ABSTRACT

Title of Dissertation: **INTERSECTIONAL STIGMA, SELF-EFFICACY, DEPRESSION, AND RESILIENCE: A RASCH ANALYSIS**

Jacqueline Rose Reuben
Doctor of Philosophy, 2022

Dissertation directed by: Dr. Hongjie Liu
Professor, Department of Epidemiology and Biostatistics

Dr. Rodman Turpin,
Assistant Research Professor, Department of Epidemiology and Biostatistics

Background: Black men who have sex with men (BMSM) are disproportionally affected by negative health outcomes associated with stigma related to both their racial and sexual minority status. Foundational to understanding stigma is the ability to correctly measure this latent construct. The Rasch model is a probabilistic model for analyzing categorical data that was developed to improve the measurement of latent traits. This study, grounded in intersectional minority stress theory, reviewed the application of Rasch analysis in the HIV/AIDS literature (Aim 1) and used the Rasch model to calibrate person measures to assess the interrelationships among internalized stigma, resilience, self-efficacy, and psychologic well-being among BMSM (Aims 2 and 3).

Methods: For Aim 1, we conducted a systematic review of the literature following PRISMA guidelines. Aims 2 and 3 used data from a cross-sectional online survey of 151 HIV-negative BMSM in 2020. For Aim 2, we conducted Rasch analysis to assess the psychometric properties of the internalized racism (IR), internalized homophobia (IH), self-efficacy, and resilience scales. For Aim 3, we used linear regression and path analysis of the Rasch-calibrated person measures to examine the mediating and modifying effects of self-efficacy and resilience on the relationship between intersectional stigma and depressive symptoms.

Results: For Aim 1, after screening 183 articles, 45 articles were included in the analysis. Strengths and weaknesses of using the Rasch approach were summarized. For Aim 2, the final IR scale had a person reliability and separation of 0.91 and 3.13, respectively, and an item reliability and separation of 0.94 and 4.01, respectively. The final IH scale had a person reliability and separation of 0.88 and 2.72, respectively, and an item reliability and separation of 0.79 and 1.95, respectively. For Aim 3, IR (β=0.296, 95% CI [0.133, 0.458]) and IH (β=0.414, 95% CI [0.204,0.623]) were independently and positively associated with depression in multivariable models controlling for age, income, and relationship status. Resilience and self-efficacy modified the relationship between IH and depression (β_{IHxRES}=-0.034, 95% CI [-0.060, -0.008] and β_{IHxSE}=-.056, 95% CI [-0.113, 0.00], respectively), but there was no evidence of effect modification by resilience or self-efficacy on the association between IR and depression.

Public health implications: Our findings suggest that interventions targeting multiply marginalized groups such as BMSM that address co-occurring forms of stigma and foster positive self-evaluation and coping skills may reduce the negative consequences of internalized stigma on mental health outcomes.

INTERSECTIONAL STIGMA, SELF-EFFICACY, DEPRESSION, AND RESILIENCE: A RASCH ANALYSIS

by

Jacqueline Reuben

Dissertation submitted to the Faculty of the Graduate School of the
University of Maryland, College Park, in partial fulfillment
of the requirements for the degree of
Doctor of Philosophy
2022

Advisory Committee:
Professor Dr. Hongjie Liu, Co-Chair
Assistant Research Professor Dr. Rodman Turpin, Co-Chair
Professor Dr. Edmond Shenassa, Dean's Representative
Assistant Professor Dr. Quynh Nguyen
Assistant Professor Dr. Tianzhou Ma

Acknowledgements

This dissertation would not have been possible without the support that I received from my dissertation advisory committee co-chairs, Dr. Hongjie Liu and Dr. Rodman Turpin. I am incredibly grateful to them for providing me with the rich dataset used in this dissertation, and their continued guidance (and patience) with each iteration of my analysis. I also would like to thank my committee members, Dr. Nguyen, Dr. Ma, and Dr. Shenassa, for their expertise and insightful feedback.

And to my family, thank you for never giving up on me. It's been a long and winding road, with several pit stops, and I couldn't have done it without you. Mom, Dad, Jamie, and Jess, I love you.

Table of Contents

Chapter 1: Introduction

1.1 Background

HIV among MSM in the U.S.

While significant progress has been made in the prevention and treatment of HIV/AIDS over the past four decades, gay and bisexual men and other men who have sex with men (MSM) remain more affected by the HIV epidemic than any other group. In 2019, MSM accounted for 69% of new HIV diagnoses in the United States (US).[1] Black MSM (BMSM) are disproportionally affected; although estimated to represent less than 1% of the US population,[2] BMSM accounted for 25% of all new HIV diagnoses, and 38% of diagnoses among MSM in 2019.[1,3]

Previous research seeking to explain the persistently higher proportion of HIV infection among BMSM has focused primarily on individual-level behavioral risk factors.[4,5] Multiple studies have identified sexual risk behaviors such as unprotected anal intercourse and drug and alcohol use as major risk factors for HIV infection among MSM.[6] A meta-analysis of HIV risk behaviors, however, found that BMSM reported fewer sexual risk behaviors and lower overall substance use compared to white MSM.[7,8] This suggests that these behavioral risk factors do not fully explain the racial disparity in HIV between BMSM and white MSM.

Beyond individual-level behavioral risk factors, the social and environmental conditions in which individuals live, many of which are rooted in the inequitable distribution of and access to socioeconomic and political resources, are increasingly recognized as key contributors to the health of individuals and populations.[9,10] In particular, a growing body of literature has begun to explore the impact of different forms of stigma and discrimination on health, and evidence

suggests self-reported experiences of stigma and discrimination are a psychosocial stress that negatively impacts a range of mental health outcomes.[11]

Stigma and Health

Since the start of the HIV epidemic, researchers have recognized stigma as a significant barrier to HIV prevention and treatment efforts for those populations at highest risk, and it continues to remain a formidable barrier today.[12] Link and Phelan (2001) conceptualized stigma as the co-occurrence of four interrelated components within a context in which power is exercised: 1) labeling, 2) stereotyping, 3) separation, and 4) status loss and discrimination.[13] Stigma can be based on multiple different aspects of an individual's actual or perceived identity or group membership, and a substantial body of literature has highlighted the detrimental impacts of HIV-related stigma, racism, and homophobia on multiple health outcomes.[14-17]

Minority stress theory, which posits that members of minority groups are subjected to unique, chronic, stressors related to their stigmatization that in turn increase their vulnerability to negative health outcomes, has framed much of the extant literature investigating the link between stigma and mental health among MSM.[18] Stigma, homophobia, and discrimination put MSM of all race/ethnicities at increased risk for multiple adverse health outcomes including depression, anxiety, and substance use disorders.[19,20] Mental health and psychosocial challenges, in turn, have been identified as major contributors to risk behaviors for HIV infection, and are also linked to a range of adverse HIV-related health outcomes.[21-24]

Until recently, the majority of the empirical research based in minority stress theory has focused on the stigma related to a single minority status.[25] BMSM, however, are faced with stigma associated with both their sexual and racial minority status.[26] Viewing characteristics

such as sexual orientation and race in isolation has impeded our understanding of how stigma related to multiple identities may intersect to affect health outcomes.

Shifting Toward Strength and Resilience

Most quantitative stigma research has taken a deficits-based approach, focusing on the processes by which stigma creates conditions for marginalization and reduced opportunities that in turn negatively impact health.[27] By focusing solely on the negative experiences of stigma, however, researchers have overlooked the positive and potentially protective factors at the intersection of race and gender, such as resilience and self-efficacy. These factors may modify the effects of stigma on the mental and sexual health of BMSM.

In recent years, strength-based conceptual frameworks have been proposed to suggest that protective factors such as resilience may help buffer the impact of homophobia and racism on HIV risk.[28] There have been increased calls for a resilience-based agenda for HIV stigma research that focuses on strength-based modifiers of the association between stigma and health.[29] Yet studies exploring stigma and resilience among BMSM have revealed conflicting results.[30-32] For example, a recent study examining condomless anal sex among BMSM in New York City that used validated scales to assess resilience, condom use self-efficacy, and internalized homophobia found that decreased condomless anal sex was associated with increased levels of condom use self-efficacy, but not associated with resilience or internalized homophobia.[33] Another study using a nationally representative sample of MSM from the National HIV Behavioral Surveillance System (NHBS) examined the association between perceived discrimination and prevalent HIV infection between BMSM and white MSM and found no association between perceived homophobia, perceived racism, and resilience with prevalent HIV.[34] Thus, there appears to be contradictory dynamics exerting influence on how stress affects

individuals who are racial and sexual minorities.[22,35] This suggests that there may be other factors at play, such as the existence of strength-based protective mechanisms from psychological stress for members of multiple marginalized minority groups.

Integrating the concept of intersectionality with minority stress theory can help shed light on the complex interplay of minority stress and positive coping factors that impact multiply marginalized groups such as BMSM. Intersectionality refers to the interaction between gender, race, and other social categories at the individual level of experience that reflect the multiple interlocking systems of privilege and oppression at the socio-structural level.[36,37] Intersectionality posits that these social identities are not independent, but rather are mutually constitutive and reinforcing, thereby creating different and distinct experiences of oppression and opportunity for the individual.[37,38] The application of intersectionality to assess multiple stigmatized identities also allows researchers to move beyond a "stigma=vulnerability" focus to identify the ways in which stigmatized groups can navigate these systems of oppression and challenge stigma via protective factors such as resilience and empowerment.[27,39,40]

Using an intersectional minority stress framework, which combines the minority stress model and intersectionality theory, this dissertation examines the risk and protective pathways generated by minority stress processes (e.g., intersectional internalized stigma) to depressive symptoms among this sexual and racial minority group.

1.2 Guiding Theoretical Frameworks

Minority Stress Theory

Minority stress theory has been the most prominent theoretical framework proposed to explain the processes that contribute to sexual orientation disparities in mental and behavioral health outcomes. The concept of minority stress is derived from several social and psychological

theories that discuss the adverse effect of social conditions, such as stigma and discrimination, on the lives of affected individuals and groups.[13,41-44] Elaborating on the notion that stress is related to social structures and conditions, Meyer describes minority stress as the excess stress to which sexual minority individuals are exposed to because of their minority status.[20] Minority stress is defined by three characteristics; it is 1) unique, in that minority stress is additive to general stressors experienced by all people; 2) chronic, in that it is related to stable underlying social and cultural structures; and 3) socially based, in that it stems from external social processes, institutions, and structures rather than individual characteristics or conditions.[18]

In applying minority stress theory to gay men, Meyer postulated that the repeated exposure to this excess stress can yield 1) chronic expectations of rejection and discrimination (i.e., anticipated stigma), 2) internalization of heterosexist attitudes (i.e., internalized homophobia), and 3) actual experiences of discrimination and violence.[19] Across studies employing different methodologies, both anticipated stigma and internalized homophobia have been associated with adverse mental health outcomes. A meta-analysis conducted by Newcomb and Mustanski (2010) found higher levels of internalized homophobia were significantly associated with measures of depressive symptomatology and symptoms of anxiety.[45] A more recent systematic review assessing minority stress and biological health also found substantial evidence to support the relationship between minority stress and multiple physical outcomes, including overall physical health, immune response, cardiovascular outcomes, and metabolic and endocrine/hormonal function.[16]

In practice, most studies utilizing a minority stress framework have considered stigma as it relates to a single stigmatized social category, ignoring other characteristics or circumstances in the person's environment that may interact to determine an individual's exposure to stress and

coping resources. For example, previous studies have focused on one type of stigma (e.g., homophobia) in isolation from others (e.g., racism). Independent relationships between stigma associated with a single minority identity and outcomes such as depression, distress, substance use, and sexual risk behaviors are well established in the literature.[46,47] While important, these links do not completely account for the health inequities experienced by BMSM.[7] This highlights the importance of a framework that explicitly acknowledges multiple intersecting identities to better understand the complexities of health disparities for populations comprised of multiple stigmatized groups.[13,47]

Intersectionality

Intersectionality theory provides such a framework for understanding the impact of minority stressors on people who live at the intersection of multiple marginalized identities. Intersectionality has been defined as the mutually constitutive relations among social identities that interact to form qualitatively different meanings and experiences.[48] Stigma researchers have shown an increased interest in the application of intersectionality,[49] and a growing body of literature has sought to elucidate how minority stress impacts members of minority groups who live at the intersection of multiple marginalized identities.[50-52] Intersectional stigma refers to the "interdependent and mutually constitutive relationship between social identities and structural inequities."[15] This convergence of multiple identities produces unique experiences of inequity and opportunity for the individual.[53]

This approach highlights two important concepts. First, because social identities do not function independently, stigma and discrimination experienced by BMSM cannot be understood by simply adding up the experiences associated with being a Black person and a man who has sex with men, but rather must be understood as a distinct experience in and of itself.[38] Second,

intersectionality is conceptualized as creating conditions for both marginalization *and* opportunity, which challenges researchers to move beyond a focus solely on the harms of stigma to include a focus on strengths and empowerment. By viewing minority stress processes through an intersectional lens, we can examine how the convergence of multiple stigmatized identities, each of which is associated with unique systems of oppression and opportunity, can create both risk and protective pathways to mental health outcomes among BMSM.

Proposed Theoretical Framework: Intersectional Minority Stress

As noted previously, there appears to be contradictory dynamics exerting influence on how stress affects individuals who are racial and sexual minorities.[22,35] This suggests that there may be other factors at play, and potentially the existence of a protective mechanism from psychological stress for members of multiple marginalized minority groups. Most quantitative stigma research has taken a deficits-based approach, focusing on the processes by which stigma creates conditions for marginalization and reduced opportunities that in turn harm health outcomes.[27] By focusing solely on the negative experiences of stigma and discrimination, however, researchers have overlooked the positive and potentially protective factors at the intersection of race and gender, such as resilience and self-efficacy. These factors may modify the effects of discrimination and stigma on the mental and sexual health of BMSM.

1.3 Methodologic Challenges

A conceptual and methodologic shift is needed to better understand the relationship between minority stress and health outcomes among BMSM. Intersectional minority stress approaches have great potential to enhance public health research through improved validity and increased attention on the causal process contributing to health inequities,[49] but there has been much debate regarding the methodologic approaches to conducting this research.[54] Some researchers have

argued that qualitative methods are more compatible with intersectional approaches than quantitative methods, and until recently, researchers have relied primarily on qualitative methods.[55,56] With the expanded application of intersectionality to disciplines such as psychology,[48,54,57] there have been increased calls for the incorporation of quantitative methods to address complimentary questions and enhance the value and validity of this research.[58] A variety of quantitative methods have been proposed, but there remains much debate on the methodologic approaches, data collection instruments, and analytic methods best suited for this research.[36,52,55,56] As intersectional approaches move from primarily qualitative research methods into quantitative methods, researchers must pay more attention to the implications of the methodological variability across studies on the interpretation of empirical tests of study hypotheses, and ultimately the quality of results and validity of conclusions.[55,56,59]

Aside from the debate surrounding the most appropriate statistical approaches to use, the conclusions that can be drawn from these studies remain restricted first and foremost by the quality of the scales used to capture these latent constructs and the measures derived from them.[60] Researchers frequently rely on scales, indices, and other item-based quantitative instruments to derive variables to quantify respondents' knowledge, abilities, attitudes, and other latent traits that cannot be directly observed. Foundational to the validity of study conclusions is the validity of the measurement of these constructs. Multiple scales have been developed to measure underlying latent constructs such as stigma and resilience, which cannot be directly measured or captured by a single variable or item. The summary measures created from these scales play a central role in subsequent analyses as both predictor variables and the outcome of interest, yet the process by which these scores are obtained is often overlooked.[61] For example, even though evidence for the internal structure of a scale is a key component of scale

development,[62,63] a recent review of the health psychology literature by Crutzen and Peters (2017) found that less than 3% of studies reported information about the validity of the scale used.[62]

Researchers typically evaluate the psychometric properties of instruments using methods based in classical test theory (CTT), also referred to as the "true score model," whereby raw scores, typically a linear combination of responses that have an ordinal scale, essentially are considered interval-scale data. However, as rating scales are ordinal, not interval or continuous, the calculation of such scores is not only unjustified but also difficult to interpret. Unjustified sum scoring can have adverse effects on the validity and reliability of the measurement, and depending on how the researcher creates a summary score from the scale items, the use of these scores in subsequent statistical analyses can lead to very different conclusions, even if derived from the same data.[64,65]

Despite its widespread use, CTT is subject to several limitations. First, it assumes a linear relationship between the latent variable and observed scores. Second, it assumes each item contributes equally to the final score, irrespective of how much each item correlates with the underlying construct. It is highly unlikely that each scale item is of equal "difficulty" or "agreeability," so a sum of raw scores cannot be used to accurately compare differences between sample participants. Third, it assumes that each item is measured on the same interval scale, and that the distance between response options is equivalent from one item to the next.[66] Fourth, parameters such as reliability, discrimination, location, and factor loading depend on the sample being used. Fifth, and related to the previous problem, there is a circular dependency between the sample and items, in that scores on a scale depend on "how much of the trait" the sample has, while "how much they have" depends on the scale norms.[67]

Novel Analytical Approach Proposed

To overcome the limitations of using conventional analytic techniques based on CTT, Rasch measurement models have been increasingly used in the development and validation of instruments measuring physical, social, and mental health constructs.[68-73] The Rasch model is a probabilistic model that is used to describe the relationship between a participant's response to a categorical question (i.e., item) and the underlying construct being measured by the scale as a function of the difference between two parameters, the person's "ability" and the item "difficulty."[74] A person's latent trait level thereby determines their probability of answering a question of a specific difficultly in a particular way. Thus, the focus in Rasch measurement is on a person's latent trait level, rather than their "true score."

The Rasch model provides a framework for formally testing empirical data to determine an instrument's capacity to emulate the properties of fundamental measurement (invariance and unidimensionality), which are required for an instrument to serve as a useful measurement tool. That is, the Rasch model is prescriptive, and if the data fit the model, the model converts nonlinear, raw data into linear, interval-level measures that can then be evaluated using parametric statistical tests.[75] Person and item measures are calibrated onto a single unidimensional latent trait continuum that accounts for unequal "difficulties" across scale items and facilitates direct comparisons between person measures and item difficulties. Rasch analysis also has the advantage of providing in-depth evaluation of individual item and person patterns of responses to provide more thorough evaluation of the instrument's psychometric properties.[66] The Rasch model has been applied extensively in educational research, but its use in the public

health sciences remains scant.[76] This dissertation brings a Rasch approach to the analysis of the complex ways in which intersectional stigma may affect the psychological well-being of BMSM.

1.4 Objectives of the Study

The objectives of this study are to (1) systematically review the application of Rasch analysis in the HIV/AIDS literature (Chapter 2), (2) use the Rasch model to evaluate the psychometric properties of scales developed to measure internalized stigma, resilience, self-efficacy, and depressive symptoms among BMSM (Chapter 3), and (3) conduct linear regression and path analysis of the Rasch-derived person measures to assess the interrelationships among intersectional stigma, self-efficacy, depression, and resilience (Chapter 4).

