5-2-2022

(Statistical) Power to the People: Testing for Measurement Invariance Across Integrated Radicalism Research

Ari Fodeman

(Statistical) Power to the People: Testing for Measurement Invariance Across Integrated Radicalism Research

by

Ari Fodeman

Under the Direction of Dr. John Horgan

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

Doctor of Philosophy

in the College of Arts and Sciences

Georgia State University

2022

ABSTRACT

Radicalism, extremism, and related phenomena have been measured myriad ways, with little standardization. The most widely used metric—the Activism and Radicalism Intentions Scales (the ARIS: Moskalenko & McCauley, 2009)—has been translated, rescaled, reworded, reorganized, and used with populations never originally tested or necessarily intended for, with little scrutiny. To support the ARIS's use across the past decade of research, I tested for Measurement Equivalence/ Invariance (ME/I) via Integrated Data Analysis (IDA) using ordinal logistic regression. The 13 harmonizable bodies of ARIS research that observed the same three RIS items using the same ordinal scale demonstrated Configural, Metric, Threshold, and Scalar Invariance; Decker and Pyrooz's (2019) latent RIS mean was the only parameter that needed to be freed to establish Partial Latent Invariance. Decker and Pyrooz's significantly higher latent RIS scores were not unexpected, as they were the only cohort to study criminals, and prior criminality is a common positive correlate of radicalism. While this work gives some credence to the use of the ARIS across multiple study contexts, more in-depth analyses with larger sample sizes will have to test for ME/I between cross-classified cohorts (e.g., by translation, country, age group, general vs. specific vs. at risk populations, etc.), AIS, and the other two RIS items. When advanced statistical techniques such as Moderated Non-Linear Factor Analysis (MNLFA) are further developed, future studies will also have to test for ME/I across rescaling of ARIS items, likely requiring a bridging study in which multiple scales are given to the same participants. It is this type of intensive, rigorous data collection and statistical analysis found in most other content areas to which we radicalism researchers can aspire.

INDEX WORDS: Radicalism, ARIS, Violent extremism measurement, Measurement invariance, Integrated data analysis, Ordinal indicators

(Statistical) Power to the People: Testing for Measurement Invariance Across Integrated Radicalism Research

by

Ari Fodeman

Committee Chair: John Horgan

Committee: Anthony Lemieux

Dominic Parrott

Elizabeth Tighe

Therese Pigott

Electronic Version Approved:

Office of Graduate Services

College of Arts and Sciences

Georgia State University

May 2022

DEDICATION

I dedicate this dissertation to my loving family and friends, who have supported me all these years while I sought the highest level of education. I could not have gotten here and become the man I am today were it not for you guys—or, as we say down here in GA, “y’all.” Also as we say down here, bless your hearts for everything you have done for me, most especially my parents, who instilled in me the importance of education and a love for science from the youngest age possible—not to mention how much you saved on my behalf so that I could, between your savings and my scholarships (in no small part due, once again, to the importance you placed on education, dedication, and hard work), obtain my bachelors and masters with few to no loans. With the ever-rising price of education in this country, and the therefore incredible savings that parents need to amass for their children, I count myself as incredibly lucky to have been afforded these educational opportunities. I know that I stand in an incredibly privileged position, and I hope, as you two have both shown me by your own professional choices and damn big hearts, that I can use that privilege to give back and do some good in this world with all this schooling. In the pursuit of knowledge for the good of others, I therefore dedicate this dissertation first and foremost to you two.

ACKNOWLEDGEMENTS

I want to thank everyone who has helped me in preparing this dissertation, and all the steps leading up to it. I want to give a special thank you first and foremost to my dissertation 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. He has also pointed me to fellowships and upheld a standard of performance to maintain them, by which I have been able to sustain myself over all these years. This dissertation would also not be possible without Drs. Anthony Lemieux, Dominic Parrott, Elizabeth Tighe, and Therese Pigott providing their range of content and methodological expertise on my committee. I greatly appreciate the time, energy, and attention that they have spent to help me attain this last milestone to my doctorate. On that note, I have to also acknowledge Dr. Lee Branum-Martin, whose incredible encouragement of my love for learning about measurement directly fueled my progression to this point of statistical aptitude, by which this dissertation would not be possible. So, too, I have to thank Mr. Jeremy Walker and Dr. Halley Riley from the Research Data Services team at GSU's Library who dove with me into the minutia of statistical problem solving and software conundrums. I also should thank Dr. Patrick Curran and his colleagues Drs. Andrea Hussong and Daniel Bauer who, as the progenitors of MNFLA and all-around statistical gurus, illuminated to me some of the statistical pitfalls I was facing in tackling problems unique to this particular IDA. Lastly, Drs. Yves Rosseel and Terrence D. Jorgensen made this dissertation possible not only for creating Lavaan and semTools, but publicly answering myriad questions and cries for help from those before me who faced similar software problems.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

1 SPECIFIC AIMS

This research project tested the theoretical structure and potential measurement bias of radicalism across a decade of archived data to pave a path for future data collection and integration in the field. Radicalism has thus far been subject to measurement with myriad metrics, but surprisingly little scrutiny. Even the Activism and Radicalism Intentions Scales (ARIS)—one of the most highly regarded and widely used instruments for measuring radicalism—has been used with limited accounting for, and tempering of, its adaptations, translations, and applications to diverse populations. The sensitive nature of the ARIS's subject matter necessitates more nuanced and precise tests of scale reliability and equitable comparisons than have been employed in the literature. Such threats to internal and external validity, as well as reliability, are harmful to radicalism research as a scientific pursuit and the ARIS as a scientific tool. With an integrated data analysis (IDA) approach using multiple group structural equation modeling (MGSEM), I tested for measurement equivalence/invariance (ME/I, i.e. differential item functioning (DIF)) of the ARIS across 13 harmonizable bodies of research from different sampled populations[1] and survey translations. Establishing ME/I will allow future IDA to test for regression invariance of theoretical radicalism covariates (e.g., risk and protective factors) and directly compare models of radicalism, its causes and consequences. Direct comparisons via IDA are more statistically powerful and reliable than meta-analysis, which already is more objective than literature or realist (i.e., qualitative) reviews that are more commonly found in radicalism research.

[1] e.g., those "at [greater] risk" (or those thought to likely experience higher risk factors and lower protective factors) vs. general populations, by gender, ethnicity, country, religion, etc. See section 3.1.3 for a discussion.

Integrating this data lays the groundwork for future planned experimental complementarity[2] (Fischer & Dinklage, 2007), and therein facilitates radicalism research collaboration. IDA's increased power over traditional meta-analysis and single studies to detect and broaden research findings is particularly advantageous given that social science researchers often observe small, nuanced samples and effects—especially radicalism researchers, given radicalism's low base rate and skew. As one of the first studies employing IDA to radicalism data, this dissertation helps radicalism research approach levels of statistical power and scrutiny promoted in fields like medicine, economics, and education. This is especially important as the use of poor statistical techniques, if any, in our field have been criticized for decades (Silke, 2001; Rich & Hoffman, 2004; Ross, 2004; LaFree & Ackerman, 2009; Neumann & Kleinmann, 2013; Sageman, 2014; Schuurman, 2018; Stampnitzky, 2010; Wolfowitz et al., 2020b), and there has been a call for not only more data collection, but more *structured* collection, with standardized tools (Veldhuis & Kessels, 2013)—i.e., more data complimentarity.

2 IMPACT

This dissertation helps bring radicalism research to the modern statistical standards used in many other content areas. The 'push to publish' felt across most fields, as well as a similar impetus for practitioners and policy makers to put programs into action (even when the evidence base is nascent and scant, if present at all), may be particularly acute in violent extremism prevention. Plagued by ever-changing existential "dread risks" (Gigerenzer, 2004), violent extremism prevention often puts reaction above rigor and replicability. Research using the ARIS

[2] That is, the similarity, and therein comparability if not integrability, of research design, variable choice and measurement, as contrasted with experimental heterogeneity (Fischer & Dinklage, 2007).

and similar tools have tapped into myriad theories of radicalism's structure, nature, emergence, causes, consequences, and therein prevention, without strongly demonstrating statistically the generalizability and reliability of their findings. This dissertation work not only provides nuanced evidence, using cutting-edge methods, that many of these studies are, in fact, discussing the same outcomes (i.e., comparing apples to apples), but sets the groundwork for subsequent studies to directly compare different theories of radicalism's covariates and predictors via integrated data. This is particularly apt given that competing theories about radicalism's origin clamor for practitioner's and policy maker's attention, though they have few tools to test the mettle of one theory over another. Testing for ME/I on a measure as prolific as the ARIS will be a huge step towards the kind of fastidiousness and scrupulousness we have come to expect from other preventive fields, such as public health.

3 BACKGROUND & SIGNIFICANCE

3.1 Radicalism: A Brief Primer

3.1.1 Activism, Radicalism, & Mobilization

Radicalism and related behaviors have been defined many ways. A well founded and accepted paradigm is McCauley and Moskalenko's Political Mobilization definitions (2009) and Two Pyramid Model (2017). The ARIS was built upon the former. They define political mobilization as, "increasing extremity of beliefs, feelings, and actions in support of intergroup conflict" (2009). This umbrella term includes the subcategories 'activism' and 'radicalism.' Activism, or "legal and non-violent political action" (Moskalenko & McCauley, 2009), includes various behaviors like volunteering, voting, protesting, lobbying, political campaigning and financing[3]. Radicalism, conversely, is the class of political mobilization that is *il*legal, if not violent. This includes terrorism, as well as other types of political violence (e.g., war or insurgency). In terrorism research[4], extremism is often used synonymously with radicalism, although the modifier "violent" may be attached. Some behaviors are *non*-violent, but illegal (e.g., civil disobedience or guerilla protest), and fall *between* activism and radicalism. Moskalenko and McCauley (2017) model actors along a spectrum of political mobilization extremity that, in terms of frequency, befits a pyramid (see Figure 1 below from Fodeman, 2020)—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 two-headed arrows in Figure 1 below). The Pyramid Model (see also Wolfowicz et al, 2019) is juxtaposed to linear models of terrorism engagement, such as Moghaddam's Staircase (2005; see Hafez & Mullins, 2015 for discussion) or the conveyor belt metaphor.

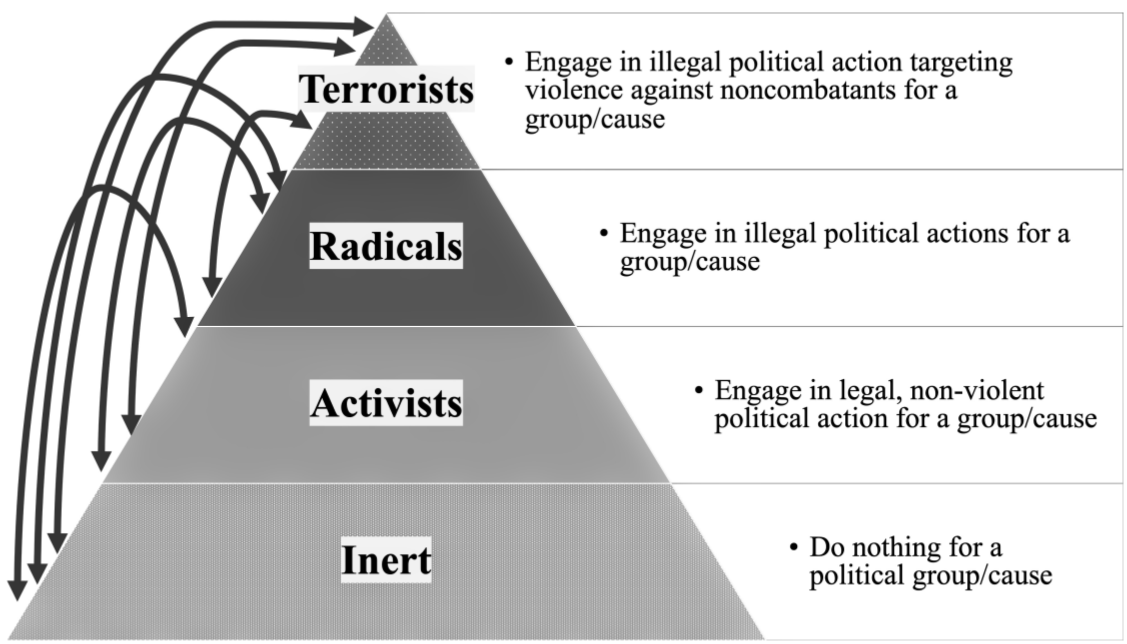

Figure 1. *Political Action Pyramid (adapted from Moskalenko & McCauley, 2017).*

[3] You also see scholars 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.

