conners Comprehensive Behavior Rating Scale: Manual*. Multi-Health Systems Inc.

Copeland, W. E., Angold, A., Costello, E. J., & Egger, H. (2013). Prevalence, Comorbidity, and Correlates of DSM-5 Proposed Disruptive Mood Dysregulation Disorder. *American Journal of Psychiatry*, *170*(2), 173–179.

<https://doi.org/10.1176/appi.ajp.2012.12010132>

Davidson, R. J., Ekman, P., Saron, C. D., Senulis, J. A., & Friesen, W. V. (1990).

Approach-withdrawal and cerebral asymmetry: Emotional expression and brain physiology: I. *Journal of Personality and Social Psychology*, *58*(2), 330–341.

<https://doi.org/10.1037/0022-3514.58.2.330>

de Jong, M. G., & Steenkamp, J.-B. E. M. (2010). Finite Mixture Multilevel Multidimensional Ordinal IRT Models for Large Scale Cross-Cultural Research.

*Psychometrika*, *75*(1), 3–32. <https://doi.org/10.1007/s11336-009-9134-z>

De Jong, M. G., Steenkamp, J.-B. E. M., & Fox, J.-P. (2007). Relaxing Measurement Invariance in Cross-National Consumer Research Using a Hierarchical IRT Model. *Journal of Consumer Research*, *34*(2), 260–278.

<https://doi.org/10.1086/518532>

del Barrio, V., Aluja, A., & Spielberger, C. (2004). Anger assessment with the STAXI-

CA: Psychometric properties of a new instrument for children and adolescents.

*Personality and Individual Differences*, *37*(2), 227–244.

<https://doi.org/10.1016/j.paid.2003.08.014>

Deng, L., & Chan, W. (2017). Testing the Difference Between Reliability Coefficients

Alpha and Omega. *Educational and Psychological Measurement*, 77(2), 185–203.  
<https://doi.org/10.1177/0013164416658325>

DiGiuseppe, R., & Tafrate, R. C. (2007). *Understanding anger disorders*. Oxford University Press.

DiGiuseppe, R., & Tafrate, R. C. (2011). *Anger Regulation and Expression Scale, Technical Manual*. Multi-Health Systems Inc.

Eckhardt, C. I., Samper, R. E., & Murphy, C. M. (2008). Anger Disturbances Among Perpetrators of Intimate Partner Violence: Clinical Characteristics and Outcomes of Court-Mandated Treatment. *Journal of Interpersonal Violence*, 23(11), 1600–1617. <https://doi.org/10.1177/0886260508314322>

Eckhardt, C., Norlander, B., & Deffenbacher, J. (2004). The assessment of anger and hostility: A critical review. *Aggression and Violent Behavior*, 9(1), 17–43.  
[https://doi.org/10.1016/S1359-1789\(02\)00116-7](https://doi.org/10.1016/S1359-1789(02)00116-7)

Ekman, P. (2005). Basic Emotions. In T. Dalgleish & M. J. Power (Eds.), *Handbook of Cognition and Emotion* (pp. 45–60). John Wiley & Sons, Ltd.  
<https://doi.org/10.1002/0470013494.ch3>

Fernandez, E., Day, A., & Boyle, G. J. (2015). Measures of Anger and Hostility in Adults. In *Measures of Personality and Social Psychological Constructs* (pp. 74–100). Elsevier. <https://doi.org/10.1016/B978-0-12-386915-9.00004-8>

Fernandez, E., & Johnson, S. L. (2016). Anger in psychological disorders: Prevalence, presentation, etiology and prognostic implications. *Clinical Psychology Review*, 46, 124–135. <https://doi.org/10.1016/j.cpr.2016.04.012>

Halle, T. G., & Darling-Churchill, K. E. (2016). Review of measures of social and