1.5 Significance and Innovation

A conceptual and methodologic shift is needed to better understand the relationship between minority stress and mental health outcomes among BMSM. Using an intersectional minority stress framework, which combines the minority stress model and intersectionality theory, this dissertation examines the risk and protective pathways generated by minority stress processes (e.g., intersectional stigma) to depressive symptoms among BMSM, a dual sexual and racial minority group. This dissertation seeks to further the research by first utilizing Rasch analysis to develop and refine high-quality measurement scales and generate Rasch-calibrated person measures, which subsequently will be analyzed by both linear regression analysis with interaction terms and path analysis to explore the relationships among intersectional stigma, self-efficacy, resilience, and depressive symptoms among BMSM. This research is an important step in reconceptualizing the statistical analysis of epidemiologic research focused on multiply marginalized groups. These data also will inform interventions targeting psychosocial health

outcomes among this population, to ultimately reduce health disparities and improve population health.

Chapter 2: Systematic Review of the Application of Rasch Analysis in the HIV/AIDS Literature

2.1 Introduction

Background

Since the start of the HIV epidemic over 40 years ago, researchers have identified stigma as a significant barrier to HIV prevention and treatment efforts for those populations at highest risk, and it continues to remain a formidable barrier today.[12] Stigma can be based on multiple different aspects of an individual's actual or perceived identity or group membership, and a substantial body of literature has highlighted the detrimental impacts of HIV-related stigma, racism, and homophobia on multiple health outcomes.[14-17] In recent years, strength-based conceptual frameworks have suggested that protective factors such as resilience may help buffer the impact of homophobia and racism on HIV risk.[28] There have been increased calls for an HIV research agenda that focuses on modifiable, strength-based protective factors that may moderate or mediate the association between stigma and negative health outcomes.[29] Since these variables are not directly measurable, researchers must rely on directly observable indicators believed to represent the underlying construct. Valid and reliable measures that effectively capture and distinguish the complexity of risk and protective pathways to health outcomes is essential to our ability to develop effective intervention and prevention programs.

Studies exploring stigma and resilience among Black men who have sex with men (BMSM), however, have revealed conflicting results.[30-32] These conflicting results may be attributed in part to a lack of clarity in the conceptualization and measurement of these constructs.[25,29,77,78] Such inconsistencies not only compromise the validity of these measures, but also may bias the

outcomes of studies and limit the utility of the public health actions informed by this research.[29,79,80]

To better understand stigma and resilience among BMSM, a psychometrically sound measure is essential. Several scales have been developed for measuring stigma and resilience, although not all are widely adopted and no one scale is preferable over others. Different approaches to measuring these constructs across studies have led to inconsistencies relating to our understanding of the nature of potential risk factors and protective processes, and in estimates of prevalence. This creates difficulties in comparing results across studies and raises questions about the extent to which researchers are measuring the constructs they are intending to measure, or an entirely different experience.

Measurement Challenges

Unlike the physical sciences, in which fundamental and derived measurements for quantifying attributes of objects are widely available, measurements in the human sciences capable of capturing unobservable quantities of beliefs or behaviors is far less straightforward.[66] Researchers must rely on scales, indices, and other item-based quantitative instruments to derive variables to quantify respondents' knowledge, abilities, attitudes, and personality traits. The primary goal in developing high-quality psychosocial health measures is to create a valid and reliable measurement of an underlying construct, also referred to as a latent trait, that cannot be observed directly.[81]

Most studies evaluating the psychometric properties of scales used to measure latent traits have used classical test theory (CTT), also referred to as the "true score model." CTT assumes that each person has a true score on an item or a scale that quantifies the person's location on the underlying construct, trait, or ability of interest, which is equal to the combination of the

observed score (X) and an error term (E). Despite its widespread use, this approach suffers from several limitations. First, CTT assumes a linear relationship between the underlying latent variable and the observed score. Most commonly, a total score is obtained by adding the scores for each individual item. However, raw survey data collected through Likert-type scales are inherently ordinal in nature. By imposing linearity, a scale item that is discrete and restricted (for example, to 5 values) is treated as if it were continuous, ranging from negative infinity to positive infinity. Data visualizations of categorical measures, however, tend to show a very different behavior, such as accumulation at certain values, multiple peaks, or gaps between values. In such cases, the continuous assumption can only serve as an approximation at best. And by imposing linearity, each item measured is assumed to be measured on the same interval scale, and the "psychological distance" between each of the response options of an item is assumed to be the same.[82] In addition, summing total scores assumes that each item contributes equally to the final score, irrespective of how much each item correlates with the underlying construct. Missing data therefore presents a problem for computing overall scores.

In CTT, item and scale statistics such as reliability, discrimination, location, and factor loading, apply only to the specific group of subjects to whom the survey or instrument was administered. That is, the psychometric properties are sample- and item-dependent; if the scale is administered to a different group of people, in a different way, or if an item is altered or deleted, the psychometric properties of the scale must be reestablished.[67] It is therefore impossible to separate the properties of the scale and item characteristics from the characteristics of the people taking it.[83] Thus, it is difficult to compare people who are assessed using different scales, and difficult to compare scale items whose psychometric properties were established using different groups of people. Constant replication and revalidation of results is needed to gauge scale and

item validity across different populations, making results difficult to generalize and inferences subject to potential bias.

Researchers must take advantage of modern psychometric procedures to develop measures to characterize these complex factors more accurately. Rasch measurement refers to a family of probabilistic models based on the original logistic model that was developed to calibrate dichotomous data by Georg Rasch in 1957.[76] The Rasch model assumes that each person is characterized by an "ability," and each item by a "difficulty," that can be expressed by numbers along a single common line, and that from the distance between these numbers, the probability of observing any particular scored response can be computed.[66] Thus, the model is built on the assumption that the most parsimonious and effective predictor of a trait is the relationship between the "difficulty" of an item and the "ability" of a person. The Rasch model assumes unidimensionality, which refers to the assumption that each scale or subscale represents a single construct or characteristic, and the invariance of the model across samples, both of which are critical criterion of fundamental measurement.[84] The model is prescriptive, demanding a data-to-model fit,[85] rather than descriptive, and thus the data are fit to the Rasch model to test how well the observed data fit the expectations of the measurement model.

Rasch measurement overcomes several of the limitations of CTT. If the data fit the model, the model converts ordinal data (from a set of items that define a single trait) into linear (equal interval) measures that have a constant interval meaning and provide objective measurement from ordered category responses. Once the interval metric is established, person measures and item difficulties are calibrated onto a single unidimensional latent trait continuum, which facilitates direct comparisons between person measures and item difficulties.[74] The Rasch-derived, linear person measure is a value that can then be used in subsequent parametric

statistical calculations, even though the analysis began with ordinal data. As Rasch models can be used to evaluate whether each item performs differently across different groups of respondents, Rasch measurement also provides a technique by which sample-independent item measures and item-independent respondent measures can be computed, such that the findings from the sample extrapolate to the population.[66]

Specific Aims

Rasch analysis was developed to improve the precision with which instruments are constructed and validated. Rasch is particularly attractive because 1) it is believed to be superior to traditional psychometric methods and 2) it can produce ordinal-to-interval conversions. While Rasch has been increasingly used within health care research in general, a comprehensive overview of the use of Rasch analyses within the HIV/AIDS field is lacking. A review of the use and application of Rasch analysis within the HIV/AIDS literature may give insight into future research directions and highlight possibilities for the improvement of our measurements.

The specific aims of this chapter are to 1) systematically review the HIV/AIDS literature to identify the frequency with which Rasch analysis has been used in the development or refinement of measurement scales, 2) characterize the use and assess the quality of reporting of the Rasch model in these studies, and 3) provide recommendations on the application of Rasch analysis in the field of HIV/AIDS.

2.2 Methods

This review followed the Preferred Reporting Items for Systematic Reviews and Meta-analyses (PRISMA) statement, which is a critical appraisal tool that helps authors improve the reporting of systematic reviews and meta-analyses.[86]

2.2.1 Search Strategy

Electronic database searches of PubMed, Web of Science, CINHAL, and Cochrane Library were conducted to identify all eligible studies published up to and including July 2022. To capture as many articles as possible, search terms were broad and included the following: (Rasch) AND (HIV/AIDS). All English-language articles published in a scientific journal that applied Rasch analysis to the development or refinement of a health scale were included. Reviews, letters, editorials, opinion papers, abstracts, posters, and purely descriptive studies were excluded. **Figure 2.1** presents an overview of the stages followed during the search process.

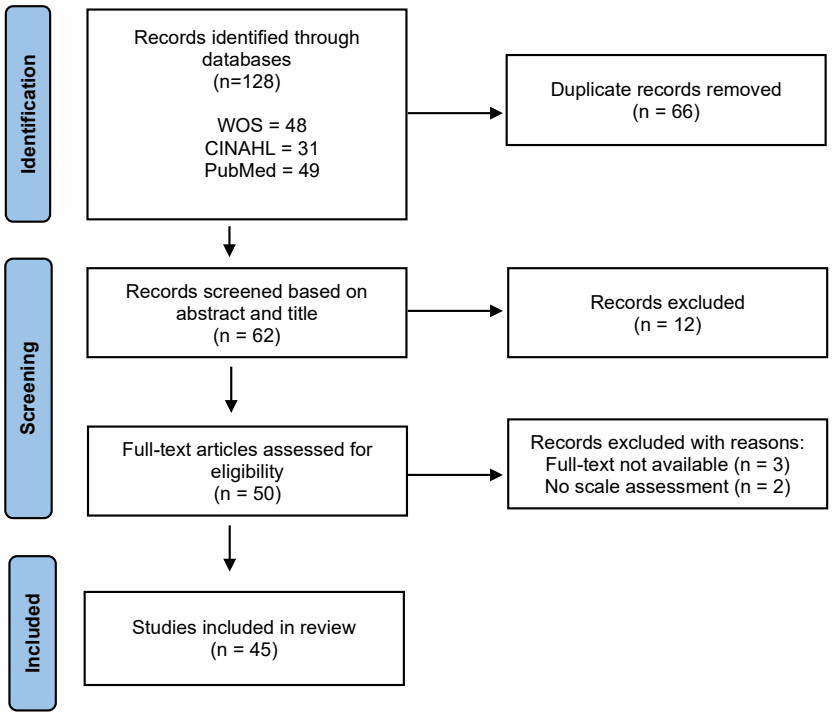


Figure 2.1 PRISMA diagram for selection of studies

2.2.2 Data Extraction

Data extraction was performed in two stages. First, the following information was extracted from all studies: author, country of origin, year of publication, aim of study, study type, sample

population, measurement instrument, instrument items, measurement construct, mode of administration, and scoring system.

The next phase involved the extraction of data related to the Rasch analysis to assess the methodological quality of each article. A checklist of quality indicators based on the guidelines developed by Tennant and Conaghan (2007)[87] on the essential points to be reported in Rasch analysis was utilized. The checklist includes:

1. *Statement of the software used for Rasch analysis*. While the basic premise is to test if the observed response pattern in the data matches the theoretical pattern expected by the model, each software reports findings in a slightly different way.
2. *Specification of the mathematical derivation of the Rasch model*. The dichotomous model is used for scales with two response options. For polytomous items (i.e., three or more response options), the Andrich Rating Scale Model (RSM) or the Master's Partial Credit Model (PCM) can be used. The main difference between these two models is the specification of the distance between thresholds; the RSM expects the distance between thresholds to be equal across all items, whereas the PCM makes no such assumption.
3. *Evaluation of threshold ordering (for polytomous items).* The category structure for polytomous scales must be examined to ensure that the responses to items are consistent with the metric estimate of the underlying construct. That is, the transition from one response category to the next should reflect, and increase with, the underlying latent construct that is being measured. When this does not happen, thresholds are disordered and collapsing of categories may be necessary.
4. *Test of item fit* to the Rasch model. Many of the fit statistics are chi-square based, given the emphasis on the difference between the observed response and that expected by the

model. WINSTEPS[88] software produces “INFIT” and “OUTFIT” statistics, which give slightly different information about the difference between the observed and expected response.[89] Infit statistics are weighted statistics that place more emphasis on unexpected responses near a person’s or item’s measure, whereas outfit statistics place more emphasis on unexpected responses far from a person’s or item’s measure.

5. *Test of person fit* to the Rasch model. Respondents with atypical response patterns may affect the fit at the item level and can be identified by high positive residuals.
6. *Test for differential item functioning (DIF)*, or item bias. DIF occurs when different groups of persons within the sample respond differently to an item, despite equal levels of the underlying latent trait being measured.
7. *Assessment of reliability*. Measures of reliability are provided by Rasch analysis software packages but differ slightly in their interpretation. Whichever software package is used, the values for reliability metrics should be reported and explained.
8. *Assessment of response dependency*. The assumption of local independence implies that once the Rasch factor has been extracted, there should be no remaining patterns in the residuals. Violation of this assumption can be identified by response dependency, which occurs when items are linked in some way such that the response to one item will determine the response to another item. Dependent items can be identified by assessing the residual correlation matrix, which should show no significant associations after taking the latent construct into account.
9. *Assessment of unidimensionality*. The Rasch model assumes items form a unidimensional scale. Rasch software packages provide a principal component analysis of the residuals to detect unidimensionality. The first remaining contrast is an important indicator and

should not be greater than two. If the first contrast has an eigenvalue greater than two, another dimension may exist.

2.3 Results

2.3.1 General Information on Included Studies

A total of 279 papers were identified through database searching. After removing 96 duplicates, 183 unique articles were identified. Of these, 128 articles did not meet inclusion criteria based on abstract and title and were excluded. The full text of 55 articles were assessed, and 45 articles were included in this analysis.

Figure 2.2 shows the increasing trend in the number of studies utilizing Rasch over the past two decades. This is consistent with a recent scientometric review of the trends and evolution of Rasch measurement research conducted by Aryadoust et al. (2019), which identified increased adoption of Rasch measurement between 1972 and 2019, primarily in the fields of medicine, neurology, and psychology.[76] Systematic reviews of Rasch analysis in the fields of

Figure 2.2 Number of published articles reporting application of Rasch within the HIV/AIDS field

rheumatology[90] and nursing[91] have reported similar results, as have scoping reviews of specific instruments, primarily of patient reported outcome measures.[92]

Table 2A.1 (see **Appendix 2A**) presents general characteristics of selected studies including author, publication year, country, primary aim of the study, sample population and size, measurement construct, instrument used, number of items, and scoring system. Most research was conducted among populations in the U.S. (n=19, 42%) and Canada (n=5, 11%). Almost all studies were cross-sectional (n=43, 96%). Two longitudinal studies were conducted; Lerdal et al. (2011) used a longitudinal design to assess DIF by day and time of test administration,[93] and Brouillette et al. (2019) sought to identify if an item-specific practice effect occurred over repeated administrations of the instrument to study participants.[94] Study sample size ranged from as few as 47 in the study conducted by Mayers et al.[95] to as many as 414,407 in the study conducted by Fendrich et al.[96] The contents of the studies and instruments used varied, but primarily focused on quality of life and psychosocial constructs. A variety of different instruments were assessed, but the Lee Fatigue and Energy Scales,[68,93,97,98] the Montreal Cognitive Assessment,[99-101] and the Medical Outcomes Survey[102-104] were all assessed by two or more studies.

2.3.2 Rasch Analyses

The reporting of Rasch parameters as per the Tennant and Conaghan (2007) guidelines is shown in **Table 2A.2** (see **Appendix 2A**). Most studies (n=41, 91%) reported the name of the software used for Rasch analysis. WINSTEPS/Bigsteps was the most commonly reported software program (n=28, 62%), followed by RUMM (n=7, 16%) and R (n=2, 4%). Rascal, jMetrik, Quest, and ACER ConQuest 2.0 were each reported by one study.

Almost one-third of studies (n=14, 31%) did not specify the Rasch model used. The dichotomous model was used in 7 studies (16%), and the multi-facet Rasch model was used in 2 studies (4%). Of the 22 studies (49%) that used a polytomous Rasch model, 13 (59%) used the partial credit model (PCM) and nine (41%) used the Andrich rating scale model (RSM). Analysis of thresholds and category function is important for instruments with polytomous response options. Of the studies using the RSC, PCM, or multi-facet Rasch, 15 (63%) reported on thresholds and category responses, while 9 (38%) made no mention of category functioning. Of the 14 studies that did not specify the Rasch model used, 4 (29%) mentioned analysis of category functioning.

Unidimensionality was tested in 33 (73%) of the studies. Methods used for this purpose were primarily some type of traditional factor analysis (confirmatory/exploratory factor analysis) or Rasch principal components analysis (PCA) of residuals. Notably, six of the 33 (18%) studies reporting on dimensionality presented evidence using both CTT and Rasch methods. The 12 studies (27%) that did not report dimensionality also failed to report local independence. A possible violation of the assumption of local independence was evaluated in only 13 (29%) of the studies.

A total of 41 (91%) studies reported item fit, but only 13 (29%) studies assessed person fit statistics. Nine (20%) studies reported no reliability measure, whereas some metric of either Rasch reliability (e.g., person/item separation and person/item reliability) or CTT reliability (e.g., Cronbach's alpha) was reported in 36 (80%) studies. Eight of these studies (22%) reported a CTT reliability metric only, while sixteen (44%) reported both a CTT and Rasch measure of reliability. Type of Rasch reliability measure (e.g., person separation index, item/person reliability, and item/person separation) depended on the type of software used. Over half (n=26,

58%) of the studies examined some form of DIF, including DIF across gender and disease type, and longitudinal DIF analysis over time.

2.4 Discussion

This study provides a systematic review of the application of Rasch analysis in psychometric assessment studies in the field of HIV/AIDS. CTT, which provides evidence of scale-level properties, has traditionally been the most commonly used approach to assess the psychometric properties of instruments measuring latent constructs. Rasch analysis can provide more detailed information of both item- and person-level characteristics, thereby providing researchers with more specific information to support the development and refinement of a measurement instrument. This review shows that use of the Rasch model in the development of new scales and the validation of existing scales in this field is scant but has increased in recent years.

The reporting of quality metrics, however, is inconsistent. Three principal requirements of the Rasch model that reflect the fundamental properties of measurement should be evaluated when conducting Rasch analysis: unidimensionality, local independence, and model fit. The Rasch model assumes all scale items collectively represent a unidimensional tool measuring a single trait (unidimensionality), and that all correlation between items is accounted for by the latent trait (local independence). That is, once the Rasch factor has been extracted, there should be no remaining patterns in the residuals. Unidimensionality was tested in the majority (73%) of the studies, but the methods used to establish dimensionality varied between traditional methods based in CTT (e.g., exploratory and confirmatory factor analysis, exploratory structural equation modeling) and Rasch PCA of the residuals. Traditional factor analysis identifies proximity of items to the underlying construct, but not their location on it, and is easily confused by ordinal variables and highly correlated factors. Rasch PCA of the residuals, on the other hand, provides

both item and person location on the underlying variable, which is considered superior in constructing linearity out of ordinal variables and facilitates the development of a construct theory.[105] Some studies relied on traditional CTT methods only to determine unidimensionality rather than utilizing the superior Rasch methods. Also concerning was the limited number of studies examining local independence; only 13 (29%) of studies examined response dependence. Notably, the 12 studies that did not report dimensionality also failed to report local independence.

