8-11-2020

Measuring the Thresholds of Extremism: Testing for Measurement Invariance Between Muslim Converts and Muslim Non-Converts of Radicalism With an Ordinal Model

Ari Fodeman

MEASURING THE THRESHOLDS OF EXTREMISM: TESTING FOR MEASUREMENT INVARIANCE BETWEEN MUSLIM CONVERTS AND MUSLIM NON-CONVERTS OF RADICALISM WITH AN ORDINAL MODEL

by

ARI FODEMAN

Under the direction of John Horgan, Ph.D.

ABSTRACT

Assessment of the risk of radicalism remains underdeveloped by terrorism researchers in part due to difficulties in standardized measurement. The Activism and Radicalism Intention Scales (ARIS) by Moskalenko and McCauley (2009)— an already robust and widely applicable tool—will be further improved upon and standardized in this proposed project using data previously collected via a grant-funded research project. Specifically, the Measurement Equivalence/Invariance of the latent factor Radicalism is tested for using two radicalism-relevant groups—Muslim converts and Muslim non-converts. Ordinal logistic regression, rather than a typical linear regression, is used since it is a more accurate statistical estimation of Likert survey data, like that found in the ARIS. These methods are the first of many steps to set a standard for more accurate measurement for surveying intervention outcomes and group comparisons in Countering Violent Extremism (e.g., at risk versus control or pre- versus post-treatment).

INDEX WORDS: Muslim, Radicalism, ARIS, CFA, Ordinal logistic, Measurement invariance

MEASURING THE THRESHOLDS OF EXTREMISM: TESTING FOR MEASUREMENT INVARIANCE BETWEEN MUSLIM CONVERTS AND MUSLIM NON-CONVERTS OF RADICALISM WITH AN ORDINAL MODEL

by

ARI FODEMAN

A Thesis Submitted in Partial Fulfillment of the Requirements for the Degree of

Master of the Arts

in the College of Arts and Sciences

Georgia State University

2020

MEASURING THE THRESHOLDS OF EXTREMISM: TESTING FOR MEASUREMENT INVARIANCE BETWEEN MUSLIM CONVERTS AND MUSLIM NON-CONVERTS OF RADICALISM WITH AN ORDINAL MODEL

by

ARI FODEMAN

Committee Chair: John Horgan

Committee: Anthony Lemieux

Kevin Swartout

Lee Branum-Martin

Electronic Version Approved:

Office of Graduate Studies

College of Arts and Sciences

Georgia State University

August 2020

ACKNOWLEDGEMENTS

I want to thank my family, friends, and mentors for supporting me personally and academically while preparing this thesis proposal. I want to give a special thank you first and foremost to my thesis committee chair, supervisor, and role model Dr. John Horgan; He has given me the room to grow into the kind of scientist I want to be, and he models a level of tenacity, productivity, and professionalism that I aspire to. He has helped me above all else to focus on honing my writing abilities, especially my wordiness that I have clearly forgone in this Acknowledgement. I cannot continue without saying that he graciously allowed me to use his lab's data for this thesis, without which I could not practice these more advanced statistical techniques. On that note, I want to thank Dr. Lee Branum-Martin profusely as well, not only for his excellent classes in Structural Equation Modeling (SEM) through which I was able to learn the skills for this thesis analysis, but his advice beyond that as a statistician in far too many an email or meeting. His patience, knowledge, and deep love of teaching is something I wish to emulate as well in my career. My two other committee members, Drs. Kevin Swartout and Anthony Lemieux, I asked to sit on my committee as well because of their exemplary work as scientists and relevant knowledge to this thesis work; I trust that I can learn a great deal from them both as well. I trust their familiarity in researching difficult topics like radicalism and violence to aid me in this thesis review. Lastly, I could not have sculpted this manuscript to its present state without the help of both the Writing Studio and the Research and Data Services teams—most notably Dr. Raeda Anderson whose knowledge of CFA and statistical writing generally was an invaluable supplement to that provided by Dr. Branum-Martin.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF TERMS & ACRONYMS

LIST OF SYMBOLS

There are many different symbol conventions for parameters in CFA between fields, statistical approaches (e.g., Structural Equation Modeling (SEM) versus Item Response Theory), and even textbooks. The figures in this manuscript utilize a combination of conventions. Below is a table adapted from Dr. Lee Branum-Martin's "Counting Degrees of Freedom" handout for his Stats III class on SEM. These symbols will be used in the proposed study thereof.

Type	Sub-type	Parameter	Name	Definition	Notes
Regression (single-headed arrows)	Means	ν	nu	Observed mean	τ used for ordinal Thresholds
		α	alpha	Latent mean	
	Factor loadings	λ, Λ	lambda	observed on latent regression: measurement	
		β, B	beta	latent on latent regression	
Covariances (double-headed arrows)		θ, Θ	theta	Observed variance-covariance	
		ψ, Ψ	psi	Latent variance-covariance	ϕ often used instead
Scores (boxes & circles)		X		Observed variable score	
		η	eta	latent factor score (implicit)	Capitalized & italicized letters (e.g., *A* or *B*) often used to distinguish different factors
		ε	epsilon	observed variable residual score (implicit)	*E* often used instead

1 INTRODUCTION

The psychology of terrorism is not well understood (Victoroff, 2005). Terrorism prevention remains particularly underdeveloped due in part to the difficulty in measuring individual risk of engagement and therefore Countering/Preventing Violent Extremism (C/PVE) programs' risk reduction effectiveness (Scarcella, Page, & Furtado, 2016; Gielen, 2017; Mirahmadi, 2016; Sarma, 2017; van den Berg, van Hemert, & van Vliet, 2018; Veldhuis & Kessels, 2013). An important part of risk assessment is its accuracy. This proposed thesis project aims to improve the measurement of the Activism and Radicalism Intention Scales (ARIS), developed by Moskalenko & McCauley (2009), one of the more robust and widely applicable tools used to assess radicalism (Scarcella, Page, & Furtado, 2016). Specifically, Measurement Invariance is tested between two groups—Muslim converts and Muslim non-converts—who are over- and under-represented respectively among Muslim terrorists in the United States (U.S.) and some Western European nations (Schuurman, Grol, & Flower, 2016). While there may be other explanations for Muslim converts' overrepresentation among Muslim terrorists than greater intentions of engaging in radicalism, the link between radicalism intentions and terrorism warrants this cautious, caveated analysis. It will probably be a long time, if ever, before science can uncover the necessary, sufficient, and discrete causes and pathways to radicalism generally and terrorism particularly; however, by adopting these contemporary measurement methods, researchers may more accurately compare populations' radicalism intentions. Using methods like these, researchers and practitioners interested in radicalism risk assessment can work towards more rigorous standards of measurement. This is particularly apt considering how the use of poor statistical techniques, if any, has been criticized in the field (Silke, 2001; Rich & Hoffman, 2004;

Ross, 2004; LaFree & Ackerman, 2009; Neumann & Kleinmann, 2013; Sageman, 2014; Schuurman, 2018; Stampnitzky, 2010).

2 BACKGROUND

2.1 Clarifying Terms

2.1.1 Terrorism

Terrorism research has had its fair share of definitional differences, even among widely used databases[1] (Schmid, 2004; Young & Findley, 2011). While the definition of terrorism is contested around the world, both legally and academically (Ganor, 2002; Schmid, 2011, 2012), it is most widely defined as violence, or the threat thereof, against noncombatants[2] (Richardson, 2006) to send a message to engender political change (Hoffman, 2006). The political motivation in terrorism can be difficult to operationalization. Motive is particularly difficult to ascertain, let alone classify, since it is a question not of *action*—something tangible and therefore more verifiable—but rather of *intention*. Regardless of what qualifies as terrorism, scholars agree that it constitutes the most extreme ends of what some call the political mobilization spectrum.

[1] Selectivity bias is a major concern in terrorism datasets due to different operationalizations (Mahoney, 2018) and ex/inclusion criteria fidelity (Chermak, Freilich, Parkin, & Lynch, 2012).

[2] Includes both civilians and, arguably, military personnel who are not, when attacked, engaged in military operations (e.g., Nidal Hasan's mass shooting at Fort Hood), or only indirectly (e.g., IEDs against military convoys). The latter distinction is difficult to tease apart from guerrilla warfare, however, depending on the law (Ganor, 2002).

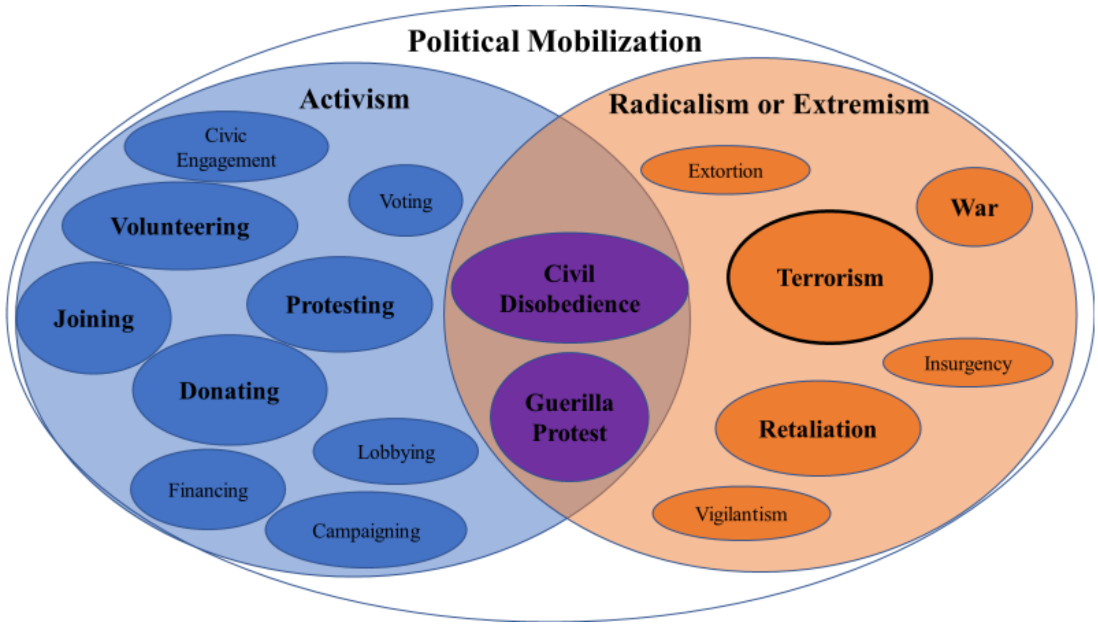

Figure 1. Venn Diagram of Political Mobilization

2.1.2 Contextualizing Terrorism: Activism, Radicalism, & Mobilization

Terrorism and related behaviors have been classified a number of different ways. One particularly helpful conception is McCauley and Moskalenko's Political Mobilization definitions (2009) and Two Pyramid Model (2017). Political mobilization, as defined by McCauley and Moskalenko, entails "increasing extremity of beliefs, feelings, and actions in support of intergroup conflict" (2009). Behaviors captured by this definition can be classified as activism and/or radicalism (see Figure 1 above for a visualization). Scholars use the term activism, or "legal and non-violent political action" (Moskalenko & McCauley, 2009), to describe various behaviors, including volunteering, voting, protesting, and even lobbying, campaigning, or financing[3]. Radicalism, conversely, is the class of political mobilization that is both illegal *and* violent. This includes terrorism, as well as other types of political violence (e.g., war or insurgency). In the fields of terrorism research[4], extremism is often used

[3] Scholars also refer to these behaviors as civic or political engagement (Abdi et al., 2015).
[4] See Bötticher (2017) for a discussion of use, misuse, disagreements and consensus in the field.

synonymously with radicalism, although the modifier “violent” is often given to emphasize violent action (i.e., “violent extremism” or VE). There are also some behaviors that are non-violent, but illegal, such as civil disobedience or guerilla protest that fall between activism and radicalism. Moskalenko and McCauley (2017) arrange those who engage in political mobilization along a spectrum of extremity that, in terms of frequency, befits a pyramid (see Figure 2 below). That is to say that, the more extreme the behavior, the fewer people engage in it. Levels of extremism can be skipped: engaging at one level is not dependent upon engaging at any other levels (visualized via purple two-headed arrows in Figure 2 below). The Pyramid Model is juxtaposed to linear models of terrorism engagement, such as Moghaddam’s Staircase (2005; see Hafez & Mullins, 2015 for discussion). It is this movement from inertia or activism into radicalism or terrorism that C/PVE programs attempt to prevent, and movement in the opposite direction that they attempt to promote.

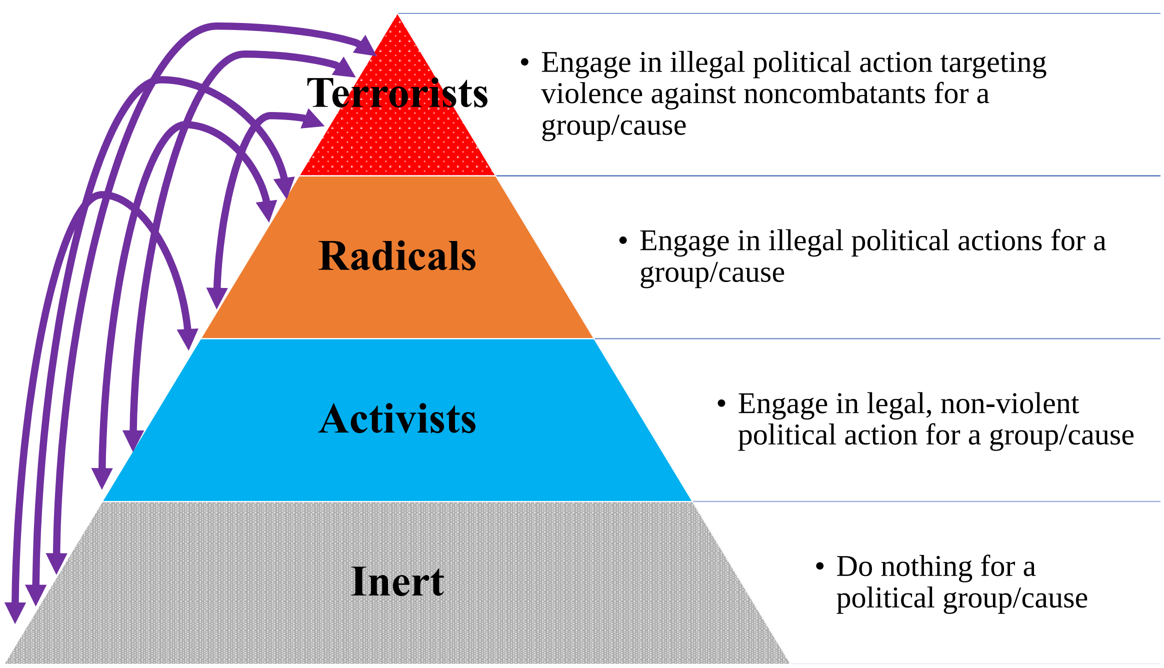

Figure 2. Political Action Pyramid (adapted from Moskalenko & McCauley, 2017).

2.2 Measurement Problems & Solutions in Radicalism Research

2.2.1 Lack of Radicalism Data

C/PVE often focuses on identifying at-risk individuals based on factors assumed relevant to radicalism engagement (Weine, 2016; van den Berg, van Hemert, & van Vliet, 2018). These factors, however, are understudied, and an evidence base for the field is still being developed (Zeiger, et al., 2015). For many years, there was little research into VE risk assessment (Scarcella, Page, & Furtado, 2016; Altier, Thoroughgood, & Horgan, 2014). Most risk assessments have been conducted on known terrorism offenders to inform their sentencing, rehabilitation, and reintegration (Sarma, 2017). However, some risk assessment is aimed at preventing movement into radicalism or terrorism from inertia, activism, or radicalism. All of these risk assessments are either quantitative, discrete checklists with actuarial potential or qualitative, structured professional judgements (SPJ). In either case, these assessments largely focus on risks of imminent and acute threats (i.e., risk of committing terrorism within months or weeks of assessment) rather than solely radical attitudes, intentions of engaging in radicalism more broadly, or the more innocuous activism. There is a dearth of data on these risk assessments and correlated outcomes (Gielen, 2017; van den Berg, van Hemert, & van Vliet, 2018), particularly non-terrorism outcomes. This is unsurprising given that these assessmenets are rarely transparent (Horgan & Braddock, 2010). Indeed, Veldhuis and Kessels (2013) bemoaned the need for not only more data, but more *structured* collection, with standardized tools. This would allow comparisons across individuals, programs, time, and studies (i.e., meta-analysis). Even if the data were available, however, a medical-model approach may not be possible given what we know (or, more importantly, what we do *not*) about radicalism.