3.1.2 Mobilization vs. Radicalization

When radicalism scholars discuss risk and protective factors (Wolfowitz et al., 2020), they refer to factors that make one more or less likely, respectively, to move up the pyramid (see the discussion below Table 2 of factors tested specifically using the ARIS). While not depicted above, Moskalenko and McCauley (2017) distinguish radical beliefs from radical behaviors—like many other scholars (see Stern, 2016 or Borum, 2017 for discussions)—with two distinct but related pyramids. Like nonlinear movement within a pyramid, movement on one pyramid does not necessitate movement on the other (e.g., one might remain behaviorally Inert even if one's beliefs have become Radical). When juxtaposed with *radicalization*, *mobilization* then refers exclusively to increasing extremity of *actions*, while the former refers to *beliefs* and *attitudes*. Radicalization and mobilization are not prerequisites for one another (e.g., an individual can be conscripted into mobilization or otherwise mobilize for non-radical reasons, such as monetary incentives or family ties), though radicalization may be one among many risk factors for mobilization (Horgan & Braddock, 2010).

3.1.3 To be "At [Greater] Risk"

No group of people are exempt from risk and protective factors to both radicalization and mobilization, and those factors therefore may be detected across any general population (Rousseau, Hassan, & Oulhote, 2017). While there is baseline general population susceptibility, many scholars, including those in research using the ARIS[5], often refer to some populations of interest they study as "at [greater] risk"—likely experiencing higher risk factors and lower protective factors. These factors are usually not individual-level factors (e.g., personality or beliefs), but community-level factors (e.g., group discrimination or oppression). Note that to be

[5] While a complete list of all studies would be too long, the aforementioned reviews and meta-analyses discuss them.

at greater risk of radicalization or mobilization is not any kind of pathology that would make one inherently different from general or 'normal' populations (Horgan, 2008). As Rousseau and colleagues note (2019a), we ought to be weary of "at risk" categorization engendering profiling and stigmatization of the very populations we would hope to help protect with our research. Such cautions are quite evidently taken in studies using the ARIS. The term "at risk" will remain in quotation marks throughout this document as a reminder of the grain of salt to take with the term.

3.2 The ARIS

3.2.1 ARIS Structure

The ARIS includes 10 items (see

Table **1** below) measuring surveyee's intentions of engaging in activism (i.e., the Activism Intentions Scale, AIS) and radicalism[6] (i.e., the Radicalism Intentions Scale, RIS). No previous scale captured both legal and illegal political behaviors. This is particularly important as activism and radicalism are intimately tied (Moskalenko & McCauley, 2009). These items cover a spectrum of political mobilization. The RIS items are phrased with less specificity than the AIS items to counteract social desirability bias. Each ARIS survey item references "my group" (see

Table **1** below). Depending on the study, participants either A) state via open response the group that is most important to them, B) choose from a list of options (e.g., national, ethnic, religious, or political), or C) are referred to an already acknowledged group membership (see Table 2 below for sampled populations). Moskalenko and McCauley (2009) posed each item on a 7-point Likert from "Strongly Disagree" to "Strongly Agree." They found that past activism and

[6] Moskalenko and McCauley (2009) say that they devised their items from the literature, but do not provide specific information on item criteria, creation, or selection. However, the ARIS has garnered clear acceptance in the field.

radicalism predicted both AIS and RIS scores respectively. Moskalenko and McCauley recommend the past actions version of the ARIS for avoiding social desirability bias that might artificially lower scores, particularly on the RIS, versus intentions.

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).

The ARIS largely, but not exclusively, captures behaviors in the context of a political group or movement (e.g., campaign or protest). Its latter two items, War and Retaliation, are intentionally much vaguer than the previous items as they pertain to more extreme behaviors with theoretically greater potential floor effects due to social desirability. The ARIS thus contrasts with other established radicalism measures like the Sympathies for Violent Radicalization (SyfoR) scale (Bhui, Warfa, & Jones, 2014) that measure a larger and more specific, illicit swath of radical and terrorism-related behaviors (e.g., bomb or weapon use). Therein, the ARIS is perhaps more appropriate for surveying general populations or populations that may be particularly sensitive to social desirability effects in their responses (e.g., Muslims or other political minorities who have been stigmatized in association with terrorism).

3.2.2 ARIS Use

The ARIS has, at the time of writing, been used in 69+ studies—published, in press or preparation—over two thirds of which have been published in just the past three years (see Table 2 below). These studies collected samples from 24 countries[7] from either general populations, or populations theoretically "at risk" of radicalization and/or mobilization to violent extremism as political minorities in current conflicts (e.g., Somali immigrants and white Southerners in the US, French-speaking Quebecois in Canada, Catalans and Muslims in Castilian/Catholic Spain, Yellow Vest protestors in France, or extradition protestors in Hong Kong). These sample sizes range from $n < 100$ to $n > 6{,}000$ (see Table 2 below). The ARIS has been translated from English into 12 other languages[8]. The ARIS is recurringly ranked as one of the top radicalism measures in systematic reviews (Scarcella, Page, & Furtado, 2016; Misiak, et al., 2018) and meta-analyses (Emmelkamp, Asscher, Wissink, & Stams, 2020; Wolfowicz, Litmanovitz, Weisburd, & Hasisi, 2020) based on methodological markers like theory, methods, and sampling, as well as psychometric properties like readability, cultural translation, construct and internal validity. The ARIS is used both independently and in tandem with other radicalism measures, such as the

[7] Austria, Belarus, Belgium, Brazil, Canada, China, Croatia, Denmark, Egypt, France, Germany, Hong Kong, Hungary, Italy, Morocco, Nicaragua, South Africa, Spain, Sweden, Turkey, Ukraine, the UK, the US, & Venezuela.
[8] Arabic, Chinese, Danish, Dutch, French, German, Hungarian, Italian, Portuguese, Spanish, Turkish, & Ukrainian.

SyfoR (Frissen, 2019; Rousseau et al., 2019ab), the Support for Political Violence Scale (Adam-Troian et al., 2019ab), and individual original or borrowed items (Bartusevicius, Leeuwen, & Petersen, 2020a; Pfundmair, Paulus, & Wagner, 2020; Schumann, et al., 2021).

Table 2. *ARIS Publications*

AUTHORS	YEAR	DATA	*n*	COUNTRY	SAMPLING
Adam-Troian, et al.	2019a	Received	1,240	Brazil; Belgium; France; Turkey	Students
Adam-Troian, et al.	2019b	Received	249	France	Students
Anastasio, Perliger, & Shortland	2021	Received	1,202	USA	Representative
Atari, et al.	2021	Received	843	USA	Electorate
Bartolo, et al.	2020	No Reply	200	Italy	Students
Bartusevicius	2020	Received	6,283	Nicaragua; South Africa; USA; Venezuela	Representative; WEIRD[9] vs. non-WEIRD
Bartusevicius, et al.	2020	Received	6,000	Denmark; Italy; Hungary; USA	Representative
Bartusevicius, Leeuwen, & Petersen	2020a	Received	3,398	South Africa; USA	Electorate
Bartusevicius, Leeuwen, & Petersen	2020b	Received	2,960	Venezuela; Nicaragua; Belarus	Electorate
Becker	2020	Received	503	USA	Students
Choi & Yoon	2021	No Reply	103	USA	Representative
Chui, et al.	2020	No Reply	120	Hong Kong	Students
Costabile, et al.	2020	No Reply	328	Italy	Students
Decker & Pyrooz	2019	Received	802	USA	Criminals
Ellis	2020ab	Received	232	USA; Canada	Somalis
Ellis, et al.	2014	Received	79	USA	Somalis
Ellis, et al.	2015	Received	374	USA; Canada	Somalis
Ellis, et al.	2019	Received	213	USA; Canada	Somalis
Ellis, et al.	2021	Received	498	USA; Canada	Somalis
Filho & Modesto	2019	Received	226	Brazil	Social media users
Fodeman, Snook, & Horgan	2020ab	Received	356	USA	Muslims (converts vs. non-converts)

[9] Western, Educated, Industrialized, Rich, and Democratic (Schultz, Bahrami-Rad, Beauchamp, & Henrich, 2018).

Frissen	2019	Received	3,378	Belgium; Canada	Students
Frounfelker, et al.	2019	Received	2,037	Belgium	Stratified
Frounfelker, et al.	2021	Received	3,364	Belgium; Canada	Stratified
Gøtzsche-Astrup	2019	Postponed	5,000	USA	Electorate
Gøtzsche-Astrup	2020	Postponed	2,488	USA; Denmark	Electorate
Gøtzsche-Astrup	2021	Postponed	1,500	USA	Representative
Jahnke, et al.	2020	Received	303	Germany	Politically-active youth
Kendrali	2020	Received	447	UK	Representative
Lemieux, et al.	2017	Received	979	Egypt; Morocco	Muslims
Levinsson, et al.	2021	Postponed	6,003	Canada	Students
Lobato	2018	Received	259	Spain	Students vs. Representative; Muslims vs. non-Muslims from at-risk neighborhood
Lobato, et al.	2018	Received	524	Spain	Students vs. Representative; Muslims vs. non-Muslims from at-risk neighborhood
Lobato, Moya, & Truijillo	2020	Received	214	Spain	Spaniards vs. Catalans
Loughery	2018	No Reply	77	Sweden	Students; Muslims
Mahfud & Adam-Troian	2020	Received	776	France	Yellow Vest supporters; Online social network
Miconi, et al.	2020	Received	1,765	Canada	Students; Québécois
Morales, et al.	2020	Received	677	France	Yellow Vest supporters; Online social network
Moreira, et al.	2018	Received	452	Brazil; Spain	Students
Moskalenko & McCauley	2009	Received	882	USA; Ukraine	Students; Electorate
Moyano & Trujillo	2014	Received	115	Spain	Students; Muslims vs. Christians from at-risk neighborhood
Pavlović & Franc	2021	Received	661	Croatia	Convenience & Quota
Pavlović, Moskalenko, & McCauley	2021	Received	443	Spain & Croatia	Representative
Pavlović, et al.	2021	Postponed	TBD	TBD	TBD

Petersen, Osmundsen, & Arceneaux	2020	Received	2,533	USA	Convenience vs. Representative vs. Diverse Convenience
Pfundmair, Paulus, & Wagner	2020	Received	110	Austria	Students
Ramos	2018	No Reply	483	USA	Students; Latino
Rottweiler & Gill	2020	Postponed	1,502	Germany	Representative
Rousseau, et al.	2020	Received	3,454	Canada	Students; Québécois
Rousseau, et al.	2019*ab*	Received	1,190	Canada	Students; Québécois
Schumann, Salmon, Clemmow, & Gill	2021	Received	1,378	UK	Representative
Shortland & McGarry	2021	Postponed	479	USA	Representative
Smith	2016	No Reply	576	USA	Students
Soliman, Bellaj, & Khelifa	2016	No Reply	662	Egypt	Students; Muslims
Trujillo, Prados, & Moyano	2015	Received	115	Spain	Students; Muslims vs. Christians from at-risk neighborhood
Villen, et al.	2022	2022	300	Italy; Spain	Football Hooligans
Wagoner, Rinella, & Barreto	2021	Pending	293	USA	Conservatives
Wong, Khiatani, & Chui	2019	No Reply	454	China	Students
Wright, Cheung, & Esses	2019	Received	559	USA	Southern Whites

Note. Studies included in the final analysis are denoted by italicized author names.

ARIS scores have been correlated with a variety of social-psychological risk factors that are theoretically relevant to the development of activism and radicalism. First and foremost are different types and levels of ingroup identification, if not Identity Fusion[10] (Atari, et al., 2021). These include birth and host country, national or municipal identity, religion (specifically Christian or Muslim), political party or a single issue, race, ethnicity, clan, tribe, or even family

[10] Identity Fusion (Gómez, Brooks, et al., 2011) is the enmeshing of the personal self (characteristics usually individualizing and unique, such as height, age, or personality) with the social self (characteristics that associate oneself with a group, such as ethnicity, nationality, or political cause), like overlapping Venn diagram circles.

(Moyano & Truijilo, 2014; Ellis, et al., 2014, 2015, 2016, 2019; Gøtzsche-Astrup, 2019, 2020; Lobato, 2018; Lobato, et al., 2018; Lobato, Moya, & Trujilo, 2020; Mahfud & Adam-Troian, 2020; Miconi et al., 2020; Morales et al., 2018; Moskalenko & McCauley, 2009; Ramos, 2018; Rousseau, et al, 2019ab, 2020; Soliman, Bellaj, & Khelifa, 2016; Wright, Cheung, & Esses, 2019). A wide array of other correlates have been found, such as relative deprivation (Chikhi, 2017); perceived discrimination (Adam-Troian, et al., 2019b; Ellis et al., 2019; Frounfelker, et al., 2019; Rousseau, et al., 2019ab); perceived oppression, religious fundamentalism (Lemieux et al., 2017; Lobato, Moya, & Truijillo, 2020; Loughery, 2018; Moyano & Truijilo, 2014; Rousseau, et al., 2019a); Social Dominance Orientation and Right-Wing Authoritarianism (Adam-Troian, et al., 2019; Bartusevičius, van Leeuwen, & Petersen, 2020a; Lemieux, Kearns, Asal, & Walsh, 2017; Wright, Cheung, & Esses, 2019); prejudice and intolerance (Adam-Troian, et al., 2019b; Wright, Cheung, & Esses, 2019); moral character (Chui, Khiatani, She, & Wong, 2020; Filho & Modesto, 2019; Pfundmair, Paulus, & Wagner, 2020); political ideology and worldview (Gøtzsche-Astrup, 2020; Mahfud & Adam-Troian, 2021; Rottweiler & Gill, 2020); mental health (Costabile, et al., 2021; Miconi, Calcagni, Mekki-Berrada, & Rousseau, 2020; Rousseau, et al., 2019a); and/or exposure to personal trauma or violence (Ellis, et al., 2016, 2019; Miconi, et al., 2020; Rousseau, et al., 2019a). Some of these correlates have even been manipulated experimentally to predict ARIS outcomes (Adam-Troian, et al., 2019; Chui, Khiatani, She, & Wong, 2020; Lemieux, Kearns, Asal, & Walsh, 2017; Ramos, 2018; Smith, 2016; Wright, Cheung, & Esses, 2019). The ARIS has also demonstrated convergent validity with other items and measures of political activity (Bartusevicius et al., 2020a; Chui, Khiatani, She, & Wong, 2020), civil disobedience (Adam-Troian, et al., 2020; Mahfud & Adam-Troian, 2021; Pfundmair, Paulus, & Wagner, 2020), political violence and extremism (Adam-Troian, et al., 2019b; Bartusevicius, Leeuwen, & Petersen, 2020a; Mahfud & Adam-Troian, 2020; Pfundmair, Paulus & Wagner, 2020; Ramos, 2018; Rousseau, et al., 2019ab; Smith, 2016).