The third requirement concerns the appropriateness of the model to reflect the true relationship between the underlying construct and item responses, which can be examined with item and person fit statistics. Rasch analysis has the advantage of analyzing individual item and person responses. Items that violate one or more of these assumptions can then be combined, rephrased, or deleted to improve the quality of the measurement tool. Similarly, erratic or unexpected responses from participants that may be impacting model fit can be individually assessed to identify potential data entry errors or participant "guessing."[75] While item fit was reported by 91% of the studies, person fit was reported by only 29% of the included studies. Thus, over two-thirds of studies did not take advantage of the ability of Rasch analysis to provide in-depth evaluation of both individual item and person patterns of responses to allow for a more thorough evaluation of the instrument's psychometric properties.

Despite the increased use of the Rasch model in recent years, the inconsistent reporting of Rasch quality metrics in the instrument development and validation process has impeded the analysis of the validity and reliability of the instruments used in these studies. And while Rasch analysis of measurement instruments has several advantages over CTT, including the ability to provide detailed evidence at both the item and person level so as to develop item and person

independent questionnaires, person fit was one of the less frequently reported metrics in the assessed articles.

What may be lacking is the availability of quality analysis tools and guidelines for conducing Rasch analyses. This was apparent in the different methods used to determine unidimensionality, which is a tenant of fundamental measurement, and by the various reliability indicators reported in different studies. A better understanding of Rasch parameters including person/item separation and person/item reliability, within the context of the more commonly known metrics of CTT, is needed. This confusion and lack of consistency may be due in part to the difference in statistical programs used to conduct Rasch analysis and the output that can be obtained from these programs. While this paper used the Tennent and Conaghan[87] quality indicator guidelines, the checklist is still unclear depending on the statistical software used, and not widely adopted. Agreement on the assumptions that should be met (for example, unidimensionality, local independence, model fit, DIF analysis) and the best ways of testing these assumptions will inform the development of uniform guidelines for performing and reporting Rasch analyses and advance the use of Rasch analysis in the field of HIV/AIDS.

2.5 Appendix

Table 2A.1 Study Characteristics

#	Author, Year; Country	Aim	Sample, size	Study Design	Construct	Instrument	Instrument Items	Mode of Administration	Scoring System
1	Askari et al., 2020;[106] Canada	Identify and validate a set of items to form a measure	Development sample: 204 HIV+ adults aged ≥35 years; Validation sample: 703 HIV+ adults aged ≥35 years and a community sample of 484 adults aged >40 with no major health conditions	Psychometric, cross-sectional	Cognition	Communicating Cognitive Concerns Questionnaire (C3Q)	48, reduced to 18 items	Self-report questionnaire	3-point
2	Brouillette et al., 2019;[94] Canada	Estimate the extent of change on repeat administration of three different forms of a brief computerized cognitive assessment battery	102 HIV+ adults aged >18 years	Psychometric, longitudinal	Cognition	Brief Cognitive Ability Measure (B-CAM) V2	6 cognitive tasks	Computerized assessment	0 to 24

3	Brouillette et al., 2015;[99] USA	Test whether the diagnostic accuracy of the MoCA as a screening tool for HIV-Associated neurocognitive disorders might be improved with a stronger scoring methodology	200 HIV+ participants aged 18-65 years	Psychometric, cross-sectional	Cognition	Montreal Cognitive Assessment (MoCa)	30 items	Interviewer administered	0 to 30
4	Brouillette et al., 2015;[107] Canada	Determine the extent to which performance on cognitive tests and self-reported cognitive symptoms form a unidimensional construct	203 HIV+ adults aged 18-65 years	Psychometric, cross-sectional	Cognition	Montreal Cognitive Assessment (MoCa), computerized cognitive tasks, and the Patient Deficit Questionnaire (PDQ)	62, reduced to 37 items (28 performance-based, 9 patient-reported)	Multi-mode (interviewer administered, computerized assessment, and self-report questionnaire)	MoCa: score range 0-30; computerized tasks: reaction time, digit span; PDQ: ordinal, score range from 0-80
5	Chang et al., 2007;[108] USA	Assess the extent to which the RAND-36/SF-36 items measure physical and mental health as implied by existing summary scoring systems	1,714 cancer and HIV/AIDS patients	Psychometric, cross-sectional	QoL (physical and mental health)	SF-36	35 items	Self-report questionnaire	Mixed rating scales with 2-6 points
6	Chang et al., 2014;[109] Taiwan	Use Rasch models to evaluate the psychometric properties of the WHOQOL-BREF for heroin-dependent patients	236 patients with a diagnosis of opioid dependence	Psychometric, cross-sectional	QoL (health-related)	Brief version of World Health Organization Quality of Life assessment (WHOQOL-BREF)	28 items	Self-report questionnaire	5-point

7	Chen et al., 2013;[110] Taiwan	Examine the reliability and validity of the MS-C when applied to three groups diagnosed with major depression, schizophrenia, or HIV/AIDS using both classical test theory and Rasch analysis	237 patients with depressive disorders, 160 with schizophrenia, and 1,612 with HIV/AIDS (n=2,009)	Psychometric, cross-sectional	Sense of mastery	Mastery Scale-Chinese version (MS-C)	7 items	Self-report questionnaire	4-point
8	Conrad et al., 2012;[111] USA	Examine key psychometric properties of the BCS using Rasch measurement model criteria and traditional psychometric strategies	7,435 cases pooled from 77 substance abuse treatment studies conducted in a variety of institutional settings	Psychometric, cross-sectional	Externalizing disorders	Behavioral Complexity Scale (BCS)	33 items	Self-report, face-to-face interview	Binary
9	Fendrich et al., 2009;[96] USA	Develop an HIV transmission risk scale and examine its psychometric properties with Rasch methods	Probability sample of 407 adult MSM in Chicago (Urban Men's Health Study 1997 cohort)	Psychometric, cross-sectional	Sexual risk behavior	HIV transmission risk scale	18 items	Self-report, face-to-face interview	Binary
10	Gay et al., 2016;[112] USA	Evaluate the CES-D's validity and reliability in relation to key demographic and clinical variables in adults with HIV/AIDS	347 adults aged ≥18 years with HIV/AIDS	Psychometric, cross-sectional	Depression	CES-D	20, reduced to 15 items	Self-report questionnaire	Score range from 0-60, with scores of 16 or higher indicating major depression

11	Gunardi et al., 2020;[113] Indonesia	Evaluate the dental profession's experience based on knowledge and attitude toward HIV patients	218 private dentists	Cross-sectional	Knowledge and attitudes toward HIV patients	"a validated questionnaire"	36 items (20 knowledge, 16 attitude statements)	Self-report questionnaire	Yes/no/unsure
12	Hernández et al., 2000;[114] Puerto Rico	Develop a scale to assess patients' satisfaction with pharmacists and determine equivalence of the English and Spanish versions	1,124 patients with cancer or HIV/AIDS	Psychometric, cross-sectional	Patient satisfaction	Satisfaction with Pharmacist (SWiP) scale	7 items	Self-report, face-to-face interview	5-point
13	Hlavinkova et al., 2014;[115] Slovakia	Evaluate results of a questionnaire given to young people before and after the prevention campaign to determine whether the prevention campaign improved HIV/AIDS-related knowledge	533 students aged 15-19 years completed the pre-test and 496 students completed the post-test (analyzed as independent groups)	Cross-sectional	HIV/AIDS knowledge	HIV/AIDS knowledge questionnaire	19 items	Self-report questionnaire	Binary
14	Hu et al., 2019;[116] USA	Provide a psychometric assessment of a resilience scale with a sample of women living with HIV	124 HIV+ women aged 18-63 years	Psychometric, cross sectional	Resilience	Resilience scale (R-PLA)	12, reduced to 11 items	Self-report questionnaire	5-point, recoded as binary
15	Hu et al., 2017;[117] USA	Provide psychometric assessment of an	338 HIV+ MSM aged ≥18 years	Psychometric, cross sectional	Disclosure beliefs	HIV disclosure belief scale (DBS)	14 items	Self-report questionnaire, ACASI	4-point, recoded as binary

		HIV disclosure belief scale for validation purposes							
16	Huang et al., 2022;[118] Myanmar	Adapt the English version of the 12-item CAMS-R and examine its psychometric properties with both CTT and Rasch analysis	159 PLHIV aged ≥18 years	Psychometric, cross-sectional	Mindfulness	Cognitive and Affective Mindfulness Scale-Revised in Myanmar (CAMS-R-M-2)	12, reduced to 9 items	Self-report, online survey	4-point
17	Huang et al., 2021;[119] Myanmar	Adapt the English version of the 40-item Berger's HIV stigma scale and the 7-item Indian HIV stigma scale into a 47-item HIV stigma scale and evaluate the scale's psychometric properties	156 PLWHA aged ≥18 years	Psychometric, cross-sectional	HIV stigma	HIV Stigma Scale in Myanmar (HIVSS-M-1)	47, reduced to 35 items	Self-report, online survey	4-point
18	Huang et al., 2021;[104] Myanmar	Investigate the psychometric properties of the English version of the 19-item MOS-SSS adapted for people living with HIV (PLHIV) in Myanmar	194 PLHIV aged ≥18 years	Psychometric, cross-sectional	Perceived social support	Medical Outcomes Study Social Support Survey Myanmar version (MOS-SSS-M)	19 items	Self-report, online survey	5-point
19	Johansson et al., 2014;[97] Sweden, Norway USA	Evaluate whether the FSS-7 demonstrates similar item hierarchy across people with	224 adults aged ≥18 years with multiple sclerosis, 104 with stroke, and	Psychometric, cross-sectional	Fatigue	Fatigue Severity Scale 7-item version (FSS-7)	7 items	Self-report, face-to-face interview	7-point

		multiple sclerosis, stroke or HIV/AIDS to ensure valid comparisons between groups, and provide further evidence of internal scale validity	316 with HIV/AIDS						
20	Kim et al., 2017;[102] USA	Use the Rasch model to validate the MOS-SSS for incarcerated individuals with HIV infection and compare item responses between men and women	411 incarcerated adults aged ≥18 years enrolled in the EnhanceLink project in 5 urban jails	Psychometric, cross-sectional	Social support	Medical Outcomes Study Social Support Survey (MOS-SSS)	19 items	Self-report, face-to-face interview	5-point
21	Kim and Mazza, 2014;[103] USA	Use the Rasch model to validate the MOS-SSS for incarcerated women with HIV infection	271 incarcerated women aged ≥18 years	Psychometric, cross-sectional	Social support	Medical Outcomes Study Social Support Survey (MOS-SSS)	19 items	Self-report, face-to-face interview	5-point
22	Koski et al., 2011;[100] Canada	Use Rasch analysis to assess the psychometric properties of the MoCA alone, and in conjunction with computerized cognitive tests, as a method of measuring	75 non-demented HIV+ patients aged 18-70 years	Psychometric, cross-sectional	Cognition	Montreal Cognitive Assessment (MoCA), 7 computerized items, 3 noncomputerized tests	24 items	Multi-mode (interviewer administered, computerized assessment)	MoCa: score range 0-5 based on sum of correct responses; computerized tests: 2-3 points

		cognition in HIV-infected patients with mild neurocognitive deficits							
23	Leplège et al., 1997;[120] France	Develop an assessment of the QoL of HIV-infected individuals and provide a preliminary description of its psychometric properties	102 HIV+ patients	Psychometric, cross-sectional	QoL (HIV/AIDS-related)	HIV-QL31	118, reduced to 31 items	Self-report, face-to-face interview	2-point
24	Lerdal et al. (2013);[121] USA	Examine the psychometric properties of the Lee Fatigue and Energy Scales using a Rasch model application	107 women with HIV/AIDS aged ≥18 years	Psychometric, cross-sectional	Fatigue and energy	Lee Fatigue and Energy Scales	18 items (13 fatigue, 5 energy)	Self-report questionnaire	10-point
25	Lerdal et al. (2013);[98] USA	Develop a psychometrically sound short version of the 13-item Lee Fatigue Scale using a Rasch model application	102 women with HIV/AIDS	Psychometric, cross-sectional	Fatigue	Lee Fatigue Scale	13, reduced to 5 items	Self-report questionnaire	10-point
26	Lerdal et al. (2011);[122] USA	Examine the psychometric properties of the 9-item FSS in a sample of HIV-infected adults	316 HIV+ adults aged ≥18 years	Psychometric, longitudinal	Fatigue	9-item Fatigue severity scale (FFS-9); 7-item FFS (FFS-7)	9, reduced to 7 items	Self-report questionnaire	7-point

		and propose a revised version with better psychometric properties							
27	Li et al. (2011);[123] China	Apply the Rasch model on psychometric assessment of an HIV/STI sexual risk scale among men who have sex with men (MSM)	351 MSM aged 18-45 years	Psychometric, cross-sectional	Sexual risk behavior	HIV/STI sexual risk scale	9 items	Self-report, face-to-face interview	2-point
28	Mayers et al. (2002);[95] USA	Describe the development of the initial item pool and evaluate the psychometric properties of the items using methods based on the Rasch model, and examine the convergent and discriminative validities of the ELQ	47 HIV+ women	Psychometric, cross-sectional	Existential loneliness	Existential Loneliness Questionnaire (ELQ)	40, reduced to 22 items	Self-report questionnaire	Mixed rating scales with 4-6 points
29	McCowan and Johnson (1991);[124] USA	Develop a Rasch-scaled measure to ascertain basic HIV disease knowledge	1,092 subjects visiting one of three hospitals in two cities	Psychometric, cross-sectional	HIV knowledge	Basic HIV Disease Knowledge	38, reduced to 13 items	Written, open-ended response	2-point, scored by 2 raters
30	Medeiros et al. (2017);[125] Brazil	Evaluate the internal validity of the Brazilian Food Insecurity (FI) Scale and	796 PLHA aged ≥18 years	Psychometric, cross-sectional	Food insecurity	Brazilian Food Insecurity Scale (EBIA)	14-items for households w/adults and minors; 8 items for	Self-report, face-to-face interview	2-point

		measure the prevalence of FI					adult-only households		
31	Merritt et al. (2013);[126] International database	Gather preliminary evidence of the validity of using the AMPS with person who are living with HIV/AIDS	137 data records with primary or secondary diagnosis of HIV	Psychometric, cross-sectional	Activities of daily living (ADL) motor and process ability	Assessment of Motor and Process Skills (AMPS)	16 ADL motor and 20 ADL process skill items	Rater scored	4-point
32	Mpofu et al., 2006;[127] South Africa	Investigate the presence of a syndrome of health risk behavior in South African adolescents using Rasch analysis	2,186 in-school adolescents aged 12-16 years	Psychometric, cross-sectional	Health risk behavior	South Africa Youth Health Survey (SAYHS)	50 items	Self-report questionnaire	Mixed rating scales with 3-6 points
33	Mugendawal and Hagedorn, 2017;[128] Uganda	Use ordinal logistic and Rasch scaling to predict the strength and direction of the associations between school and community-based initiatives and the perceived risk of HIV infection reported by headteachers	264 headteachers drawn from 264 primary schools in Uganda	Cross-sectional	HIV/AIDS awareness and support	School and community-based HIV/AIDS awareness (SCBHA) and school-based HIV/AIDS support (SBHS)	36 items (SCBHA); 19 items (SBHS)	Self-report questionnaire	2-point

34	Munambah et al., 2021;[129] Zimbabwe	Compare the level of playfulness of children with HIV/AIDS to that of age and gender matched typically developing children living in the same low resourced community	Children aged 4-7 years who are typically developing children (n = 52) and children with HIV/AIDS who receive antiretroviral treatment (n = 44)	Psychometric, cross-sectional	Playfulness	Test of Playfulness (ToP)	29 items	Rater scored	4-point
35	Naar et al., 2021;[130] USA	Systematically develop the MI-CRS and complete an initial evaluation of its psychometric performance, assess the performance of the revised MI-CRS in two ethnically diverse samples, and use an objective standard-setting procedure to define criterion scores for MI competence	(a) recordings (n=99) of intervention sessions in a clinical trial of African American adolescents with obesity; (b) standard patient interactions (n=370) in an implementation science study with youth living with HIV; and (c) standard patient interactions (n=172) in a diverse community sample	Psychometric, cross-sectional	Motivational interviewing	Motivational Interviewing Coach Rating Scale (MI-CRS)	12 items	Rater scored	4-point

36	O'Brien et al., 2021;[131] Canada and Ireland	Develop a short-form version of the HIV Disability Questionnaire (SF-HDQ) to facilitate use in clinical and community-based practice among adults living with HIV	An existing dataset of adults living with HIV in Canada (n = 941) and Ireland (n = 96) who completed the HDQ (n = 1037)	Psychometric, cross-sectional	Disability	HIV Disability Questionnaire (HDQ)	69, reduced to 35 items	Self-report questionnaire	5-point
37	Packham et al., 2017;[132] USA	Use Rasch analysis to assess whether the scoring and measurement properties of painDETECT would support its use as an outcome measure	Data from 624 persons with neuropathic pain	Psychometric, cross-sectional	Neuropathic pain	painDETECT (PD-Q)	9, reduced to 7 items	Self-report questionnaire	6-point
38	Pan et al., 2007;[133] Taiwan	Examine the reliability and validity of the SIS when applied to three groups of individuals diagnosed with major depression, schizophrenia, or HIV/AIDS	580 patients, including 237 patients with depressive disorder, 119 with schizophrenia, and 224 with HIV/AIDS	Psychometric, cross-sectional	Stigmatization	Social Impact Scale (SIS)	24 items	Self-report questionnaire	4-point
39	Peterman et al., 1997;[134] USA	Revise the FAHI QoL instrument to provide more complete and accurate coverage of HIV/AIDS-related QoL	3 sources: 110 patients in the BIOQoL project, 39 HIV+ individuals in a pilot study at a university, and	Psychometric, cross-sectional	QoL (HIV/AIDS-related)	Functional Assessment of Human Immunodeficiency Virus Infection (FAHI) quality of life (QoL) instrument	44 items	Self-report questionnaire	5-point

			212 subjects from a large community-based multi-center HIV ART trial						
40	Pilatti et al., 2014;[135] Argentina	Analyze the psychometric properties of the Spanish Brief YAACQ in a sample of Argentinean college students applying the Item Response Theory	302 undergraduate college students	Psychometric, cross-sectional	Negative consequences of alcohol drinking	Brief Young Adult Alcohol Consequences Questionnaire (B-YAACQ)	24 items	Self-report questionnaire	2-point
41	Rose et al., 2020;[136] USA	Create and validate an instrument to measure mental health professionals' attitudes towards people living with HIV/AIDS	454 mental health professionals	Psychometric, cross-sectional	Attitudes toward people living with HIV/AIDS	Mental Health Professionals' Attitude Towards People Living with HIV/AIDS Scale (MHP-PLHIV-AS)	25 items	Self-report questionnaire	4-point
42	Selman et al., 2012;[137] South Africa and Uganda	Describe the dimensionality of a measure of spiritual well-being (SWB) (the "Spirit 8") in palliative care patients in South Africa and Uganda, and to determine SWB in this population	285 patients receiving specialist palliative care	Psychometric, cross-sectional	Spiritual well-being	Missoula Vitas Quality of Life Index (MVQOLI)	8 items	Self-report, face-to-face interview	5-point