2.2.2 Difficulty of Measuring Radicalism

Few of the already rare quantitative radicalism measures and risk assessments thereof that have been conducted are based on scientific research, and those that are have an insufficient evidence base to predict who will commit VE (Sarma, 2017; Horgan & Taylor, 2011). The difficulty in evidence gathering and modeling arises from issues of causality and sample size (*n*). Radicalization and VE are the products of a multitude of neither sufficient nor necessary risk factors that vary widely across individuals. Measuring and modeling these complex, dynamic, multi- and equi-final processes (Horgan, 2008) is difficult given the low base rate of radicalism generally, let alone terrorism specifically (Gill, Horgan, Corner, & Silver, 2016)—not to mention the even rarer recidivism rates among former terrorists. Low base rates make actuarial science and establishing discriminant validity of these assessments a Herculean task in comparison to the already difficult assessment of risk of violence generally (Pressman, 2009). Regardless of evidence, however, intervention programs operate across the globe, affecting thousands[5]. Given that reality, researchers ought to guide practitioners towards standardized tools for measuring radicalism starting with the best candidate tools for improvement.

2.2.3 Measuring Radicalism Better: The ARIS

In order to find the best and most relevant tools, Scarcella et al. (2016) conducted a systematic review of articles concerning measures of extremism, terrorism, fundamentalism, radicalization, and authoritarianism. They aggregated 835 articles from 20 databases across 5 subject areas (law, medicine, psychology, sociology and politics). Following PRISMA[6] standards

[5] E.g., The UK Channel Project had >8,000 referrals in its first 10 years, of which ~43% were assessed to be 'at risk of radicalization' and received support (Sarcella, Page, & Furtado, 2016).
[6] The "Preferred Reporting Items for Systematic Reviews & Meta-Analysis" (PRISMA) guidelines as described by the organization of the same name.

(Moher, Liberati, Tetzlaff, Altman, & Group, 2009), they narrowed this pool of material down to 37 articles covering 16 different tools measuring the identified concepts. Among those 16 tools, Scarcella and colleagues ranked the Activism and Radicalism Scales (ARIS; Moskalenko & McCauley, 2009) as one of the best measures. This was based on methodological markers such as explanation of theory, methods, and sample selection, as well as psychometric properties including readability, cultural translation, and construct and internal validity. The ARIS is a survey of 8 to 10 questions (see Table 1 below), depending on the version (see section 2.3.3 below) gauging people's intentions of engaging in activism (i.e., the Activism Intentions Scale, AIS) and radicalism (i.e., the Radicalism Intentions Scale, RIS). No previous scale captured both legal and illegal political behaviors together.

Table 1. ARIS Item Names & Survey Questions

AIS	**Join**	I would join/belong to an organization that fights for my group's political & legal rights
	Donate	I would donate money to an organization that fights for my group's political & legal rights
	Volunteer	I would volunteer my time working (i.e., write petitions, distribute flyers, recruit people, etc.) for an organization that fights for my group's political & legal rights
	Protest	I would travel for one hour to join in a public rally, protest, or demonstration in support of my group
RIS	**Illegal Group**	I would continue to support an organization that fights for my group's political & legal rights even if the organization sometimes resorts to violence
	Violent Group	I would continue to support an organization that fights for my group's political & legal rights even if the organization sometimes breaks the law
	Violent Protest	I would participate in a public protest against oppression of my group even if I thought the protest might turn violent
	Police Defense	I would attack police or security forces if I saw them beating members of my group
	War	I would go to war to protect the rights of my group

	Retaliation	I would retaliate against members of a group that had attacked my group, even if I couldn't be sure I was retaliating against the guilty party

Note. Response scale 7-point Likert-type: "Strongly Disagree" (1) to "Strongly Agree" (7).

As seen in Table 1 above, each ARIS survey item references "my group." Depending on the context of the study, participants can either A) choose the group that they report is most important to them—whether familial, collegiate, national, ethnic, religious, or political—or B) a group can be assigned to them based on an already acknowledged group membership. These items cover a range of activism and radicalism activities to varying degrees of extremity. The radicalism items are phrased with less specificity than the activism questions in order to counteract social desirability bias in respondents. Respondents answer each item on a 7-point Likert-type scale, from "Strongly Disagree" (1) to "Strongly Agree" (7), from which a composite score for each factor is calculated. Specifically, the AIS is composed of the first 4 questions (resulting in a range from 4 ("Strongly Disagree" for every item) to 28 ("Strongly Agree" for every item), and the RIS is composed of the latter 4 or 6, depending on the version (see discussion in section 2.3.3 below). In their three initial trials of the ARIS, Moskalenko and McCauley (2009) found that past activism and radicalism behaviors predicted both AIS and RIS scores respectively, connecting intentions to behaviors. After establishing this relationship, Moskalenko and McCauley created both a past actions and future intentions version of the survey questions. The past actions version, they argue, is useful for avoiding social desirability bias that might artificially lower scores, particularly on the RIS, versus future intentions.

While published research thus far on the ARIS has primarily been used in only laboratory settings, ARIS levels have been correlated with a variety of social-psychological factors that are theoretically relevant to the development of activism and radicalism. First and foremost are different types and levels of ingroup identification. These include birth and host country, religion (specifically Christian or Muslim), political affiliation, clan, tribe, or even family (Moyano & Truijilo, 2014; Ellis, et al., 2016; Moskalenko & McCauley, 2009). A wide array of other correlates have been found, such as relative deprivation (Chikhi, 2017); perceived oppression and religious extremism (Moyano & Truijilo, 2014); Social Dominance Orientation and Right-Wing Authoritarianism (Lemieux, Kearns, Asal, & Walsh, 2017); and/or exposure to personal trauma and strong social bonds (Ellis, et al., 2016). Some of these correlates have even been manipulated experimentally to predict ARIS outcomes (Lemieux, Kearns, Asal, & Walsh, 2017). However, none of these studies reconsidered factor strength or measurement bias after Moskalenko and McCauley's original paper (see section 2.3.3 below for their analysis). Researchers should as the ARIS demonstrates several peculiarities that warrant more careful consideration—namely, non-normal, positively skewed Likert survey responses. This research project aims to address those peculiarities and measure them appropriately. In order to do so, we will first review, in brief, the primary means of developing survey index scales: Factor Analysis.

2.3 Evaluating Survey Measures: Factor Analysis and Bias

2.3.1 Factor Analysis

Factor Analysis (FA) is a technique used to determine if a set of indicators measure a single cohesive concept. More specifically, FA is a statistical method for testing whether variation in observed, correlated variables (indicators, such as individual survey questions) can be better explained by a fewer number of latent variables (factors, such as activism or radicalism). For example, FA is a way of modeling correlations between several indicators such as salary, assets, job skill, and education as a single factor of socioeconomic status. In Confirmatory Factor Analysis (CFA), factors constructed *a priori* are modeled by constraining indicators (e.g., X_3 or Volunteering from the ARIS in Figure 3 below) to be predicted by their predetermined factors (e.g., Activism or η_1), means (e.g., ν_3) and variance or error (e.g., θ_3).

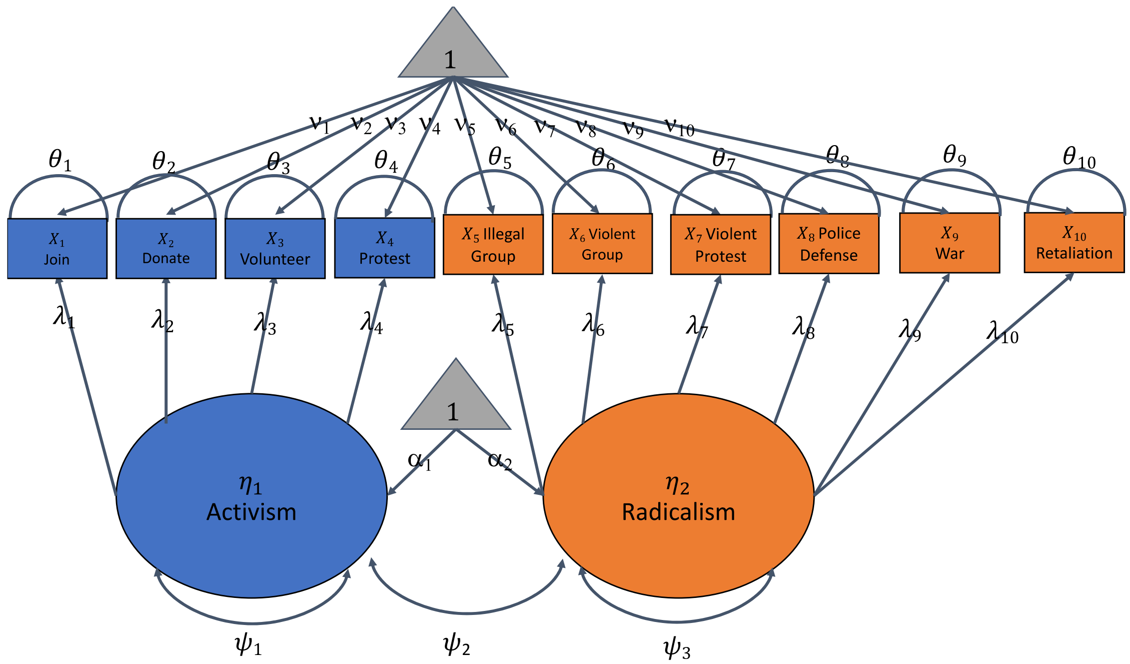

Figure 3. ARIS Structural CFA Model Diagram.
Key: Squares = manifest variables, i.e. indicators ($X's$); circles = latent variables ($\eta's$).
Key: λ = factor loadings; θ = indicator variances; ψ = latent factor (co)variances;
ν = indicator means; α = latent means.

CFA improves accuracy and specificity over typical composite factor scoring (i.e., adding up indicator scores), as the latter makes several often false assumptions: A) the theorized factor structure is real as modeled and detectable as measured; B) each item has equal weight[7] (λ's) or different loadings will not affect factor means (α); and C) individual items' variances (θ's) are equal or will not affect factor variances (ψ's). CFA, on the other hand, assumes none of the above, and can be used to test those assumptions. Violations of these assumptions threaten the internal validity of the latent construct. Save for Decker and Pyrooz's 2019 study (discussed in section 2.3.3 below), all studies employing the ARIS subsequent to Moskalenko and McCauley's

[7] i.e., items are often added linearly unless one item is given different weight than another item *a priori;* the standard in ARIS research thus far has been to composite scores linearly.

2009 development of it have made those assumptions. While Decker and Pyrooz tested the factor structure in a more robust way than Moskalenko and McCauley originally conducted, they still assumed distributional qualities of the ARIS that would bias estimates of λ and ψ—namely, that the ARIS items are continuous and normally-distributed, when in fact they are ordinal with typically positive skew. Failing to meet these assumptions threatens the internal validity of the ARIS as measured in all of the aforementioned studies, let alone any ARIS score comparisons.

When comparing groups, there is an additional assumption called Measurement Equivalence/Invariance (ME/I) that, if violated, threatens *external* validity. ME/I is the condition that groups respond to all parts of a measure the same way (i.e., that the test is not biased against one group or another), such that all the parameters just discussed are equal across groups (e.g., that Protesting is as good a predictor of Activism in an autocracy as it is in a democracy with the right to assemble). Researchers ought to concern themselves with ME/I to avoid biased survey results when comparing groups on factor scores. ME/I is explained in 2.3.2 below.

2.3.2 Measurement Equivalence/Invariance.

ME/I is the condition under which groups are measured (e.g., via survey) in the same way. Often, surveyors are interested in determining whether there are significant differences between groups on a factor score such as RIS, but only the factor score—not on any individual part of the survey. In order to accurately compare factor scores, surveyors test for ME/I between groups on all of the aforementioned model parameters—i.e., indicator means ($\boldsymbol{\nu}$) and (co)variances (θ), plus factor loadings ($\boldsymbol{\lambda}$), means (α) and (co)variances (ψ). If only factor means ($\boldsymbol{\alpha}$'s) significantly vary between groups, then those groups can be said to truly differ; if, however, there is not ME/I—if one or more other parameters significantly differ between groups—then the factor mean differences may be biased.

The order and type of ME/I tests are largely agreed upon in the literature for continuous indicators (Vandenberg & Lance, 2000). First, Configural Invariance: the factor structure for one group is equivalent for another ($\boldsymbol{Model\ structure_{Group1} = \ Model\ structure_{Group2}}$). For example, protest may be a sign of activism in a democracy, but a sign of radicalism in an autocracy; thus, the configuration of the ARIS is noninvariant across countries. Second, Metric Invariance: the factor loadings for one group equal those of another (Configural Invariance & $\boldsymbol{\lambda_{Group1} = \ \lambda_{Group2}}$). For example, donating might be a stronger predictor of activism for the middle class than for the lower or upper class because the former does not have the capitol to donate, and the latter may donate for status or tax breaks rather than for activism; thus, the metrics of activism (the indicators) are noninvariant across class. Third, Scalar Invariance: the indicator means for one group are equal to those of another ($\boldsymbol{Metric\ Invariance}$ & $\boldsymbol{\nu_{Group1} = \ \nu_{Group2}}$). For example, the average number of violent protests may be less for environmentalist radicals than for anti-abortion radicals because the former group does not have an equally fervent opposition with whom to clash; therefore, the scale of violent protests is noninvariant across political movements. The last widely agreed upon step is testing for Factor Means Invariance: whether the factor means are equal across groups (Scalar Invariance & $\boldsymbol{\alpha_{Group1} \neq \ \alpha_{Group2}}$). Factor means often are assumed to be unequal, as in this study where the literature would suggest that Muslim converts ought to demonstrate higher intentions of radicalism engagement than Muslim non-converts. Note that at each ME/I stage, the previously equality-constrained parameters continue to be restricted, resulting in increasingly stringent models (i.e., more and more parameters are modeled as equivalent across groups). If at any stage restricted models fit significantly poorer than the previously less restricted one, then there is not complete ME/I, but rather Configural, Metric, Scalar, or Factor Mean Noninvariance. Noninvariance suggests that groups may respond differently to one or more parts of a survey (i.e., there is Differential Item Functioning, or DIF). The parameters with DIF can be estimated separately between groups for Partial Invariance (PI); however, testing for PI is beyond the scope of this thesis.

2.3.3 ARIS Measurement Specifications

When Moskalenko and McCauley created the ARIS (2009), they tested 10 items—four for the AIS, *six* for the RIS. In their first sample (140 American college students), the researchers found poor fit for two RIS items—“War” and “Retaliation,” items nine and 10 in

Table 2 below—via either Principal Components Analysis (PCA) or[8] Exploratory Factor Analysis (EFA). PCA and EFA test specifically how many factors, if any, most strongly correlate[9] with a respective set of indicators (see Kline (2016), pp. 191-194 for more explanation of EFA). Moskalenko and McCauley (2009) found that, as expected, the two-factor model of AIS and RIS fit better than a one or three factor model, and that the indicators aligned with the AIS versus RIS dimensions as hypothesized (see λ values outside parentheses in

[8] PCA and EFA are different approaches, but Moskalenko and McCauley (2009) refer in their article to "*Exploratory* Principal Components Analysis" (emphasis added), which is unclear. Both PCA (Snook, Branum-Martin, & Horgan, In Review) and EFA assume normal distributions, which the ARIS items do not have.
[9] i.e., as in the factor loadings ($\lambda's$) of indicators regressed onto the latent factor(s).

Table 2 below), with the exception of items nine and 10. They excluded those items, ran another PCA or EFA (see λ's in parentheses below), and still found good fit for the remaining items. They therefore chose to include those eight in their final scales (see bolded items below). They found that the index scales (composite means) were moderately correlated (r = .42). They also found the scales to be reliable (Cronbach's alphas of 0.86 and 0.83 each).