- emotional development. *Journal of Applied Developmental Psychology*, 45, 8–18.  
<https://doi.org/10.1016/j.appdev.2016.02.003>
- Handbook of Psychological Assessment*. (2019). Elsevier. <https://doi.org/10.1016/C2014-0-01970-3>
- Harris, M. (2014). *A Hebrew Translation of the Anger Regulation and Expression Scale (ARES): A Cross-Cultural Study of Anger in Israel and the United States Youth* [Unpublished doctoral dissertation]. St. John's University.
- Hart, E. J., & Ostrov, J. M. (2013). Functions of aggressive behavior and future functional impairment. *Early Childhood Research Quarterly*, 28(4), 683–691.  
<https://doi.org/10.1016/j.ecresq.2013.05.005>
- Her, D., & Joo, H. (2018). Individualism-Collectivism as the Dimension Distinguishing the Western from the Eastern Culture. *International Journal of Social Science Studies*, 6(7), 31. <https://doi.org/10.11114/ijsss.v6i7.3324>
- Hofstede, G., & Bond, M. H. (1984). Hofstede's Culture Dimensions: An Independent Validation Using Rokeach's Value Survey. *Journal of Cross-Cultural Psychology*, 15(4), 417–433. <https://doi.org/10.1177/0022002184015004003>
- Holodynski, M. (2013). The Internalization Theory of Emotions: A Cultural Historical Approach to the Development of Emotions. *Mind, Culture, and Activity*, 20(1), 4–38. <https://doi.org/10.1080/10749039.2012.745571>
- Igna, C. V., Julkunen, J., & Vanhanen, H. (2009). Anger expression styles and blood pressure: Evidence for different pathways. *Journal of Hypertension*, 27(10), 1972–1979. <https://doi.org/10.1097/HJH.0b013e32832f4f8f>
- Irwin, D. E., Stucky, B. D., Langer, M. M., Thissen, D., DeWitt, E. M., Lai, J.-S., Yeatts,

- K. B., Varni, J. W., & DeWalt, D. A. (2012). PROMIS Pediatric Anger Scale: An item response theory analysis. *Quality of Life Research, 21*(4), 697–706.  
<https://doi.org/10.1007/s11136-011-9969-5>
- Jasinski, M. J., Lumley, M. A., Latsch, D. V., Schuster, E., Kinner, E., & Burns, J. W. (2016). Assessing Anger Expression: Construct Validity of Three Emotion Expression-Related Measures. *Journal of Personality Assessment, 98*(6), 640–648. <https://doi.org/10.1080/00223891.2016.1178650>
- Johnson, T. R. (2003). On the use of heterogeneous thresholds ordinal regression models to account for individual differences in response style. *Psychometrika, 68*(4), 563–583. <https://doi.org/10.1007/BF02295612>
- Kankaras, M., & Moors, G. (2010). Researching measurement equivalence in cross-cultural studies. *Psihologija, 43*(2), 121–136.  
<https://doi.org/10.2298/PSI1002121K>
- Kerr, M. A., & Schneider, B. H. (2008). Anger expression in children and adolescents: A review of the empirical literature. *Clinical Psychology Review, 28*(4), 559–577.  
<https://doi.org/10.1016/j.cpr.2007.08.001>
- Kidwell, K. M., Van Dyk, T. R., Guenther, K. D., & Nelson, T. D. (2016). Anger and children's health: Differentiating role of inward versus outward expressed anger on sleep, medical service utilization, and mental health. *Children's Health Care, 45*(3), 342–358. <https://doi.org/10.1080/02739615.2015.1038680>
- King, G., Murray, C. J. L., Salomon, J. A., & Tandon, A. (2004). Enhancing the Validity and Cross-Cultural Comparability of Measurement in Survey Research. *American Political Science Review, 98*(1), 191–207.



<https://doi.org/10.1017/S000305540400108X>

Levenson, R. W. (2006). Blood, Sweat, and Fears: The Autonomic Architecture of Emotion. *Annals of the New York Academy of Sciences*, 1000(1), 348–366.

<https://doi.org/10.1196/annals.1280.016>

Matsumoto, D., Yoo, S. H., & Chung, J. (2010). The Expression of Anger Across Cultures. In M. Potegal, G. Stemmler, & C. Spielberger (Eds.), *International Handbook of Anger* (pp. 125–137). Springer New York.

[https://doi.org/10.1007/978-0-387-89676-2\\_8](https://doi.org/10.1007/978-0-387-89676-2_8)

Mauss, I. B., Levenson, R. W., McCarter, L., Wilhelm, F. H., & Gross, J. J. (2005). The Tie That Binds? Coherence Among Emotion Experience, Behavior, and Physiology. *Emotion*, 5(2), 175–190. <https://doi.org/10.1037/1528-3542.5.2.175>

Mooney, J. E. (2014). *Validation of the Conners Comprehensive Behavior Rating Scale (CBRS) and Conners Early Childhood (EC) with a Vietnamese population* [Unpublished doctoral dissertation]. St. John's University.

Moscoso, M. S., & Spielberger, C. D. (2011). Cross-cultural assessment of emotions: The expression of anger. *Revista de Psicología*, 29(2), 343–360.

Muris, P., Meesters, C., Morren, M., & Moorman, L. (2004). Anger and hostility in adolescents: Relationships with self-reported attachment style and perceived parental rearing styles. *Journal of Psychosomatic Research*, 57(3), 257–264.