43	Shim et al., 2018;[138] Korea	Examine the psychometric properties of the K-MOS-HIV in a sample of people living with HIV/AIDS	201 outpatients with HIV/AIDS from four teaching hospitals	Psychometric, cross-sectional	QoL (health-related)	Korean version of the Medical Outcomes Study HIV Health Survey (K-MOS-HIV)	35 items	Self-report questionnaire	Mixed rating scales with 2-, 3-, 5-, and 6-points
44	Thandi et al., 2017;[139] UK	Investigate the unidimensional structure of the WSAS across different disease populations	554 patients with HIV, breast cancer, and inflammatory conditions	Psychometric, cross-sectional	Perceived functional impairment	Work and Social Adjustment Scale (WSAS)	5 items	Self-report questionnaire	9-point
45	Yen et al., 2021;[140] Taiwan	Use advanced psychometric testing to evaluate the traditional Chinese version of the MISS-LG	1,000 LGB individuals aged 20-30 years	Psychometric, cross-sectional	Internalized sexual stigma	Measure of Internalized Sexual Stigma for Lesbians and Gay Men (MISS-LG)	17 items	Self-report, face-to-face interview	5-point

Table 2A.2 Rasch Parameters

#	Instrument	Main Goal of Study	Applied Rasch Model	Software	Unidimension-ality Metric	Local Ind.	Item Fit	Person Fit	Reliability Metric	DIF	Rating scale analysis
1	C3Q	Dev/eval of new instrument	Polytomous - PCM	RUMM 2030	PCAR	Y	Y	Y	PSI	Y	Y
2	B-CAM Version 2	Eval of existing instrument	Not specified	Not specified	None	N	N	N	None	Y	N
3	MoCA	Scoring of instrument	Not specified	RUMM 2030	PCAR	Y	Y	N	PSI	Y	Y
4	MoCA, computerized tasks, PDQ	Dev/eval of new instrument	Polytomous - PCM	RUMM 2030	PCAR	N	Y	Y	Cronbach's alpha, PSI	Y	Y
5	SF-36	Eval of existing instrument	Polytomous - RSM	Winsteps	PCAR	N	Y	N	None	N	N
6	WHOQOL-BREF	Eval of existing instrument	Polytomous - RSM	Winsteps	PA	Y	Y	Y	Item and person reliability; item and person separation	Y	Y
7	MS-C	Eval of existing instrument	Polytomous - PCM	Winsteps	PCAR, CFA	N	Y	N	Cronbach's alpha, person separation	Y	Y
8	BCS	Eval of existing instrument	Dichotomous	Winsteps	PCAR	N	Y	N	Cronbach's alpha, person reliability	Y	NA
9	HIV transmission risk scale	Dev/eval of new instrument	Dichotomous	Winsteps	PCAR	Y	Y	Y	Item and person reliability	N	NA
10	CES-D	Comparison of psychometric properties of a single instrument in various patient groups	Polytomous - PCM	Winsteps	PCAR	N	Y	Y	Cronbach's alpha, person separation	Y	Y

11	Dentists' knowledge and attitudes toward HIV patients	Eval of existing instrument	Not specified	Not specified	None (item fit only)	N	Y	N	Cronbach's alpha; item and person reliability; item and person separation	N	NA
12	SWiP	Dev/eval of new instrument	Polytomous - RSM	Bigsteps	None (item fit only)	N	Y	N	Cronbach's alpha	Y	N
13	HIV knowledge and risk estimation	Scoring of instrument	Not specified	ACER ConQuest 2.0	FA	N	N	N	None	N	NA
14	R-PLA	Dev/eval of alt/short form version	Dichotomous	R	FA	Y	Y	Y	Item and person reliability	Y	NA
15	DBS	Dev/eval of new instrument	Dichotomous	R	ESEM	Y	Y	Y	Cronbach's alpha, item and person reliability	Y	NA
16	CAMS-R	Cross-cultural validation	Polytomous - PCM	Winsteps	PCAR, CFA	N	Y	N	Cronbach's alpha; item and person reliability; item and person separation	Y	N
17	Myanmar HIV stigma scale	Cross-cultural validation	Polytomous - RSM	Winsteps	PCAR, EFA	N	Y	N	Cronbach's alpha; item and person reliability; item and person separation	Y	Y
18	MOS-SSS-M	Cross-cultural validation	Polytomous - RSM	Winsteps	PCAR, CFA	N	Y	N	Cronbach's alpha; person reliability and person separation	Y	Y
19	FSS	Comparison of psychometric properties of a single instrument in various patient groups	Polytomous - PCM	Winsteps	None	N	N	N	None	Y	N

20	MOS-SSS	Comparison of psychometric properties of a single instrument in various patient groups	Polytomous - RSM	Winsteps	FA	Y	N	N	Item and person reliability	Y	Y
21	MOS-SSS	Eval of existing instrument	Polytomous - RSM	Winsteps	PCAR, FA	N	Y	N	Cronbach's alpha, item reliability	N	Y
22	MoCA	Eval of existing instrument	Polytomous - PCM	RUMM2020	PCAR	Y	Y	N	Cronbach’s alpha	N	Y
23	HIV-QL31	Dev/eval of new instrument	Dichotomous	Bigsteps	PCAR	Y	Y	N	Cronbach’s alpha	N	NA
24	Lee Fatigue and Energy Scale	Eval of existing instrument	Polytomous - RSM	Winsteps	PCAR	N	Y	N	None	Y	Y
25	Lee Fatigue Scale	Dev/eval of alt/short form version	Polytomous - PCM	Winsteps	PCAR	N	Y	N	Cronbach’s alpha, person separation	Y	N
26	FSS	Dev/eval of alt/short form version	Polytomous - PCM	Winsteps	PCAR	N	Y	Y	Cronbach’s alpha, person separation	Y	Y
27	HIV/STI sexual risk scale	Dev/eval of new instrument	Dichotomous	Winsteps	PCAR	Y	Y	Y	Person and item reliability	Y	NA
28	ELQ	Dev/eval of new instrument	Polytomous - PCM	Quest	None	N	Y	N	Cronbach’s alpha	N	N
29	HIV Knowledge	Dev/eval of new instrument	Not specified	RASCAL	None	N	Y	N	Test-retest	N	NA
30	Brazilian Food insecurity Scale	Eval of existing instrument	Not specified	Winsteps	None	N	Y	N	None	N	NA

31	AMPS	Grader consistency	Multi-facet Rasch	Winsteps - FACETS	None	N	Y	Y	None	N	N
32	SAYHS	Dev/eval of new instrument	Polytomous - PCM	Winsteps	PCAR	N	Y	N	Item and person reliability; item and person separation	Y	N
33	SCBHA and SBHS	Eval of existing instrument	Dichotomous	Not specified	None	N	Y	N	Cronbach's alpha	N	NA
34	ToP	Eval of existing instrument	Not specified	Winsteps	None	N	Y	Y	None	N	N
35	MI-CRS	Dev/eval of new instrument	Multi-facet Rasch	Winsteps - MFRM	PCAR	Y	Y	Y	Cronbach's alpha; person reliability and person separation , coder reliability, coder separation	Y	Y
36	HDQ	Eval of existing instrument	Polytomous - RSM	RUMM2030	PCAR	N	Y	N	Cronbach's alpha, PSI	Y	Y
37	painDETECT	Eval of existing instrument	Not specified	RUMM2030	PCAR	Y	Y	Y	PSI	Y	Y
38	SIS	Comparison of psychometric properties of a single instrument in various patient groups	Not specified	Winsteps	None (item fit only)	N	Y	N	Person reliability, item separation	N	Y
39	FAHI	Eval of existing instrument	Not specified	Bigsteps	FA	N	Y	N	Cronbach's alpha	N	N
40	B-YAACQ	Cross-cultural validation	Not specified	Not specified	None	N	Y	N	Item and person reliability	Y	NA
41	MHP-PLHIV-AS	Dev/eval of new instrument	Not specified	Winsteps	PCAR	N	Y	N	Person reliability and person separation	N	Y

42	MVQOLI	Dev/eval of alt/short form version	Not specified	RUMM2020	FA	Y	Y	N	Cronbach's alpha, person reliability	N	N
43	K-MOS-HIV	Cross-cultural validation	Polytomous - PCM	jMetrik	ESEM	N	Y	N	Cronbach’s alpha	N	Y
44	WSAS	Comparison of psychometric properties of a single instrument in various patient groups	Not specified	Winsteps	PCAR, FA	N	Y	N	Cronbach's alpha; item and person reliability; item and person separation	Y	N
45	MISS-LG	Cross-cultural validation	Polytomous - PCM	Winsteps	FA	N	Y	N	None	N	N

Chapter 3: Application of the Rasch Model to the Psychometric Assessment of Intersectional Stigma Measures

3.1 Introduction

Despite continued advances in HIV prevention and treatment, gay and bisexual men and other men who have sex with men (MSM) remain more affected by the HIV epidemic than any other group, accounting for 70% of new HIV infections in the United States (US) in 2019.[141] Black MSM (BMSM), in particular, are disproportionally affected. Although estimated to represent less than 1% of the US population,[2] BMSM accounted for 25% of all new HIV diagnoses, and 38% of diagnoses among MSM in 2019.[1,3] In addition, epidemiologic studies and meta-analyses have consistently documented sexual orientation disparities in mental health, with MSM reporting higher overall prevalence of mental health disorders such as depression and anxiety, substance abuse, and deliberate self-harm, as compared to heterosexual men.[142-145]

Stigma is a complex social phenomenon that may contribute to health disparities for MSM. It can include labeling, stereotyping, separation into groups representing "us" and "them," status loss, and discrimination, all operating within the context of differential social, economic, and political power that enables one group to devalue another.[13] Stigma has been operationalized by three fundamental mechanisms: enacted stigma, or past experiences of prejudice, stereotyping, and discrimination; anticipated stigma, or expectations of experiencing prejudice, stereotyping, and discrimination in the future; and internalized stigma, or personally adopting negative beliefs and feelings associated with societal stigma.[80] Enacted, anticipated, and internalized stigma

reflect the ways in which individuals respond to stigma, and how stigma operates to negatively impact health and perpetuate disparities in health outcomes.[146]

Stigma can be based on multiple different aspects of an individual's actual or perceived identity or group membership, and a substantial body of literature has highlighted the detrimental health impacts of HIV-related stigma, racism, and homophobia on several mental and physical health outcomes.[14-17] This stigmatization process, borne from interlocking systems of power and oppression, is referred to as intersectional stigma.[37] It is now widely recognized that multiple stigmatized attributes result in unique experiences that supersede the experience of any one stigmatized attribute. Although measures of the three aforementioned stigma mechanisms—enacted, anticipated and internalized stigma—have been developed, they have not addressed multiple co-occurring stigmatized attributes and have not encompassed the concept of intersectionality.

BMSM face stressors associated with both racism and heterosexism. BMSM have been observed to have higher levels of internalized homophobia and rarer disclosure of homosexual orientation than white MSM.[147] Greater internalized homophobia has been associated with lower awareness of HIV-prevention services and with fewer changes in the perception of one's ability to use condoms.[148] Higher levels of psychosocial distress and gay identity have also been associated with higher levels of sexual risk-taking among Black gay and bisexual men.[149]

While there is recognition of the importance of stigma to health outcomes, the complexity of the construct has led to difficulties and disagreements about how to define and measure stigma.[25,29,77,78] These challenges have hindered our ability to adequately assess and address the role of stigma as a social determinant of health, including its impact on the significantly heightened risk and suboptimal HIV care and treatment outcomes observed in key populations.

And while substantial attention has focused on the assessment of HIV-related stigma, limited measurement research has been conducted regarding other intersecting stigmas such as those related to sexual orientation and race/ethnicity.

Psychometric methods: classical test theory vs. Rasch analysis

To better understand stigma among BMSM, a psychometrically sound measure is essential. Numerous scales have been developed to measure stigma,[77,80,150-152] but there is no consensus or "gold" standard. Researchers frequently rely on such scales, indices, and other quantitative instruments comprised of ordinal-level items on a Likert scale to derive variables to quantify respondents' knowledge, abilities, attitudes, and other latent traits that cannot be directly observed. Foundational to the validity of study conclusions is the validity of the measurement of these constructs. The psychometric concepts of reliability and validity are often assessed within the framework of classical test theory (CTT). Typically, researchers sum or average the responses of all items to form a composite summary score that is intended to represent the person's "level" of the latent trait, and then use these scores as the foundational unit of statistical analyses.

There are several errors, however, in analyzing raw scores. First, it assumes that each item contributes equally to the final score, irrespective of how much each item correlates with the underlying construct. In addition, it assumes each item is measured on the same interval scale, and that the "psychological distance" between each response option from an item is assumed to be the same.[82] However, rating scales are ordinal, not interval or continuous, so the calculation of such scores is not only unjustified but also difficult to interpret. Unjustified sum scoring can have adverse effects on the validity and reliability of the measurement, and depending on how the

researcher creates a summary score from the scale items, the use of these scores in subsequent statistical analyses can lead to very different conclusions, even if derived from the same data.[64,65]

Rasch Analysis

To overcome these measurement challenges, the Rasch model has been increasingly used in the development and validation of instruments that measure latent traits. Rasch analysis is the formal testing of a rating scale against a mathematical measurement model that operationalizes the formal axioms of additive conjoint measurement, which underpin the fundamental construct of measurement. The measurement model assumes that person-level responses to an individual item estimate their actual position on the underlying continuum of the latent construct, and that their position on the latent construct should be estimable by only their responses to each individual item. The Rasch model is prescriptive, and shows what should be expected in the responses to items if interval scale measurement is to be achieved, which takes the form of probabilistic Guttman scaling.[66] The Rasch model relaxes the deterministic pattern of Guttman scaling, which expects a strict hierarchical ordering of items, to assume if a harder task is affirmed, then there is a high probability that easier tasks will also be affirmed. If the empirical data fit the model, the probability of a given respondent affirming an item is a logistic function of the relative distance between the item location and the respondent location on a linear scale.[87] For the dichotomous Rasch model, the probability (P) of success for the person n on item i is a function of the difference between the person's ability (B) and the item's difficulty (D), or:

$$(B_n - D_i) = ln\left(\frac{P_{\mathrm{ni}}}{1-P_{\mathrm{ni}}}\right).$$

To extend this model to analyze items with more than two response categories, a threshold parameter (τ) is included to represent the equal probability point between any two adjacent

response options within an item. P_{nik} is now the probability of person *n* affirming category *k* of item *i,* and τ_k is the item's difficulty level of the *k* threshold, or:

$$(B_n - D_i - \tau_{\mathrm{k}}) = ln\left(\frac{P_{nik}}{1-P_{nik}}\right).$$

Rasch analysis provides a unified approach to the testing of several measurement challenges that are required for the validity of the transformation from ordinal to interval scaling. These requirements are 1) testing the internal construct validity of the scale for unidimensionality, which is the basic concept in scientific measurement that only one attribute of an objects is measured at a time and is required for a valid raw summed score, 2) testing the invariance of items, which refers to the maintenance of the value of item and person parameters across all appropriate contexts and is required for interval-level scaling, and 3) appropriate category ordering, which indicates if polytomous response categories are working as expected.[87]

Utilization of the Rasch model to assess the measurement of intersectional stigma presents an opportunity to improve the validity and reliability of scales developed to measure this construct, and ultimately enhance our understanding of the role of intersectional stigma on health outcomes. The objective of this chapter is to utilize Rasch analysis to evaluate the measurement properties of instruments to measure stigma associated with internalized racism and internalized homophobia among BMSM and propose appropriate modification of the scales to address any concerns that are identified. The specific aims of this chapter are to 1) utilize Rasch analysis to assess selected psychometric properties of the internalized racism and internalized homophobia scales by examining the construct validity and internal consistency reliability of each scale in accordance with the Rasch paradigm, and 2) examine the interval properties of the internalized racism and internalized homophobia scales by examining fit to the Rasch model and the ordering of item thresholds.

3.2 Methods

3.2.1 Study Design and Participants

Cross-sectional data were collected from a pilot sample of HIV-negative BMSM (n=151) recruited through BMSM-specific social media (e.g., Jack'd, Grindr) and BMSM community-serving organizations. Eligibility criteria included being at least 18 years of age, identifying as Black, African American, African, Afro-Caribbean (including multiracial Black identities), identifying as male, having had a male sexual partner in the last 6 months, and being HIV negative. All participants were also assigned male at birth. Data were collected between July 2nd and September 3rd, 2020, using online surveys. The University of Maryland Institutional Review Board granted approval for the study (IRB # 1486118), and all participants provided written informed consent.

3.2.2 Measures

Multi-item scales were used to measure stigma, resilience, self-efficacy, and depression. The items of each scale are listed in **Table 3A.1**.

Stigma

Internalized Racism was measured using a 7-item scale adapted from The Appropriated Racial Oppression Scale.[153] Internalized Homophobia was measured using a modified version of Herek and Glunt's internalized homophobia scale.[154] Participants were asked the extent to which they agreed or disagreed to each statement on a 5-point scale ranging from *strongly disagree* to *strongly agree*.

Self-efficacy

Self-efficacy was measured with the 10-item Generalized Self-efficacy Scale (GSE).[155] Participants were asked the extent to which they agreed or disagreed on a 5-point scale ranging from *strongly disagree* to *strongly agree* to statements such as "I can always manage to solve difficult problems if I try hard enough" and "I can usually handle whatever comes my way."

Resilience

Resilience was measured by 6 items from the Connor-Davidson Resilience scale (CD-RISC).[156] Participants were asked the extent to which they agreed or disagreed on a 5-point scale ranging from *strongly disagree* to *strongly agree* to statements such as "I tend to bounce back quickly after hard times" and "I have a hard time making it through stressful events." Negatively worded items were reverse coded.

Depressive Symptoms

Depressive symptoms were measured with the 9-item Patient Health Questionnaire (PHQ).[157] Participants were asked the frequency with which they had been bothered by symptoms of depression such as feeling sad and having restless sleep in the past two weeks. Responses were indicated on a 4-point scale ranging from *not at all* to *nearly every day*.

3.2.3 Sample Size

Linacre has prepared guidelines for appropriate sample size for Rasch rating scale analysis; he suggests a sample size of only 30 is needed to achieve 99% confidence that no item calibration is more than 1 logit from its stable value.[158] For polytomous scales, at least 10 observations per category are recommended to ensure that responses are appropriately distributed across the categories, and a sample of 50 well-targeted participants is sufficient to obtain useful, stable estimates.[158] With the current study sample of 151, the Rasch analysis can

expect to have an appropriate degree of precision irrespective of the targeting of the group, or of the distribution across the response options of each item.