*Table 2. Original ARIS Likert Scale Items & EFA/PCA Factor Loadings (**λ**)*

Questions	Item	AIS λ	RIS λ
1. I would join/belong to an organization that fights for my group's political and legal rights	**Join**	**.84 (.89)**	.00 (.00)
2. I would donate money to an organization that fights for my group's political and legal rights	**Donate**	**.86 (.88)**	.00 (.00)
3. I would volunteer my time working (i.e. write petitions, distribute flyers, recruit people, etc.) for an organization that fights for my group's political and legal rights	**Volunteer**	**.90 (.88)**	.00 (.00)
4. I would travel for one hour to join in a public rally, protest, or demonstration in support of my group	**Protest**	**.76 (.81)**	.15 (.38)
5. I would continue to support an organization that fights for my group's political and legal rights even if the organization sometimes breaks the law	**Illegal Group**	.24 (.44)	**.67 (.73)**
6. I would continue to support an organization that fights for my group's political and legal rights even if the organization sometimes resorts to violence	**Violent Group**	.00 (.00)	**.85 (.82)**
7. I would participate in a public protest against oppression of my group even if I thought the protest might turn violent	**Violent Protest**	.12 (.43)	**.78 (.86)**
8. I would attack police or security forces if I saw them beating members of my group	**Police Defense**	.21 (.00)	**.83 (.88)**
9. I would go to war to protect the rights of my group	War	0.00 (n.a.)	0.48 (n.a.)
10. I would retaliate against members of a group that had attacked my group, even if I couldn't be sure I was retaliating against the guilty party	Retaliation	0.00 (n.a.)	0.39 (n.a.)

Note.

Table 2 above is adapted from Moskalenko and McCauley's Table 2 (2009).
Note. n = 140; bold items selected for 8-item version; 10-item EFA loadings (λ) outside parentheses, 8-item inside.

However, far more similar loadings for all *six* RIS factors, and indeed higher loadings overall, emerged in a sample collected by Dr. John Horgan's lab (Fodeman, Snook, & Horgan, In Press; Fodeman, Snook, & Horgan, 2020) than in Moskalenko and McCauley's (2009) original sample (see Figure 4 below). While perhaps items nine and 10 capture more extreme intentions than the other four RIS items, theoretically they fall no less under radicalism, which might explain why Moskalenko and McCauley originally included these items[10]. They might not have detected much endorsement of those items as there may have been a floor effect[11]. The Horgan lab sample might have higher and more varied endorsement of those items, and higher correlations thereof with the other RIS items, for two reasons: 1) while Moskalenko and McCauley sampled only 140 individuals, the Minerva *n* was 356, providing more opportunities for detecting endorsement of extreme/rare RIS item thresholds[12]; 2) while the groups that Moskalenko and McCauley's participants chose to identify with have not historically been defended by radicalism in the U.S. (the most common identity was "Women," followed by "Catholics," and then a plethora of groups as irrelevant as "Runners" or "Honors Students"), the Horgan lab's participants did—that is, they identified as Muslims who, as a group, have historically defended themselves via both activism and radicalism.

10 They do not explain individual items, saying only that items were adapted from the literature.
11 i.e., items have a lower variance limit than can reliably be detected in the general population.
12 McCauley and Moskalenko did not provide ν's or response rates to compare to.

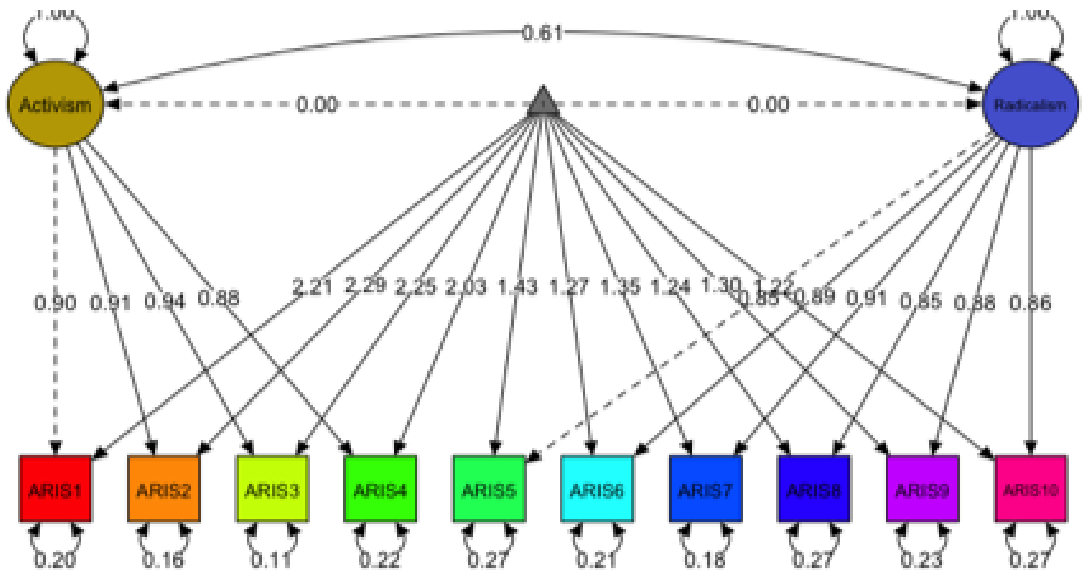

Figure 4. Horgan lab sample ARIS 10-item Estimated CFA Model
Note. Estimates are standardized, and dashed lines represent fixed parameters.
Note. Double-headed arrows refer to variances while single-headed arrows refer to regressions.

Similarly, Decker and Pyrooz (2019) found in their sample of 680 male prisoners that a CFA of the full 10-item ARIS fit within some GoF[13] indices' cutoffs (following Kline's cutoff recommendations (2005); see

Table 3 below). While they provided neither the correlations for the 10 items, nor their factor loadings (λ's) and intercepts (ν's), they did provide the means and Standard Deviations (SDs). Decker and Pyrooz's sample and the Horgan lab sample demonstrate similar mean indicator responses amongst AIS and RIS survey items respectively, with lower endorsement on the 7-point Likert-scale of the RIS items than the AIS items. No other studies have tested the ARIS's model fit[14], and none have tested for ME/I. Furthermore, all studies previously have failed to address issues of ARIS item score distribution and variable type.

[13] For an explanation of goodness-of-fit (GoF) indices, see section **Error! Reference source not found.** below.
[14] While Soliman and colleagues (2016) conducted a SEM with the RIS predicted by a set of personality factors related to radicalism, they did not model the RIS as a latent factor itself.

Table 3. Comparing Mean & SD of All 10 ARIS Item Scores Between Studies

Factor	Item Name	**Decker & Pyrooz, 2019** (n = 680)	**Cronbach's α**	**Fodeman, Snook & Horgan, 2020** (n = 356)	**Cronbach's α**
AIS	1. Join	5.34 (1.98)	.875	4.63 (2.09)	.947
	2. Donate	5.51 (1.80)		4.59 (2.01)	
	3. Volunteer	5.18 (1.91)		4.55 (2.03)	
	4. Protest	4.83 (2.10)		4.23 (2.07)	
RIS	5. Illegal Group	2.91 (2.03)	.840	3.17 (2.22)	.927
	6. Violent Group	2.39 (1.95)		2.79 (2.19)	
	7. Violent Protest	2.83 (2.12)		2.97 (2.21)	
	8. Police Defense	2.54 (2.12)		2.59 (2.09)	
	9. War	4.53 (2.47)		2.90 (2.22)	
	10. Retaliation	2.35 (1.98)		2.54 (2.08)	
Goodness of Fit Indices					
χ^2		270.400*		232.282***	
RMSEA		.101		.129	
CFI		.928		.947	

Note. Mean scores on 7-point Likert scale (low to high) unstandardized (SDs in parentheses); *p<0.05, ***p<0.001.

The ARIS presents non-normally distributed results that ought not to be treated as ordinal rather than continuous. Consider the Horgan lab sample results (Fodeman, Snook, & Horgan, In Press; Fodeman, Snook, & Horgan, 2020). The respondents received the 10-item version of the ARIS, which showed strong reliability[15], as in Moskalenko and McCauley's original study. The composite RIS and AIS factor scores shared significant, positive, moderate correlations (Pearson's r = .59, p < 0.001; Spearman's ρ = .60, p < 0.001), as did Moskalenko and McCauley's[16]. One must consider any analysis of the ARIS assuming a normal distribution carefully, however, as the scores tend to be positively skewed, i.e. few people would even *gently* endorse any RIS items. For example, consider the following histograms (Figure 5 below) of the RIS. Even removing the "Strongly Disagree" category (the second figure) still

[15] The total ARIS Cronbach's α was .944, the AIS's was .947, and the RIS's was .927—all good fit by any standard.
[16] Moskalenko and McCauley reported a significant r (no p-value) of .42 between the AIS and RIS.

presents a positively skewed distribution. The AIS presents similar skew, as do each individual ARIS item. This skew might be expected given the aforementioned rare endorsement of AIS and especially RIS, but sample distribution has not been previously discussed in the ARIS literature. While all previous analyses of the ARIS have used the assumption that responses fit a normal distribution, normality should not have been expected given that Likert-type survey items are ordinal. Ordinal data warrants another type of analysis to not bias estimates, as discussed below.

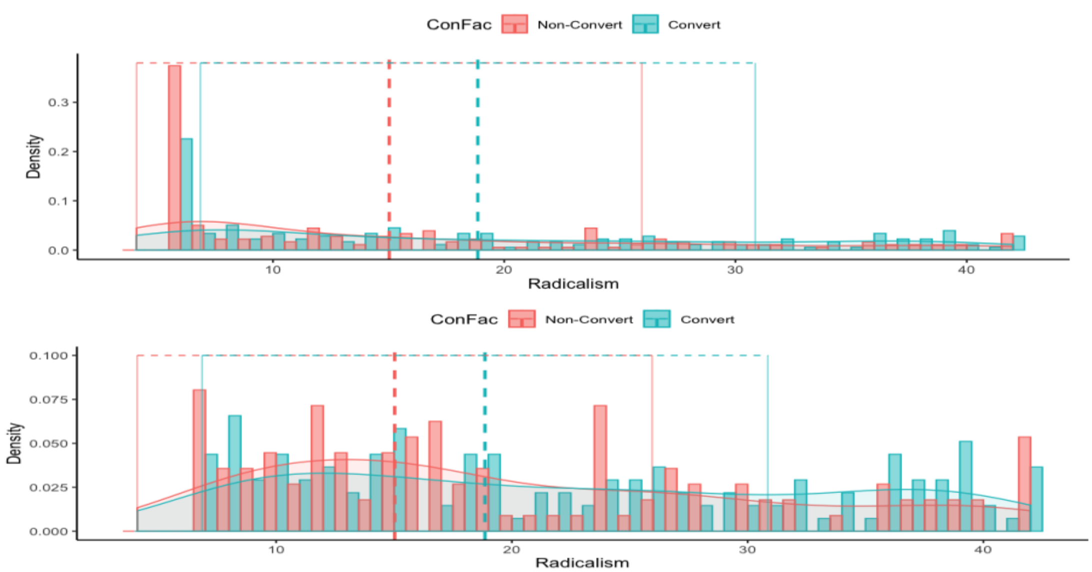

Figure 5. Stacked RIS Histograms, Means, and SDs of RIS With & Without 'Zeroes'
Note. Dashed lines refer to means, solids to SDs. Curved lines are kernel density approximations.
Note. Graphed with R package “ggplot2” (Wickham, 2016), as are all of the following figures.

2.4 Estimating Ordinal Survey Items With CFA For ME/I

2.4.1 Comparing OLS & Ordinal Logistic Regression for CFA

There is a long history of estimating ordinal Likert-type survey responses to scales with five or more response options, like those of the ARIS, via Maximum Likelihood (ML)—the estimation method used for OLS regression (see Appendix Section 7.2 for a discussion of estimators); however, to do so violates the assumptions of multivariate normality (Lubke & Muthén, 2004) and can lead to inappropriate inferences (Rutkowski & Svetina, 2017). Lubke and Muthén note that 1) a high number of response categories, 2) low skewness, and 3) relatively equal and large λ's warrant estimation of latent factors via ML *assuming* the data are from a single homogeneous population. In a *multi*group context, however, like in this study between Muslim converts and Muslim non-converts), ordinal data estimated with ML and OLS regression may distort the factor structure in FA differently across groups. Lubke and Muthén note that ME/I is therefore difficult to detect and interpret with OLS for ordinal data. Indeed, the ν's and θ's estimated by OLS regression are meaningless for ordinal variables whose data are integer rankings with no meaningful distances (Jöreskog, 2005). For example, coding "Very Unlikely," "Somewhat Unlikely," and "Neutral" as 1, 2, and 3 does not mean that there is some distance of '1 unit' equally between the 1st and 2nd category and the 2nd and 3rd category. In order to avoid these assumption violations, measurement inaccuracies, and group comparison biases, researchers can use ordinal logistic instead of OLS regression for CFAs and MCCFAs.

2.4.2 Multiple-Group Categorical CFA

We can instead compare groups via Mutiple-Group *Categorical* CFA or MCCFA (Kim & Yoon, 2011)—that is, CFAs estimated for multiple groups simultaneously with ordinal indicators. Differences of ordinal thresholds (τ's) between groups indicate ME/I (Vandenberg & Lance, 2000; Millsap & Yun-Tein, 2004). The underlying latent, continuous response variable is estimated by the likelihood, given the sample, of an individual choosing response category 1 as compared to 0, 2 as compared to 1, and so on. In MCCFA, we can compare the individual thresholds between groups. Therein, a true score of 1 on the latent response variable might lead differentially to a response of "Strongly Disagree" on, say, the Police Defense item for Muslim non-converts in our sample, but "Agree" for Muslim converts. We can test if this difference between τ's is significant. In order to avoid the disadvantages of any one ordinal ME/I testing pathway strategy (see Appendix section 7.1 for discussion), all paths are tested in this thesis.

That is, τ's and λ's are tested both independently from one another (i.e., separate Metric and Threshold Invariance models) as well as combined (i.e., the Scalar Invariance model). Below is a depiction of the order of model testing (Figure 6). Note that Threshold Invariance refers only to the latent response variable for each indicator, rather than individual thresholds tested for PI.

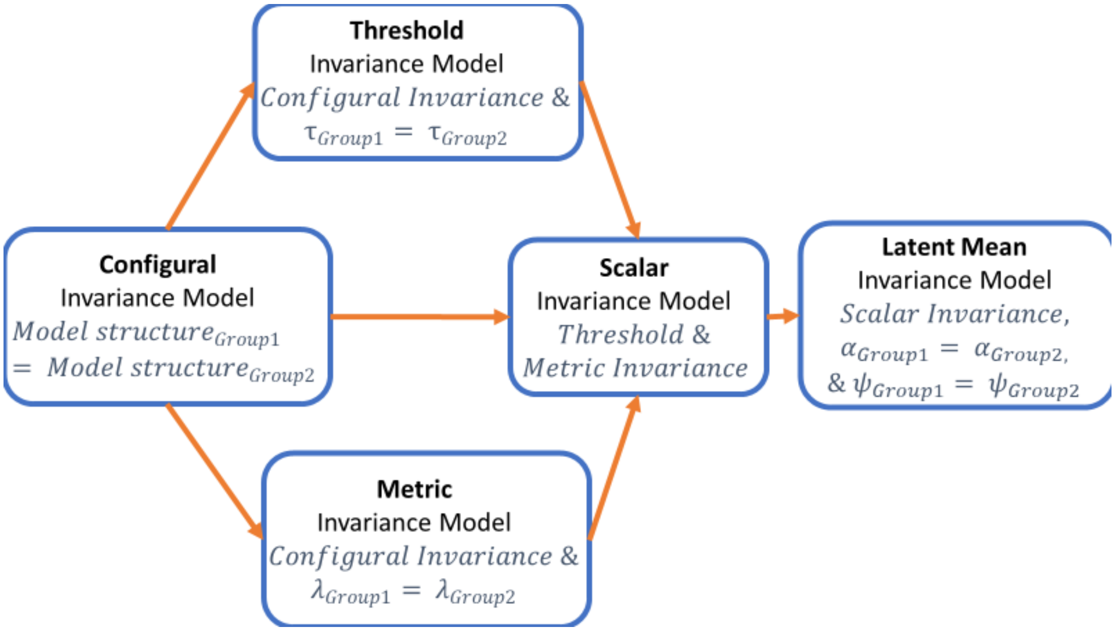

Figure 6. Proposed ME/I Thesis Testing Steps.