3.2.3 ARIS Measurement & Study Comparison Issues

All these studies may be subject to untreated measurement error that could skew their results. Most treat the ARIS scales as not latent factors, but composites[11]. When researchers use composite scores, they assume that there is no measurement error in their estimation of scores from their samples (Kline, 2016), and thus measurement error is bound to the factor mean, whereas it is separated in latent factor analysis (FA). While it is reasonable to use composite scores with scales whose measurement has been well established for the particular version, populations, and contexts being measured, composite ARIS scores are used for previously untested populations[12], as well as new scale translations, item re-scales, and other adaptations. Furthermore, most researchers treat the ARIS indicators as continuous, even though they are discrete Likert items, which can also bias results and miss crucial distributional information about thresholds of activism and radicalism (e.g., how most respondents will never endorse radicalism items to any degree). Prior to my Masters thesis (Fodeman, 2020), no studies had tested for measurement equivalence/invariance (ME/I) on the full suite of items, i.e. testing whether there is any bias or differential item functioning (DIF) in ARIS mean estimation when comparing any two or more populations (e.g., between translations, ARIS factor remodeling or

[11] A minority model the ARIS as latent factors (Costabile et al., 2020; Decker & Pyrooz, 2019; Ellis et al., 2014; Frissen, 2019; Frounfelker, et al., 2019; Gøtzsche-Astrup, 2019; Miconi et al., 2020; Smith, 2016; Soliman, Belaj, & Khelifa, 2016; Wagoner, Rinella, & Barreto, 2021), but none have tested for ME/I. Decker and Pyrooz (2019) tested multigroup ARIS models, akin to testing for Structural Invariance, but with a convict population, which is hard to compare to other studies' populations; in fact, I found them be Latently Invariant from other studies' samples.

[12] While many of these studies *do* report scale reliability measures like Cronbach's Alpha and McDonald's Omega (Adam-Troian et al., 2019; Bartusevicious, Leeuwen, & Petersen, 2020ab; Becker, 2020; Decker & Pyrooz, 2019; Ellis et al., 2015, 2019; Filho & Frissen, 2019; Frounfelker et al., 2019; Jahnke et al., 2020; Lemieux et al., 2017; Loughery, 2018; Modesto, 2019; Morales et al., 2020; Moriera et al., 2018; Moyano & Trujillo, 2014; Ramos, 2018; Rousseau et al., 2019a; Soliman, Belaj, & Khelifa, 2016; Smith et al., 2016, 2020; Wong, Khiatani, & Chui, 2019), these measures do not test for scale unidimensionality or other elements of factor structure and functionality (McNeish, 2018; Peters, 2014). These scores also are not useful for comparing scale reliability between samples.

item re-scaling, political affiliations, religious or ethnic groups, age ranges, etc.). This dissertation provides a starting point to address those issues using an integrated data analysis (IDA) approach with multiple group structural equation modeling (MGSEM)—a technique yet unemployed in ARIS studies, and indeed a novel improvement on more traditional and limited meta-analyses (MA). While I had originally proposed standardizing item scores across rescaling with proportion of maximum scoring (POMS: Little, 2013), and using multiple imputation (MI) to account for excluded ARIS items, discussion with IDA experts revealed that this was not currently possible—leading to evaluating ME/I of only 13 collections of studies on only RIS items #2-4. Before discussing the specific design and methods for this study, and what could and could not be done, I will first briefly review each of these concepts.

3.3 Measurement Advances

3.3.1 Meta-Analysis (MA)

Meta-analysis (MA) is a widely used statistical approach whereby sets of results are sampled from several independent studies to draw broader conclusions. It is a special case of multilevel modeling where the focus of analysis is on studies themselves over individual results. With MA, you can test to see what proportion in variation of an outcome (e.g., average levels of radicalism) is due to random variation between studies or study-level moderators (e.g., publication year, population sampled, measurement method, etc.) rather than participant-level variables. While standard MA can only test for one outcome at a time[13] (e.g., a particular mean or effect size), meta-analytic Structural Equation Modeling (MASEM) can estimate multiple

[13] Note that there are multilevel and generally multivariate forms of MA as well that can test for multiple outcome variables at once (see Pustejovsky and Tipton, 2021).

outcomes simultaneously while accounting for the family-wise error (i.e., inflated Type I error from testing multiple dependent variables)—such as in testing multiple survey scale items.

3.3.2 Factor Analysis (FA)

Another technique to consider with multiple survey scale items (as in the case of this dissertation) is factor analysis (FA), a technique for determining if a set of items measure a single cohesive concept. More specifically, FA is a statistical method for testing whether variation in observed, correlated variables (indicators, such as the ARIS's 10 survey items) can be better explained by a fewer number of latent variables (factors, such as activism or radicalism). FA treats those indicators as separate outcomes predicted by one or more factors, rather than as composites of a total score. FA 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[14] or differences will not affect factor means; and C) individual items' variances (e.g., measurement errors) are equal or will not affect factor variances. FA, 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 construct and the reliability of its measurement.

3.3.3 Measurement Equivalence/Invariance (ME/I)

When comparing groups (e.g., Muslims versus Christians, Québécois versus other Canadians, Catalonian versus Spanish nationals, or simply one body of research versus another), 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

[14] i.e., items are often added linearly unless one item is given different weight than another item *a priori.*

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 a democracy as it is in an autocracy without the right to assemble). Put another way, if there is not ME/I, then the scale or other tool may measure something different between groups. The greater the proportion of noninvariant parameters (e.g., indicators (Chen, 2007) or regression coefficients (Guenole & Brown, 2014)), the greater the bias[15] in factor or outcome means, which are the parameters researchers typically want to compare between populations, not individual indicators’ means/intercepts, weights/loadings, variances/noise.

3.3.4 Integrated Data Analysis (IDA)

One can test for ME/I of factors across groups within the same study, but one can also test for factors *within* groups across *studies* using integrated data analysis (IDA)—that is, the analysis of multiple datasets pooled into one. By comparison, MA and MASEM rely on means, thresholds, (co)variances and/or effect sizes reported in publications, which may not be directly comparable as they are not often reported with complete information, estimated with the same techniques, and are part of divergent models including different covariates. MA and MASEM also cannot account for *individual*-level variation, *or* group data *sub*sets across studies whose subsetted information was not already reported in the original publications (e.g., controlling for demographic effects across studies). MA is a traditional approach developed when data were not as easily or readily shared as they are today. In the modern internet age, and especially with more emphasis being placed on research transparency and open access data sharing in the wake of the Replication Crisis (Maxwell, Lau, & Howard, 2015), we can access researchers’ original data files and use IDA to estimate parameters directly from those datasets (Curran et al., 2018). IDA

[15] Note that DIF can be benign (reflecting true differences between groups) or adverse (reflecting measurement bias). With scales, typically researchers want to test for differences of factors, not items, such as in ARIS research.

with raw, uncleaned data also allows us to apply the same standards of data quality (e.g., removing straight-lined answers[16]) and missing data strategies (e.g., pairwise deletion, full information maximum likelihood, multiple imputation, etc.) across studies. IDA-SEM allows us to estimate multivariate models using pooled information across studies. Furthermore, it allows us to include and account for variation across studies in indicator use, such as item inclusion/exclusion and re-scaling, by estimating factors using the same scaled indicators.

These statistical approaches are particularly helpful given the nature of the ARIS and all the changes that have been made to it, if not inconsistencies in its use. When Moskalenko and McCauley created the ARIS (2009), they tested ten 7-point Likert items—four for the AIS, *six* for the RIS. Since then, researchers have changed the scales from as few as 4-point to as many as 100-point 'continuous' scales. Researchers have not only chosen to measure the AIS and the RIS separately, but they have also removed items, added supplemental ones, changed item wording and even factor structure (e.g., nesting the Civil Disobedience item under the AIS instead of the RIS (Becker, 2020; Filho & Modesto, 2019; Moyano et al., 2021; Pavlović & Franc, 2021; Smith et al., 2020) or using a bifactor model (Pavlović, Moskalenko, & McCauley, 2021)). All of those changes can be accounted for via the aforementioned statistical tools. This study will be one of the first[17] to employ IDA with research concerning radicalism and terrorism; it is therein one step for radicalism research towards the level of statistical power and scrutiny promoted in other fields (e.g., medicine, education, or economics) that we might hope to achieve.

[16] A.k.a. non-differentiation in ratings, i.e. when participants choose the same response category to a series of questions (e.g., selecting the lowest ordinal category for every item), often due to lack of engagement.
[17] Orazani (2020) integrated multiple samples in his study of radicalism, but not with any archival data or the ARIS.

4 RESEARCH DESIGN & METHODS

4.1 Procedures

4.1.1 Protection of Human Subjects

Following GSU's IRB guidelines, this study is not considered human subjects research and therefore was waived from GSU's IRB review (approved IRB #H22257; outcome letter reference #367794). This is because this dissertation is secondary data analysis for which no Personally Identified Information (PII) was collected, *and* therein disclosure of the data would not place the original subjects at risk of harm.

4.1.2 Resources

This research was conducted on R version 4.1.2, particularly with packages Lavaan (Rosseel, 2012; version 0.6-10) and semTools (Jorgensen et al., 2018; version 0.5-5). A solid state drive was used for efficient access to the large integrated dataset. I encrypted the data, and used both virtual private networking and malware protection to protect all of the datasets that I received. I used the GSU University Library's Research Data Services (RDS) staff for technical help with R. Furthermore, I attended Dr. Todd Little's Analysis Retreat where I received consultation on what at the time we thought were appropriate and viable statistical methodologies discussed in my dissertation proposal (i.e., POMS and multiple imputation or MNLFA across ARIS rescales and item exclusion sets), though ultimately these proved to be as-yet-impossible, undefendable, inappropriate, or sub-optimal strategies. Further discussion with some dissertation committee members, RDS staff, IDA's progenitors, and Lavaan and semTools's creators, as well as my own deeper reading on the subject matter, enlightened me as to what analyses were possible and appropriate.

4.1.3 Archival Data Collection

I found published, pre-printed, and pre-registered research on the ARIS via a combination of search terms: either requiring citation of Moskalenko and McCauley's 2009 article in addition to discussing radicalism or activism, or requiring mention explicitly of the ARIS even if citing another study instead that used the scales (see Figure 2 below for workflow). I entered these criteria into Google Scholar to search across all academic article databases, as well as the Open Science Foundation (OSF), covering over 360 potential articles. I systematically reviewed each article to ascertain whether the authors used a version of the ARIS, resulting in approximately 69 confirmed articles. I then contacted the authors to request access to their dataset(s) and codebook(s) if they were not already made publicly available (e.g., on OSF, GitHub, or a personal website). If it was unclear in the text whether they used a version of the ARIS, I asked the authors for clarification. I started with the points of contact recommended in the publication and, failing that, searched for other forms of contact (e.g., alternative email addresses, accounts on LinkedIn, ResearchGate, Academia.edu, OSF, etc.). Some researchers were willing to share their data in the future (denoted in Table 2 above as "Postponed"), but not in time for the scope of this dissertation work. After multiple emails and other points of contact, some researchers did not reply (noted as "No Reply"). Most researchers, however, not only replied, but were happy to share their data if it was not already publicly available (denoted as "Received"); researchers shared 43 articles' datasets with me for the purposes of this dissertation work.