3.2.4 Statistical Analysis

Rasch analysis to examine how well the observed data fit the expectations of the measurement model was carried out using WINSTEPS[88] software. The steps taken to fit the data to the model followed the recommendations by Tennant and Conaghan.[87] A description of the iterative steps in a Rasch analysis and the interpretation of the parameters from the model are provided in **Table 3A.2.**

Rating Scale Functioning

Rating scale category functioning can be examined by use statistics (i.e., category frequencies and average measures), analysis of thresholds, and category fit. Each category should have at least 10 observations to ensure a sufficient number of responses to estimate stable category threshold values. Observed average measures of the persons in the category should increase monotonically across response categories such that those with higher ability endorse progressively higher categories, whereas those with lower abilities endorse progressive lower categories.[159] Threshold calibrations are the 50:50-point difficulties estimated for choosing one response category over the adjacent category. Step thresholds are examined to identify the classification structure of response levels to polytomous scales, and to ensure the functioning of the rating scale items is consistent with the metric estimate of the underlying construct.[159] That is, the transition from one response level to the next should increase according to the underlying measure. Category probability plots, which show the probability of endorsing any given rating scale category for every agreeability-endorsability difference estimate, are generated to graphically display the category thresholds for the rating scale. Ideally, each response category

should have a distinct peak in the probability curve to confirm that each category is the most probable response category for some portion of the measured variable. Disordered thresholds occur when the categories do not match the intervals on the latent variable and collapsing categories should be considered. The magnitude of the difference between adjacent threshold estimates is also important, as each step should define a distinct position on the underlying variable. It is generally recommended that thresholds should be neither too close (<1.4) nor too far apart (>5) on the logit scale.[66] Fit statistics provide another criterion for assessing the quality of rating scales. Outfit mean squares greater than 2 indicate more misinformation than information.[160] Such categories introduce more noise than meaning into the measurement process and should be considered as candidates for collapsing.

Fit Statistics

The basic premise of Rasch analysis is that the model is predetermined, and persons and items are "fit" to this fixed model, rather than developing a model around the data points. The extent to which actual empirical responses differ from the Rasch-modeled theoretical expectations are summarized in fit statistics expressed either as unstandardized mean-squares or standardized as a Z or *t* distribution with a mean of 0 and standard deviation of 1. These two fit statistics are categorized further into a) *outfit* statistics, which place more emphasis on unexpected responses far from a person's or item's measure, or b) weighted *infit* statistics, which place more emphasis on unexpected responses near a person's or item's measure. Mean-square values less than one indicate less variation than modeled and suggests overfit of the model (i.e., a deterministic or Guttman-style response), while mean-square values greater than one indicate more variation than modeled and suggest underfit of the model (i.e., unpredictable, erratic response or noise). Mean-square values between 0.6 to 1.4 are considered productive for rating

scale measurement, whereas values greater than 2.0 may distort or degrade the measurement system.[161]

Unidimensionality and Local Independence

Unidimensionality refers to the ability of a rating scale to measure a single attribute at a time. The Rasch model assumes all scale items collectively represent a unidimensional tool measuring a single trait (unidimensionality), and that all correlation between items is accounted for by the latent trait (local independence). Rasch statistical software provides standardized person-item differences between the observed data and what is expected by the model for every person's response to every item. A Principal Component Analysis (PCA) of these residuals can be used to evaluate unidimensionality and test the model's assumption of local independence of items. That is, once the Rasch factor has been taken into account, there should be no associations between items other than random associations. Guidelines for assessing unidimensionality via Rasch PCA include variance explained by measures greater than 50%, and unexplained variance in the first contrast less than 2.0 eigenvalues.[162] Residual bivariate correlation between the items less than 0.3 provide evidence that the local independence assumption was met.

Reliability

In the Rasch model, reliability is estimated for both persons and items, and refers to the replicability of person and item ordering along the latent trait. Person reliability provides an estimate of the replicability of person placements that can be expected if a particular sample of persons were given another set of suitable items measuring the same construct. Item reliability provides an estimate of the replicability of item placement within a hierarchy of items along the measured variable if the same items were to be given to another sample of comparable ability.[66]

More usefully, person and item reliabilities can be transformed into person and item separation indices, wherein the reliability is calculated as the number of standard errors of spread of persons and items.[163] Person separation determines the number of distinct person subgroups within the sample, providing an estimate of how many strata or comparative groups can be determined within the sample. Low person separation implies that the instrument may not be sensitive enough to distinguish between high and low performers, and more items may be necessary. A person separation index of 1.5 represents an acceptable level of separation, 2.00 represents a good level of separation, and 3.00 represents an excellent level of separation. Item separation is used to verify the item hierarchy. Low item separation implies the person sample is not large enough to confirm the item difficulty hierarchy (i.e., construct validity) of the instrument. An item separation index value of 1.5 is required for analyzing at the individual level, and 2.5 is required for analysis of groups.[87]

Differential Item Functioning

The measurement requirement of invariance is that the values (measures) attributed to variables by any measurement system should be independent of the particular measurement instrument used (as long as the instrument is appropriate to the purpose). That is, the values should be invariant. Differential item functioning (DIF) occurs when different groups with different demographic characteristics in a sample who are measured at equal levels of response for the latent variable respond differently to the same item. Common-item linking is a procedure to check the invariance of item calibrations derived from splitting the person sample into two. Items with a statistically significant difference in item location across subgroups and a DIF contrast >.64 indicate potentially meaningful DIF.[164]

Analysis of construct validity

Construct validity indicates that a scale designed to measure a particular construct is actually measuring that construct. Convergent validity is a type of construct validity that refers to how closely a new scale is related to other variables and other measures of the same construct to which it should be related. Not only should the construct correlate with related variables, but it should also *not* correlate with dissimilar, unrelated ones (i.e., discriminant validity). This type of validity can be evaluated through the pattern of correlations between the final Rasch-derived measures.

3.3 Results

3.3.1 Description of Study Sample

In total, 151 eligible subjects were recruited and interviewed. Participant demographic characteristics are presented in **Table 3A.3**. Age ranged from 18 to 64 years of age, with a median age category of 25 to 34 years. Over half (62%) of participants received a college degree (undergraduate or graduate), whereas only 2% received less than a high school education. Seventy percent were currently employed full-time. More than half (64%) reported a household income below $60,000. The majority (57%) of the sample identified as gay, with 20% identifying as bisexual, and 11% identifying as heterosexual/straight. Forty-four percent of participants were single, 17% were dating, and 38% were in a relationship.

3.3.2 Internalized Racism Scale

Category function and step thresholds

Category function and threshold ordering for each iteration of the scale (**Table 3A.4**) were examined. Frequencies for all categories were greater than 10, suggesting that there was a

sufficient number of responses to estimate stable category threshold values. For the original 7-item, 5-category scale, average measures increased monotonically from -1.91 to 2.09, indicating that on average, those with higher internalized racism endorsed progressively higher response categories. Item thresholds, however, indicated disordering between the thresholds, as the second threshold (-1.51) was not greater than the first threshold (-0.75). The category probability plot (**Figure 3A.1**) also showed that category 2 was depressed (i.e., it was at no point along the measured variable the most probable response category), indicating that this category was less utilized by the respondents compared to other categories. This threshold disordering was likely due to the smaller observed count for category 2 (n=126) as compared to category 1 (n=315) and category 3 (n=243).[165,166] The outfit and infit mean-square values for each category were less than 2.0, indicating that none of the categories introduced more noise than information, so the analysis proceeded with the 5-category structure to preserve the most data as possible.

Fit of items and persons to the Rasch model

Analysis of the item fit statistics revealed that Item 7 (*White gay/bisexual men have a better community than Black gay/bisexual men*) had the least acceptable goodness of fit. The infit and outfit mean-square values of 1.8 and 1.74, respectively, suggested that this item did not tap into the same underlying construct as the other items. In addition, the item pathway or bubble plot (**Figure 3A.2**), visually confirmed that Item 7 did not fall along the same pathway as the other items. Upon further investigation, it was apparent that this was the only item referring to characteristics of the community, rather than characteristics of the individual. This item was therefore removed, and the analysis was repeated with the remaining six items, all of which demonstrated infit and outfit statistics within the acceptable range for items of 0.6 to 1.4.[161]

Person misfit refers to the degree to which a person does not act in a predictable manner regarding the difficulty ordering and spacing of items. As we expect persons to be less "well-behaved" than items, a person outfit mean-square with a cutoff of 2.0 was used to evaluate data quality in this analysis. Review of person fit for the 6-item, 5-category scale revealed a total of 11 misfitting respondents. These respondents were temporarily removed from the analysis (i.e., these observations were given a weight of 0) to obtain item difficulty estimates unaffected by the misfitting data. The difficulty calibrations and fit statistics for each set of items and category structure analyzed are presented in **Table 3A.5**. The item-person Wright maps, which display the relative person and item placement along the latent trait continuum, for the final 6-item, 5-category scale are displayed in **Figure 3A.3**, and the operating range of item difficulties are presented in **Figure 3A.22a**.

Unidimensionality and local independence

A principal components analysis (PCA) of the standardized residuals **(Table 3A.6**) was conducted to detect any substantial and meaningful variance that was not captured by the primary measure. The PCA indicated that 68.7% of the variance was explained by the final 6-item, 5-category internalized racism measure, which is greater than the guideline of 50% for assessing unidimensionality via Rasch PCA. In addition, the unexplained variance in the first contrast was only 1.34 eigenvalue units or 7.0% of the variance. Residual correlations were also assessed to identify dependent items (**Table 3A.7**). No item pair had an absolute value of correlation greater than 0.3, which suggested that the local independence requirement was met.

Person and Item Reliability

Person reliability can be interpreted similarly to the more traditional reliability indices in classical test theory (i.e., Cronbach's alpha), meaning that values closer to one indicate a more

internally consistent measure. The final 6-item, 5-category scale (**Table 3A.5**) had a person reliability of 0.91, meaning that 91% of the variance observed in person measures was systematic, and 9% was due to random error. The level of person separation (3.13) implied that the measure was sensitive enough to distinguish between those with high and low levels of stigma. Item reliability and item separation were slightly higher (0.94 and 4.01, respectively), indicating that the sample was large enough to precisely locate the items on the latent variable, thereby verifying the item hierarchy.

DIF

To assess DIF, item difficulty estimates were estimated separately for each sample subgroup, and then compared by plotting the estimates against each other. The model for invariance of item estimates is represented by a straight line with a slope of one through the mean item difficulty estimates from each sample, and control lines show which items do not display invariance within the boundaries of measurement error across the two person samples.[163] No DIF items were detected across age, education, or relationship status in split-sample analyses (**Figure 3A.4**). Item 6 (*Money management is something that people of my racial group cannot do*) showed a statistically significant, "slight to moderate"[164] difference by income (DIF contrast=.61, p=0.02). The magnitude of this difference, however, was below the suggested threshold of 0.64 to indicate a meaningful effect size (a |DIF contrast| >0.64 indicates moderate to large DIF).

3.3.3 Internalized Homophobia Scale

Category function and step thresholds

Category and threshold statistics are presented in **Table 3A.9**. The 5-category rating scale functioned as anticipated: all category frequencies were greater than 10, average measures increased monotonically from -1.48 to 1.50, and there was no disordering of item thresholds. The

distance between thresholds, however, was only 0.06 logits between the first and second thresholds, and 0.62 logits between the second and third thresholds, both of which were below the suggested ideal distance of at least 1.4 logits. The category probability plot (**Figure 3A.5**) confirmed that categories 2 and 3 spanned only a very small portion of the underlying latent variable. Collapsing categories, however, did not substantially improve category function, and degraded other fit and reliability indices (**Table 3A.10**). As the outfit and infit mean-square values for each category were less than 2.0, the 5-category structure was retained despite the close spacing of the thresholds.

Fit of items and persons to the Rasch model

Analysis of the fit statistics for the nine items of the internalized homophobia scale revealed that Item 9 (*Black straight men have a better community than Black gay/bisexual men*) had the least acceptable goodness of fit, with an infit and outfit mean-square value of 2.12 and 2.32, respectively. The bubble plot (**Figure 3A.6**) visually confirmed that Item 9 was the most misfitting. This item, like Item 7 in the internalized racism scale, appeared to be measuring a different underlying construct than the other items, as it referred to the community rather than the individual. This item was removed, and the analysis was repeated with the remaining eight items, all of which demonstrated infit and outfit statistics within the acceptable range for items.

Review of person fit revealed a total of 15 misfitting respondents with outfit mean-square values greater than 2.0. These respondents were temporarily removed from the analysis to obtain item difficulty estimates unaffected by the misfitting data. The difficulty calibrations and fit statistics for each set of items and category structure analyzed are presented in **Table 3A.10**. The item-person Wright maps for the final 8-item, 5-category scale are presented in **Figure 3A.7**, and the operational range of item difficulties are presented in **Figure 3A.22b**.

Unidimensionality and local independence

The Rasch PCA of the standardized residuals (**Table 3A.11**) indicated that the 8-item, 5-category measure accounted for 60.4% of the variance, and the eigenvalue of the first contrast was 1.76, which was below the guideline of 2.0. This supported the assumption of unidimensionality and suggested that the residuals were merely random noise. Residual correlations were assessed to identify item dependency (**Table 3A.12**). Correlations between all item pairs had an absolute value less than 0.3 except for Items 2 and 7, and Items 4 and 5, which showed marginal residual correlations of -0.35 and -0.34, respectively. These items were retained in the model, however, because item fit statistics were within the acceptable range, and the items measured distinct points along the underlying continuum.

Person and Item Reliability

The final 8-item, 5-category scale (**Table 3A.10**) had a person reliability of 0.88 and a person separation of 2.72, indicating that the sample reported a range of internalized homophobia sufficiently wide to distinguish between those with high and low levels of stigma. Item reliability and item separation were slightly lower (0.79 and 1.95, respectively), signifying that in this analysis there was better information about persons than items, so person estimates were more reliable. Item estimates, nonetheless, were still acceptable.

DIF

No DIF items were detected across age, education, income, or relationship status in split-sample analyses (**Figure 3A.8**).

3.3.4 Self-efficacy Scale

Category function and step thresholds

Category and threshold statistics are presented in **Table 3A.14**. For the 10-item, 5-category scale, category frequencies ranged from 46 to 591, and threshold calibrations showed no disordering. The observed average measures, however, did not increase monotonically between the *strongly disagree* and the *somewhat disagree* categories. The outfit mean-square value for the *strongly disagree* category was 2.06, which was slightly above the recommended maximum value of 2.0. This misfit was likely due to unpredictable responses among the small percentage (3%) of participants in this sample who chose the *strongly disagree* category. After removing 16 misfitting persons, the fit statistics were within an acceptable range, so the 5-category structure was retained.

Fit of items and persons to the Rasch model

Analysis of the item fit statistics (**Table 3A.15**) and bubble plot (**Figure 3A.10**) for the ten items of the 5-category scale revealed that Item 2 (*If someone opposes me, I can find the means and ways to get what I want*) had the least acceptable goodness of fit, with an outfit mean-square of 1.64. The infit mean-square was at the acceptable maximum value (mean-square=1.41), however, suggesting that this misfit may be due to unexpected responses from misfitting persons. Sixteen persons with outfit mean-square values greater than 2.0 were temporarily removed from the analysis. The resulting outfit mean square for Item 2 decreased 1.44, which is slightly above the recommended maximum of 1.4 for productive measurement. The corresponding standardized outfit statistic was below 2.0, however, so this item was retained in the analysis. The difficulty calibrations and fit statistics for each set of items and category structure analyzed are presented in **Table 3A.15.** The item-person Wright maps for the final 10-item, 5-category scale are presented in **Figure 3A.11**, and the operational range of item difficulties are presented in **Figure 3A.22c**.

Unidimensionality and local independence

The PCA of the standardized residuals (**Table 3A.16**) indicated that the final 10-item, 5-category self-efficacy measure accounted for 56.4% of the variance, which was greater than the guideline of 50% for assessing unidimensionality via Rasch PCA. There was no evidence of a secondary dimension, as the eigenvalue of the first contrast was below 2.0. Residual correlations between item pairs (**Table 3A.17**) suggested that the local independence requirement was sufficiently met.

Person and Item Reliability

The final 10-item, 5-category scale (**Table 3A.15**) had a person reliability of .90, indicating that the sample reported a sufficiently wide range of self-efficacy. This finding was supported by the good level of person separation (2.97), meaning the measure was sensitive enough to distinguish between those with high and low levels of self-efficacy. Item reliability and item separation were slightly lower (0.76 and 1.80, respectively). A larger person sample would have increased item reliability and further confirmed the item difficulty hierarchy.

DIF

No DIF items were detected across relationship status in split-sample analyses (**Figure 3A.12**). Item 1 (*I can always manage to solve difficult problems if I try hard enough*) indicated slight DIF by age (DIF contrast=0.72, p=.01), suggesting that this item defined the trait in a different manner for the older and younger age subgroups. Item 2 (*If someone opposes me, I can find the means and ways to get what I want*) exhibited DIF as a function of both income (DIF contrast=0.82, p=0.00) and education (DIF contrast=0.90, p=0.00). Item 3 (*It is easy for me to stick to my aims and accomplish my goals*) exhibited slight DIF as a function of education (DIF contrast=0.67, p=.01). **Figure 3A.13** depicts two Wright Maps for the final 10-item, 5-category

scale separately by high and low education level to visualize differences in the pattern of items as a function of education.

3.3.5 Resilience Scale

Category function and step thresholds

Category and threshold statistics are presented in **Table 3A.19.** The 6-item, 5-category rating scale generally functioned as anticipated: all category frequencies were greater than 10, average measures increased monotonically, and there was no disordering of item thresholds. The distance between thresholds, however, was only 0.15 logits between the second and third thresholds, which was less than the suggested distance of at least 1.4 logits. The category probability plot (**Figure 3A.14a**) confirmed that category 3 spanned only a small portion of the underlying latent variable and was at no point the most probable response category. All other rating scale diagnostics indicated acceptable functioning, so the analysis proceeded with the 5-category structure to preserve the most data as possible. In addition, the subsequently revised 3-item, 5-category scale showed improved functioning of the rating scale categories and thresholds (**Table 3A.19** and **Figure 3A.14b**).

Unidimensionality and local independence

The PCA of the standardized residuals (**Table 3A.20**) revealed that the variance explained by the 6-item, 5-category resilience scale was only 43.2%, and the eigenvalue of the first contrast was 3.1, suggesting a potential second dimension. The plot of standardized residuals (**Figure 3A.15)** visually confirmed that, based on their loadings on the first contrast, Items 4, 5 and 6 (the negatively worded and reverse-coded items) clustered towards the top of the plot (Cluster 1), separately from Items 1, 2, and 3 at the bottom of the plot (Cluster 3). A weak, negative correlation between the measures on the two clusters (Pearson's r =-0.16) was further evidence

of the presence of two dimensions. The high values observed in the residual correlation matrix also indicated a lack of local independence between items for the 6-item scale (**Table 3A.21**).