2.5 Specific Aim: Test for ME/I of the ARIS between American Muslim converts and American Muslim non-converts.

Hypothesis: American Muslim converts significantly differ in latent mean (α) activism (AIS) and radicalism (RIS) scores on the ARIS from American Muslim non-converts, but not on other parameters. As discussed in other research (Fodeman, Snook, & Horgan, In Press), 1) Muslim converts are overrepresented amongst American Muslims who engage in terrorism-related offenses related to the general American Muslim population, and 2) activism and radicalism are potentially antecedent cognitions and behaviors to terrorism-related offenses. In other words, American Muslim converts will demonstrate Configural, Metric, Threshold, and Scalar Invariance, but not Latent Mean Invariance (see Figure 6 above) on their ARIS scores.

3 METHODS

3.1 Procedures

3.1.1 Original Study's Participant Screening & Recruitment

The GSU Institutional Review Board approved data collection on June 15, 2016 (IRB number: H15619, Reference number: 339012). Between August 2016 and October 2016, a research team working under the Principal Investigator Dr. John Horgan sampled 356 Muslim participants living in the United States ages 18 to 65 years old. Of these,177 were self-identified converts. This was part of a suite of research projects on Muslim converts supported by the Office of Naval Research (Grant N00014-16-1-2-19). The participants in this study were recruited through an online Qualtrics survey panel. Qualtrics recruited participants on the team's behalf using established participant networks to produce a sample of United States residents who are Muslim. Participants were notified via email of a new study and invited to participate. Qualtrics provided non-monetary incentives (e.g., air miles, gift certificates, etc., equaling approximately $10) to participants. The survey remained open until the sampling quotas were met. All participants were English speakers for the expediency of administering the task. Originally, a random sample (not stratified) of 1,100 responded to the survey, of whom 659 were screened from participating in the study as they were not 18+ years old, Muslim, a United States resident, and/or did not agree to informed consent. This resulted in 441 Respondents (208 converts and 233 non-converts). This small *n*, resulting from a high screening gradient, was to be expected given the minority representation of Muslims in the U.S., let alone of converts amongst U.S. Muslims (see the next section 3.1.2 below for a discussion). An additional 85 of those responses were excluded from the analysis (83 at first pass and another 2 at second pass) due to survey errors, including excess time[17], data omission[18], and inattention to attention-checking questions[19]. This resulted in a final sample of 356 valid survey responses (177 converts and 179 non-converts). These rates of raw online survey data exclusion are not uncommon following standard comprehensive criteria (e.g., the Ilumeo Quality Criteria), as cohorts of respondents fitting said paradata criteria often produce poor survey responses due to inattention, lack of commitment or coherent understanding of the survey (Freire, Senise, dos Reis, & Ono, 2017). Those poor survey responses were excluded from analysis so as not to bias analyses. Theoretically, those responses should only produce noise and not have any directional effects on results. This thesis re-analyzes the ARIS comparisons between Muslim converts and Muslim non-converts as in the original article (Fodeman, Snook, & Horgan, In Press), although using estimated factor scores with ordinal logistic links rather than single composite scores.

3.1.2 Original Study's Sample Recruitment Considerations

This study required a large sample of US Muslims and US Muslim converts. Recruiting these participants was a daunting task given their relative scarcity in the US population. Currently, the United States is home to approximately 3.3 million Muslims—only about 1% of the US population (Mohamed, 2016). Of that, about 20% of Muslims in the US are converts to Islam (Pew Research Center, 2011). This percentage is much higher than other Western countries (Schuurman, Grol, & Flower, 2016), but still means that only about 0.2% of the US population were eligible for inclusion. Above that, endorsement of radicalism is extremely rare amongst Muslims (Schmid, 2017). Qualtrics recruited converts and non-converts such that the

17 For measurement reliability, participants had 3 hours to complete the survey and could leave the survey inactive for no longer than 30 minutes. Response time outliers and *a priori* thresholds like these are commonly used in online surveys (Malhotra, 2008; Greszki, Meyer, & Schoen, 2015; Matjašic, Vehovar, & Manfreda, 2018).
18 E.g., outliers or "straight-lining" answer "A" in a row (Levin, Fox, Forde, & David, 2012).
19 E.g., selecting "A" when the question says, "For this question, select C": Freire, Senise, dos Reis, & Ono, 2017; Gummer, Roßmann, & Silber, 2018; Silber, Danner, & Rammstedt, 2019).

sample would be representative of recent US census results in terms of demographics. One exception to this strategy regarded Muslim converts' race. Although little is known about US Muslim converts, there is evidence that the majority of them (59%) are African American (Pew Research Center, 2011). Therefore, in order to ensure the sample represented the US convert population, Qualtrics did not recruit converts according to general US census results for race.

3.1.3 Original Study's Materials

The study, titled, "Understanding American Muslim Converts," was conducted on the online survey platform Qualtrics. The research team used Qualtrics to administer the following surveys: The Psychological Measure of Islamic Religiousness (PMIR), the Adult Religious Conversion Experiences Questionnaire (ARCEQ), and the Activism-Radicalism Intentions Scale (ARIS). The former two were employed for analysis discussed elsewhere (Fodeman, Snook, & Horgan, 2020; Snook, 2018; Snook, Branum-Martin, & Horgan, in review; Snook, Horgan, Kleinmann, & White, in press). Those surveys were given in that order such that the PMIR and ARCEQ (skipped for non-converts) primed respondents to think of their group for the ARIS as Muslims[20]—the last questions on the PMIR regarded the Islamic Exclusivism subscale as a particular prime. Several demographic questions were also asked (e.g., gender, race, occupation, education, age, etc.). Conversion did not lose significance as a predictor of either AIS or RIS scores in stepwise linear, ordinal, or negative binomial regressions with demographic control variables (Fodeman, Snook, & Horgan, In Press). Therefore, those demographics will not be included in this thesis. The PMIR and ARCEQ are also beyond the scope of this thesis.

[20] Note that there was no explicit statement as such, but participants were repeatedly informed that the purpose of the study was regarding being a Muslim.

3.1.4 Original Study's Data Collection Participant Protection

The research team used Qualtric's Waiver of Documentation of Consent to maintain respondent anonymity. Since participants cannot give verbal consent online, they were asked to "click yes" if they agreed to proceed. The Waiver of Documentation of Consent explained 1) the risks of participation in the study, 2) that there were no direct benefits to participate, 3) participation was completely voluntary, and 4) data would be protected, anonymous and confidential. Neither Qualtrics, its panel providers, nor our lab collected any personally identifiable information about the respondents. The Informed Consent read as follows: "We would like you to participate in a study. The purpose of the study is to understand why people change to Islam as a new religion. We would like you to participate because you are a Muslim. You will complete an online survey. The surveys will ask you questions about being Muslim."

3.2 Estimation Method & Fit Indices

Weighted Least Squares with Mean-and-Variance-adjusted (WLSMV: Jöreskog, 2005) was chosen as the optimal estimation method, as compared to ML or others, to decrease bias given this sample's small *n*, asymmetric τ's, and large λ's (for a detailed explanation, see Appendix Section 7.2 below). The Goodness-of-Fit (GoF) indices χ_2, CFI, and RMSEA will be used to compare models as recommend by Rutkowski and Svetina in their ME/I GoF cutoff review (2017), although ΔCFI will be relied upon the most (see Appendix Sections 7.3-7.5).

4 RESULTS

4.1 Summary Statistics: Thresholds, Variances & Polychoric Correlations

For an explanation of the summary reporting procedures for ME/I with ordinal indicators and the graphing, table and estimate choices below, see Appendix Section 7.5. While 10 ARIS items were surveyed, the four AIS items could not be estimated for ME/I and thus are excluded from analysis. The estimation difficulties are beyond the scope of this thesis. Regardless, the RIS is the construct of greater concern and will remain the sole focus of this thesis. Figure 7 below, depicting RIS response frequencies, was graphed with the R package "sjPlot" (Lüdecke, 2019).

4.1.1 Summary Frequencies & Thresholds

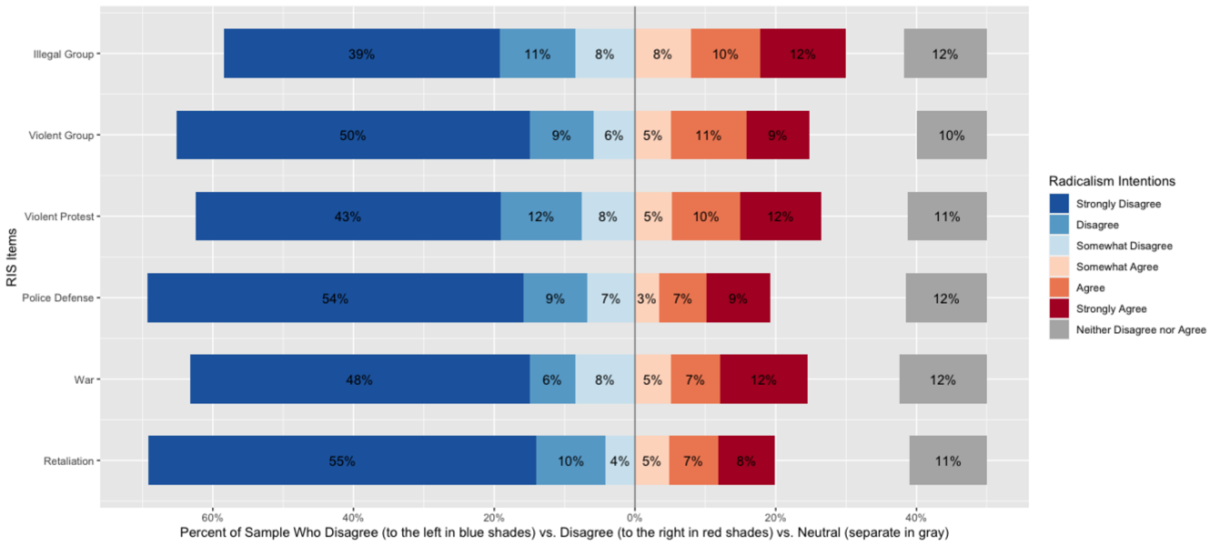

Figure 7. Response Frequencies To All ARIS Items For All Participants.
Note. Response categories refer to agreement with intentions of engagement statements.

4.1.2 Summary Correlations

Table 4. Polychoric Correlations Between RIS Indicators

Variable	1	2	3	4	5
1. Illegal Group					
2. Violent Group	.89				
3. Violent Protest	.82	.87			
4. Police Defense	.75	.78	.83		
5. War	.78	.81	.85	.87	
6. Retaliation	.79	.82	.80	.86	.84

Note. All correlations have *p* values below .01.

Table 4 above contains the polychoric correlations for the entire dataset. Polychorics are appropriate for ordinal by ordinal data. The correlations are strong enough to warrant FA.

4.2 Model Results: Data Elements, Parameter Estimates & Fit

Table 5. Data elements, parameters & specifications for ME/I models of the RIS

Data Elements	Variances 0	Covariances 30		Thresholds 72		Total 102
Parameters	**Null**	**Configural**	**Metric**	**Threshold**	**Scalar**	**Means**
θ	0	0	0	0	0	0
Ψ	0	0	1	0	1	0
λ	0	12	6	12	6	6
τ	72	72	72	36	36	36
α	0	0	0	1	1	0
Total	72	84	79	49	44	42
df	30	18	23	53	58	60

Note: The statistical notation is adapted from LISREL (Mehta, 2013); see List of Symbols above.

The six 7-point Likert ordinal RIS items per the two groups resulted in 30 covariances (θ's) and 72 thresholds (τ's) total. Those 102 data elements (see Table 5 above—15 θ's and 46 τ's per group—equal 102 *df* with which to estimate the ME/I models[21]. A Configural Invariance Model with no latent means (α) or variances (Ψ) estimated—fixed, instead, to 0 and 1 for each group respectively—results in 84 estimated parameters—12 factor loadings (λ's, 6 per group) and 72 τ's. In the Metric Invariance Model, the λ's are equality-constrained (that is, the same λ is estimated for both groups per item), while Ψ is freed for one group, resulting in six *df* gained from the λ's, but one spent to free Ψ, resulting in a net gain of 5 *df*. Similarly, τ's are equality-constrained in the Threshold Invariance Model while an α is freed. In the Scalar Invariance Model, *both* equality-constraints and latent free estimates are maintained, while in the Latent Mean Invariance Model all latent parameters are fixed— α's to 0 and Ψ's to 1—standardizing RIS scores as equivalent across groups. This way, ever more restricted models are tested to

[21] In this parameterization ("Theta"), residual variances are fixed to 1; the alternative "Delta" is beyond this thesis.

compare the parameters of the RIS between Muslim converts and non-converts for any bias.

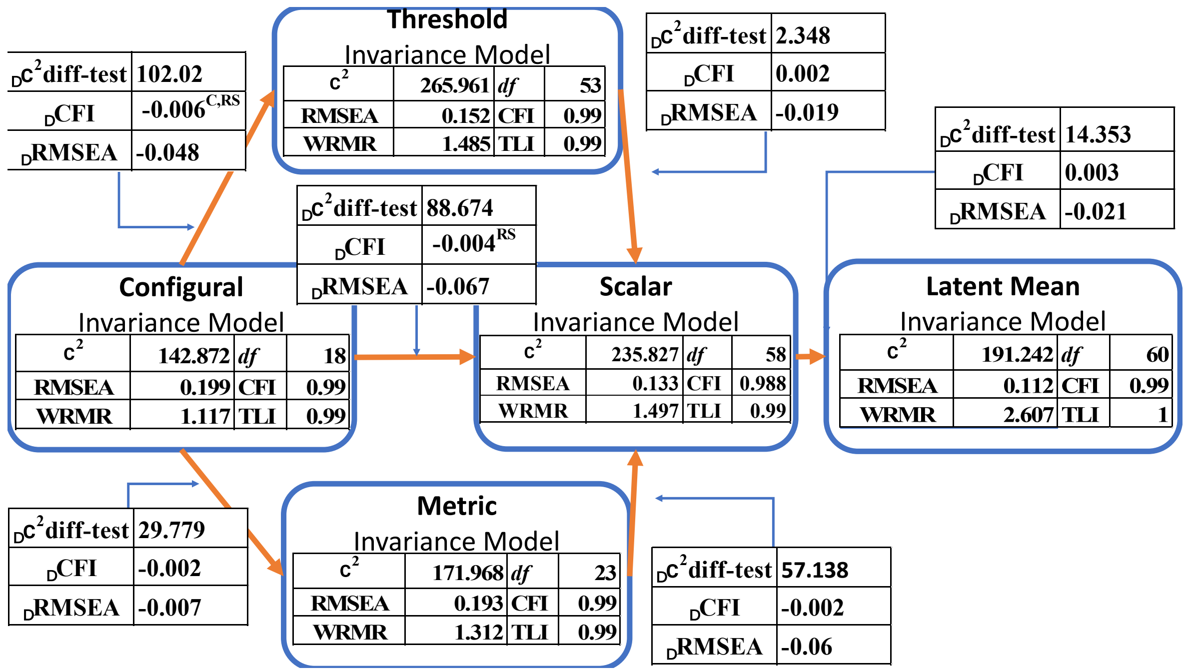

Figure 8. GoF and ΔGoF Indices For All ME/I Models and Paths.
Note. The $\Delta\chi^2$diff-tests use the T_3 method (nested model comparison) to account for non-normal distributions using a function like the MPlus DIFFTEST (Asparouhov & Muthen, 2010).
Note: ΔGoF failures noted by subscripts C (Chen, 2007) and RS (Rutkowski & Svetina, 2017).
Note. Estimated with R packages Lavaan (Rosseel, 2012) and SemTools (Jorgensen et al., 2018).

These models and their [Δ]GoF's are reported in Figure 8 above per each ME/I testing path. All models failed to pass GoF cutoffs (DiStefano, et al., 2017; West, Taylor, & Wu, 2012) for χ^2 ($p < .05$), RMSEA (≤.06), and WRMR (<1.0), though not CFI (≥.95). The poor χ^2 and RMSEA results despite positive CFI's may be due to small sample size, though the same blame cannot be attributed to WRMR which is more biased against *large* sample sizes. However, given that even the Configural Model has poor fit, it may be reasonable to proceed with comparing models. No models failed to pass all of either Rutkowski and Svetina's (2017) or Chen's (2007) ΔGoF cutoffs. Note that while the ΔCFI's from the Configural invariance model to either the Threshold or Scalar Invariance models were at the cusp of Rutkowski and Svetina's cutoff (≤-.004), they argue that *both* CFI and RMSEA cutoffs have to simultaneously be surpassed in order to qualify as noninvariance. Therefore, it is reasonable to conclude that the RIS is invariant between Muslim converts and Muslim non-converts.