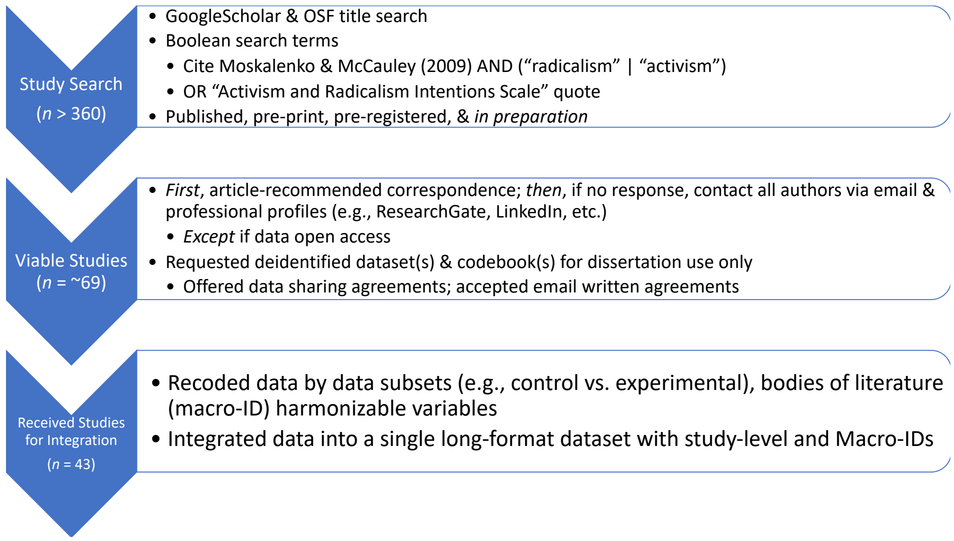

Figure 2. *Systematic Data Review & Integration*
Note. "*n*" refers only to the number of studies (e.g., not the number of authors contacted).

All the published studies under consideration collected their data with the approval and supervision of their institutions' respective Institutional Review Boards (IRBs). I requested and received all data as deidentified. Sensitive populations that were sampled included prisoners (Decker & Pyrooz, 2019) and minors (Lobato, 2019; Lobato, Moya, Moyano, & Trujillo, 2018; Moyano & Truijilo, 2014; Pfundmair, Paulus, & Wagner, 2020; Trujillo, Prados, & Moyano, 2016; in all cases, procedures to ethically collect consent were followed[18]. I informed all researchers I contacted about who I am, the graduate program I am in, and that I asked for their data specifically for the purposes of my dissertation. I signed a data sharing agreement with one researcher for use of their dataset (Becker, 2020), with Dr. Anthony Lemieux[19] as the officiant. I came to a verbal agreement via email with all other researchers. No individual participant's results are identifiable in this dissertation as analyses target the aggregate level across studies.

---

[18] i.e., both parents' and their child's consent were received for samples of minors, and prisoners' consent was given with full information and independent from instructions from a warden or other authority figure to participate.
[19] The University of Maryland, which oversees that dataset, will not transfer fair use of that dataset to a student directly, but only under supervisor of a professor; Dr. Lemieux already had received permission for the same dataset for one of his advisee's dissertations; an amendment to that original agreement was suitable to all parties.

All researchers will be informed of the dissertation results and invited to participate in subsequent ‘big team science’ projects further investigating the integrated data.

Most studies were cross-sectional in-person or online surveys. Some studies used resources such as Qualtrics or Amazon Mechanical Turk. These companies use established participant networks to produce representative stratified samples from national or more specific populations. Some studies recruited participants within established participant networks of community contexts (e.g., school systems, ethnic or religious community resources, etc.): The former were generally conducted among student populations who were recruited via classes, if not incentivized to participate via academic credit, as is common in psychology and other social science disciplines; the latter were generally conducted as community participatory research projects in which community members were invited to be part of the process of data collection planning. All studies provided either small, negligible monetary compensation to participate that was approved as non-coercive by their respective IRBs, academic credit compensation, or no compensation at all. Studies employed different screening gradients (e.g., language, age, group membership, etc.) and data exclusion criteria. I requested data in their raw, uncleaned, original format when available. I applied comprehensive data exclusion criteria, including: removing univariate outliers (Levin, Fox, Forde, & David, 2012), i.e. respondents who do not vary their item responses (a.k.a., “straight-lining”); removing incoherent responses to attention-checking questions; and removing respondents with outlier response times (Malhorta, 2008; Greszki et al., 2015). 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 O. B., Senise, dos Reis, & Ono, 2017). Such poor survey responses were excluded from analysis as they should only produce noise and not have any meaningful directional effects on results. These practices are standard for both survey companies and researchers (Freire et al., 2017).

4.1.4 Eligible Studies for Data Harmonization

Unfortunately, not all ARIS studies were eligible for IDA, based first on their ordinal scales. In IDA, one often has to harmonize data, i.e. transform data so that they can be compared across studies (Hussong et al., 2021). One can either 'logically harmonize' (i.e., equating items via face validity and expert opinion) or 'analytically harmonize' (i.e., test for ME/I). Harmonized variables are given the same meaning and metrics; for example, if one study includes exact age as an integer, while another study only asked participants to choose ordinal age ranges, then the former data would be re-categorized into the latter data's age ranges to have the same meaning and scale. However, harmonizing that condenses items onto the same scale (e.g., Proportion of Maximum Scaling (Little, 2013), median or mean splits, reducing a count model to only a binary or hurdle model, etc.) reduces item variability, and therein obfuscates potentially important information while reducing overall statistical power (Cohen, 1983; Curran & Hancock, 2021; MacCallum et al., 2002; Olsson, 1979; Preacher et al., 2005; Rucker, McShane, & Preacher, 2015; Taylor, West, & Aiken, 2006). While Moderated Non-Linear Factor Analysis (MNLFA: Hussong et al., 2021) can estimate a latent score from multiple different types of indicators with appropriate link functions (e.g., latent "age" from exact continuous number of years old (e.g., "17.35 years old," ordinal age ranges (e.g., "15-to-18-year-olds"), and binary data (e.g., "minor vs. adult"), it requires that integrated studies have enough overlapping data in order to estimate parameters and test for ME/I without imputing missing values[20] (e.g., Study A uses both scales X and Y, Study B uses both scales Y and Z, and Study C uses both scales X and Z). After

[20] Indeed, this is similar to the data overlap needs of imputing data across groups with control variables.

discussions with the progenitors of IDA and MNLFA (Drs. Andrew Hussong, Patrick Curran, and Daniel Bower), as well as other statistics experts and consultants recommended by my dissertation committee members or otherwise available at GSU, it became evident that current statistical science cannot yet support the harmonization and integrated ME/I testing of ordinal data with the same variables rescaled, as in the case of the ARIS. Therefore, only 32 studies using the original and most widely used ordinal scale (7-point Likert) were considered eligible for analysis, while 11 others were excluded (see Figure 3 below).

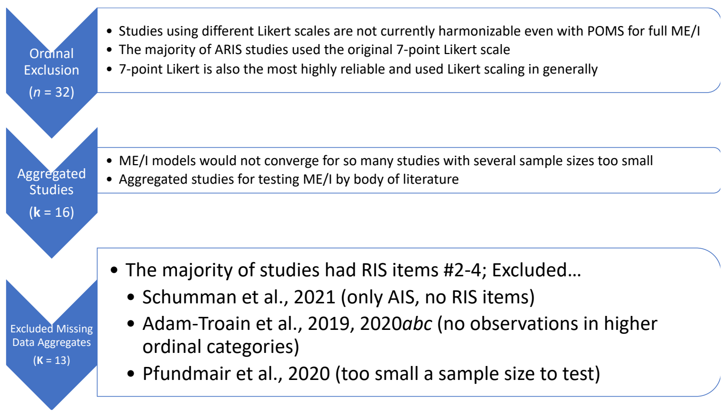

Figure 3. *Studies' Inclusion & Aggregation for IDA*

Further studies had to be excluded due to missing data, particularly missing items. Not all ARIS studies used the full original 10 item set. While we could parcel together averages of similar items (Little, 2013) so that all datasets were represented by the same set of items (e.g. ARIS items #1-4, #5-8, and #9-10), we should not do so for similar reasons to why we should not use composite scores in the first place: we would assume equal measurement error, weight, and intercepts of original items between groups, for which we cannot test for ME/I when parceled (Curran & Hancock, 2021). Furthermore, variation in one item is mixed with variation in another, and cannot be teased apart after parceling; this is particularly problematic for ordinal data, where ordinal categories (e.g., 1 = "Strongly Disagree," 2 = "Disagree," 3 = "Neither Agree nor Disagree," etc.) lose their meaning upon parceling (e.g., is a 1 from Item A and a 3 from Item B equivalent to a 2 from both items?). On top of all that, parceling assumes that all parceled items have similar meaning, or that their combined value is meaningful—for example, the parceled or composited score on a math test from individual test questions is a meaningful measure of math competency on the given math test topic, *but* adding scores from items taken on a math test to scores from items take on an *English* test would be mixing apples and oranges.

Therefore, I included studies with the maximum number of overlapping items while minimizing observation loss. Given the complicated need to maximize sample size in general for estimation, especially relative to the number of groups to compare for ME/I, relative to the proportion of missing data due to item exclusion and other sources, as well as how published "studies" could include any number of data-subsets (e.g., multiple true studies within a paper, multiple experimental or cross-sectional groups, multiple sampling sites, etc.), I therefore combined studies into bodies of literature collected by author cohorts. These author cohorts (referred to as *k* in Figure 3 above) reliably used the same ARIS translations and ordinal scales, if not always the same ARIS sets of items, and generally sampled similar populations study after study with similar study conditions. While the opposite tactic of comparing data-subsets might have been the most accurate in ME/I testing, sample sizes were too small to converge models, let alone control for data-subset relationships (e.g., from the same study and author cohorts, sampling the same population demographics, sampled in the same language, etc.).

Table 3. *Author Cohorts by Percent Missing Data Per ARIS Item*

1st Author	***n***	**AIS 1**	**AIS 2**	**AIS 3**	**AIS 4**	**RIS 1**	**RIS 2**	**RIS 3**	**RIS 4**	**RIS 5**	**RIS 6**
Adam-Troian	1,219	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%
Bartusvecius	3,535	**100%**	**100%**	**100%**	**100%**	2%	1%	2%	3%	**100%**	**100%**
Becker	617	14%	14%	**100%**	14%	14%	14%	14%	14%	14%	14%
Decker	680	**100%**	**100%**	**100%**	**100%**	**100%**	0%	0%	0%	0%	0%
Ellis	1,295	0%	0%	0%	0%	1%	2%	2%	2%	**100%**	1%

Fodeman	356	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%
Jahnke	303	2%	1%	0%	0%	1%	0%	1%	0%	**100%**	**100%**
Lemieux	979	0%	0%	0%	0%	0%	0%	0%	0%	**100%**	**100%**
Lobato	1,439	0%	0%	0%	0%	0%	0%	0%	0%	**71%**	**71%**
McGill*	569	**100%**	**100%**	**100%**	**100%**	1%	1%	2%	2%	**100%**	**100%**
Morales	1,415	0%	0%	0%	0%	0%	0%	0%	0%	**100%**	**100%**
Moreira	561	0%	0%	0%	0%	0%	0%	0%	0%	**100%**	**100%**
Moskalenko	656	0%	0%	1%	1%	0%	0%	0%	0%	12%	**79%**
Pavlovic	512	0%	0%	0%	0%	0%	0%	0%	0%	**100%**	**100%**
Pfundmair	**110**	0%	0%	0%	0%	0%	0%	0%	0%	**100%**	**100%**
Schumann	1,378	0%	0%	0%	0%	**100%**	**100%**	**100%**	**100%**	**100%**	**100%**

Note. 100% missingness due to item exclusion. **Bolded** percentages and *n*'s cannot be estimated. Highlighted author cohorts were excluded from analysis due to estimation problems. *Refers instead to many different first authors from the same cohort using the same sample location.

To maximize the number of author cohorts to be compared, the number of ARIS items had to be minimized. The majority of author cohorts excluded or were largely otherwise missing observations for RIS items #5 and #6 (see **Table *3*** above)—"War" and "Retaliation." Moskalenko and McCauley (2009) dropped those two items themselves after their first pilot study due to poor performance and fit for their pilot population[21], and many subsequent studies excluded those items, whether because those items were not relevant to the sampled group affiliations, those items are worded quite differently from the other items, those items are the most skewed and difficult to detect observations for, or it was just simpler to exclude them; therefore, RIS #5 and #6 were excluded from this IDA. Three large author cohorts ("Bartusevius," "Decker," and "McGill") did not measure the AIS; the RIS is more important to the field (evidenced, in part, by its higher use than the AIS), and therefore the AIS was excluded

[21]Moskalenko and McCauley (2009) used "*Exploratory* Principal Components Analysis" (emphasis added), which is unclear. Both PCA (related example: Snook, Branum-Martin, & Horgan, 2021) and EFA (see Kline (2016), pp. 191-194 for more explanation of EFA) assume normal distributions, which the ARIS items do not have, which might have contributed to poor fit (those two items tend to be the most skewed). Those items might also have been least relevant for their pilot study, for which students could choose via open response any group with which to identify; they chose groups who have not historically engaged in 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."

from this IDA. RIS #1 was also excluded ("Civil Disobedience"), as another large and important author cohort, "Decker," did not measure it. It is also reasonable theoretically to exclude it, as RIS #1 pertains to a mid-ground between activism and radicalism and is often modeled under AIS instead of RIS as such. RIS items #1-3 are the core, most consistently agreed upon indicators of radicalism by the ARIS: "Violent Group," "Violent Protest," and "Police Defense." Conveniently, three is the minimum number of indicators to estimate a factor and test for ME/I without losing latent or other parameter information (e.g., equality-constraining loadings).