The scale was therefore split into two different scales consisting of Items 1-3, and Items 4-6. These two scales were analyzed separately in subsequent analyses. The Rasch PCA of the reduced scales indicated that both 3-item, 5-category measures accounted for a sufficient proportion of the variance to support the unidimensionality of the scales: Items 1-3 accounted for 66.9% of the variance, and Items 4-6 accounted for 62.3% (**Table 3A.22**). Analysis of residual correlations, however, suggested that both scales failed to meet the local independence assumption, as correlations were greater than 0.3 (**Table 3A.21a**).

Fit of items and persons to the Rasch model

Analysis of the item fit statistics for both 3-item, 5-category scales demonstrated improved item fit to the Rasch model, with item infit and outfit statistics within the acceptable range for items of 0.6 to 1.4 (**Table 3A.22**). The item reliability and separation indices for the scale consisting of Items 4-6, however, decreased to 0.0, indicating that this person sample was not large enough to precisely locate the items on the underlying latent variable. Subsequent analyses therefore continued with the scale consisting of Items 1-3 only.

Review of person fit for the 3-item scale revealed a total of 14 misfitting persons with an outfit mean-square values greater than 2.0. These persons were temporarily removed from the analysis to obtain item difficulty estimates unaffected by the misfitting data. The difficulty calibrations and fit statistics for each set of items and category structure analyzed are presented in **Table 3A.22**. The item-person Wright map for the final 3-item, 5-category scale (Items 1-3) are presented in **Figure 3A.16,** and the operational range of item difficulties is presented in **Figure 3A.22d**.

Person and Item Reliability

The final 3-item, 5-category scale (Items 1-3) had a person reliability of 0.89 and person separation of 2.89. Item reliability and item separation were 0.76 and 1.80, respectively.

DIF

No DIF items were detected across age, education, income, or relationship status in split-sample analyses (**Figure 3A.17**).

3.3.6 Patient Health Questionnaire (PHQ) – Depressive Symptoms

Category function and step thresholds

The 9-item, 4-category rating scale functioned as anticipated: all category frequencies were greater than 10, average measures increased monotonically, there was no disordering of item thresholds, and outfit and infit mean-square values for all categories were less than 2.0 (**Table 3A.24)**. Visual review of the category probability plot **(Figure 3A.18**) revealed distinct peaks for each of the response categories, illustrating that each category was at some point the most probable response category for a portion of the measured variable.

Fit of items and persons to the Rasch model

Analysis of the fit statistics for the nine items of the PHQ indicated that all items fit the Rasch model, with infit and outfit statistics within the acceptable range for items of 0.6 to 1.4 (**Table 3A.25**). Person fit was also good: only nine individuals (less than 10% of participants) had an outfit mean-square value greater than 2.0. The difficulty calibrations and fit statistics for the scale are presented in **Table 3A.25**. The bubble plot, item-person Wright maps, and the operational range of item difficulties for the final scale are presented in **Figures 3A.19, 3A.20,** and **3A.22**, respectively.

Unidimensionality and local independence

The PCA of the standardized residuals (**Table 3A.26**) indicated that the 9-item, 4-category measure accounted for 63.8% of the variance, which is greater than the guideline of 50% for assessing unidimensionality via Rasch PCA. In addition, there was no evidence of a secondary dimension as the first principal component of the residuals was only 1.87 eigenvalue units (or 7.5% of the total variance). Assessment of the residual correlations between all item pairs suggested that the local independence requirement was sufficiently met **(Table 3A.27**).

Person and Item Reliability

The final 9-item, 4-category scale (**Table 3A.25**) had a person reliability of 0.89 and a person separation of 2.82, demonstrating that the sample reported a range of depressive symptoms sufficiently wide to distinguish between those with high and low levels of depression. Item reliability and item separation were of a similar magnitude (0.85 and 2.42, respectively), indicating that the sample was large enough to precisely locate the items on the latent variable, thereby verifying the item hierarchy.

DIF

No DIF items were detected across age or relationship status in split-sample analyses (**Figure 3A.21**). Item 9 (*Thoughts that you would be better off dead, or of hurting yourself*), however, showed statistically significant DIF by education (DIF contrast=1.1, p<.01) and income (DIF contrast=1.19, p<.01). Item 8 (*Moving or speaking so slowly other people could have noticed…*) also showed slight DIF by income (DIF contrast=0.75, p=.02).

3.3.7 Conversion of raw scores to logit person measures

After confirming that the data for each of these scales fit the Rasch model, interval-level person measures were calculated for this sample for use in subsequent parametric analyses. For each scale, each possible raw sum score has a corresponding Rasch person measure and standard

error in the logit unit. The range of possible Rasch measures for each scale are presented in **Tables 3A.8, 13, 18, 23** and **28**.

3.3.8 Construct Validity

To further demonstrate construct validity, a correlation analysis of the final Rasch person-measures was conducted to assess convergent and divergent validity (**Table 3A.29**). As expected, the internalized racism and internalized homophobia stigma measures were positively correlated with each other and with depressive symptoms (as measured by the PHQ), and negatively correlated with self-efficacy and resilience. In addition, self-efficacy and resilience were positively correlated with each other, and negatively correlated with depressive symptoms.

3.4 Discussion

This study used Rasch analysis to evaluate the psychometric properties of scales developed to measure internalized racism, internalized homophobia, self-efficacy, resilience, and depressive symptoms. Previous studies with a similar purpose have relied on CTT to confirm the validity and reliability of the proposed scales, which not only limits the generalizability of results but also may compromise the validity of findings due to the inherent limitations in applying CTT methods to ordinal scales. Thus, the properties of these scales have not always been sufficiently tested despite their widespread use in the field.

The initial internalized racism scale consisted of six items adapted from the Appropriated Racial Oppression Scale and one additional item, "*White gay/bisexual men have a better community than Black gay/bisexual men.*" Preliminary Rasch analysis revealed this item to be misfitting. Infit and outfit statistics were well outside the recommended ranges, which suggested that the item was not part of the same underlying unidimensional trait. Upon further investigation

of the item content and wording, it was clear that this item differed from the other items in its reference to the community, rather than the individual. In addition, it was the only item addressing both sexual and racial minority constructs simultaneously. With this item removed from the analysis, the original 6-items adapted from the Appropriated Racial Oppression scale satisfactorily fit the Rasch model overall on the domain level, and at the item- and person-levels. The Rasch PCA results confirmed the unidimensionality of the scale, and analysis of the residual correlations confirmed the local independence of items. The high item reliability (0.92) and item separation (3.31) indicated the reproducibility of item placement within the hierarchy across other samples with similar stigma levels. In terms of item fit, item-level statistics provided by the Rasch analysis verified item validity and provided further confirmation of unidimensionality. The operating range of item difficulty was from -4.27 logits to 5.88 logits, suggesting that the scale particularly suits respondents with a stigma measure within ± 4 logits, but it may be less precise when measuring respondents with extremely low (or high) internalized racism levels. The mean person measure of this sample (including extreme and non-extreme persons) was -0.69, indicating that the sample reported a slightly lower overall level of stigma than the items measured. A total of fifteen persons (9.9%) received the minimum possible score, and 6 respondents (4%) received the maximum possible score, so their level of internalized racism could not be precisely measured. To adequately assess the level of internalized racism among persons with extremely low (or high) stigma, additional items of lower (and higher) difficulty are needed. No DIF was detected by age, education, or relationship status. Item 6, "*Money management is something that people of my racial group cannot do*," showed statistically significant DIF by income status. It is entirely plausible that among individuals with the same overall level of stigma, this item would not measure internalized racism in the same way for

those with higher income compared to those with lower income. The magnitude of this difference, however, was only moderate, and may have been an artifact of the data. Additional studies are needed to verify this conclusion.

The initial internalized homophobia scale consisted of eight items adapted from the Herek and Glunt scale and one an additional item, "*Black straight men have a better community than Black gay/bisexual men*." Like the misfitting item in the internalized racism scale, this item did not fit the Rasch model, likely due to its focus on the community rather than the individual. This item was removed from the analysis, and the original eight items adapted from Herek and Glunt fit the model to form a unidimensional scale. The Rasch PCA results further confirmed the unidimensionality of the 8-item scale. Analysis of the residuals revealed a marginal residual correlation between Items 4 and 5, and Items 2 and 7. The correlation between Items 4 and 5 could be due to the wording of the items, both of which inquire about sexual attraction, or the proximity of the two items in the questionnaire. Each item, however, marked a different and distinct location along the latent variable, and item fit indices were within the recommended guidelines, so the items were retained. Item reliability (0.73) and item separation (1.62) were within acceptable range, but slightly lower than the internalized racism scale, and the operating range of item difficulties was narrower. A wider item difficulty range or a larger person sample would improve these reliability indices, and further confirm the item difficulty hierarchy (i.e., construct validity) of the scale. As the mean person measure was slightly negative (-0.22), and 14 persons (9%) obtained the minimum score, additional items of lower difficulty would improve both the person and item reliabilities. No DIF was detected across age, education, income, or relationship status, providing evidence of item invariance across these characteristics.

In addition to reviewing the overall performance of the scale, Rasch analysis allows for the comprehensive assessment of the functioning of each item, which is not possible with CTT methods. The benefit of this approach was further demonstrated in the analysis of the self-efficacy scale. Overall summary statistics of the initial 10-item scale suggested adequate fit and reliability of the scale as a whole. Analysis of the individual items, however, revealed that Item 2 ("*If someone opposes me, I can find the means and ways to get what I want*") did not fit with the other nine items. After flagging this item based on the Rasch fit statistics, a detailed assessment of the item wording and the person-response pattern could be conducted to identify possible reasons for misfit. Item 2 also exhibited DIF by income and education, indicating that this item functioned differently by subgroups of these variables. While we retained Item 2 in this analysis for illustrative purposes, future studies seeking to generate person measures for analysis should take additional steps to ensure the item does not impact the measures of respondents.[75]

The ability to conduct detailed assessment of individual items using Rasch analysis revealed concerns regarding the commonly used approach of including positive and negative items to create "balanced" measurement scales. The six resilience items adapted from the Connor-Davidson Resilience scale consisted of three positively worded items (Items 1-3), and three parallel, negatively worded items (Items 4-6) that were recoded for analysis. Negatively phrased items are typically included to check against response bias, social conformity, or dishonesty, and the reversing (or recoding) of the rating scale numbers used to code negatively phrased items is a common technique to facilitate the analysis of the items. Several researchers, however, have warned against creating scales in which some of the items must be flipped, and the potential problems associated with this practice have been documented across a wide range of studies.[66,75,109,167,168] In their text on rating scale analysis, Wright and Masters (1982) concluded

that analyzing 'for' and reverse-scored 'against' items simultaneously does not yield the best quality measures.[163] Review of item fit, the principal components analysis, and residual item-pair correlations all suggested that the initial six resilience items did not reflect a single unidimensional construct; in fact, examination of dimensionality revealed two clear dimensions, with the three positively-worded items clustering to form one group, and the three negatively-worded items forming a different group. A weak, negative correlation between the two sets of items was further evidence of the presence of two dimensions, suggesting that in this scale, "negative" was not necessarily the opposite of "positive," but rather a distinct construct. It is also plausible that the use of negatively worded items caused confusion among respondents and subsequent measurement disturbances. The final 3-item scale functioned substantially better across all fit and reliability indices, and no DIF was identified, thereby supporting the selection of the positively worded reduced scale.

This study is not without limitations. First, although participants were recruited through a variety of BMSM-specific social media and community-serving organizations, this was a convenience sample, which might not make our sample representative of all BMSM in the U.S. Second, while the anonymity of participants was enhanced through the online administration of the survey, these were self-reported measures and subject to potential reporting bias. In addition, while both scales had acceptable person and item reliability, approximately 9% of respondents on both the internalized racism and internalized homophobia scales reported the lowest possible score. This resulted in a small floor effect, skewing the distribution of scores and making it difficult to differentiate among the individuals at that low level. A scale with better targeted items would be necessary to accurately measure the level of stigma for these individuals. No

significant DIF was identified in either stigma scale across age, education, income, or relationship status, providing evidence of item invariance across these characteristics.

In conclusion, this study used advanced psychometric methods to analyze and confirm the validity and reliability of two measures of stigma. Our findings illustrate the utility of Rasch analysis for modeling these stigma mechanisms as unidimensional latent constructs, assessing individual item and rating scale functioning, establishing the hierarchy of internalizing beliefs and attitudes, and improving the reliability and validity of the measures of the magnitude of internalized racism and internalized homophobia yielded from these scales. This study also provides a useful tool to generate interval-level measurements of stigma that can subsequently be used in parametric analyses.

3.5 Appendix: Tables and Figures

Table 3A.1 Scale Items

Scale and Item #		Retained in model	Item Content
IR	1	✓	Black gay/bisexual men should stop complaining about racism.
	2	✓	Men of my racial group are irresponsible.
	3	✓	People of my racial group are lazy.
	4	✓	Most people of my racial group are on welfare.
	5	✓	Most criminals are men of my racial group.
	6	✓	Money management is something that people of my racial group cannot do.
	7	✗	White gay/bisexual men have a better community than Black gay/bisexual men.
IH	1	✓	Sometimes I wish I were not attracted to men.
	2	✓	Sometimes I think that if I were straight, I would probably be happier.
	3	✓	In order to avoid disclosure of my sexual identity, I have tried to stop being attracted to men.
	4	✓	Sometimes I wish I could become more sexually attracted to women.
	5	✓	I feel that being attracted to men is a personal shortcoming for me.
	6	✓	Sometimes, I feel ashamed of my sexual orientation.
	7	✓	I am afraid my family and friends will find out about my sexual orientation.
	8	✓	I try to look masculine in order to avoid other's rejection.
	9	✗	Black straight men have a better community than Black gay/bisexual men.
SE	1	✓	I can always manage to solve difficult problems if I try hard enough.
	2	✓	If someone opposes me, I can find the means and ways to get what I want.
	3	✓	It is easy for me to stick to my aims and accomplish my goals.
	4	✓	I am confident that I could deal efficiently with unexpected events.
	5	✓	Thanks to my resourcefulness, I know how to handle unforeseen situations.
	6	✓	I can solve most problems if I invest the necessary effort.
	7	✓	I can remain calm when facing difficulties because I can rely on my coping abilities.
	8	✓	When I am confronted with a problem, I can usually find several solutions.
	9	✓	If I am in trouble, I can usually think of a solution.
	10	✓	I can usually handle whatever comes my way.
RES	1	✓	I tend to bounce back quickly after hard times.
	2	✓	It does not take me long to recover from a stressful event.
	3	✓	I usually come through difficult times with little trouble.
	4	✗	It is hard for me to snap back when something bad happens,
	5	✗	I tend to take a long time to get over setbacks in my life.
	6	✗	I have a hard time making it through stressful events.
PHQ	1	✓	Little interest or pleasure in doing things.
	2	✓	Feeling down, depressed, or hopeless.
	3	✓	Trouble falling or staying asleep, or sleeping too much.
	4	✓	Feeling tired or having little energy.
	5	✓	Poor appetite or overeating.
	6	✓	Feeling bad about yourself – or that you are a failure or have let yourself or your family down.
	7	✓	Trouble concentrating on things, such as reading the newspaper or watching TV.
	8	✓	Moving or speaking so slowly other people could have noticed. Or the opposite-being so fidgety or restless that you have been moving around a lot more than usual.
	9	✓	Thoughts that you would be better off dead, or of hurting yourself.

Table 3A.2 Iterative Steps to Conducting Rasch Analysis

Fit to the Rasch model	Fit refers to the degree of mismatch between the pattern of actual observed responses and the Rasch-modeled theoretical expectations and is important for both persons and items. Rasch analysis provides several fit indices: Person/Item INFIT ZSTD, Person/Item OUTFIT ZSTD, Person/Item INFIT MNSQ, and Person/Item OUTFIT MNSQ. Boone (2014) recommends first using the item and person OUTFIT MNSQ fit statistics to identify misfitting items and persons because its sensitivity to outliers makes it easier to identify and correct issues of fit. For rating scales, mean squares values between 0.6 and 1.4 suggest reasonable fit of the data to the model, whereas values greater than 2.0 may distort or degrade the measurement system.
Unidimensionality	A requirement of the Rasch model is that a single latent trait is being measured. Linacre (1998) suggests a three-step process to investigate data dimensionality. First, examine any negative-point biserial correlations to identify potentially problematic items; second, diagnose misfitting persons and items using Rasch fit statistics; third, examine dimensionality using Rasch Factor Analysis to identify any common variance among the data that remains unexplained by the primary Rasch measure. A large (usually more than 2.0) eigenvalue of the first contrast suggests a second dimension besides the primary Rasch dimension.
Response dependency	Uniqueness of the information provided by the items is a requirement of the Rasch model. Items with pair-wise residual (after controlling for the latent trait) correlations greater than 0.3 could indicate lack of independence of the responses which inflates the reliability. Solutions include creating a super-item which combines the response options across items or choosing the one item that best suits the testing context.
Discrimination or person-separation	Person separation and item separation indices are additional measures of reliability. Person separation determines the number of distinct person subgroups within the sample, providing an estimate of how many strata or comparative groups can be determined within the sample. Low person separation implies that the instrument may not be sensitive enough to distinguish between high and low performers, and more items may be necessary. A person separation index of 1.5 represents an acceptable level of separation, 2.00 represents a good level of separation, and 3.00 represents an excellent level of separation. Item separation is used to verify the item hierarchy. Low item separation implies the person sample is not large enough to confirm the item difficulty hierarchy (i.e., construct validity) of the instrument. An item separation index value of 1.5 is required for analyzing at the individual level and 2.5 for analysis of groups.
Rating scale functioning: category use	The simplest way to assess category functioning is to examine category use statistics for each response option. Category frequencies indicate how many respondents actually chose any particular response category, summed for each category across all items. A recommended minimum number of responses per category is 10. Average measures are the average of the ability estimates for all persons in the sample who chose a particular response category, calculated as the average across all observations in a category. Average measures should increase monotonically across response categories, indicating that on average, those with higher ability/stronger attitudes endorse progressively higher categories, and those with lower ability/weaker attitudes endorse progressively lower categories. Disordered categories should be considered candidates for collapsing.
Rating scale functioning: threshold calibrations	Threshold calibrations are the 50:50-point difficulties estimated for choosing one response category over the adjacent category.

	Step calibrations should increase monotonically, and between-threshold distances should indicate that each step defines a distinct position on the underlying variable. Guidelines recommend that thresholds should increase by at least 1.4 logits to show empirical distinction between categories, but not more than 5 logits to avoid gaps in the variable. Category probability curves, which show the probability of endorsing any given rating scale category for every agreeability-endorsability difference estimate, graphically display the category thresholds for the rating scale. Each response category should have a distinct peak in the probability curve to confirm that each category is the most probable response category for some portion of the measured variable.
Differential item functioning (DIF)	The items should have the same ordering of difficulty across all people being measured defined by personal factors such as age, education, income, and relationship status in this study. If an item does not act in the same manner for different subgroups, there may be DIF. Items with a statistically significant difference in item location across subgroups and a DIF contrast >.64 indicate potential DIF.