Indeed, while we would expect the Latent Means model to be *non*invariant compared to the Scalar one, the Latent Mean model maintains reasonable fit below ΔGOF cutoffs. Were we to only look at GoF, the Latent Mean model would be retained, and the Null Hypothesis that Muslim Converts and Non-Converts have the same mean RIS scores could not be rejected. However, looking at the Scalar model's parameter estimates, the Converts' α and Ψ are significantly different from 0 and 1 (the fixed parameters for Non-Converts as the reference group), respectively. This indicates a significant difference between groups, which contradicts the omnibus ΔGOF tests of the Latent Mean model. The reasons for this contradiction are beyond the scope of this thesis, but may be attributable to difficulties in estimating the skewness of this data, particularly regarding specification of and sensitivity to comparing the first two thresholds, as evidenced by the non-significant 1st and 2nd threshold estimates of some of the RIS items (see

Table 6 below). There is evidence thereof for a substantive difference in latent radicalism between Muslim converts and Muslim non-converts; specifically, Muslim converts demonstrate a 4/10ths higher average RIS score than non-converts ($p < .001$), with 7/10ths the variance ($p < .001$)—i.e., less dispersed around that mean (see the left side of

Table 6 below).

Table 6. RIS Scalar vs. Latent Mean Invariance Models' Estimates

	Scalar Invariance Model		Latent Mean Invariance Model	
	Estimate	SE	Estimate	SE
Factor Loadings (λ's)				
Illegal Group	2.44***	.25	2.19***	.15
Violent Group	3.27***	.39	3.06***	.28
Violent Protest	2.54***	.26	2.38***	.18

Police Defense	2.64***	.27	2.58***	.22
War	2.74***	.27	2.68***	.23
Retaliation	2.35***	.24	2.21***	.17
		Thresholds (τ's)		
Illegal Group				
1-to-2	-0.18	.23	-0.66***	.16
2-to-3	0.52*	.24	-0.01	.16
3-to-4	1.05***	.25	0.51**	.17
4-to-5	1.84***	.26	1.28***	.18
5-to-6	2.44***	.26	1.88***	.18
6-to-7	3.42***	.27	2.86 ***	.18
Violent Group				
1-to-2	0.73*	.33	0.03	.22
2-to-3	1.48***	.35	0.78***	.24
3-to-4	1.98***	.36	1.28***	.25
4-to-5	2.89***	.39	2.22***	.28
5-to-6	3.43***	.40	2.75***	.29
6-to-7	4.98***	.46	4.28***	.35
Violent Protest				
1-to-2	0.09	.25	-0.41*	.17
2-to-3	0.85**	.25	0.30	.18
3-to-4	1.35***	.26	0.81***	.18
4-to-5	2.16***	.26	1.61***	.19
5-to-6	2.59***	.26	2.02***	.19
6-to-7	3.61***	.27	3.05***	.19
Police Defense				
1-to-2	0.80**	.26	0.23	.19
2-to-3	1.41***	.28	0.86***	.21
3-to-4	1.91***	.28	1.35***	.22
4-to-5	2.88***	.29	2.36***	.24
5-to-6	3.21***	.30	2.70***	.25
6-to-7	4.11***	.30	3.67***	.28
War				
1-to-2	0.45	.26	-0.13	.19
2-to-3	0.90**	.27	0.33	.20
3-to-4	1.51***	.26	0.96***	.21
4-to-5	2.47***	.26	1.91***	.22
5-to-6	2.93***	.26	2.39***	.23

6-to-7	3.76***	.28	3.28***	.25
Retaliation				
1-to-2	0.81**	.24	0.30	.17
2-to-3	1.43***	.25	0.93***	.18
3-to-4	1.71***	.26	1.22***	.19
4-to-5	2.53***	.26	2.02***	.20
5-to-6	2.97***	.26	2.47***	.21
6-to-7	3.85***	.28	3.41***	.23
		Residual Variances (θ's)		
Volunteer	1.00§		1.00§	
Join	1.00§		1.00§	
Donate	1.00§		1.00§	
Protest	1.00§		1.00§	
Violent Protest	1.00§		1.00§	
Illegal Group	1.00§		1.00§	
Violent Group	1.00§		1.00§	
Police Defense	1.00§		1.00§	
War	1.00§		1.00§	
Retaliation	1.00§		1.00§	
		Latent Variances (Ψ's)		
Radicalism				
Non-Convert	1.00§		1.00§	
Convert	0.73***	.17	1.00§	
		Latent Means (α's)		
Radicalism				
Non-Convert	0.00§		0.00§	
Convert	0.40***	.11	0.00§	

Note. § = fixed. All parameters equality-constrained, save for α & Ψ in the Scalar model.
Note. *p<0.05; **p<0.01: ***p<0.001. See Figure 8 above for [Δ]GoF.

RIS items' latent response thresholds are graphed in Figure 9 below. These thresholds fall on a theoretical latent continuous response distribution of likelihood of selecting higher and higher ordinal response choices. Dots represent the Scalar Invariance model's estimated thresholds (numbered by order) with SE bars. Examples of the Likert response choices are noted as brackets between thresholds' SE bars. The estimates are graphed relative to RIS factor scores (the *x*-axis), such that a RIS factor score of *x* would likely result in the ordinal responses falling within the items' respective thresholds. Note that α's are depicted, demonstrating the higher mean RIS score for Muslim Converts than for Non-Converts (whose α is fixed to 0 and Ψ to 1). Note, however, that some thresholds are not, by their SEs, distinguishable for some of the items. There does not seem to be any pattern as such, though the last threshold (choosing Strongly Agree over Agree) is always unique across all items, as is the first threshold (choosing Disagree over Strongly Disagree) for all but the War item. That may be due to the aforementioned floor effects, and a substantive difference between those who choose the lowest, highest, and any middle responses.

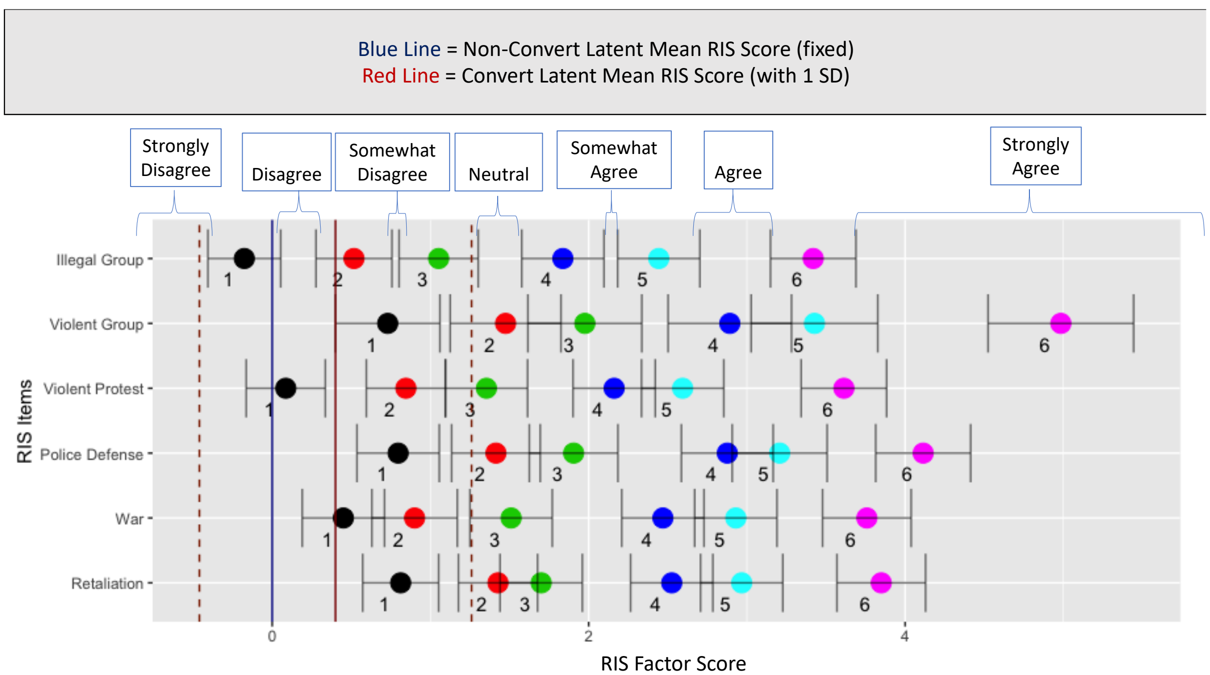

Figure 9. RIS Scalar Invariance Thresholds & Factor Scores
Note. Based on Branum-Martin & co.'s Figure 3 (2013); plotted with ggplot2 (Wickham, 2016).

5 CONCLUSION

5.1 Discussion

Marsh and colleagues (Marsh, Hau, & Wen, 2004) remind us about the dangers of blindly following GoF cutoff values. While Marsh and colleagues (2004) harken to the dangers of earlier ME/I guidelines meant for continuous indicators (Hu & Bentler, 1999), the aforementioned modern caveats and mere differences in GoF cutoff recommendations for ordinal data highlight the same issues that Marsh and colleagues discuss. Svetina and colleagues (2019) similarly make note that GoF cutoffs are always limited in their generalizability and are guidelines rather than absolute rules[22]. The better question for researchers to ask is not whether any particular cutoff is met, but what that degree of misfit means clinically. Unfortunately for this thesis, no published studies have correlated the ARIS with clinically-relevant outcomes. Given that there are no precedents from which to appropriately determine clinically significant GoF cutoffs or ranges of clinically impactful estimated parameter differences, this thesis will stop short at determining which parameters present DIF, and to what degree, if any, for future clinical use concerns. This thesis analysis is, first and foremost, a methodological exercise, if not a means to buttress previous treatment of the ARIS as a tenable composite score for comparing groups. The only parameters for which there is a theoretical basis for noninvariance with this thesis sample are latent means; that is, Muslim converts are expected have significantly higher overall (i.e., latent) radicalism scores than Muslim non-converts per the literature (see Fodeman, Snook, & Horgan[a,b]), but not significantly differ on any other individual parameter (i.e., factor loadings or response thresholds). As to how large a difference in radicalism scores on the ARIS is clinically significant, the literature has yet to comment. Future studies may demonstrate this and look back to comparative studies like these to see how large a difference in radicalism on the ARIS is demonstrated in groups thought to be clinically relevant like Muslim converts.

Were [$_\Delta$]GoF tests the only tools by which to accept or reject models and hypotheses, we would likely conclude complete invariance between Muslim converts and Muslim non-converts. While all models only demonstrated strong GoF scores by CFI, this is the most reliable GoF

[22] To treat GoF cutoffs as rules is akin to treating *p*-value cutoffs as rules and not guidelines.

indicator for this study (see Appendix Sections 7.3-7.5). Indeed, GoF indicators like RMSEA counterintuitively improved with more restrictive models—a phenomenon known to sometimes occur with ordinal data, but whose explanation is beyond the scope of this thesis. No models demonstrated poor ΔGoF, either, which would lead to acceptance of all models. However, given the significant latent factor score differences between groups in the Scalar Invariance model and the findings in the literature suggesting such a difference, the Latent Mean invariance model is rejected, as well, therein, as the Null Hypothesis.

Part of the difficulty in assessing by omnibus tests the significant latent factor score differences in the Latent Mean invariance model may be due to misspecification of the skewed ordinal responses. Ordinal logistic regression assumes a normally-distributed latent response variable underlying responses to an ordinal item, but a non-normal latent response distribution may be more appropriate given the data's high positive skew. This skew is evidenced in the spacing of the ordinal thresholds, with largely increasing distance from lowest to highest thresholds. It is not uncommon to see such skew in similar data on radical cognitions and behaviors, but latent response distribution adjustment is neither the norm within such research nor within the scope of this thesis. Suffice it to say, both groups demonstrate this skew, and it is reasonable by omnibus and individual parameter estimates Wald statistics to treat item thresholds as the same across groups for at least the higher four thresholds if not all six. The aforementioned floor effect may be affecting the groups differently (as evidenced by the lower Wald statistic scores for the lower two thresholds—Disagree over Strongly Disagree and Somewhat Disagree over Disagree), but this might only give further credence to the findings of a significant difference in so far as Muslim converts may require less or be more likely to pass the hurdle of giving any endorsement of radicalism at all. Testing for such partial invariance and/or mixture modeling, however, is also beyond the scope of this thesis.

The RIS clearly demonstrates skew found not only in this study but all the aforementioned reported findings in the literature. This ought to be taken into account for future measurements. This skew and floor affect might be ameliorated with a larger pool of items that cover the lower end of radicalism endorsement not currently captured by the RIS (i.e., less controversial radicalism behaviors). Doing so will only increase scale sensitivity. What can be said for the surveyed population (Muslims), like other populations surveyed, is that endorsement of radical behaviors is rare, both across several types (i.e., individual RIS items) and collectively in terms of a global factor. The findings would suggest that this skewed distribution may be shifted for Muslim converts relative to Muslim non-converts towards higher endorsement.

5.2 Strengths and Limitations

This analysis is both a unique comparison of Muslim convert versus non-convert radicalism and a technical buttress to other research products on the subject of either Muslim converts or the ARIS. This type of analysis, while more accurate than typical t-tests of survey composite scores, is more difficult to conduct and, therefore, for other researchers or practitioners to replicate. Indeed, some of the particulars as to the exact statistical procedures are still being debated in the field. However, the use of CFA, ordinal estimation, and ME/I is more appropriate for the data, overcoming the assumptions of past studies as to survey structure, response distributions, and invariance. This study’s top-down modeling approach is appropriate in so far as the literature has little to say thus far about the structure, function, and accuracy of the ARIS or radicalism measurement generally. Future research can apply and report the results of these more nuanced modeling strategies, which will help inform any theoretical basis behind ARIS functionality or radicalism assessment.

This study is limited in many respects, however, by the nature of its sample. Its *n* is small compared to what is commonly used for these types of analyses (see section 3.2 above on fit indices, particularly the simulation studies mentioned in Appendix Section 7.2 below), although within viable ranges. Indeed, the strong correlations within the RIS's theorized indicators are perhaps the saving grace that has enabled testing for ME/I with such a small *n*. Neither participant group is from a clinical population, although Muslim converts are identified in the literature as at a higher risk than their non-convert counterparts of engaging in radicalism. Note that these increased rates are only marginally higher and are more likely to lead to false positive diagnoses (White, 2019). Unfortunately, no use of the ARIS with clinical populations (i.e., those that have engaged in terrorism-related offenses) has been reported publicly for comparison. Therefore, instead of drawing on clinical significance of model misfit or noninvariance, this study instead will be a benchmark for future studies to call back to with clinical populations.

Even with ordinal estimation, this study may still be limited by the skew of its results. Some of the RIS questions, such as Violent Protest, Police Defense, War or Retaliation, might not be as relevant to Muslims as to other groups; therein, their results might capture less variance due to irrelevance (e.g., the plurality of respondents choosing "Strongly Disagree" compared to all other ordinal categories; see Figure 5 above). For example, the War item might be construed by respondents as being more relevant to active asymmetric conflict zones rather than in a democratic political space, or the Retaliation item might be construed in the context of apolitical inter-group vendettas. There is no theoretical reason to believe that these items should be any less relevant for Muslim converts as for Muslim non-converts, however, and therefore any irrelevance thereof ought not to bias comparisons between the two groups. Similarly, the RIS is framed in the context of the AIS, which entails more socially desirable intentions than those in the RIS, but there is no theoretical reason to believe that converts would be more affected by this framing than non-converts. All in all, the limits of skew should not differently affect groups.

Since no GoF cutoffs have been simulated that fit the conditions of this exact study (i.e., ordinal distributions of six seven-point Likert-type indicators for one factor between two groups of ~175 participants each), it is difficult to use them. In any case, GoF cutoffs are more like guidelines, fraught with their own inherent inaccuracies if used as absolute rules. One should instead consider the *degree* of misfit between models and what that means clinically. Since publications on the ARIS are still relatively few, with no clinical outcomes as of yet, it is difficult to accept or reject any particular level of invariance outright. While the AIS could not be tested, the RIS does not surpass any of the noninvariance criteria proposed in the literature—even for latent means, which was expected between the two groups. To run more exact simulations than those from the field is beyond the scope of this thesis. Judging, then, by those criteria based on simulations from conditions most closely matching this thesis, as well as relative changes in fit and general knowledge about the ARIS, it is assumed at this time to be invariant for Muslim converts and Muslim non-converts, save for their overall radicalism per the significant latent factor parameter estimates in the Scalar Invariance model.