In order to focus on RIS items #1-3, three of the 16 author cohorts were excluded from this IDA. The "Schumann" cohort did not measure the RIS. The "Pfundmair" cohort collected too small a sample (*n* = 110). While the "Adam-Troian" cohort did not have any missing data the items overall, none of their respondents chose any of the latter two ordinal categories ("Agree" and "Strongly Agree"), meaning that those latter ordinal thresholds could not be estimated and tested for ME/I along with those in all of the other cohorts.

4.1.5 Testing for Study Exclusion Bias

I did not find any potential bias for IDA by the studies excluded from analysis or that were never received, except for planned missing items (ARIS item exclusion and Likert scaling) and an overrepresentation of studies with known specific samples (e.g., sampled ethnic, religious, political, and other groups, especially those "at risk"). I ran chi-squared and related tests for differences between not just author cohorts, but individual studies, by whether they were included in the IDA (*n* = 33, *k* = 13), excluded (*n* = 10, *k* = 3), or the study data was never received in the first place[22] (*n* = 17; see this link for a webpage of the crosstabs, as the table was

[22] While I have no evidence for bias in authors' non-response, I should consider why they may not have responded. As noted previously, there is perhaps a particularly high 'publish or perish' mentality in radicalism research given that often aim to address new, high impact 'flashbulb' events. Given that 2/3 of ARIS studies were published in only

too large for this document). There is no significant difference between these three groups of studies in sample size, age target, representative sampling, sampling of students, voters, and general populations, nor sampling in English, adapting, or supplementing the ARIS. By design, studies using non-7-point Likert scales were excluded from analysis, and an additional two 7-point studies were excluded for reasons just discussed above, while 10 7-point studies' data were never received (24% of all 7-point studies). The IDA included the *only* studies that specifically sampled ethnic groups (n = 10), *almost all* that targeted religious groups (n = 10, with 1 never received), and 5/8th of those that targeted political groups (n = 8 total, the other three were excluded), biasing the IDA towards greater specificity—though included studies did not significantly include fewer general population samples. Therein, the IDA included 77% of studies that targeted theoretically at-risk groups (n = 22 total). While included studies did not themselves on average included higher counts of RIS items, they *did* include on average *lower* counts of AIS items, which is not surprising given that RIS items were the focus of the IDA.

4.2 Statistical Methodologies

4.2.1 Estimation Method

ARIS data are not normally distributed. ARIS items are ordinal Likert items (i.e., bounded integers on a 7-point or similar scale)—not only bounded, but positively skewed. It is not surprising that most respondents to the ARIS hardly endorse any level of intention of engaging in activism and especially radicalism. There are several estimation methods we can choose to use

---

the past few years, it is not unreasonable to think that many of these authors were reticent to share their data. They may have also not deigned to share their data with a 'non-entity' graduate student, whom they did not know from Adam, as opposed to someone established in the field. Furthermore, burnout has been so high during the COVID-19 pandemic that they simply may have not had the energy and attention to respond to the data sharing request (or completely missed all of those emails). I also did not receive responses from several researchers who are not currently in academia, to my knowledge, or research altogether, and therein may no longer have access to that data, or are no longer inclined to be a part of the scientific process. I may also simply have had incorrect contact information, as it was not uncommon for researchers to have changed institutional emails or no longer use professional platforms.

for this kind of data, with varying strengths and weaknesses. Some use Maximum Likelihood (ML), which is more typically used for Ordinary Least Squares (OLS) regression (i.e., with normal, continuous data). However, for ordinal data, most researchers recommend utilizing *Weighted* Least Squares (WLS) and its robust 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 some convergence problems when modeling ordinal data (Bovaird & Koziol, 2012). Even robust ML (i.e., with relaxed normality assumptions) is inferior to WLS in controlling for Type I error, save for in large *n's* (i.e., $n > 1{,}000$, which is true for only four out of the 13 included author cohorts) for $\Delta\chi^2$ tests (Li, 2016). (Robust) WLS also provides more accurate factor loadings (λ), standard errors (SEs), 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 regularly with ARIS data (see Fodeman, 2020 for a discussion). Many researchers recommend using WLSMV in particular (Flora & Curran, 2004; DiStefano & Morgan, 2014; Sass, Schmitt, & Marsh, 2014; Bovaird & Koziol, 2012). WLSMV yields better fit and convergence likelihood than WLS (DiStefano & Morgan, 2014), especially with smaller *n's* (Flora & Curran, 2004) like those of several ARIS author cohorts (Jahnke et al., 2020 has as few as 303 observations). 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 (Jöreskog, 2005) for small *n's*, asymmetric τ's, and large λ's—like those in archived ARIS studies and this dissertation.

4.2.2 Goodness of Fit (GoF) Indices for Ordinal Indicators

There are many different model goodness-of-fit (GoF) indices to consider. GoF indices measure discrepancies between expected and observed outcomes. They are useful for not only determining how good a single model fits the data, but also comparing models. Higher versus lower degrees-of-freedom (*df*)—i.e., more restricted or fewer estimated parameters—lead to poorer fit. In ME/I testing, more stringent models (i.e., Configural, Metric/Threshold, or Scalar) subsequently increase *df* and reduce GoF. Researchers disagree, however, as to how dramatically reduced GoF between models signifies noninvariance (i.e. a cutoff score for $_{\Delta}$GoF), as well as which GoF indices are most appropriate, reliable, or sensitive for different model conditions (e.g., model complexity, *n*, data type or distribution). The GoF indicators that, based on the literature (Kline, 2016; Chen, 2007; Svetina, Rutkowski, & Rutkowski, 2020), are appropriate, reliable, and sensitive for comparing ME/I models with ordinal indicators are chi-squared (χ^2) test values, the Root Mean Square Error of Approximation (RMSEA), the Comparative Fit Index (CFI), and the Weighted and Standardized Root Mean Square Residuals (WRMR & SRMR). χ^2 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.” RMSEA 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). CFI 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. WRMR is 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. SRMR—the average standardized residual covariance—is similar to WRMR, though with some tradeoffs (see Table 4 below). I will follow

Hu and Bentler's (1999) recommendation to apply a two-index GoF strategy—that is, using both absolute (χ^2 & CFI) and incremental (RMSEA, WRMR, & SRMR) fit indices. Each of these GoF indicators present different strengths and weaknesses considered below in section 4.2.3.

Table 4. *GoF Index Comparisons*

GoF Index	Developer	Better Fit Direction	Range	Cutoff Criteria	*n* Size Type I/II Error Rate Inflation	Model Complexity Penalty?
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]** **≤ .05[d]** ***≤ .02[e]***	Small *n*[a]	Yes[a]
CFI	(Bentler, 1990)	Higher	0 - 1	**≥ .95[a]** **≥ .96[d]** ***≥ .98[e]***	No[a]	Yes[a]
WRMR	(Muthén, 1998-2004)	Lower	> 0	**< 1.0[b]** **≤ .95[d]**	Large *n*[b]	Yes[a]
SRMR	(Bentler, 1995)	Lower	> 0	< .08[c] ***< .05[d]***	Large *n*[c]	No[c]

Note. This table is based on Table 13.1 by West, Taylor & Wu (2012). **Bolded** criteria specified for ordinal data, and ***bolded italicized*** for ordinal multilevel data (for fewer than 100 groups).
Note. Superscripts refer to sources a) West, Taylor & Wu (2012), b) DiStefano, Liu, Jiung, & Shi (2017), c) Hu and Bentler (1999), d) Yu (2002), and e) Padgett & Morgan (2021).
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*.

4.2.3 Choosing GoF Indices

Statisticians propose different GoF index cutoffs at which a model may have a reliably good fit relative to a null or baseline model (see Table 4 above). χ^2 tests, while ubiquitously reported across SEM studies regardless of conditions, assumes that 1) manifest variables are normally distributed and 2) *n*'s are large (West, Taylor, & Wu, 2012); the former is not true for ARIS data, and the latter is not true for all ARIS studies. 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 ≥ .95 and RMSEA

≤ .06 to be considered[23] (2012), though the latter is sensitive to small *n*'s. However, RMSEA may not be as meaningful with WLS estimators as with *un*weighted least squares estimators, and so may not be useful for this analysis (Lai, 2020). WRMR was designed for ordinal data with a cutoff of < 0.90[24]. WRMR is especially useful for comparing samples with "widely varying variances" (Muthén & Muthén, 1998-2012), which is likely to be the case as such a diverse sample of studies with different populations, survey settings, and exclusion criteria is likely to be Residually Noninvariant. Shi and colleagues (2019) have specifically compared SRMR with RMSEA for ordinal FA, finding that RMSEA is unlikely to reject models with five or more categories (such as 7-point Likert), few variables (like the only three RIS items for this IDA), and little misfit; conversely, they found that SRMR is far less susceptible to these Type I Error risks and is generally powerful for these types of data. SRMR is uniquely useful for ME/I as it can be computed both within- and between-models.

For comparing models, statisticians also propose cutoffs for GoF *differences* (Δ) between *more* restricted models and *less* restricted models (e.g., $M_{Configural}$ - M_{Metric}). Increasingly equality-constrained models inherently worsen model fit due to increased *df*. Methodologists debate which measure or degree of ΔGoF indicates noninvariance (see We can also look at Modification Indices (MI), or measures of expected ΔGoF improvements if one parameter versus another is freed; while MI can hint at which parameters, if any, might need to be freed in a Partially Invariant model, if MI recommend freeing parameters that do not make sense to free until later invariance stages are established (e.g., latent means and variances), then we can proceed cautiously to the next invariance testing step.

23 Raykov et al. (2012) note baselines need not meet fit criteria before testing Configural Invariance.
24 Though DiStefano, Liu, Jiung, and Shi (2017) argue a cutoff of < 1.0 is sufficient, as above.

Table *5* below). While researchers widely use ΔCFI $\leq -.010$ as indicative of noninvariance for ordinal data[25] (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[26] (Chen, 2007). ΔRMSEA and especially $\Delta\chi^2$ perform well for testing ME/I with ordinal data regardless of the degree 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 increased error rates (Chen, 2007; Hu & Bentler, 1999). ΔSRMR shows promise, but is relatively untested for this study's conditions beyond work by Sokolov (2019). Therefore, as Rutkowski and Svetina recommend in their ME/I GoF cutoff review (2021), all four indicators ($\Delta\chi^2$, ΔCFI, ΔRMSEA, & ΔSRMR) will be considered together, each making up for potential weaknesses in the other. We can also look at Modification Indices (MI), or measures of expected ΔGoF improvements if one parameter versus another is freed; while MI can hint at which parameters, if any, might need to be freed in a Partially Invariant model, if MI recommend freeing parameters that do not make sense to free until later invariance stages are established (e.g., latent means and variances), then we can proceed cautiously to the next invariance testing step.

Table 5. *Relevant GoF Cutoff Values Indicating Noninvariance for Ordinal Data*

Source	*Data & Model Conditions*			*ΔGoF Cutoffs Indicating Noninvariance*				
	Groups	*n*/group	Factors	$\Delta\chi^2$ *p*	ΔCFI	ΔRMSEA	ΔSRMR	Model
(French & Finch, 2006)	2	150-**500**	2	.05				All
(Rutkowski & Svetina, 2017)	**10-20**	**600-6K**	**1**	$< .05$	$\leq$ -.004	$\geq$ .005		Metric

[25] The same standards are confirmed for multivariate normal models (French & Finch, 2006).
[26] Note that most simulations thereof largely only use ML, not WLS.

(Sokolov, 2019) (cont.)	**10**-50	**1k-2k**	**1**		≤ -.005		≤ .01*	Metric
					≤ -.005	≥ .000	≤ .01*	Scalar
				< .05	≤ -.004	≥ .001		Scalar
(Svetina & Rutkowski, 2017) (cont.)	**10-20**	**750-6K**	2-5	< .05		≥ .005		Metric
				< .05	≤ -.002	≥ .001		Scalar

Note. This table is adapted from Svetina, Rutkowski & Rutkowski's Table 1 (2020). $_{\Delta}$GoF is *more* restricted minus *less*. **Bolded** conditions match those in this project. *MLR not WLSMV.

5 RESULTS

5.1.1 Ordinal Indicators' Summary Statistics

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 10.2. Three RIS items were estimated. Figure 4 below, depicting RIS response frequencies irrespective of author cohorts, was graphed with the R package "sjPlot" (Lüdecke, 2019). These ordinal response frequencies are typical for RIS items: similar to zero-inflated count data, in that the majority of participants said they "Strongly Disagree" (level 1 of 7) with the radicalism statements. Frequencies, especially 'Agree' categories, are quite similar between items.