Table 3A.3 Participant Demographics

Variable		n (%)
Age	18 – 24	31 (21)
	25 – 34	65 (43)
	35 – 44	44 (29)
	45+	11 (7)
Received a college degree		93 (62)
Employed full-time		105 (70)
Annual household income	Less than $30,000	39 (26)
	$30,000 - $59,999	58 (38)
	$60,000 or more	54 (36)
Region of the U.S.	Northeast	83 (55)
	West	8 (5)
	Midwest	10 (7)
	South	47 (31)
	Outside the U.S.	3 (2)
Sexual identity	Bisexual	30 (20)
	Gay	86 (57)
	Heterosexual/straight	17 (11)
	Other	18 (12)
Relationship status	Single	67 (44)
	Dating	26 (17)
	In a relationship	58 (38)

Table 3A.4 IR Scale – Category Function and Threshold Ordering

Category Label	Score	Obs. Count	%	Obs. Avg	Ex. Avg	Infit MNSQ	Outfit MNSQ	Andrich Threshold	Category Measure
				7 Item, 5 Category Scale					
Strongly disagree	1	315	30	-1.91	-1.77	0.70	0.74	NONE	(-2.52)
Somewhat disagree	2	126	12	-0.99	-1.18	1.18	1.29	-0.75	-1.33
Neither agree nor disagree	3	243	23	-0.42	-0.50	0.65	0.67	-1.51	-0.35
Somewhat agree	4	216	20	0.78	0.64	1.04	1.11	0.12	1.21
Strongly agree	5	157	15	2.09	2.42	1.88	1.53	2.14	(3.33)
				7 Item, 5 Category Scale, without 3 unexpected responses to Q7					
Strongly disagree	1	314	30	-1.99	-1.85	0.71	0.74	NONE	(-2.57)
Somewhat disagree	2	124	12	-1.08	-1.23	1.05	1.08	-0.81	-1.37
Neither agree nor disagree	3	243	23	-0.43	-0.51	0.68	0.72	-1.56	-0.37
Somewhat agree	4	216	20	0.85	0.69	1.05	1.13	0.14	1.25
Strongly agree	5	157	15	2.18	2.53	1.94	1.55	2.23	3.42
				7 Item, 4 Category Scale					
Strongly/Somewhat disagree	1	441	42	-2.16	-2.07	.94	1.03	NONE	(-2.59)
Neither agree nor disagree	2	243	23	-1.03	-1.16	0.70	0.57	-1.24	-0.91
Somewhat agree	3	216	20	0.33	.15	.93	0.93	-0.45	0.74
Strongly agree	4	157	15	1.67	1.97	1.63	1.48	1.68	(2.87)
				6 Item, 5 Category Scale					
Strongly disagree	1	297	33	-2.26	-2.17	0.78	0.79	NONE	(-2.89)
Somewhat disagree	2	113	12	-1.41	-1.49	1.01	1.02	-1.23	-1.60
Neither agree nor disagree	3	205	23	-0.57	-0.64	0.69	0.72	-1.69	-0.44
Somewhat agree	4	169	19	1.09	.96	1.15	1.14	0.26	1.50
Strongly agree	5	122	13	2.59	2.87	1.89	1.44	2.66	(3.82)
				6 Item, 5 Category Scale, without 11 misfitting persons[1]					
Strongly disagree	***1***	***266***	***32***	***-3.07***	***-2.93***	***0.71***	***0.71***	***NONE***	***(-3.56)***
Somewhat disagree	***2***	***108***	***13***	***-1.96***	***-2.04***	***0.94***	***0.96***	***-2.00***	***-2.14***
Neither agree nor disagree	***3***	***198***	***24***	***-0.75***	***-0.91***	***0.72***	***0.72***	***-2.12***	***-0.66***
Somewhat agree	***4***	***156***	***19***	***1.62***	***1.60***	***1.38***	***1.20***	***0.41***	***2.08***
Strongly agree	***5***	***112***	***13***	***3.82***	***4.07***	***1.46***	***1.30***	***3.72***	***(4.84)***
				6 Item, 4 Category Scale					
Strongly/Somewhat disagree	1	410	45	-2.40	-2.33	0.93	0.95	NONE	(-2.97)
Neither agree nor disagree	2	205	23	-1.27	-1.32	0.75	0.73	-1.71	-1.07
Somewhat agree	3	169	19	0.55	0.39	.96	1.00	-0.35	0.92
Strongly agree	4	122	13	2.03	2.26	1.54	1.39	2.07	(3.23)

[1]Final model

Table 3A.5 IR Scale – Properties at Each Stage of Analysis

Item	Measure	MnSq		Pt. Ms	Reliability		Separation		Variance
		Infit	Outfit	Corr	Person	Item	Person	Item	Explained
Scale 1: 7 items, 5 categories									
1	0.74	1.1	1.22	0.81					
2	-0.13	0.91	0.88	0.79					
3	0.39	0.59	0.61	0.85					
4	0.21	0.71	0.74	0.83	0.87	0.96	2.57	4.67	65%
5	0.10	0.82	0.85	0.81					
6	-0.26	0.83	0.81	0.8					
7	-1.05	1.8*	1.74*	0.62					
Scale 2: 7 items, 5 categories, without 3 unexpected responses to Q7									
1	0.78	1.14	1.27	0.81					
2	-0.12	0.91	0.88	0.80					
3	0.41	0.58	0.60	0.85					
4	0.23	0.72	0.75	0.84	0.87	0.96	2.64	5.0	68.8%
5	0.11	0.86	0.89	0.81					
6	-0.25	0.87	0.86	0.80					
7	-1.17	1.69*	1.59*	0.64					
Scale 3: 6 items, 5 categories									
1	0.70	1.22	1.23	0.84					
2	-0.38	0.98	0.96	0.83					
3	0.27	0.69	0.69	0.87	0.87	0.92	2.58	3.31	68.4%
4	0.04	0.77	0.78	0.87					
5	-0.10	1.02	1.00	0.83					
6	-0.54	1.16	1.10	0.81					
Scale 4: 6 items, 5 categories, without 11 misfitting persons[1]									
1	***1.04***	***0.95***	***0.93***	***0.89***					
2	***-0.47***	***0.98***	***0.99***	***0.86***					
3	***0.21***	***0.73***	***0.77***	***0.89***	***0.91***	***0.94***	***3.13***	***4.01***	***68.7%***
4	***0.18***	***0.67***	***0.69***	***0.90***					
5	***-0.24***	***0.98***	***0.94***	***0.86***					
6	***-0.72***	***1.33***	***1.26***	***0.83***					
Scale 5: 6 items, 4 categories									
1	0.54	1.03	0.94	0.84					
2	-0.32	1.11	1.05	0.84					
3	0.32	0.80	0.75	0.87	0.82	0.81	2.10	2.09	63.4%
4	0.05	0.85	0.83	0.87					
5	-0.06	1.07	1.12	0.84					
6	-0.51	1.13	1.15	0.83					

*Misfitting item

[1]Final model

Table 3A.6 Final IR Scale – Standardized Residual Variance in Eigenvalue Units

	Eigenvalue	Observed		Expected
Total raw variance in observations	19.16	100.0%		100.0%
Raw variance explained by measures	13.16	68.7%		72.5%
Explained by persons	11.77	61.5%		64.8%
Explained by items	1.38	7.2%		7.6%
Raw unexplained variance (total)	6.00	31.3%	(100%)	27.5%
Unexplained variance in 1st contrast	1.34	7.0%	(22.3%)	
Unexplained variance in 2nd contrast	1.29	6.7%	(21.5%)	
Unexplained variance in 3rd contrast	1.22	6.4%	(20.3%)	
Unexplained variance in 4th contrast	1.11	5.8%	(18.5%)	
Unexplained variance in 5th contrast	1.01	5.3%	(16.9%)	

Table 3A.7 Final IR Scale – Residual Correlation Matrix

1	2	3	4	5	6	Item
1	-0.17	-0.15	-0.27	-0.24	-0.19	Q1_BMSMShouldStopComplaining
-0.17	1	-0.13	-0.11	-0.24	-0.30	Q2_MenOfRacialGroupIrresponsible
-0.15	-0.13	1	-0.15	-0.16	-0.26	Q3_PplOfRacialGroupLazy
-0.27	-0.11	-0.15	1	-0.17	-0.15	Q4_PplOfRacialGroupOnWelfare
-0.24	-0.24	-0.16	-0.17	1	-0.22	Q5_MostCriminalsAreFromRacialGroup
-0.19	-0.30	-0.26	-0.15	-0.22	1	Q6_PplOfRacialGroupCantManageMoney

Table 3A.8 Conversion of raw IR composite scores to Rasch person measures

Score	Measure	S.E.	Score	Measure	S.E.	Score	Measure	S.E.
6	-5.2	1.78	**15**	-1.44	0.5	**24**	2.09	0.75
7	-4.08	0.96	**16**	-1.18	0.52	**25**	2.66	0.76
8	-3.43	0.7	**17**	-0.9	0.54	**26**	3.24	0.76
9	-3.02	0.59	**18**	-0.59	0.57	**27**	3.84	0.79
10	-2.7	0.54	**19**	-0.25	0.6	**28**	4.52	0.87
11	-2.43	0.51	**20**	0.13	0.63	**29**	5.46	1.11
12	-2.18	0.5	**21**	0.55	0.67	**30**	6.83	1.89
13	-1.93	0.49	**22**	1.02	0.7			
14	-1.69	0.5	**23**	1.54	0.73			

Table 3A.9 IH Scale – Category Function and Threshold Ordering

Category Label	Score	Obs. Count	%	Obs. Avg	Exp. Avg	Infit MNSQ	Outfit MNSQ	Andrich Threshold	Category Measure
			9 Item, 5 Category Scale						
Strongly disagree	1	331	24	-1.48	-1.42	0.91	1.04	NONE	(-2.28)
Somewhat disagree	2	187	14	-0.71	-0.67	0.73	0.72	-0.74	-1.02
Neither agree nor disagree	3	253	19	0.10	-0.10	0.92	1.26	-0.68	-0.16
Somewhat agree	4	334	25	0.53	0.58	1.08	1.07	-0.06	0.92
Strongly agree	5	254	19	1.50	1.56	1.22	1.15	1.48	(2.72)
			9 Item, 4 Category Scale						
Strongly/Some-what disagree	1	518	38	-1.48	-1.38	0.85	0.92	NONE	(-2.11)
Neither agree nor disagree	2	253	19	-0.38	-0.63	1.15	1.30	-0.61	-0.70
Somewhat agree	3	334	25	0.15	0.19	0.98	0.94	-0.51	0.55
Strongly agree	4	254	19	1.17	1.23	1.18	1.19	1.12	(2.35)
			8 Item, 5 Category Scale						
Strongly disagree	1	307	25	-1.56	-1.57	1.03	1.17	NONE	(-2.52)
Somewhat disagree	2	175	14	-0.93	-0.81	0.72	0.76	-1.07	-1.16
Neither agree nor disagree	3	212	18	0.00	-0.15	0.78	0.89	-0.68	-0.22
Somewhat agree	4	312	26	0.63	0.67	1.13	1.15	-0.16	1.05
Strongly agree	5	202	17	1.88	1.89	1.10	1.06	1.90	(3.09)
		8 Item, 5 Category Scale, without 15 misfitting persons[1]							
Strongly disagree	***1***	***254***	***23***	***-2.03***	***-2.04***	***1.03***	***1.03***	***NONE***	***(-3.01)***
Somewhat disagree	***2***	***169***	***16***	***-1.16***	***-1.01***	***0.75***	***0.80***	***-1.67***	***-1.43***
Neither agree nor disagree	***3***	***203***	***19***	***0.02***	***-0.17***	***0.83***	***0.91***	***-0.77***	***-0.26***
Somewhat agree	***4***	***289***	***27***	***0.88***	***0.91***	***1.12***	***1.13***	***-0.02***	***1.34***
Strongly agree	***5***	***173***	***16***	***2.50***	***2.53***	***1.09***	***1.06***	***2.47***	***(3.63)***
			8 Item, 4 Category Scale						
Strongly/Some-what disagree	1	482	40	-1.65	-1.60	0.93	0.99	NONE	(-2.33)
Neither agree nor disagree	2	212	18	-0.59	-0.70	0.98	1.00	-0.86	-0.84
Somewhat agree	3	312	26	0.20	0.23	1.05	1.04	-0.64	0.62
Strongly agree	4	202	17	1.52	1.52	1.03	1.10	1.50	(2.68)

[1]Final model

Table 3A.10 IH Scale – Properties at Each Stage of Analysis

Item	Measure	MnSq		Pt. Ms Corr	Reliability		Separation		Variance Explained
		Infit	Outfit		Person	Item	Person	Item	
Scale 1: 9 items, 5 categories									
1	-0.19	0.90	0.85	0.79					
2	-0.12	1.06	1.15	0.74					
3	0.11	0.76	0.89	0.79					
4	0.06	0.82	0.90	0.78					
5	0.33	0.75	0.89	0.79	0.84	0.86	2.30	2.47	58.2%
6	0.09	0.78	0.77	0.79					
7	0.30	0.94	0.85	0.78					
8	0.02	0.91	0.90	0.77					
9	-0.61	2.12*	2.32*	0.60					
Scale 2: 9 items, 5 categories, without 7 unexpected responses to Q9									
1	-0.19	0.94	0.94	0.78					
2	-0.12	1.10	1.20	0.75					
3	0.12	0.79	0.92	0.80					
4	0.07	0.86	0.94	0.79					
5	0.35	0.79	0.97	0.79	0.85	0.89	2.40	2.86	58.4%
6	0.10	0.81	0.80	0.80					
7	0.33	0.98	0.89	0.78					
8	0.03	0.95	0.94	0.78					
9	-0.69	1.80*	1.65*	0.66					
Scale 3: 9 items, 4 categories (11234)									
1	-0.29	0.81	0.83	0.78					
2	-0.09	1.07	1.09	0.72					
3	0.14	0.80	0.84	0.77					
4	0.11	0.86	0.98	0.76					
5	0.39	0.78	0.78	0.77	0.80	0.84	1.98	2.31	51.6%
6	0.13	0.75	0.76	0.78					
7	0.26	0.90	0.86	0.76					
8	0.05	0.88	0.81	0.77					
9	-0.69	2.24*	2.54*	0.55					
Scale 4: 8 items, 5 categories									
1	-0.31	0.98	0.97	0.80					
2	-0.23	1.29	1.41	0.75					
3	0.04	0.85	0.96	0.81					
4	-0.02	0.93	0.99	0.80	0.85	0.73	2.36	1.62	60.4%
5	0.30	0.81	0.95	0.81					
6	0.02	0.91	0.92	0.81					
7	0.27	1.11	1.02	0.79					
8	-0.06	1.02	0.97	0.80					
Scale 5: 8 items, 5 categories, without 15 misfitting persons[1]									
1	***-0.32***	***0.98***	***0.96***	***0.84***					
2	***-0.41***	***1.19***	***1.29***	***0.80***					
3	***-0.04***	***0.80***	***0.79***	***0.86***					
4	***-0.02***	***1.01***	***1.09***	***0.83***	***0.88***	***0.79***	***2.72***	***1.95***	***60.4%***
5	***0.41***	***0.72***	***0.79***	***0.85***					
6	***0.06***	***0.98***	***0.98***	***0.83***					
7	***0.37***	***1.20***	***1.12***	***0.82***					
8	***-0.04***	***0.97***	***0.94***	***0.84***					
Scale 6: 8 items, 4 categories (11234)									
1	-0.46	0.90	0.95	0.82	0.80	0.70	2.02	1.54	54.5%
2	-0.21	1.35	1.48*	0.74					

3	0.07	0.90	0.93	0.81					
4	0.03	0.98	1.07	0.79					
5	0.37	0.86	0.88	0.80					
6	0.05	0.88	0.94	0.81					
7	0.21	1.08	1.11	0.78					
8	-0.04	0.98	0.89	0.80					
Scale 7: 8 items, 4 categories (12334)									
1	-0.27	0.98	0.97	0.83	0.87	0.65	2.64	1.36	60.6%
2	-0.38	1.26	1.38	0.74					
3	0.01	0.85	0.88	0.84					
4	-0.07	0.90	0.84	0.82					
5	0.38	0.86	0.98	0.84					
6	0.08	0.83	0.73	0.84					
7	0.34	1.20	1.05	0.82					
8	-0.09	0.96	0.89	0.82					
Scale 8: 8 items, 4 categories (12234)									
1	-0.47	0.92	0.99	0.83	0.86	0.67	2.52	1.42	56.9%
2	-0.19	1.30	1.33	0.75					
3	0.00	0.86	0.90	0.82					
4	0.06	0.92	0.95	0.81					
5	0.34	0.83	0.90	0.82					
6	0.06	0.89	0.94	0.81					
7	0.31	1.16	1.11	0.79					
8	-0.11	1.05	1.03	0.79					

*Misfitting item
[1]Final model

Table 3A.11 Final IH Scale – Standardized Residual Variance in Eigenvalue Units (8 items, 5 categories)

	Eigenvalue	Observed		Expected
Total raw variance in observations	20.18	100.0%		100.0%
Raw variance explained by measures	12.18	60.4%		64.8%
Explained by persons	11.64	57.7%		61.9%
Explained by items	0.55	2.7%		2.9%
Raw unexplained variance (total)	8.00	39.6%	(100%)	35.2%
Unexplained variance in 1st contrast	1.76	8.7%	(22.0%)	
Unexplained variance in 2nd contrast	1.26	6.3%	(15.8%)	
Unexplained variance in 3rd contrast	1.19	5.9%	(14.9%)	
Unexplained variance in 4th contrast	1.17	5.8%	(14.6%)	
Unexplained variance in 5th contrast	0.99	4.9%	(12.4%)	

Table 3A.12 Final IH Scale – Residual Correlation Matrix

1	2	3	4	5	6	7	8	ITEM
1	-0.06	-0.20	-0.07	-0.07	-0.28	-0.26	-0.05	Q1_WishNotAttractedToMen
-0.06	1	-0.18	0.05	-0.26	-0.26	-0.35	-0.11	Q2_WouldBeHappierIfStraight
-0.20	-0.18	1	-0.19	-0.06	-0.19	-0.02	-0.12	Q3_StopAttractionToAvoidDisclosure
-0.07	0.05	-0.19	1	-0.34	-0.08	-0.20	-0.16	Q4_WishMoreAttractedToWomen
-0.07	-0.26	-0.06	-0.34	1	0.02	0.04	-0.26	Q5_AttractionToMalesIsShortcoming
-0.28	-0.26	-0.19	-0.08	0.02	1	-0.02	-0.11	Q6_AshamedOfSexualOrientation
-0.26	-0.35	-0.02	-0.20	0.04	-0.02	1	-0.18	Q7_AfraidFamilyFriendsWillFindOut
-0.05	-0.11	-0.12	-0.16	-0.26	-0.11	-0.18	1	Q8_TryToLookMasculineToAvoidRejection