5.3 Conclusion

Per the results of this thesis, American Muslim converts and Muslim non-converts can be unbiasedly compared on the radicalism portion of the Activism and Radicalism Intentions Scale (ARIS). Therein, American Muslim converts do demonstrate significantly higher intentions of engaging in radicalism behaviors than non-converts. This is demonstrated in this thesis by the successful tests of measurement equivalence or invariance (ME/I)—the condition that different groups respond the same way to the same test. That is, in this case, the same survey questions indicate intentions of engaging in radicalism the same way they do for American Muslim converts as for non-converts. The factor loadings—how strongly those indicators each contribute to measuring overall radicalism—can reasonably be treated as equivalent across as well. So, too, can the seven-point Likert-type question thresholds—the estimated likelihood of choosing a "2" over a "1," a "3" over a "2," and so forth—be treated as equivalent across groups. The ARIS, or at least the radicalism portion, can be tested the same way for other group comparisons in future studies and applied settings. It is especially important to demonstrate the unbiasedness of measures like these in terrorism research given how difficult it can be to obtain samples, how sensitive that data is, how difficult establishing at risk populations are, and how little quantitative work has been done.

All in all, our findings modestly support the literature that would assume Muslim American converts to demonstrate higher intentions of engaging in radicalism than Muslim American non-converts due to disproportionate terrorism activity. Therein, if conversion is a risk factor for terrorism engagement, radicalism itself may be an intervening variable; however, given the contradictory Scalar Invariance and Latent Mean Invariance model findings, this finding should be taken with a grain of salt. Future studies ought to test individual thresholds of the ARIS for noninvariance—particularly the lower thresholds that may be part of a floor effect that varies by group surveyed—and/or consider non-normal latent continuous response variable distributions that might better model such expected skewed responses. Future studies ought also to consider testing additional radicalism items that cover less extreme behaviors and therein lower hurdles to endorse; in doing so, a larger variation of radicalism endorsement can be captured. In so doing, an expanded ARIS might not suffer the same estimation difficulties at the latent factor level, much like how an IQ score can be reliably estimated across populations even if some individual IQ indicators have strong floor or ceiling effects (i.e., questions that capture some of the highest and lowest levels of intelligence). It is not, therein, that the ARIS is unreliable per-se, but that it may not measure a sufficiently broad spectrum of radicalism to reliably capture endorsement within non-clinical (i.e., non-terrorist) populations.

6 REFERENCES

Agnew, R. (2010). A general strain theory of terrorism. *Theoretical Criminology, 14*(2), 131-153.

Altier, M., Thoroughgood, C. N., & Horgan, J. G. (2014). Turning away from terrorism: Lessons from psychology, sociology, and criminology. *Journal of Peace Research, 51*(5), 647-661.

Asparouhov, T., & Muthen, B. (2010, May 3). *Simple Second Order Chi-Square Correction.* Retrieved from Stat Model: statmodel.com/download/WLSMV_new_chi21.pdf

Bötticher, A. (2017). *Radikalismus und Extremismus: Konzeptualisierung und Differenzierung zweier umstrittener Begriffe in der deutschen Diskussion.* Doctoral dissertation, Leiden University.

Bailey, G. M., & Edwards, P. (2017). Rethinking 'radicalisation': microradicalisations and reciprocal radicalisation as an intertwined process. *Journal for Deradicalization, 10,* 255-281.

Bartlet, J., & Miller, C. (2012). The edge of violence: Towards telling the difference between violent and non-violent radicalization. *Terrorism and Political Violence, 24*(1), 1-21.

Bentler, P. M. (1990). Comparative fit indexes in structural models. *Psychological Bulletin, 107,* 238-246.

Boer, M. D., Hillebrand, C., & Nölke, A. (2008). Legitimacy under pressure: the European web of counter-terrorism networks. *JCMS: Journal of common market studies, 46*(1), 101-124.

Borum, R. (2012). Radicalization into Violent Extremism I: A Review of Social Science Theories. *Journal of Strategic Security, 4*(4), 7-36.

Borum, R. (2014). Psychological Vulnerabilities and Propensities for Involvement in Violent Extremism. *Behavioral Sciences and the Law, 32*(1), 286-305.

Borum, R. (2015). Assessing risk for terrorism involvement. *Journal of Threat Assessment and Management, 2*(2), 63.

Bouhana, N., Thornton, A., Corner, E., Malthaner, S., Lindekilde, L., Schuurman, B., & Perry, G. (2013). *Preventing, Interdicting and Mitigating Extremist events: Defending against lone actor extremism.* PRIME Risk Analysis Framework, Deliverable D3.1.

Bovaird, J. A., & Koziol, N. A. (2012). Measurement models for ordered-categorical indicators. In R. H. Hoyle (Ed.), *Handbook of structural equation modeling.* New York, NY: Guilford Press.

Bowen, N. K., & Masa, R. D. (2015). Conducting Measurement Invariance Tests with Ordinal Data: A Guide for Social Work Researchers. *Journal of the Society for Social Work and Research, 6*(2), 119-248.

Brand, P. A., & Anastasio, P. A. (2006). Violence-Related Attitudes and Beliefs Scale Construction and Psychometrics. *Journal of Interpersonal Violence, 21*(7), 856-868.

Brooke, N. (2015). *The Dog that Didn't Bark: Political Violence and Nationalism in Scotland, Wales and England.* Thesis, The University of St. Andrews.

Byrne, B. M., Shavelson, R. J., & Muthén, B. O. (1989). Testing for equivalence of factor covariance and mean structures: The issue of partial measurement invariance. *Psychological Bulletin, 105,* 456-466.

Chen, F. (2007). Sensitivity of goodness of fit indexes to lack of measurement invariance. *Structural Equation Modeling, 14*(3), 464-504.

Chermak, S. M., Freilich, J. D., Parkin, W. S., & Lynch, J. P. (2012). American Terrorism and Extremist Crime Data Sources and Selectivity Bias: An Investigation Focusing on Homicide Events Committed by Far-Right Extremists. *Journal of Quantitative Criminology, 28*, 191-218.

Cheung, G. W., & Rensvold, R. B. (2002). Evaluating goodness-of-fit indexes for testing measurement invariance. *Structural Equation Modeling: A Multidisciplinary Journal, 9*(2), 233-255.

Chikhi, S. (2017). Non-violence or Violent Extremism: Young Refugees' Propensities Pending the Resolution of the Conflict in Western Sahara. *Journal of Peacebuilding & Development*, *12*(2), 51-65.

Christmann, K. (2012). *Preventing Religious Radicalisation and Violent Extremism: A Systematic Review of the Research Evidence.* London: Youth Justice Board.

Decker, S. H., & Pyrooz, D. C. (2019). Activism and Radicalism in Prison: Measurement and Correlates in a Large Sample of Inmates in Texas. *Justice Quarterly, 36*(5), 787-815.

Della Porta, D. (2013). *Clandestine Political Violence.* Cambridge University Press.

Dimitrov, D. (2010). Testing for factorial invariance in the context of construct validation. *Measurement and Evaluation in Counseling and Development, 43*(2), 121-149.

DiStefano, C., & Morgan, G. B. (2014). A Comparison of Diagonal Weighted Least Squares Robust Estimation Techniques for Ordinal Data,. *Structural Equation Modeling: A Multidisciplinary Journal, 21*(3), 425-438.

DiStefano, C., Liu, J., Jiang, N., & Shi, D. (2017). Examination of the Weighted Root Mean Square Residual: Evidence for Trustworthiness? *Structural Equation Modeling: A Multidisciplinary Journal*, 453-466.

Dono, M., Azate, M., Seoane, G., & Sabucedo, J. M. (2018). Development and validation of the Monopoly on Truth Scale. A measure of political extremism. *Psicothema, 30*(3), 330-336.

Dunwoody, P. T., & Funke, F. (2016). The Aggression-Submission-Conventionalism scale: Testing a new three factor measure of authoritarianism. *Journal of Social and Political Psychology, 4*(2), 571-600.

Ekström, J. (2011). *On the relation between the polychoric correlation coefficient and Spearman's rank correlation coefficient.* Retrieved from https://escholarship.org/uc/item/7j01t5sf

Ellis, B. H., Abdi, S. M., Horgan, J., Miller, A. B., Saxe, G. N., & Blood, E. (2016). Trauma and openness to legal and illegal activism among Somali refugees. *Terrorism and Political Violence, 27*(5), 857-883.

Flora, D. B., & Curran, P. J. (2004). An Empirical Evaluation of Alternative Methods of Estimation for Confirmatory Factor Analysis With Ordinal Data. *Psychol Methods, 9*(4), 466-491.

Fodeman, A. (2015). Safety and danger valves: functional displacement in American anti-abortion terrorism. *Behavioral Sciences of Terrorism and Political Aggression, 7*(3), 169-183.

Fodeman, A. D., Snook, D., & Horgan, J. (In Press). Picking Up and Defending the Faith: Activism and Radicalism among Muslim Converts in the United States. *Political Psychology*.

Fodeman, A. D., Snook, D., & Horgan, J. (2020). Pressure to prove for fear of faith: Muslim converts' activism and radicalism mediated by religious struggle and punishing Allah reappraisal. *Manuscript in preparation*.

Fox, J. (2016, 08 26). *Polychoric and Polyserial Correlations.* Retrieved 06 12, 2019, from http://CRAN.R-project.org/package=polycor

Freilich, J. D., Chermak, S. M., & Gruenewald, J. (2015). The future of terrorism research: a review essay. *International Journal of Comparative and Applied Criminal Justice, 39*(4), 353-369.

Freire, O. B., Senise, D. D., dos Reis, W. B., & Ono, H. S. (2017). # Step@ by# Step: recommendations for the development of high quality online research. *Revista Gestão & Tecnologia*, *17*(3), 10-35.

French, B. F., & Finch, H. (2016). Factorial Invariance Testing under Different Levels of Partial Loading Invariance within a Multiple Group Confirmatory Factor Analysis Model. *Journal of Modern Applied Statistical Methods*, *15*(1).

French, B. F., & Finch, W. H. (2006). Confirmatory factor analytic procedures for the determination of measurement invariance. *Structural Equation Modeling: A Multidisciplinary Journal, 13*(3), 378-402.

Funke, F. (2005). The dimensioality of right-wing authoritarianism: Lessons from the dilemma between theory and measurement. *Political Psychology, 26*(2), 195-218.

Ganor, B. (2002). Defining Terrorism: Is One Man's Terrorist Another Man's Freedom Fighter? *Police Practice and Research, 3*(4), 287-304.

Gielen, A. (2017). Countering Violent Extremism: A Realist Review for Assessing What Works, for Whom, in What Circumstances, and How? *Terrorism and Political Violence*, 1-19.

Gill, P., Horgan, J., Corner, E., & Silver, J. (2016). Indicators of lone actor violent events: The problems of low base rates and long observational periods. *ournal of Threat Assessment and Management, 3*(3-4).

Goodwin, J. (2012). Introduction to a Special Issue on Political Violence and Terrorism: Political Violence as Contentious Politics. *Mobilization: An International Journal, 17*(1), 1-5.

Gunning, J. (2009). Social Movement Theory and the Study of Terrorism. In R. Jackson, M. B. Smyth, & J. Gunning, *Critical Terrorism Studies: A New Research Agenda* (pp. 156-177). London: Routledge.

Hafez, M., & Mullins, C. (2015). The radicalization puzzle: a theoretical syntehsis of empirical approaches to hmegrown extremism. *Studies in Conflict and Terrorism*, *38*, 958-975.

Hamm, M. (2005). *Crimes committed by terrorist groups: theory, research and prevention.* Final Report to the National Institute of Justice.

Hart, S. D., Cook, A. N., Pressman, D. E., Strang, S., & Lim, Y. L. (2017). *A concurrent evaluation of threat assessment tools for the individual assessment of terrorism.* Waterloo, Ontario: Canadian Network for Research on Terrorism, Security and Society.

Hirschfeld, G., & von Brachel, R. (2014). Multiple-Group confirmatory factor analysis in R – A tutorial in measurement invariance with continuous and ordinal indicators. *Practical Assessment, Research & Evaluation, 19*(7).

Hoffman, B. (2006). *Inside Terrorism: revised and exanded edition.* Columbia University Press.

Horgan, J. (2003). Chapter 2: The Case for Firsthand Research. In A. Silke (Ed.), *Research on Terrorism: Trends, Achievements and Failures.*

Horgan, J. (2008). From Profiles to Pathways and Roots to Routes: Perspectives from Psychology on Radicalization into Terrorism. *The ANNALS of the American Academy of Political and Social Science*, *618*(1), 80-94.

Horgan, J. (2012). Interviewing the terrorists: reflections on fieldwork and implications for psychological research. *Behavioral Sciences of Terrorism and Political Aggression, 4*(3), 195-211.

Horgan, J. (2017). Psychology of Terrorism: Introduction to the Special Issue. *American Psychologist*, *72*(3), 199-204.

Horgan, J. G., & Boyle, M. J. (2008). A case against 'Critical Terrorism Studies'. *Critical Studies on Terrorism, 1*(1), 51-64.

Horgan, J., & Braddock, K. (2010). Rehabilitating the terrorists?: Challenges in assessing the effectiveness of de-radicalization programs. *Terrorism and Political Violence, 22*(2), 267-291.

Horgan, J., & Taylor, M. (2011). Chapter 13: Disengagement, de-radicalization, and the arc of terrorism: Future directions for research. In R. Coolsaet (Ed.), *Jihadi Terrorism and the Radicalisation Challenge: European and American Experiences* (2nd Edition ed.).

Hu, L. T., & Bentler, P. M. (1999). Cutoff criteria for fit indexes in covariance structure analysis: Conventional criteria versus new alternatives. *Structural equation modeling: a multidisciplinary journal*, *6*(1), 1-55.

Jöreskog, K. (1969). A general approach to confirmatory maximum likelihood factor aanalysis. *Psychometrika*, *34*, 183-202.

Jöreskog, K. (1994). On the estimation of polychoric correlations and their asymptotic covariance matrix. *Psychometrika, 59*, 381-389.

Jöreskog, K. (2005). *Structural equation modeling with ordinal variables using LISREL.* Lincolnwood, IL: Scientific Software International.

Jackson, R. (2007). Constructing Enemies: "Islamic Terrorism" in Political and Academic Discourse. *Government and Opposition, 42*(3), 394-426.

Kerodal, A. G., Freilich, J. D., & Chermak, S. M. (2016). Commitment to extremist ideology: using factor analysis to move beyond binary measures of extremism. *Studies in Conflict & Terrorism, 39*(7-8), 687-711.

Kim, E. S., & Yoon, M. (2011). Testing Measurement Invariance: A Comparison of Multiple-Group Categorical CFA and IRT. *Structural Equation Modeling, 18,* 212–228.

Kline, R. B. (2016). *Principles and Practice of Structural Equation Modeling.* (T. D. Little, Ed.) New York & London: The Guilford Press.

Knoke, D. (2015). Emerging trends in social network analysis of terrorism and counterterrorism. *Emerging Trends in the Social and Behavioral Sciences: An Interdisciplinary, Searchable, and Linkable Resource*, 1-15.

Kunreuther, H., Michel-Kerjan, E., & Porter, B. (2005). Extending Catastrophe Modeling to Terrorism and Other Extreme Events. In *Catastrophe modeling: A new approach to managing risk* (pp. 209-333). Boston, MA: Springer.

Kurzman, C., Kamal, A., & Yazdiha, H. (2017). Ideology and Threat Assessment: Law Enforcement Evaluation of Muslim and RIght-Wing Extremism. *Socius: Sociological Research for a Dynamic World, 3*, 1-13.

Lüdecke, D. (2019). *_sjPlot: Data Visualization for Statistics in Social Science_*. Retrieved from http://doi.org/10.5281/zenodo.1308157

LaFree, G., & Ackerman, G. (2009). The Empirical Study of Terrorism: Social and Legal Research. *Annual review of law and social science., 5*(1), 347-374.