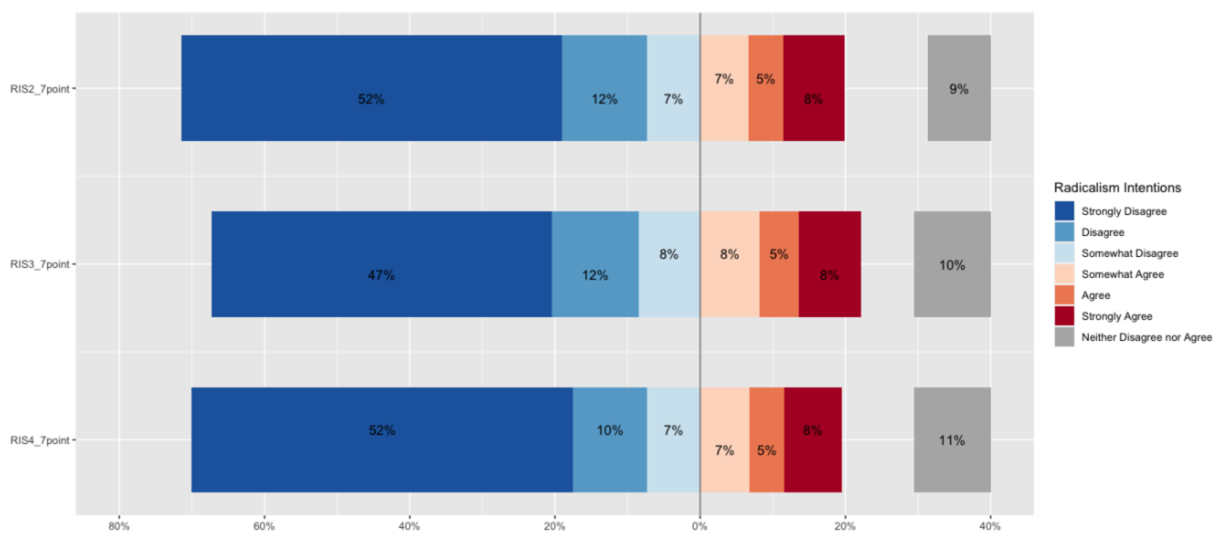

Figure 4. *Response Frequencies To RIS Items #2-4 For All Participants*

Note. Response categories refer to agreement with intentions of engagement statements.

Table 6 below contains the polychoric correlations for the entire dataset (i.e., irrespective of the 13 author cohorts). Polychoric correlations are appropriate for ordinal by ordinal data. The correlations are strong enough to warrant factor analysis. As would be expected based on the ordinal categories' frequencies' similarity, the ordinal thresholds across items are quite similar, especially the latter two thresholds between "Somewhat Agree, "Agree," and "Strongly Agree." Polychoric correlations and thresholds looked quite similar across the 13 author cohorts, but their equitability will be tested within establishing ME/I.

Table 6. *Polychoric Correlations Between, & τ's of, RIS Items #2-4 For All Participants*

Variable	1.	2.	τ_1	τ_2	τ_3	τ_4	τ_5	τ_6
1. Violent Protest	1.00		0.06	0.36	0.57	0.84	1.1	1.4
2. Violent Group	0.79	1.00	-0.08	0.22	0.45	0.77	1.1	1.4
3. Police Defense	0.67	0.71	0.06	0.32	0.52	0.86	1.1	1.4

5.1.2 ME/I Testing

The ME/I testing step results are reported in Figure 5 below (for a discussion of testing steps, see section 10.2 in the Appendix below). While every non-partial invariance test did not meet the literature's recommended cutoff scores for three out of four $_{\Delta}$GoF indicators (save for Metric Invariance), all MI recommendations with significant predicted $_{\Delta}$GoF changes were not viable or useful. Specifically, MI recommendations were to: free latent means, which is only viable at the Latent Mean Invariance testing stage; free indicator covariances, which not useful for a latent model with only three indicators; or free scale parameters or variances, which involves a level of fastidiousness beyond the scope of this dissertation and most agreed upon ME/I testing procedures, especially with ordinal data. Therefore, I proceeded with non-partial invariance stages until the Latent Mean Invariance Model, for which I could appropriately free MI-recommended latent means and variances. Only freeing Decker and Pyrooz (2019) was necessary to fit a model that fit all $_{\Delta}$GoF cutoff recommendations.

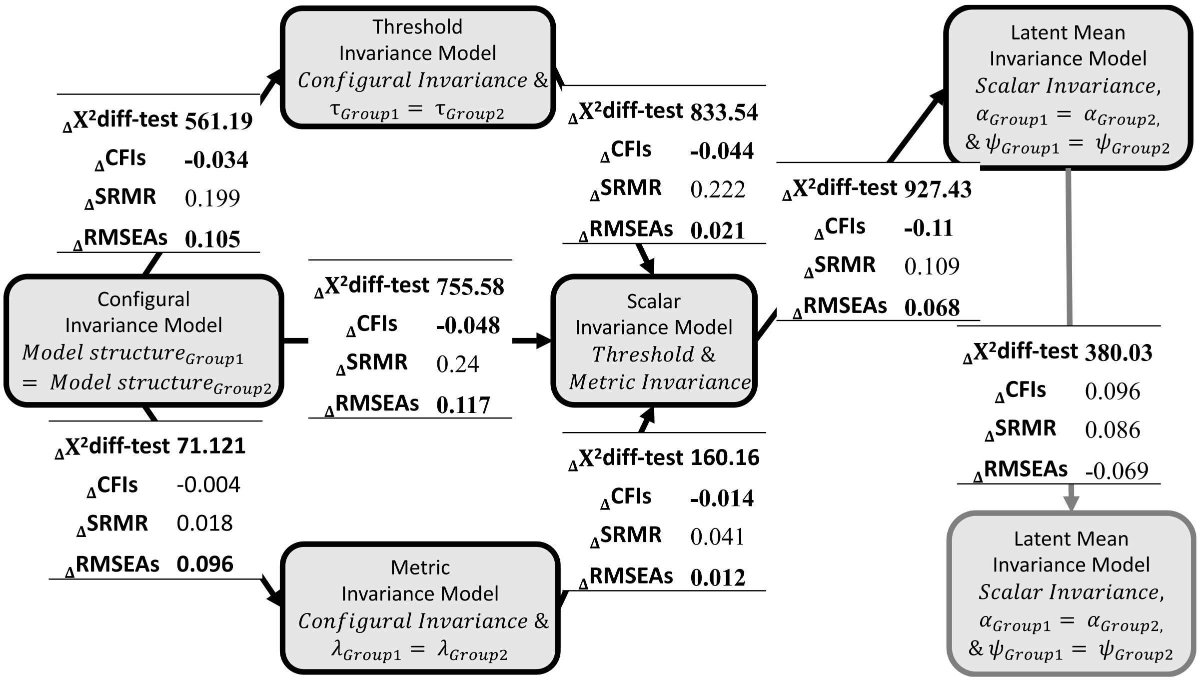

Figure 5. *ME/I Testing Outcomes*

Note. Bolded ΔGoF scores[27] are non-invariant per at least one of the cutoff recommendations.

6 DISCUSSION

These results provide evidence for the measurement equivalence of the RIS items #2 (Violent Group), #3 (Violent Protest), and #4 (Police Defense) across 13 author cohorts—specifically, metric, threshold, and scalar invariance. Furthermore, the latent RIS means and variances across all of these studies may be invariant, save for Decker and Pyrooz's (2019) inmate sample. Specifically, their RIS score was estimated at 1.610 ($p < 0.001$) in the partially latent invariant compared to the standardized 0.000 in all other groups, suggesting that participants in their sample had, on average, chosen ordinal scales 1-to-2 categories higher than participants in all other studies (e.g., choosing "Agree" or "Strongly Agree" when an participant in another study would likely have, all things being equal, chosen "Somewhat Agree" or "Neither Agree nor Disagree"). This is not surprising as Decker and

[27] All ΔGoF scores are scaled or use Bentler corrections to be robust to non-normal distributions.

Pyrooz (2019) are the only authors to have surveyed inmates, who categorically have already committed crimes, and therefore might be more comfortable endorsing illegal, violent (i.e., criminal) behaviors. It is commonly found in criminology that those who have committed crimes previously are likelier than those who have not to commit a future crime.

While many of these models fit poorly, and significantly poorer upon fitting equality constraints, the models' MIs' did not suggest changes for Partial Invariance or any changes that made sense within the model (e.g., freeing covariances between two or more of the three indicators)—except for the Latent Invariance Model. That model's MI's suggested changes to free Decker and Pyrooz's sample's latent Radicalism means and variances ultimately ameliorated all of the indications of significantly poor fit between models—suggesting that the poor fit all along might have been due to missing that group distinction. Furthermore, those models' poor GoF scores should be taken with a grain of salt, as the 'offending' GoF indicators—RMSEA and CFI—are not built for non-continuous data; even robust or scaled, they still do not have well developed cutoffs or 'guidelines' for testing ME/I with ordinal data—unlike for SRMR, which was designed for ordinal data. Furthermore, there may be some localized poor fit within bodies of literature or individual studies, as the ARIS may fit better for more appropriate populations sampled—that is, politically-relevant samples, rather than a general population of students, for example. Local GoF cannot be extrapolated from a single overall model, however—each study or cohort of studies would have to be modeled completely separately to identify a statistical 'culprit.'

What is surprising is that Decker and Pyrooz's (2019) latent RIS mean and variance were the *only* parameters I had to model as Partially Invariant in order to fit an equality-constrained model within ME/I testing $_{\Delta}$GoF cutoffs. This means that all *other* parameters can be treated as equal across all 13 ME/I-tested author cohorts, *and* that latent RIS means and variances are expected to be equal (at least based on items #2-4) across all groups *other* than Decker and Pyrooz's (2019). In other words, based on this most widely used subset of ARIS items from over 50% of all ARIS studies, there is initial evidence that not only is the ARIS being used unbiasedly despite differences in translations, populations sampled, etc., but that most bodies of literature collect samples with approximately the same latent radicalism scores—regardless of country, ethnicity, religion, political affiliation, or any other sampling focus. This evidence is limited, however, to 'parcels' or author cohorts of those studies using RIS items #2-4 with 7-point ordinal scales. While RIS items #2-4 are the most commonly used in research, they usually are not used in a vacuum—often included with RIS item #1, if not items #5 and #6. RIS items #2-4 are likely most widely used because, psychometrically, they are the most sound—RIS item #1 theoretically fits under both Activism and Radicalism, while RIS items #5 and #6 have a very different format from all other ARIS items, and refer to categorically broader and more extreme behaviors than the other items. This research therefore establishes groundwork for further study based on the most widely used and statistically comparable ARIS data available.

7 STRENGTHS & LIMITATIONS

This analysis is a unique systematic comparison of ARIS radicalism research and therein a technical buttress to prior findings and use of the ARIS. This type of analysis, while more accurate than typical meta-analyses or systematic reviews alone, 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, and ultimately limited the number of studies that could be compared—particularly across item exclusion subsets (i.e., systematically missing variables) and studies that rescaled the ARIS from the original 7-point. However, the use of MCCFA, IDA, 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. It includes a majority, but not all, of ARIS research; while study inclusion criteria do not seem to biasedly represent the broader ARIS literature (both received and overall), it does significantly, however slightly, overrepresent several specific population samples (e.g., sampling specific ethnic, religious, or political groups). It is possible that the ARIS may 'work' better for these more targeted studies (e.g., consider the ARIS's poor performance with McCauley and Moskalenko's pilot study of a general population of college students), but, as a measure of political behavioral intentions in support of a specific group, the ARIS might inherently be best suited for specific population samples—If not specific populations that are politically relevant.

The bigger issue with this study is that it cannot be extrapolated to other ARIS items or to ordinal rescales of its items. While the ARIS items are reliably highly correlated from study to study with similar response functionality, we cannot assume without testing that the remaining ARIS items, and the scales in their entirety, will necessarily hold up to the same tests of statistical rigor. Indeed, the more parameters one tests for ME/I on, the likelier one is to detect ME/I—though this can be assuaged by improving model fit with appropriate and strong items to a scale, which we would expect with the ARIS, based on typical individual study-level performance. Unfortunately, after consulting with several researchers far more statistically knowledgeable than myself, including IDA and MNLFA's progenitors, current statistical methodologies cannot yet be used to test the entire ARIS literature for ME/I nor multilevel mixed effects modeling (not to mention some software limitations). Futhermore, even among the harmonizable/integratable studies that I tested, unfortunately several studies' sample sizes were too small to be tested for ME/I between studies (let alone study data-subset (e.g., control vs. experimental conditions)), requiring me to 'parcel' studies into author cohorts. While testing for ME/I at the level of study subsets would be most important, as that is the level of statistical comparison in research (e.g., between one population thought to be at "greater risk" than another), current statistical modeling techniques cannot fit a model to so many groups with so few minimal observations, particularly for ordinal data, let alone for different sets of item scales and systematic missingness/exclusion. Along those lines, methods are still being established by which, via MNLFA, we could test for participant-, study-, and author cohort-level potential moderators of measurement across author cohorts, such as publication year, country or language sampled, population type, and participant demographics, though the nested nature of the data could make estimation difficult. These are problems for the future when MNLFA is developed to tackle data like in this dissertation—ordinal, with multiple ranges, and sets of missing items.