[https://doi.org/10.1016/S0022-3999\(03\)00616-0](https://doi.org/10.1016/S0022-3999(03)00616-0)

Muthén, L. K., & Muthén, B. O. (2017). *Mplus User's Guide. Eighth Edition* (Vol. 7). Muthén & Muthén.

Neuhaus, R. (2013). *An Item Response Theory Analysis of the Anger Regulation and*

- Expression Scale* [Unpublished doctoral dissertation]. St. John's University.
- Panksepp, J., & Biven, L. (2012). *The archaeology of mind: Neuroevolutionary origins of human emotions* (1st ed). W. W Norton.
- Park, J., Kitayama, S., Markus, H. R., Coe, C. L., Miyamoto, Y., Karasawa, M., Curhan, K. B., Love, G. D., Kawakami, N., Boylan, J. M., & Ryff, C. D. (2013). Social status and anger expression: The cultural moderation hypothesis. *Emotion, 13*(6), 1122–1131. <https://doi.org/10.1037/a0034273>
- Petras, H., Buckley, J. A., Leoutsakos, J.-M. S., Stuart, E. A., & Ialongo, N. S. (2013). The Use of Multiple Versus Single Assessment Time Points to Improve Screening Accuracy in Identifying Children at Risk for Later Serious Antisocial Behavior. *Prevention Science, 14*(5), 423–436. <https://doi.org/10.1007/s11121-012-0324-z>
- Potegal, M., Stemmler, G., & Spielberger, C. D. (Eds.). (2010). *International handbook of anger: Constituent and concomitant biological, psychological, and social processes*. Springer.
- Pzena, R. (2018). *Validation of Two American Anger Scales with a Vietnamese Youth Population* [Unpublished doctoral dissertation]. St. John's University.
- Quinn, C. A., Rollock, D., & Vrana, S. R. (2014). A test of Spielberger's state-trait theory of anger with adolescents: Five hypotheses. *Emotion, 14*(1), 74–84. <https://doi.org/10.1037/a0034031>
- Revelle, W. R. (2017). *psych: Procedures for Personality and Psychological Research*. <https://www.scholars.northwestern.edu/en/publications/psych-procedures-for-personality-and-psychological-research>
- Rowe, R., Costello, E. J., Angold, A., Copeland, W. E., & Maughan, B. (2010).

- Developmental pathways in oppositional defiant disorder and conduct disorder. *Journal of Abnormal Psychology*, 119(4), 726–738.  
<https://doi.org/10.1037/a0020798>
- Schieffelin, E. L. (1983). Anger and Shame in the Tropical Forest: On Affect as a Cultural System in Papua New Guinea. *Ethos*, 11(3), 181–191. JSTOR.
- Sell, A. (2005). *Regulating Welfare Tradeoff Ratios: Three Tests of an Evolutionary-computational Model of Human Anger*. University of California, Santa Barbara.
- Smajlaj, A. (2014). *Albanian Translations of the Strengths and Difficulties Questionnaire (SDQ) and the Anger Regulation and Expression Scale (ARES): A Cross-Cultural Study of Albania and the United States* [Unpublished doctoral dissertation]. St. John's University.
- Spielberger, C. D. (1999). *State-trait anger expression inventory-2*. Psychological Assessment Resources.
- Steele, R. G., Legerski, J.-P., Nelson, T. D., & Phipps, S. (2007). The Anger Expression Scale for Children: Initial Validation among Healthy Children and Children with Cancer. *Journal of Pediatric Psychology*, 34(1), 51–62.  
<https://doi.org/10.1093/jpepsy/jsn054>
- Stucky, B. D. (2009). *Item response theory for weighted summed scores* [Master's Thesis, University of North Carolina at Chapel Hill].  
<https://cdr.lib.unc.edu/concern/dissertations/7m01bk77g?locale=en>
- van Lier, H. G., Siemons, L., van der Laar, M. A. F. J., & Glas, C. A. W. (2018). Estimating Optimal Weights for Compound Scores: A Multidimensional IRT Approach. *Multivariate Behavioral Research*, 53(6), 914–924.

<https://doi.org/10.1080/00273171.2018.1478712>

Wolfe, R., & Firth, D. (2002). Modelling subjective use of an ordinal response scale in a many period crossover experiment. *Journal of the Royal Statistical Society Series C*, 51, 245–255. <https://doi.org/10.1111/1467-9876.00267>

## Vita

Name	<i>Tovah Weinrib</i>
Baccalaureate Degree	<i>Bachelor of Arts, Case Western Reserve University, Cleveland Major: Psychology</i>
Date Graduated	<i>May, 2013</i>
Other Degrees and Certificates	<i>Master of Arts, St. John's University, Queens, Major: Clinical Psychology</i>
Date Graduated	<i>May, 2019</i>