Lederberg, A. R., Branum-Martin, L., Webb, M.-Y., Schick, B., Antia, S., Easterbrooks, S. R., & Connor, C. M. (2019). Modality and Interrelations Among Language, Reading, Spoken Phonological Awareness, and Fingerspelling. *Journal of Deaf Studies and Deaf Education*, 1-16. doi:10.1093/deafed/enz011

Lee, D. (2019, June 7). *G: The Miseducation of Larry P.* Retrieved from Radiolab: https://www.wnycstudios.org/podcasts/radiolab/articles/g-miseducation-larry-p

Lemieux, A. F., & Asal, V. H. (2010). Grievance, social dominance orientation, and authoritarianism in the choice and justification of terror versus protest. *Dynamics of Asymmetric Conflict*, *3*(3), 194-207.

Lemieux, A. F., Kearns, E. M., Asal, V., & Walsh, J. I. (2017). Support for political mobilization and protest in Egypt and Morocco: an online experimental study. *Dynamics of Asymetric Conflict*, *10*(2-3), 124-142.

Li, C.-H. (2016). The Performance of ML, DWLS, and ULS Estimation With Robust Corrections in Structural Equation Models With Ordinal Variables. *Psychological Methods*, *21*(3), 369–387.

Li, Z., Sun, D. Y., Li, B., Lin, Z. H., & Tang, M. (2015). Detecting Changes in Terrorism Networks: A Perspective of Multi-Meta-Network Modeling. *2015 European Intelligence and Security Informatics Conference* (p. 179). IEEE.

Logan, C., & Lloyd, M. (2018). Violent extremism: A comparison of approaches to assessing and managing risk. *Legal and Criminological Psychology*, 1-21.

Lubke, G. H., & Muthén, B. O. (2004). Applying Multigroup Confirmatory Factor Models for Continuous Outcomes to Likert Scale Data Complicates Meaningful Group Comparisons. *Structural Equation Modeling*, *11*(4), 514-534.

Lugtig, P., Boeije, H., & Lensvelt-Mulders, G. L. (2011). Change, what change? *Methodology*.

Lum, C., Kennedy, L. W., & Sherley, A. J. (2006). The Effectiveness of Counter-Terrorism Strategies: A Campbell Systematic Review.

Lum, C., Kennedy, L. W., & Sherley, A. J. (2006). *The effectiveness of counter-terrorism strategies: a campbell systematic reviewThe effectiveness of counter-terrorism strategies: a campbell systematic review.* The Campbell Collaboration, Crime and Justice Coordinating Group, Washington, DC.

Mahoney, C. W. (2018). More Data, New Problems: Audiences, Ahistoricity, and Selection Bias in Terrorism and Insurgency Research. *International Studies Review, 20*(4), 589-614.

Marsh, H. W., Hau, K.-T., & Wen, Z. (2004). In Search of Golden Rules: Comment on Hypothesis=Testing Approaches to Setting Cutoff Values for Fit Indexes and Dangers in Overgeneralizing Hu and Bentler's (1999) Findings. *Structural Equation Modeling, 11*(3), 320-341.

Masten, A. (2011). Resilience in children threatened by extreme adversity: Frameworks for research, practice, and translational synergy. *Development and Psychopathology, 23*(2), 493-506.

McArdle, J. J. (1988). Dynamic but Structural Equation Modeling of Repeated Measures Data. In J. R. Nesselroade, & R. B. Cattell (Eds.), *Handbook of Multivariate Experimental Psychology* (pp. 561-614). New York: Plenum Press.

McCauley, C., & Moskalenko, S. (2008). Mechanisms of political radicalization: Pathways toward terrorism. *Terrorism and Political Violence, 20*(3), 415-433.

McCauley, C., & Moskalenko, S. (2014). Toward a Profile of Lone Wolf Terrorists: What Moves an Individual From Radical Opinion to Radical Action. *Terrorism and Political Violence, 26*(1), 69-85.

McCauley, C., & Moskalenko, S. (2016). *Friction: How conflict radicalizes them and us.* Oxford University Press.

McCauley, C., & Moskalenko, S. (2017). Understanding Political Radicalization: The Two-Pyramids Model. *American Psychologist, 72*(3), 205-216.

Mehta, P. (2013). n-level structural equation modeling. In Y. Petsccher, C. Schatschneider, & D. L. Compton (Eds.), *Applied quantitative analysis in the social sciences* (pp. 329-362). New York: Routledge.

Millsap, R. E., & Olivera-Aguilar, M. (2012). Investigating measurement invariance using confirmatory factor analysis. In R. H. Hoyle (Ed.), *Handbook of structural equation modeling.* New York, NY: Guilford Press.

Millsap, R. E., & Yun-Tein, J. (2004). Assessing Factorial Invariance in Ordered-Categorical Measures. *Multivariate Behavioral Research, 39*(3), 479-515.

Mirahmadi, H. (2016). Building resilience against violent extremism: A community-based approach. *The ANNALS of the AMerican Academy of Political and Social Science, 668*(1), 129-144.

Mohamed, B. (2016). *A new estimate of the U.S. Muslim population.* Washington, DC: Pew Research Center.

Moher, D., Liberati, A., Tetzlaff, J., Altman, D., & Group, T. P. (2009). Preferred Reporting Items for Systematic Reviews and Meta-Analyses: The PRISMA Statement. *PLoS Medicine, 6*(7).

Morris, N. A., & Slocum, L. (2012). Estimating Country-Level Terrorism Trends Using Group-Based Trajectory Analyses: Latent Class Growth Analysis and General Mixture Modeling. *Journal of Quantitative Criminology, 28*, 103-139.

Moskalenko, S., & McCauley, C. (2009). Measuring political mobilization: The distinction between activism and radicalism. *Terrorism and political violence, 21*(2), 239-260.

Moyano, M., & Truijilo, H. M. (2014). ntention of activism and radicalism among Muslim and Christian youth in a marginal neighbourhood in a Spanish city/Intención de activismo y radicalismo de jóvenes musulmanes y cristianos residentes en un barrio marginal de una ciudad española. *Revista de Psicologia Social, 29*(1), 90-120.

Muthén, B. (1984). A general structural equation model with dichotomous, ordered categorical, and continuous latent variables indicators. *Psychometrika, 49*, 115-132.

Muthén, B., & Asparouhov, T. (2002). Latent Variable Analysis With Categorical Outcomes: Multiple-Group And Growth Modeling In Mplus. *Mplus web notes, 4*(5), 1-22. Retrieved from Mplus web notes.

Neumann, P., & Kleinmann, S. (2013). How Rigorous Is Radicalization Research? *Democracy and Security, 9*(4), 360-382.

Odgers, C. L., Moffitt, T. E., Broadbent, J. M., Dickson, N., Hancox, R. J., Harrington, H., & Caspi, A. (2008). Female and male antisocial trajectories: From childhood origins to adult outcomes. *Development and psychopathology, 20*(2), 673-716.

Office, U.S. Government Accountability. (2017). Countering Violent Extremism: Actions Needed to Define Strategy and Assess Progress of Federal Efforts.

Pew Research Center. (2011). *A demographic portrait of Muslim Americans.* Washington, DC.

Polaschek, D. L., Collie, R. M., & Walkey, F. H. (2004). Criminal attitudes to violence: Development and preliminary validation of a scale for male prisoners. *Aggressive Behavior, 30*(6), 484-503.

Pressman, D. (2009). *Risk Assessment Decision for Violent Political Extremism, 2009–02.* Ottawa: Government of Canada: SPJC Protocol.

Putra, I. E., & Sukabdi, Z. A. (2014). Can Islamic Fundamentalism Relate to Nonviolent Support? The Role of Certain Conditions in Moderating the Effect of Islamic Fundamentalism on Supporting Acts of Terrorism. *Peace and Conflict: Journal of Peace Psychology, 20*(4), 583-589.

Qureshi, A. (2016). *The 'Science' of Pre-Crime: The Secret 'Radicalisation' Study Underpinning Prevent.* CAGE.

Raykov, T., Marcoulides, G. A., & Li, C. (2012). Measurement invariance for latent constructs in multiple populations: A critical view and refocus. *Educational and Psychological Measurement*, *72*, 954-974.

Revelle, W. (2018). *psych: Procedures for Personality and Psychological Research.* Retrieved from https://CRAN.R-project.org/package=psych

Rhemtulla, M., Brosseau-Liard, P. É., & Savalei, V. (2012). When Can Categorical Variables Be Treated as Continuous? A Comparison of Robust Continuous and Categorical SEM Estimation Methods Under Suboptimal Conditions. *Psychological Methods, 17*(3), 354–373.

Rich, P., & Hoffman, B. (2004). *Research on Terrorism: Trends, Achievements & Failures.* Psychology Press.

Richardson, L. (2006). *What Terrorists Want: Understanding the Enemy, Containing the Threat.* New York, NY: Random House Inc.

Rokeach, M. (1956). Political and religious dogmatism: An alternative to the authoritarian personality. *Psychological Monographs: General and Applied*, *70*(18), 1-43.

Ross, J. I. (2004). Taking stock of research methods and analysis on oppositional political terrorism. *The American Sociologist, 35*(2), 26-37.

Rosseel, Y. (2012). Lavaan: An R Package for Structural Equation Modeling. *48*(2), 1-36.

Rutkowski, L., & Svetina, D. (2017). Measurement Invariance in International Surveys: Categorical Indicators and Fit Measure Performance. *Applied Measurement in Education, 30*(1), 39-51.

Sageman, M. (2014). The Stagnation in Terrorism Research. *Terrorism and Political Violence, 26*(4), 565-580.

Sarma, K. (2017). Risk Assessment and the Prevention of Radicalization from Nonviolence Into Terrorism. *American Psychologist, 72*(3), 278-288.

Sass, D. (2011). Testing measurement invariance and comparing latent factor means within a confirmatory factor analysis framework. *Journal of Psychoeducational Assessment., 29*, 347–363.

Sass, D. A., Schmitt, T. A., & Marsh, H. W. (2014). Evaluating Model Fit With Ordered Categorical Data Within a Measurement Invaiance Framework: A Comparison of Estimators. *Structural Equation Modeling: A Multidisciplinary Journal*, *21*, 167-180.

Saucier, G., Akers, L. G., Shen-Miller, S., Knezevic, G., & Staknov, L. (2009). Patterns of Thinking in Militant Extremism. *Perspectives on Psychological Science, 4*(3), 256-271.

Saucier, G., Stankov, L., & Knezevic, G. (2008). *Measurement of a culturally decentered construct of fanatical thinking.* University of Oregon, Eugene.: Unpublished manuscript.

Scarcella, A., Page, R., & Furtado, V. (2016). Terrorism, Radicalisation, Extremism, Authoritarianism and Fundamentalism: A Systematic Review of the Quality and Psychometric Properties of Assessments. *PLoS One, 11*(12).
Schmid, A. (2004). STATISTICS ON TERRORISM: THE CHALLENGE OF MEASURING TRENDS IN GLOBAL TERRORISM. *Forum on Crime & Society, 4*(1), 49-69.
Schmid, A. P. (2011). Glossary. In *The Routledge handbook of terrorism research* (pp. 598-706). Routledge.
Schmid, A. P. (2011). The Definition of terrorism. In T*he Routledge handbook of terrorism research.* (pp. 57-116). Routledge.
Schmid, A. P. (2012). The Revised Academic Consensus Definition of Terrorism. *Perspectives on Terrorism*, *6*(2).
Schmid, A. P. (2013). *Radicalisation, de-radicalisation, counter-radicalisation: A conceptual discussion and literature review.* International Center for Counter-Terrorism - The Hague.
Schmid, A. P. (2014). *Violent and Non-Violent Extremism: Two Sides of the Same Coin? Research Paper.* The Hague: ICCT.
Schmid, A. P. (2017). *Public Opinion Survey Data to Measure Sympathy and Support for Islamist Terrorism: A Look at Muslim Opinions on Al Qaeda and IS.* The Hague: ICCT Research Paper.
Schuurman, B. W., Grol, P., & Flower, S. (2016). Converts and Islamist terrorism: An introduction. *International Centre for Counter-Terrorism Policy Brief, 7*(21).
Silke, A. (2001). The Devil You Know: Continuing Problems with Research on Terrorism. *Terrorism and Political Violence, 13*(4), 1-14.
Silke, A. (2003). Becoming a terrorist. *Terrorists, victims and society: Psychological perspectives on terrorism and its consequences.*, 29-53.
Silke, A. (2006). The impact of 9/11 on research on terrorism. In *Mapping terrorism research* (pp. 90-107). Routledge.
Silke, A. (2008). Research on terrorism: a review of the impact of 9/11 and the global QWar on terrorism. In H. e. Chen (Ed.), *Terrorism informatics: knowledge management and data mining for homeland security* (pp. 27-50). New York, NY: Springer.
Snook, D. (2018). *Zeal of the Convert? Comparing the Structure of Islamic Religiousness Between Convert and Non-Convert Muslims.* (thesis, in press), Georgia State University.
Snook, D. W., Branum-Martin, L., & Horgan, J. G. (In Review). Zealous or Just Different? Comparing the Structure of Islamic Religiousness Between Convert and Non-convert Muslims.
Snook, D., Branum-Martin, L., & Horgan, J. (2019). Exploring the Structure of Islamic Religiousness Between Convert and Non-Convert Muslims in the United States. *Manuscript in review.*
Snook, D., Horgan, J., Kleinmann, S., & White, G. (In Press). Conversion Motifs Among U.S. Muslim Converts. *Psychology of Religion and Spirituality*.
Soliman, A., Bellaj, T., & Khelifa, M. (2016). An integrative psychological model for radicalism: Evidence from structural equation modeling. *Personality and Individual Differences, 95*, 127-133.
Stampnitzky, L. (2010). Disciplining an Unruly Field: Terrorism Experts and Theories of Scientific/Intellectual Production. *Qualitative Sociology, 34*(1), 1-19.

Stampnitzky, L. (2014). *Disciplining Terror: How Experts Invented 'Terrorism'.* Cambridge: Cambridge University Press.

Stankov, L., Higgins, D., Saucier, G., & Knežević, G. (2010). Contemporary Militant Extremism: A Linguistic Approach to Scale Development. *Psychological Assessment, 22*(2), 246-258.

Stankov, L., Knezevic, G., Saucier, G., & Radovic, B. (2018). Militant Extremist Mindset and the Assessment of Radicalization in the General Population. *Journal of Individual Differences, 39*(2), 88-98.

Stankov, L., Saucier, G., & Knežević, G. (2010). Militant extremist mind-set: Proviolence, Vile World, and Divine Power. *Psychology assessment, 22*(1), 70-86.

Steiger, J. H., & Lind, J. C. (1980). Statistically-based tests for the number of common factors.

Steinmetz, H. (2013). Analyzing Observed Composite Differences Across Groups: Is Partial Measurement Invariance Enough? *Methodology, 9*(1), 1-12.

Strang, K. D., & Zhaohao, S. (2017). Analyzing Relationships in Terrorism Big Data Using Hadoop and Statistics. *Journal of Computer Information Systems, 57*(1), 67-75.

Svetina, D., & Rutkowski, L. (2017). Multidimensional measurement invariance in an international context: Fit measure performance with many groups. *Journal of Cross-Cultural Psychology, 48*(7), 991-1008.

Svetina, D., Rutkowski, L., & Rutkowski, D. (2019). Multiple-Group Invariance with Categorical Outcomes Using Updated Guidelines: An Illustration Using Mplus and the Lavaan/semTools Packages. *Structural Equation Modeling: A Multidisciplinary Journal, 0*, 1-20.

Swartout, K. M., Thompson, M. P., Koss, M. P., & Su, N. (2015). What is the Best Way to Analyze Less Frequent Forms of Violence? The Case of Sexual Aggression. *Psychology of Violence, 5*(3), 305-313.

Taiyun, W., & Viliam, S. (2017). *R package "corrplot": Visualization of a Correlation Matrix.* Retrieved from https://github.com/taiyun/corrplot

Timofeeva, A. (2017). Data Type Detection for Choosing an Appropriate Correlation Coefficient in the Bivariate Case. *YSIP*, 188-194.

U.S. Military. (2014, October 24). *3-26*. Retrieved from http://www.dtic.mil/doctrine/new_pubs/jp3_26.pdf

Vajjhala, N. R., Strang, K. D., & Sun, Z. (2015). Statistical modeling and visualizing open big data using a terrorism case study. *2015 3rd International Conference on Future Internet of Things and Cloud* (pp. 489-496). IEEE.

van de Schoot, R., Lugtig, P., & Hox, J. (2012). A checklist for testing measurement invariance. *Euroepan Journal of Developmental Psychology, 9*(4), 486-492.

van den Berg, H., van Hemert, D. A., & van Vliet, A. J. (2018). Creating a Knowledge Base for Interventions Countering (Violent) Extremism: Intervention Goals and Mechanisms for Different Ideologies. *Journal for Deradicalization, 15*, 43-75.