Another limitation is that no GoF cutoffs have been simulated under conditions that fit this IDA exactly (i.e., ordinal indicators of a single factor for 13 groups with n's from 300 to 3,000) for which statisticians have agreed upon. 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. To run more exact simulations than those from the field is beyond the scope of this dissertations. Judging, then, by those criteria based on simulations from conditions most closely matching this IDA, as well as relative changes in fit and general knowledge about the ARIS, it is assumed at this time to be invariant for all tested groups, save for Decker and Pyrooz's (20019) sample's latent means and variances.

8 CONCLUSION

This dissertation provides evidence for the unbiased measurement of radicalism using the ARIS across over a decade of research since its conception (Moskalenko & McCauley, 2009). Specifically, I found that a comparable subset of ARIS author cohorts present measurement equivalence or invariance (ME/I)—the condition that different groups respond the same way to the same test, i.e. without statistical bias or differential item responses. That is, in this case, the same survey questions indicate intentions of engaging in radicalism the same way across many different studied populations, survey translations, and other differences. The factor loadings—how strongly those indicators each contribute to measuring overall radicalism—can reasonably be treated as equivalent across studies 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, my findings modestly support the field's use of the ARIS beyond the original populations for which it was translated and tested. However, future studies will have to test the ARIS for ME/I between a broader, if not the full, suite of ARIS items, and with other Likert scales. As MNFLA advances, future studies will also have to test for ME/I between cross-classified study or data subset groups (e.g., by translation, country, age group, general vs. specific vs. at risk populations, different ordinal scales, etc.). Some may groan to hear an old refrain—“Further research is necessary”—but it is a common truth that holds no less than in this case. ARIS use is growing rapidly since it’s conception in 2009—with 2/3 of publications coming out since 2018, and more on the way. I would caution any researcher who uses the ARIS to continue to test for the ARIS’s GoF for their new samples, if not to test for ME/I of their sample with any publicly shared ARIS datasets that overlap in item set inclusion and scaling, until we can more thoroughly establish the ARIS’s measurement soundness and reliability.

9 REFERENCES

Adam-Troian, J., Baidada, O., Arciszewski, T., Apostolidis, T., Celebi, E., & Yurtbakan, T. (2019). Evidence for indirect loss of significance effects on violent extremism: The potential mediating role of anomia. *45*(6), 691-703.

Adam-Troian, J., Bonetto, E., Araujo, M., Baidada, O., Celebi, E., Dono Martin, M., . . . Yurtbakan, T. (2020). Positive associations between anomia and intentions to engage in political violence: Cross-cultural evidence from four countries. *Peace and Conflict: Journal of Peace Psychology*, *26*(2), 217.

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.

Anastasio, N., Perliger, A., & Shortland, N. (2021). How Emotional Traits and Practices Lead to Support in Acts of Political Violence. *Studies in Conflict & Terrorism*, 1-21.

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

Atari, M., Davani, A. M., Kogon, D., Kennedy, B., Saxena, N. A., Anderson, I., & Dehghani, M. (2021). Morally Homogeneous Networks and Radicalism. *Manuscript in review*.

Bartolo, M. G., Servidio, R., Musso, P., Palermiti, A. L., Iannello, N. M., Perucchini, P., & Costabile, A. (2020). Identità multiple, connessione sociale e attivismo sociale: un modello esplicativo in adolescenti immigrati e italiani. *Ricerche Di Psicologia*.

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.

Bartusevičius, Henrikas (2021) Physical strength predicts political violence, *Evolution and Human Behavior* 42(5): 423–430.

Bartusevičius, H., van Leeuwen, F., & Petersen, M. B. (2020a). Dominance-driven autocratic political orientations predict political violence in western, educated, industrialized, rich, and democratic (weird) and non-weird samples. *Psychological science, 31*(12), 1511-1530.

Bartusevicius, H., van Leeuwen, F., & Petersen, M. (2020b). Political repression motivates antigovernment violence. https://doi.org/10.31234/osf.io/z5b89

Bartusevičius H, Bor A, Jørgensen F, Petersen MB. The Psychological Burden of the COVID-19 Pandemic Is Associated With Antisystemic Attitudes and Political Violence. Psychol Sci. 2021 Sep;32(9):1391-1403. doi: 10.1177/09567976211031847. Epub 2021 Aug 9. PMID: 34369207.

Becker, M. H. (2020). Deciding to support violence: An empirical examination of systematic decision-making, activism, and support for political violence. *Criminology & Criminal Justice*, *21*(5), 669-686.

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

Bhui, K., Warfa, N., & Jones, E. (2014). Is violent radicalisation associated with poverty, migration, poor self-reported health and common mental disorders?. *PloS one*, *9*(3), e90718.

Birrell, C. L. (2017). andling Missing Categorical Data in a Multi-level Structure. *Frontiers in Social Statistics Methodology.* Sydney: UTS.

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.

Borum, R. (2017). The etiology of radicalization. *The handbook of the criminology of terrorism*, 218-219.

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.

Bovaird, J. A., & Koziol, N. A. (2012). Measurement models for ordered-categorical indicators. In R. H. Hoyle, *Handbook of structural equation modeling* (pp. 495-511). 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.

Cardeli, E., Sideridis, G., Lincoln, A. K., Abdi, S. M., & Ellis, B. H. (2020). Social bonds in the diaspora: The application of social control theory to somali refugee young adults in resettlement. *Psychology of Violence, 10*(1), 18.

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.

Choi, H., & Yoon, Y. (2021). Finding Myself Fast and Furiously: The Role of Agency-Communion Orientation and Self-Concept Clarity in Support for Radicalism. *Sustainability*, *13*(5), 2764.

Chui, W. H., Khiatani, P. V., She, M. H., & Wong, C. C. (2020). Using a modified version of SIMSOC to promote active learning in crime causation and response in an unjust society. *Innovations in Education and Teaching International, 57*(1), 16-28.

Costabile, A., Musso, P., Iannello, N. M., Servidio, R., Bartolo, M. G., Palermiti, A. L., & Scardigno, R. (2021). Adolescent Psychological Well-being, Radicalism, and Activism: The Mediating Role of Social Disconnectedness and the Illegitimacy of the Authorities. *Journal of Research on Adolescence, 31*(1), 25-33.

Curran, P. J., Cole, V., Giordano, M., Georgeson, A. R., Hussong, A. M., & Bauer, D. J. (2018). Advancing the study of adolescent substance use through the use of integrative data analysis. *Evaluation & the health professions*, *41*(2), 216-245.

Curran, P.J., & G. Hancock (Hosts) (2021, October 12th). Item Parcels: What Could Possibly Go Wrong?! (Season 3, Episode 6). [Audio podcast episode].In *Quantittude.* QuantitudePod. https://quantitudepod.org/s3e06-item-parcels-what-could-possibly-go-wrong/

Curran, P.J., & G. Hancock (Hosts) (2021, November 16th). The Cons and the Cons of Median Splits (Season 3, Episode 11). [Audio podcast episode].In *Quantittude.* QuantitudePod. https://quantitudepod.org/s3e11-the-cons-and-cons-of-median-splits/

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. (2015). Trauma and openness to legal and illegal activism among Somali refugees. *Terrorism and Political Violence*, *27*(5), 857-883.

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.

Ellis, B. H., Sideridis, G., Miller, A. B., Abdi, S. M., & Winer, J. (2019). Trauma, trust in government, and social connection: How social context shapes attitudes related to the use of ideologically or politically motivated violence. *Studies in Conflict & Terrorism*, 1-18.

Ellis, B. H., Decker, S. H., Abdi, S. M., Miller, A. B., Barrett, C., & Lincoln, A. K. (2020a). A qualitative examination of how Somali young adults think about and understand violence in their communities. *Journal of interpersonal violence*, 0886260520918569.

Cardeli, E., Sideridis, G., Lincoln, A. K., Abdi, S. M., & Ellis, B. H. (2020b). Social bonds in the diaspora: The application of social control theory to somali refugee young adults in resettlement. *Psychology of Violence*, *10*(1), 18.

Ellis, B. H., Miller, A. B., Sideridis, G., Frounfelker, R. L., Miconi, D., Abdi, S., ... & Rousseau, C. (2021). Risk and Protective Factors Associated With Support of Violent Radicalization: Variations by Geographic Location. *International Journal of Public Health, 66*, 20.

Emmelkamp, J., Asscher, J. J., Wissink, I. B., & Stams, G. J. (2020). Risk factors for (violent) radicalization in juveniles: a multilevel meta-analysis. *Aggression and violent behavior*(101489).

Filho, G. M., & Modesto, J. G. (2019). Morality, activism and radicalism in the brazilian left and the brazilian right. *Trends in Psychology, 27*(3), 763-777.

Fischer, R., & A. DInklage (2007). The concept of Integrated Data Analysis of complementary experiments. AIP Conference Proceedings 954, 195 (2007); https://doi.org/10.1063/1.2821262

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. (2020). *Measuring the Thresholds of Extremism: Testing for Measurement Invariance Between Muslim Converts and Muslim Non-Converts of Radicalism With an Ordinal Model.* Thesis, Georgia State University.

Fodeman, A. D., Snook, D., & Horgan, J. (2020). Picking Up and Defending the Faith: Activism and Radicalism among Muslim Converts in the United States. *Political Psychology, 41*(4), 679-698.

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.

Freire, O. D., Senise, D. S., 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.

Frissen, T. (2019). (Hard) Wired for Terror: Unraveling the Mediatized Roots and Routes of'Radicalization'.

Frounfelker, R. L., Frissen, T., Vanorio, I., Rousseau, C., & d'Haenens, L. (2019). Exploring the discrimination–radicalization nexus: empirical evidence from youth and young adults in Belgium. *International journal of public health*, *64*(6), 897-908.

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

Gómez, A., Brooks, M. L., Buhrmester, M. D., Vázquez, A., Jetten, J., & Swann Jr, W. B. (2011). On the nature of identity fusion: insights into the construct and a new measure. *Journal of personality and social psychology*, *100*(5), 918.

Gøtzsche-Astrup, O. (2019). Personality moderates the relationship between uncertainty and political violence: Evidence from two large US samples. *Personality and individual differences*, *139*, 102-109.

Gøtzsche-Astrup, O. (2020). Pathways to violence: do uncertainty and dark world perceptions increase intentions to engage in political violence? *Behavioral Sciences of Terrorism and Political Aggression*, 1-18.

Gøtzsche-Astrup, O. (2021). *Manuscript in Preparation.*

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.

Gigerenzer, G. (2004). Dread risk, September 11, and fatal traffic accidents. *Psychological science*, *15*(4), 286-287.

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.

Greszki, R., Meyer, M., & Schoen, H. (2015). Exploring the effects of removing "too fast" responses and respondents from web surveys. *Public Opinion Quarterly, 79*(2), 471-503.

Guenole, N., & Brown, A. (2014). The consequences of ignoring measurement invariance for path coefficients in structural equation models. *Frontiers in psychology*, *5*, 980.

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.

Hussong, A. M., Curran, P. J., & Bauer, D. J. (2013). ntegrative data analysis in clinical psychology research. *Annual review of clinical psychology*, *9*, 61-89.

Hussong, A. M., Bauer, D. J., Giordano, M. L., & Curran, P. J. (2021). Harmonizing altered measures in integrative data analysis: A methods analogue study. *Behavior Research Methods*, *53*(3), 1031-1045.

Jahnke, S., Schröder, C. P., Goede, L. R., Lehmann, L., Hauff, L., & Beelmann, A. (2020). Observer Sensitivity and Early Radicalization to Violence Among Young People in Germany. *Social justice research*, *33*(3).

Jöreskog, K. (1969). A general approach to confirmatory maximum likelihood factor analysis. *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.

Lai, K. (2020). Correct point estimator and confidence interval for RMSEA given categorical data. *Structural Equation Modeling: A Multidisciplinary Journal, 27*(5), 678-695.

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.

Levinsson, A., Miconi, D., Li, Z., Frounfelker, R. L., & Rousseau, C. (2021). Conspiracy theories, psychological distress, and sympathy for violent radicalization in young adults during the CoViD-19 pandemic: a cross-sectional study. *International journal of environmental research and public health*, *18*(15), 7846.

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.

Little, T. D. (2013). *Longitudinal structural equation modeling*. Guilford press.

Lobato, R. M. (2019). En busca de los extremos: tres modelos para comprender la radicalización. *Revista de Estudios en Seguridad Internacional, 5*(2), 107-125.

Lobato, R. M., Moya, M., & Trujillo, H. M. (2020). Minority-versus Majority-Status Group Intentions to Transgress the Law When Oppression Is Perceived. *nalyses of Social Issues and Public Policy*.

Lobato, R. M., Moya, M., Moyano, M., & Trujillo, H. M. (2018). From oppression to violence: The role of oppression, radicalism, identity, and cultural intelligence in violent disinhibition. *Frontiers in psychology, 9*, 1505.

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

Loughery, Z. (2018). *Reconstructing Radicalisation–A Risk Analysis Perspective of Radical Attitudes in a Swedish Muslim Sample.* Master's Thesis, Lund University.

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.

Mahfud, Y., & Adam-Troian, J. (2021). “Macron demission!”: Loss of significance generates violent extremism for the Yellow Vests through feelings of anomia. *Group Processes & Intergroup Relations*, *24*(1), 108-124.

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.

Malhotra, N. (2008). Completion time and response order effects in web surveys. *Public opinion quarterly, 72*(5), 914-934.

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.

Maxwell, S. E., Lau, M. Y., & Howard, G. S. (2015). Is psychology suffering from a replication crisis? What does “failure to replicate” really mean? American Psychologist, 70, 487–498.

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

McNeish,D. (2018).Thanks coefficient alpha, we'll take it for here. *Psychological Methods, 23*, 412–433. DOI:https://doi.org/10.1037/met0000144

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.

Miconi, D., Calcagni, A., Mekki-Berrada, A., & Rousseau, C. (2020). Are there Local Differences in Support for Violent Radicalization? A Study on College Students in the Province of Quebec, Canada. *Political Psychology.*

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.

Misiak, B., Samochowiec, J., Bhui, K., Schouler-Ocak, M., Demunter, H., Kuey, L., . . . Dom, G. (2018). A systematic review on the relationship between mental health, radicalization and mass violence. *56*(1), 51-59.

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).

Morales, A., Ionescu, O., Guegan, J., & Tavani, J. L. (2020). The Importance of Negative Emotions Toward the French Government in the Yellow Vest Movement. *nternational Review of Social Psychology, 33*(1).

Moreira, P. D. L., Rique Neto, J., Sabucedo, J. M., & Camino, C. P. D. S. (2018). Moral judgment, political ideology and collective action. *Scandinavian journal of psychology*, *59*(6), 610-620.

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 radiaclism. *Terrorism and political violence, 21*(2), 239-260.

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.

Moyano, M., Villen, M., et al. (*Manuscript in Progress*).

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.

Navruz, B. (2016). *The behaviors of robust weighted least squares estimation techniques for categorical/ordinal data in multilevel CFA models.* Dissertation, Texas A&M University.

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

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

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.

Orazani, S. N. (2020). *Shadows of the Past: The Effects of Movements' Past Strategy on Third-Parties' Support for Its Current Strategy.* [Doctoral dissertation, Carleton University].

Padgett, R. N., & Morgan, G. B. (2020). Multilevel CFA with ordered categorical data: A simulation study comparing fit indices across robust estimation methods. *Structural Equation Modeling: A Multidisciplinary Journal*, 1-18.

Pavlović, T., & Franc, R. (2021). Antiheroes fueled by injustice: dark personality traits and perceived group relative deprivation in the prediction of violent extremism. *Behavioral Sciences of Terrorism and Political Aggression*, 1-26.

Pavlović, T., Moskalenko, S., & McCauley, C. (2021). Bifactor analyses provide uncorrelated measures of activism intentions and radicalism intentions. *Dynamics of Asymmetric Conflict*, 1-18.

Pavlović, T., et al. (*Manuscript in Progress*).

Peters, Y. Gjalt-Jorn (2014). The alpha and the omega of scale reliability: Why and how to abondon Cronbach's alpha and the route towards more comprehensive assessment of scale quality. *The European Health Psychologist, 16*(2): 56-69.

Petersen, M. B., Osmundsen, M., & Bor, A. (2021). The Psychology of Status-Seeking and Extreme Political Discontent. *The Psychology of Populism: The Tribal Challenge to Liberal Democracy*.

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

Pfundmair, M., Paulus, M., & Wagner, E. (2020). Activism and radicalism in adolescence: an empirical test on age-related differences. *Psychology, Crime & Law*, 1-16.

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.

Pustejovsky, J. E., & Tipton, E. (2021). Meta-analysis with Robust Variance Estimation: Expanding the Range of Working Models. Prevention Science. https://doi.org/10.1007/s11121-021-01246-3

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.

Ramos, C. (2018). *El Pueblo Unido: Analyzing Group-Based Activism.* Dissertation, The University of Texas at El Paso.

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.

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.

Rottweiler, B., & Gill, P. (2020). Conspiracy Beliefs and Violent Extremist Intentions: The Contingent Effects of Self-efficacy, Self-control and Law-related Morality. *errorism and Political Violence*, 1-20.

Rousseau, C., & Miconi, D. (2020). Vivere insieme in contesti di polarizzazione sociale: fattori di rischio e di protezione in un campione di giovani studenti canadesi. *Educational reflective practices, 1*.

Rousseau, C., Hasan, G., Miconi, D., Lecompte, V., Mekki-Berrada, A., El Hage, H., & Oulhote, Y. (2019b). From social adversity to sympathy for violent radicalization: the role of depression, religiosity and social support. *Archives of public health, 77*(1), 1-12.

Rousseau, C., Hassan, G., & Oulhote, Y. (2017). And if there were another way out? Questioning the prevalent radicalization models. *Canadian Journal of Public Health*, *108*(5), e633-e635.

Rousseau, C., Oulhote, Y., Lecompte, V., Mekki-Berrada, A., Hassan, G., & El Hage, H. (2019a). Collective identity, social adversity and college student sympathy for violent radicalization. *Transcultural psychiatry*.

Rubin, D. B. (1976). Noniterative Least Squares Estimates, Standard Errors and F-Tests for Analyses of Variance with Missing Data. *Journal of the Royal Statistical Society: Series B (Methodological)*, *38*(3), 270-274.

Rubin, D. B. (2004). *Multiple imputation for nonresponse in surveys* (Vol. 81). John Wiley & Sons.

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.

Sass, D. A., Schmitt, T. A., & Marsh, H. W. (2014). Evaluating Model Fit With Ordered Categorical Data Within a Measurement Invariance Framework: A Comparison of Estimators. *Structural Equation Modeling: A Multidisciplinary Journal, 21*(2), 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 *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.

Schultz, J., Bahrami-Rad, D., Beauchamp, J., & Henrich, J. (2018). The Origins of WEIRD Psychology. *Available at SSRN 3201031*.

Schumann, S., Salman, N. L., Clemmow, C., & Gill, P. (2021). Does cognitive inflexibility predict violent extremist behaviour intentions? A registered direct replication report of Zmigrod, Rentfrow, & Robbins, 2019. *Legal and Criminological Psychology*, *26*(2), 145-157.

Schuurman, B. W., Grol, P., & Flower, S. (2016). Converts and Islamist terrorism: An introduction. *International Centre for Counter-Terrorism Policy Brief*, *7*(21).

Shortland, N., & McGarry, P. (2021). The personality and propaganda puzzle: Exploring the effect of personality on exposure to extremist content online. *Psychology of Violence*.

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.

Smith, B. A. (2016). *Religious priming and moral reasoning as a manipulation for supporting violence.* Dissertation, The University of Texas at El Paso.

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. W., Branum-Martin, L., & Horgan, J. G. (2021). Zeal of the convert? Comparing religiousness between convert and nonconvert Muslims. *Psychology of Religion and Spirituality.* Advance online publication. https://doi.org/10.1037/rel0000421

Snook, D. W., Kleinmann, S. M., White, G., & Horgan, J. G. (2021). Conversion motifs among Muslim converts in the United States. *Psychology of Religion and Spirituality*, *13*(4), 482–492. https://doi.org/10.1037/rel0000276

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.

Stern, J. (2016). Radicalization to extremism and mobilization to violence: What have we learned and what can we do about it? The Annals of the American Academy of Political and Social Science, 668(1), 102–117.

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. (2020). Multiple-Group Invariance with Categorical Outcomes Using Updated Guidelines: An Illustration Using Mplus and the Lavaan/semTools Packages. *Structural Equation Modeling: A Multidisciplinary Journal, 27*(1), 111-130.

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.

Trujillo, H. M., Prados, M., & Moyano, M. (2016). Psychometric properties of the Spanish version of the activism and radicalism intention scale/Propiedades psicométricas de la versión española de la escala de intención de activismo y radicalismo. *Revista de Psicología Social*, *31*(1), 157-189.

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 de Schoot, R., Lugtig, P., & Hox, J. (2012). A checklist for testing measurement invariance. *European 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.

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.

Vandenplas, C., Loosveldt, G., Beullens, K., & Denies, K. (2018). Are interviewer effects on interview speed related to interviewer effects on straight-lining tendency in the European Social Survey? An interviewer-related analysis. *Journal of Survey Statistics and Methodology*, *6*(4), 516-538.

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.

Wagoner, J. A., Rinella, M. J., & Barreto, N. B. (2021). “It was rigged”: Different types of identification predict activism and radicalism in the US 2020 election. *Analyses of Social Issues and Public Policy*.

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

Web survey paradata on response time outliers: A systematic literature review. (2018). *Metodološki zvezki*, *15*(1), 23-41.

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.

Wolfowicz, M., Litmanovitz, Y., Weisburd, D., & Hasisi, B. (2020). A field-wide systematic review and meta-analysis of putative risk and protective factors for radicalization outcomes. *Journal of quantitative criminology, 36*(3), 407-447.

Wolfowicz, M., Litmanovitz, Y., Weisburd, D., & Hasisi, B. (2020). What Is the State of the Quantitative Literature on Risk Factors for Radicalization and Recruitment to Terrorism?. In *Understanding Recruitment to Organized Crime and Terrorism* (pp. 25-53). Springer, Cham.

Wong, M. Y., Khiatani, P. V., & Chui, W. H. (2019). Understanding youth activism and radicalism: Chinese values and socialization. *The Social Science Journal*, *56*(2), 255-267.

Wright, J. D., Cheung, I., & Esses, V. M. (2019). Is Support for Confederate Symbols Motivated by Southern Pride or Racism? Further Distinctions Between "Traditionalists" and "Supremacists". *Manuscript in review*.

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.

10 APPENDIX

10.1 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 **Error! Reference source not found.**). 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[28], 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, Multiple-Group Invariance with Categorical Outcomes Using Updated Guidelines:

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

An Illustration Using Mplus and the Lavaan/semTools Packages, 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 (Millsap & Yun-Tein, 2004). This contradicts standards for continuous data ME/I testing[29]. In order to avoid the disadvantages of any one ordinal ME/I testing pathway strategy, all paths are tested in this dissertation (see **Figure *6*** below for a depiction). 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).

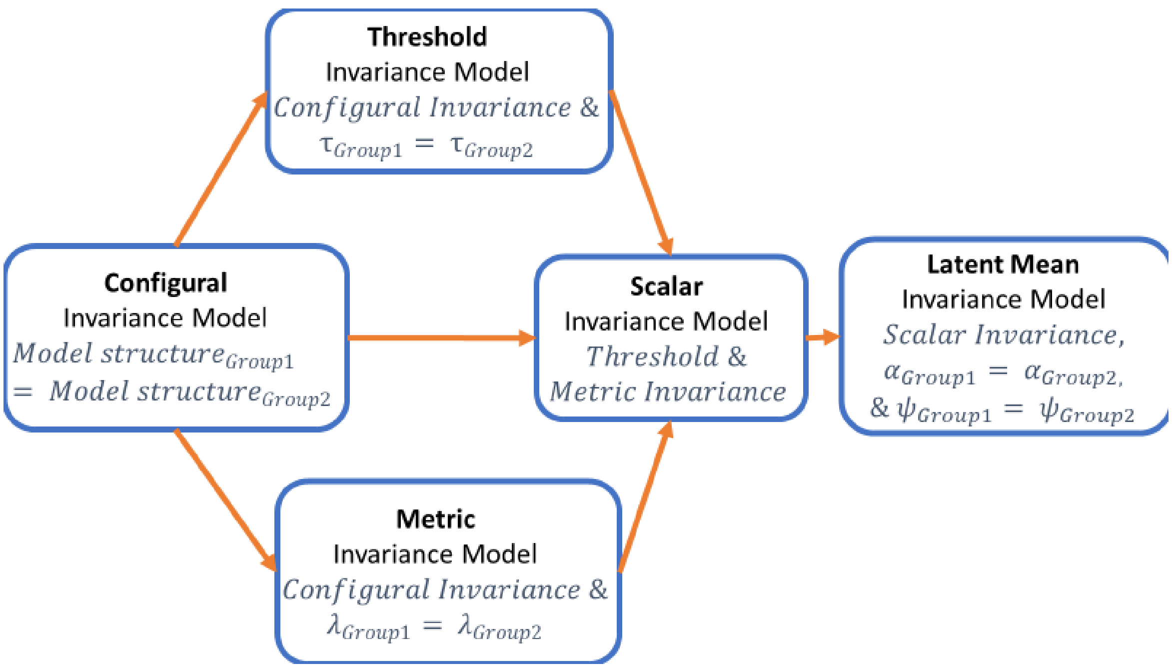

Figure 6. *Ordinal ME/I Testing Steps (Fodeman, 2020)*

10.2 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

[29] 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).

"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 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 5 (Results) are many of their visualizations—more succinct and clear ways of reporting that information.

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