Vandenberg, R. J., & Lance, C. E. (2000). A review and synthesis of the measurement invariance literature: Suggestions, practices, and recommendations for organizational research. *Organizational research methods, 3*(1), 4-70.

Veldhuis, T. M., & Kessels, E. J. (2013). *Thinking before leaping: The need for more and structural data analysis in detention and rehabilitation of extremist offenders.* The Hague: The International Centre for Counter-Terrorism.

Victoroff, J. (2005). The Mind of the Terrorist: A Review and Critique of Psychological Approaches. *Conflict Resolution, 49*(1), 3-42.

Wall, M. (2019, August 27th). *Exploratory Factor Analysis.* Retrieved from Popualtion Health Methods: https://www.mailman.columbia.edu/research/population-health-methods/exploratory-factor-analysis

Webber, K. (2014). *School engagement of rural early adolescents: Examining the role of academic relevance and optimism across racial/ethnic groups.* Unpublished doctorl dissertation, The University of North Carolina at Chapel Hill.

Wegmann, K.M. (2014). *A mixed-methods exploration of stereotype threat in middle childhood.* Unpublished doctoral dissertation, The University of North Carolina at Chapel Hill.

Weine, S. (2016). Resilience and countering violent extremism. In U. Kumar (Ed.), *The Routledge International Handbook of Psychosocial Resilience.* London, United Kingdom: Routledge.

Weine, S., & Younis, A. (2015). Developing CVE Programs through Building Community Policing Capacities. In S. Zeiger, & A. Aly (Eds.), *Countering Violent Extremism: Developing an Evidence-base for Policy and Practice.* Perth, Western Australia: Curtin University.

West, S. G., Taylor, A. B., & Wu, W. (2012). Model fit and model selection in structural equation modeling. In R. H. Hoyle (Ed.), *Handbook of structural equation modeling* (pp. 209-231). New York, NY: Guilford Press.

White, G. (2019). *Over-representation or Over-reaction? Converts to Islam in Jihadist Terrorism.* Manuscript in preparation.

Wickham, H. (2016). *ggplot2: Elegant Graphics for Data Analysis.* New York: Springer-Verlag.

Williams, M. J., Horgan, J. G., & Evans, W. P. (2016). The critical role of friends in networks for countering violent extremism: toward a theory of vicarious help-seeking. *Behavioral Sciences of Terrorism and Political Aggression, 8*(1), 45-65.

Wu, H., & Estabrook, R. (2016). Identification of COnfirmatory Factor Analysis Models of Different Levels of Invariance for Ordered Categorical Outcomes. *Psychometrika, 81*(4), 1014-1045.

Young, J. K., & Findley, M. G. (2011). Promise and Pitfalls of Terrorism Research. *International Studies Review, 13*, 411-431.

Zeiger, S., Aly, A., Neumann, P. R., El Said, H., Zeuthen, M., Romaniuk, P., . . . Koehler. (2015). *Countering violent extremism: Developing an evidence-base for policy and practice.* Hedayah and Curtin University.

Zeng, P., Wei, Y., Zhao, Y., Liu, J., Liu, L., Zhang, R., & et al. (2014). Variable selection approach for zero-inflated count data via adaptive lasso. *Journal of Applied Statistics, 41*(4), 879-894.

7 APPENDIX

7.1 Ordinal ME/I Conundrum: Scalar vs. Threshold Invariance

There is far more disagreement with regards to the appropriate ME/I testing procedure for ordinal data, as compared to continuous data, because there are many issues with model identification (Wu & Estabrook, 2016). As Vandenberg and Lance note (2000), ordinal data do not have true means (ν's), but thresholds for each response interval (τ, i.e., log-likelihoods of an individual choosing ordinal item response 2 over 1, 3 over 2, etc.). Therefore, τ's replace the ν's tested in Scalar Invariance (review section 2.3.2 above). However, factor loadings (λ's) for ordinal data are inextricably linked to the underlying τ's of the observed ordinal responses. As a result, there is disagreement as to whether ordinal indicators' λ's and τ's should be constrained and freed simultaneously or separately (Bowen & Masa, 2015)—i.e., if it is possible to test for ordinal Metric ($\lambda_{Group1} = \lambda_{Group2}$) and Scalar Invariance ($\tau_{Group1} = \tau_{Group2}$) independently. Some researchers (e.g., Sass, 2011), including Muthén and Muthén in the MPlus User's Guide (1998-2012), argue for joint constraints, as λ's and τ's *jointly* define item functioning. There would be no separate test for invariant λ's, instead excluding the Metric Invariance step. Other researchers[23], however (Webber, 2014; Wegmann, K.M., 2014), argue that because loadings (λ's) and thresholds (τ's) contribute different information about item functioning, they should be constrained and freed *separately* so as to pinpoint and interpret specific sources of noninvariance (e.g. τ_{42}). In fact, some researchers (Wu & Estabrook, 2016; Svetina, Rutkowski, & Rutkowski, 2019) recommend testing for "Scalar"—otherwise known in this case as "Threshold" Invariance ($\tau_{Group1} = \tau_{Group2}$)—*before* Metric Invariance ($\lambda_{Group1} = \lambda_{Group2}$) given that individual τ's within ordinal items might be invariant and could be freed separately

[23] Plus the Muthéns *outside* of the MPlus Manual (Lubke & Muthén, 2004; Muthén & Asparouhov, 2002).

(Millsap & Yun-Tein, 2004). This contradicts standards for continuous data ME/I testing[24]. In order to avoid the disadvantages of any one ordinal ME/I testing pathway strategy, all paths are tested in this thesis. That is, τ's and λ's are tested both independently (i.e., separate Metric and Threshold Invariance models) as well as combined (i.e., the Scalar Invariance model).

7.2 Statistical Estimator Choice

There are several estimation methods, each with their own strengths and weaknesses. The most common for ME/I is Maximum Likelihood (ML), more typically used for Ordinary Least Squares (OLS) regression. However, for ordinal and other categorical data, most researchers recommend utilizing *Weighted* Least Squares (WLS) and its variants (e.g., Mean-and-Variance-adjusted WLS or WLSMV: Jöreskog, 2005). Diagonally-weighted matrices like those in robust WLS reduce *n* requirements and prevent certain convergence problems when modeling ordinal data (Bovaird & Koziol, 2012). Even robust ML (i.e., with relaxed assumptions of normality) is inferior to WLS in controlling for Type I error, save for in large *n's* (i.e., $n > 1{,}000$) for $\Delta\chi^2$ tests (Li, 2016). (Robust) WLS also provides more accurate factor loadings (λ), standard error, and inter-factor correlation estimates than (robust) ML regardless of simulation conditions (Li, 2016). This is especially true for large λ's or asymmetric thresholds (τ's: Rhemtulla, Brosseau-Liard, & Savalei, 2012; Sass, Schmitt, & Marsh, 2014) like those found in this thesis sample[25]. Many researchers recommend using WLSMV in particular for estimation (Flora & Curran, 2004; DiStefano & Morgan, 2014; Sass, Schmitt, & Marsh, 2014; Bovaird & Koziol, 2012). WLSMV estimations yields better fit and convergence likelihood generally than WLS estimation (DiStefano & Morgan, 2014), especially with smaller *n's* (Flora & Curran, 2004) like in this

[24] However, some researchers do recommend testing continuous data for intercept invariance separately from loading invariance before the typical combined Scalar Invariance model (van de Schoot, Lugtig, & Hox, 2012).
[25] The λ's were high (.85-.94); the positive skew and inflated 0's produce asymmetric thresholds.

thesis (n = 356 for two groups). While (robust) ML often displays greater power to detect *Scalar* noninvariance compared specifically to WLSMV, ML demonstrates lower power to identify *Metric* noninvariance (Sass, Schmitt, & Marsh, 2014). Overall, WLSMV is the optimal estimation method for small *n's*, asymmetric τ's, and large λ's—like those found in this thesis.

7.3 GoF Indices for Ordinal Indicators

There are many different model goodness-of-fit (GoF) indices for researchers to consider. GoF indices measure discrepancies between expected and observed outcomes. They are useful not only to determine how good a fit a single model is, but to compare models. Higher versus lower degrees-of-freedom (*df*)—i.e., more restricted or fewer estimated parameters—lead to poorer fit. In ME/I testing, ever more stringent models (i.e., Configural, Metric/Threshold, or Scalar) subsequently increase *df* and typically reduce GoF. Researchers disagree, however, as to how dramatic a reduction in GoF between more and less stringent models signifies noninvariance (i.e. a cutoff score for change in GoF), as well as which GoF indices are most appropriate, reliable, or sensitive for different model conditions (i.e., model complexity, *n*, data type and distribution). The GoF indicators that, based on the literature (Kline, 2016; Chen, 2007; Svetina, Rutkowski, & Rutkowski, 2019), are appropriate, reliable, and sensitive for comparing ME/I models with ordinal indicators are: chi-squared (χ^2), which assesses the degree of discrepancy between the sample and fitted covariance matrices, with a p-value based on an H_0 of, "The model fits perfectly;" the Root Mean Square Error of Approximation (RMSEA), which measures the same discrepancy, but relative to *df* and *n,* and for which 0 represents a perfect fit, but there is no hypothesis test of significantly poorer fit than the null model (though cutoffs are recommended); the Comparative Fit Index (CFI), which demonstrates incrementally superior fitting models as compared to the null model (manifest covariance matrix) from 0 (poorest fit) to

1 (perfect fit), with recommended cutoffs; and the Weighted Root Mean Square Residual (WRMR), designed specifically for ordinal data modeled with robust WLS estimators or non-normal continuous data with robust ML estimators by weighting the average differences in sample versus fitted covariances, and for which, like RMSEA, lower values represent better fit. Each of these GoF indicators present different strengths and weaknesses considered below.

Table 7. GoF Index Comparisons

GoF Index	Developer	Better Fit Direction	Range	Cutoff Criteria	*n* Size Type I/II Error Rate Inflation	Complexity Sensitivity
χ^2	(Jöreskog, 1969)	Lower	≥ 0	$p < .05$[a]	Both[a]	No[a]
χ^2/df	(Jöreskog, 1969)	Lower	≥ 0	< 5.0[a]	Both[a]	Yes[a]
RMSEA	(Steiger & Lind, 1980)	Lower	> 0	**≤ .06[a]**	Small *n*[a]	Yes[a]
CFI	(Bentler, 1990)	Higher	0 - 1	**≥ .95[a]**	No[a]	Yes[a]
WRMR	(Muthén, 1998-2004)	Lower	> 0	**< 1.0[b]**	Large *n*[b]	Yes[a]

Note. This table is based on Table 13.1 by West, Taylor & Wu (2012). **Bolded** criteria specified for ordinal data.
Note. Superscripts refer to sources a) West, Taylor & Wu (2012) and b) DiStefano, Liu, Jiung, & Shi (2017).
Note. "Small *n*" refers to increased Type II error rate with small *n*'s, "Large *n*" refers to increased Type I error rate with large *n*'s, "Both" refers to risks heightened at either *n* extremes, while "No" refers to no risks relative to *n*.

7.4 Choosing GoF Indices & Cutoffs

To establish GoF, statisticians have proposed various GoF index cutoff values by which a model can be judged to have a reliably good fit relative to a null or baseline model (see Table 7 above) χ^2 tests, while reported for every model test, assume that 1) manifest variables are normally distributed and 2) the *n* is large (West, Taylor, & Wu, 2012), neither of which are true of this thesis. χ^2 and its derivative χ^2/df serve better as descriptive indicators of relative model fit rather than absolute benchmarks. For ordinal estimation, West, Taylor and Wu recommend that all models must have CFI $\geq .95$ and RMSEA $\leq .06$ to be considered[26] (2012), though the latter is sensitive to small *n*’s like in this thesis. WRMR was designed for ordinal data with a cutoff of < 0.90[27]. However, WRMR is particularly useful to compare models whose, “sample statistics have widely varying variances… [or] are on different scales” (Muthén & Muthén, 1998-2012), unlike this thesis with all similarly skewed 7-point Likert items.

For comparing models, statisticians propose another set of GoF cutoffs as to whether the GoF *difference* (Δ) between the *more* restricted model and the *less* restricted model (e.g., $M_{Configural}$ - M_{Metric}) is significant. Increasingly equality-constrained models inherently worsen model fit due to increased degrees of freedom. Methodologists debate what measure and degree of ΔGoF indicates noninvariance (see Table 8 below). While researchers widely use ΔCFI $\leq -.010$ as indicative of noninvariance for ordinal data[28] (Cheung & Rensvold, 2002), some statisticians have shown by simulation that optimal cutoffs for ΔCFI, or $\Delta\chi^2$ for that matter, are strongly biased by model complexity[29] (Chen, 2007). ΔRMSEA and especially $\Delta\chi^2$ perform well for ME/I testing of MCCFA regardless of data type, degree of DIF, and source of noninvariance (Kim & Yoon, 2011; Sass, Schmitt, & Marsh, 2014), although both are subject to increased risk of Type II error rates with small *n*’s and Type I error rates with large *n*’s (West, Taylor, & Wu, 2012). Conversely, CFI is relatively independent from *n* and therefore avoids those increased error rates (Chen, 2007; Hu & Bentler, 1999). Therefore, as Rutkowski and Svetina recommend in their ME/I GoF cutoff review (2017), all three indicators ($\Delta\chi^2$, ΔCFI, and ΔRMSEA) will be considered together, although ΔCFI will be relied upon the most.

26 Raykov et al. (2012) note baselines need not meet fit criteria before testing Configural Inv.
27 Though DiStefano, Liu, Jiung, and Shi (2017) argue a cutoff of < 1.0 is sufficient, as above.
28 The same standards are confirmed for multivariate normal models (French & Finch, 2006).
29 Note that most simulations thereof largely only use ML, not WLS.

Table 8. Relevant GoF Cutoff Values Indicating Noninvariance

Source	*Data & Model Conditions*				*ΔGoF Cutoffs Indicating Noninvariance*			
	Groups	*n*/group	Factors	Data Type	$\Delta\chi^2$ *p*	ΔCFI	ΔRMSEA	Model
(Chen, 2007)	**2**	**150-** 300	**1**	Cont.		≤ -.005	≥ .010	All
(French & Finch, 2006)	**2**	**150-** 500	***2***	**Ordinal**	< .05			All
(Rutkowski & Svetina, 2017)	10	600-6K	**1**	**Ordinal**	< .05	≤ -.004	≥ .005	Metric
(cont.)					< .05	≤ -.004	≥ .001	Scalar
(Svetina & Rutkowski, 2017)	10	750-6K	***2***	**Ordinal**	< .05		≥ .005	Metric
(cont.)					< .05	≤ -.002	≥ .001	Scalar

Note. **Bolded** conditions match this thesis, while ***bolded italicized*** ones might not (see AIS issues in section 4.2).

Note. Table 3 adapted from Svetina, Rutkowski & Rutkowski's Table 1 (2019). ΔGoF is *more* restricted minus *less*.

7.5 ME/I Summary Statistics Reporting Procedures for Ordinal Indicators

The literature recommends the following reporting procedures for FA (Jöreskog, 1994; Muthén B. , 1984): "first order statistics," i.e. frequencies, thresholds, means, and variances, then "second order statistics," i.e. polychoric correlations[30] between those ordinal variables, followed by the parameters of the structural part of the model. Polychoric correlations are estimated with the *polychoric* function from the "psych" package (Revelle, 2018), which is based on the package *polycor* (Fox, 2016). The two-step method is employed, estimating thresholds separately from the marginal distribution of each variable before calculating ρ (see Fox, 2016 for details). Note that polychoric correlations are better suited for statistical inferences from ordinal response

[30] Correlations of latent response variables, not ordinal outcomes directly (Timofeeva, 2017).

categories than Spearman's rank coefficient (Ekström, 2011), reported previously above, and therefore will be used for analysis instead. Note as well that response frequencies and thresholds are reported, but neither means, SDs, nor variances are. While standard practice reporting for continuous data (Jöreskog, 1994; Muthén B. , 1984), means and SDs are arguably not appropriate to report for ordinal data as they do not have true means. Similarly, no indicator variances are estimated with ordinal logistic regression, only latent response variance. Summary statistics tables are relegated to the appendix below, but above in section 4 (Results) are many of their visualizations—more succinct and clear ways of reporting that information.