Table 3A.13 Conversion of raw IH composite scores to Rasch person measures

Score	Measure	S.E.	Score	Measure	S.E.	Score	Measure	S.E.
8	-4.98	1.83	**19**	-0.97	0.38	**30**	0.85	0.48
9	-3.76	1.01	**20**	-0.83	0.38	**31**	1.09	0.5
10	-3.04	0.73	**21**	-0.68	0.38	**32**	1.35	0.53
11	-2.6	0.6	**22**	-0.54	0.38	**33**	1.64	0.55
12	-2.28	0.53	**23**	-0.4	0.38	**34**	1.96	0.57
13	-2.03	0.48	**24**	-0.25	0.39	**35**	2.3	0.6
14	-1.81	0.45	**25**	-0.09	0.4	**36**	2.68	0.63
15	-1.62	0.43	**26**	0.07	0.41	**37**	3.11	0.69
16	-1.44	0.41	**27**	0.24	0.42	**38**	3.65	0.79
17	-1.28	0.4	**28**	0.42	0.44	**39**	4.47	1.06
18	-1.12	0.39	**29**	0.63	0.46	**40**	5.76	1.86

Table 3A.14 Self-efficacy Scale – Category Function and Threshold Ordering

Category Label	Score	Obs. Count	%	Obs. Avg	Exp. Avg	Infit MNSQ	Outfit MNSQ	Andrich Threshold	Category Measure
				10 Item, 5 Category Scale					
Strongly disagree	1	46	3	-0.20	-0.97	1.75	2.06	NONE	(-3.46)
Somewhat disagree	2	114	8	-0.45*	-0.45	1.16	1.39	-2.20	-1.65
Neither agree nor disagree	3	237	16	0.17	0.35	0.73	0.77	-0.82	-0.23
Somewhat agree	4	522	35	1.75	1.74	.82	0.81	0.21	1.59
Strongly agree	5	591	39	3.36	3.31	1.01	0.97	2.81	(3.96)
				10 Item, 5 Category Scale, without 16 misfitting persons[1]					
Strongly disagree	***1***	***38***	***3***	***-0.74***	***-1.58***	***1.74***	***1.79***	***NONE***	***(-4.11)***
Somewhat disagree	***2***	***92***	***7***	***-1.04****	***-0.88***	***0.93***	***0.98***	***-2.88***	***-2.16***
Neither agree nor disagree	***3***	***222***	***16***	***0.07***	***0.15***	***0.89***	***0.91***	***-1.29***	***-0.41***
Somewhat agree	***4***	***479***	***35***	***2.30***	***2.28***	***0.86***	***0.84***	***0.34***	***2.12***
Strongly agree	***5***	***519***	***38***	***4.47***	***4.45***	***1.11***	***1.01***	***3.83***	***(4.95)***
				10 Item, 4 Category Scale					
Strongly/Somewhat disagree	1	160	11	-1.05	-1.36	1.44	1.62	NONE	(-2.90)
Neither agree nor disagree	2	237	16	-0.66	-0.43	0.74	0.78	-1.60	-1.10
Somewhat agree	3	522	35	1.03	1.04	0.84	0.83	-0.53	0.88
Strongly agree	4	591	39	2.71	2.65	1.00	0.98	2.13	(3.28)
				9 Item, 5 Category Scale					
Strongly disagree	1	41	3	-0.43	-1.05	1.61	1.84	NONE	(-3.56)
Somewhat disagree	2	104	8	-0.52*	-0.53	1.17	1.44	-2.31	-1.70
Neither agree nor disagree	3	199	15	0.04	0.28	0.63	0.62	-0.81	-0.27
Somewhat agree	4	465	34	1.85	1.79	0.77	0.84	0.12	1.64
Strongly agree	5	550	40	3.56	3.55	1.13	1.03	3.00	(4.14)
				9 Item, 5 Category Scale, without 15 misfitting persons					
Strongly disagree	1	32	3	-1.11	-1.79	1.6	1.62	NONE	(-4.45)
Somewhat disagree	2	85	7	-1.20*	-1.05	0.94	0.97	(-3.25)	-2.36
Neither agree nor disagree	3	190	16	-0.03	0.08	0.79	0.78	-1.34	-0.43
Somewhat agree	4	420	34	2.63	2.55	0.83	0.87	0.39	2.32
Strongly agree	5	497	41	4.82	4.86	1.26	1.07	4.20	(5.32)
				9 Item, 4 Category Scale					
Strongly/Somewhat disagree	1	145	11	-1.19	-1.46	1.41	1.61	NONE	(-2.94)
Neither agree nor disagree	2	199	15	-0.83	-0.52	0.65	0.65	-1.63	-1.17
Somewhat agree	3	465	34	1.10	1.04	0.80	0.84	-0.65	0.89
Strongly agree	4	550	40	2.85	2.84	1.10	1.04	2.28	(3.41)

[1]Final model

Table 3A.15 Self-efficacy Scale – Properties at Each Stage of Analysis

Item	Measure	MnSq		Pt. Ms Corr	Reliability		Separation		Variance Explained
		Infit	Outfit		Person	Item	Person	Item	
Scale 1: 10 items, 5 categories									
1	-0.19	1.16	1.00	0.78					
2	0.52	1.41*	1.64*	0.72					
3	0.09	1.18	1.18	0.77					
4	0.28	0.80	0.93	0.81					
5	-0.38	0.86	0.87	0.79	0.86	0.73	2.46	1.63	56.4%
6	-0.07	0.89	0.79	0.80					
7	0.07	1.25	1.32	0.75					
8	-0.03	1.00	1.01	0.76					
9	-0.38	0.62	0.64	0.82					
10	0.09	0.88	0.79	0.80					
Scale 2: 10 items, 5 categories, without 4 unexpected responses to Q2									
1	-0.18	1.20	1.03	0.78					
2	0.38	1.10	1.18	0.77					
3	0.11	1.21	1.22	0.78					
4	0.31	0.84	0.96	0.82					
5	-0.38	0.88	0.90	0.79	0.86	0.67	2.52	1.41	57.3%
6	-0.05	0.91	0.81	0.80					
7	0.09	1.29	1.36	0.75					
8	-0.01	1.04	1.03	0.76					
9	-0.38	0.65	0.69	0.82					
10	0.11	0.92	0.86	0.80					
Scale 3: 10 items, 5 categories, without 16 misfitting persons[1]									
1	***-0.23***	***1.28***	***1.14***	***0.82***					
2	***0.66***	***1.35***	***1.44****	***0.79***					
3	***0.11***	***1.20***	***1.17***	***0.83***					
4	***0.36***	***0.81***	***0.87***	***0.86***	***0.90***	***0.76***	***2.97***	***1.77***	***56.4%***
5	***-0.56***	***0.71***	***0.66***	***0.85***					
6	***-0.15***	***1.02***	***0.90***	***0.83***					
7	***-0.04***	***0.96***	***0.95***	***0.83***					
8	***0.03***	***1.09***	***1.07***	***0.80***					
9	***-0.39***	***0.68***	***0.68***	***0.86***					
10	***0.21***	***0.93***	***0.81***	***0.85***					
Scale 4: 10 items, 4 categories									
1	-0.34	0.94	0.86	0.81					
2	0.61	1.41*	1.58*	0.72					
3	0.12	1.23	1.15	0.78					
4	0.36	0.84	0.93	0.82					
5	-0.39	0.93	0.91	0.79	0.86	0.76	2.47	1.80	56.1%
6	-0.07	0.98	0.89	0.80					
7	0.10	1.25	1.32	0.76					
8	-0.01	1.06	1.06	0.77					
9	-0.41	0.64	0.65	0.83					
10	0.03	0.75	0.72	0.83					
Scale 5: 9 items, 5 categories									
1	-0.14	1.18	1.07	0.79					
3	0.16	1.19	1.21	0.79					
4	0.36	0.92	1.08	0.82	0.86	0.59	2.43	1.21	58.7%
5	-0.34	0.89	0.92	0.80					
6	-0.01	0.88	0.81	0.81					
7	0.14	1.32	1.40	0.76					

8	.003	1.06	1.05	0.77					
9	-0.34	0.65	0.69	0.83					
10	0.16	0.88	0.83	0.82					
Scale 6: 9 items, 5 categories, without 15 misfitting persons									
1	-0.12	1.41	1.28	0.84					
2	0.11	1.09	1.06	0.85					
3	0.35	0.73	0.83	0.89					
4	-0.51	0.75	0.71	0.86					
5	-0.01	0.94	0.86	0.86	0.90	0.54	2.97	1.08	58.7%
6	0.02	1.23	1.17	0.83					
7	0.19	1.12	1.10	0.82					
8	-0.30	0.75	0.75	0.87					
9	0.27	0.93	0.87	0.86					
Scale 7: 9 items, 4 categories									
1	-0.29	0.96	0.93	0.82					
3	0.19	1.24	1.17	0.80					
4	0.45	0.95	1.07	0.82					
5	-0.34	0.97	0.96	0.80					
6	0.0	0.99	0.90	0.81	0.85	0.65	2.42	1.37	58%
7	0.17	1.33	1.42*`	0.77					
8	0.07	1.12	1.10	0.78					
9	-0.36	0.68	0.71	0.84					
10	0.11	0.76	0.76	0.84					

*Misfitting item
[1]Final model

Table 3A.16 Final SE Scale – Standardized Residual Variance in Eigenvalue Units (10 items, 5 categories)

	Eigenvalue	Observed		Expected
Total raw variance in observations	22.95	100.0%		100.0%
Raw variance explained by measures	12.95	56.4%		62.9%
Explained by persons	11.11	48.4%		54.0%
Explained by items	1.84	8.0%		8.9%
Raw unexplained variance (total)	10.00	43.6%	(100%)	37.1%
Unexplained variance in 1st contrast	1.85	8.1%	(18.5%)	
Unexplained variance in 2nd contrast	1.52	6.6%	(15.2%)	
Unexplained variance in 3rd contrast	1.30	5.7%	(13.0%)	
Unexplained variance in 4th contrast	1.14	4.9%	(11.4%)	
Unexplained variance in 5th contrast	1.10	4.8%	(11.0%)	

Table 3A.17 Final SE Scale – Residual Correlation Matrix

1	2	3	4	5	6	7	8	9	10	Item
1	-0.20	0.04	-0.25	-0.21	0.03	-0.16	-0.28	0.05	0.08	Q1
-0.20	1	-0.23	0.06	-0.17	-0.27	-0.12	-0.15	-0.09	-0.22	Q2
0.04	-0.23	1	-0.14	-0.20	-0.17	-0.14	-0.07	-0.16	-0.01	Q3
-0.25	0.06	-0.14	1	-0.04	-0.27	-0.10	0.00	-0.24	-0.08	Q4
-0.21	-0.17	-0.20	-0.04	1	0.16	-0.18	0.08	0.04	-0.29	Q5
0.03	-0.27	-0.17	-0.27	0.16	1	-0.16	-0.16	-0.05	0.11	Q6
-0.16	-0.12	-0.14	-0.10	-0.18	-0.16	1	0.00	-0.21	-0.12	Q7
-0.28	-0.15	-0.07	0.00	0.08	-0.16	0.00	1	-0.09	-0.35	Q8
0.05	-0.09	-0.16	-0.24	0.04	-0.05	-0.21	-0.09	1	0.07	Q9
0.08	-0.22	-0.01	-0.08	-0.29	0.11	-0.12	-0.35	0.07	1	Q10

Table 3A.18 Conversion of raw SE composite scores to Rasch person measures

Score	Measure	S.E.	Score	Measure	S.E.	Score	Measure	S.E.
10	-6.42	1.84	**24**	-1.46	0.42	**38**	1.45	0.56
11	-5.17	1.03	**25**	-1.29	0.42	**39**	1.77	0.58
12	-4.41	0.75	**26**	-1.11	0.42	**40**	2.12	0.6
13	-3.93	0.64	**27**	-0.94	0.42	**41**	2.48	0.6
14	-3.57	0.57	**28**	-0.76	0.42	**42**	2.85	0.6
15	-3.27	0.52	**29**	-0.59	0.42	**43**	3.21	0.6
16	-3.02	0.49	**30**	-0.4	0.43	**44**	3.56	0.6
17	-2.78	0.47	**31**	-0.22	0.44	**45**	3.92	0.61
18	-2.57	0.46	**32**	-0.02	0.45	**46**	4.3	0.63
19	-2.36	0.44	**33**	0.18	0.46	**48**	5.27	0.79
20	-2.17	0.43	**34**	0.4	0.47	**49**	6.08	1.06
21	-1.99	0.43	**35**	0.63	0.49	**50**	7.37	1.86
22	-1.81	0.42	**36**	0.88	0.51			
23	-1.63	0.42	**37**	1.15	0.53			

Table 3A.19 Resilience Scale – Category Function and Threshold Ordering

Category Label	Score	Obs. Count	%	Obs. Avg	Exp. Avg	Infit MNSQ	Outfit MNSQ	Andrich Threshold	Category Measure
				6 Item, 5 Category Scale					
Strongly disagree	1	83	9	-0.53	-0.67	1.20	1.43	NONE	(-2.48)
Somewhat disagree	2	165	18	-0.35	-0.27	0.81	0.75	-1.15	-0.97
Neither agree nor disagree	3	178	20	0.03	0.13	0.88	0.88	-0.15	0.03
Somewhat agree	4	254	28	0.72	0.60	0.78	0.82	0.00	0.94
Strongly agree	5	225	25	1.18	1.24	1.12	1.14	1.30	(2.58)
				3 Item, 5 Category Scale (Items 1-3)[1]					
Strongly disagree	***1***	***20***	***4***	***-2.05***	***-3.04***	***3.26***	***2.40***	***NONE***	***(-4.71)***
Somewhat disagree	***2***	***68***	***15***	***-1.24***	***-1.02***	***0.79***	***0.81***	***-3.58***	***-2.07***
Neither agree nor disagree	***3***	***82***	***18***	***0.21***	***0.47***	***0.66***	***0.62***	***-0.43***	***0.11***
Somewhat agree	***4***	***136***	***30***	***2.17***	***1.96***	***0.69***	***0.74***	***0.69***	***2.08***
Strongly agree	***5***	***147***	***32***	***3.33***	***3.41***	***1.16***	***1.11***	***3.33***	***(4.48)***
				3 Item, 5 Category Scale (Items 4-6rec)					
Strongly disagree	1	63	14	-2.29	-2.78	2.14	1.88	NONE	(-4.32)
Somewhat disagree	2	97	21	-1.46	-1.20	0.69	0.71	-3.17	-1.92
Neither agree nor disagree	3	96	21	0.07	0.11	0.64	0.58	-0.50	-0.01
Somewhat agree	4	118	26	1.46	1.28	0.77	0.76	0.48	1.92
Strongly agree	5	78	17	2.33	2.56	1.41	1.31	3.20	(4.35)

[1]Final model

Table 3A.20 Resilience Scale – Standardized Residual Variance in Eigenvalue Units (6 items, 5 categories)

	Eigenvalue	Observed		Expected
Total raw variance in observations	10.57	100.0%		100.0%
Raw variance explained by measures	4.57	43.2%		42.6%
Explained by persons	3.28	31.1%		30.6%
Explained by items	1.29	12.2%		12.0%
Raw unexplained variance (total)	6.00	56.8%	(100%)	57.4%
Unexplained variance in 1st contrast	3.10	29.3%	(51.6%)	
Unexplained variance in 2nd contrast	1.00	9.5%	(16.7%)	
Unexplained variance in 3rd contrast	.81	7.7%	(13.5%)	
Unexplained variance in 4th contrast	.60	5.7%	(10.0%)	
Unexplained variance in 5th contrast	.48	4.5%	(8.0%)	

Table 3A.21 Residual Correlation Matrix for 6-item, 5-category Resilience Scale

1	2	3	4	5	6	Item
1	0.38	0.28	-0.47	-0.60	-0.52	Q1
0.38	1	0.17	-0.51	-0.48	-0.43	Q2
0.28	0.17	1	-0.46	-0.58	-0.54	Q3
-0.47	-0.51	-0.46	1	0.26	0.08	Q4r
-0.60	-0.48	-0.58	0.26	1	0.48	Q5r
-0.52	-0.43	-0.54	0.08	0.48	1	Q6r

Table 3A.21a. Residual Correlation Matrix for 3-item, 5-category Resilience Scales

Scale	Item	Item	Correlation
Items 1-3	1	2	-0.32
	1	3	-0.49
	2	3	-0.67
Items 4r-6r	4r	5r	-0.54
	4r	6r	-0.72
	5r	6r	-0.20

Table 3A.22 Resilience Scale – Properties at Each Stage of Analysis

Item	Measure	MnSq		Pt. Ms	Reliability		Separation		Variance
		Infit	Outfit	Corr	Person	Item	Person	Item	Explained
Scale 1: 6 items, 5 categories									
1	-0.54	0.90	0.82	0.59					
2	-0.24	0.83	0.79	0.65					
3	-0.25	1.31	1.58*	0.46	0.70	0.94	1.54	3.80	43.2%
4	0.40	1.09	1.26	0.61					
5	0.30	0.82	0.78	0.71					
6	0.33	0.98	0.94	0.67					
Scale 2: 6 items, 5 categories, without 19 misfitting persons									
1	-0.63	0.98	0.90	0.64					
2	-0.21	0.85	0.84	0.71					
3	-0.30	1.21	1.26	0.57	0.79	0.92	1.92	3.41	42.8%
4	0.40	0.88	0.87	0.73					
5	0.34	0.90	0.87	0.74					
6	0.40	1.11	1.09	0.69					
Scale 3: 3 items (1-3), 5 categories									
1	-0.46	0.77	0.77	0.87					
2	0.24	0.98	0.93	0.87	0.77	0.81	1.83	2.04	66.9%
3	0.22	1.19	1.12	0.84					
Scale 4: 3 items (1-3), 5 categories, without 14 misfitting persons[1]									
1	***-0.60***	***0.90***	***0.84***	***0.92***					
2	***0.34***	***0.91***	***0.83***	***0.92***	***0.89***	***0.76***	***2.89***	***1.80***	***66.7%***
3	***0.26***	***1.15***	***1.09***	***0.91***					
Scale 5: 3 items (4-6rec), 5 categories									
4r	0.12	1.31	1.27	0.83					
5r	-0.10	0.66	0.65	0.91	0.81	0.00	2.06	0.00	62.3%
6r	-0.03	1.01	0.95	0.87					

*Misfitting item

[1]Final model

Table 3A. 23 Conversion of raw Resilience composite scores to Rasch person measures

Score	Measure	S.E.	Score	Measure	S.E.
3	-12.53	1.94	**10**	2.34	1.06
4	-10.98	1.25	**11**	3.53	1.16
5	-9.53	1.25	**12**	5.31	1.49
6	-5.32	4.67	**13**	7.15	1.2
7	-1.15	1.21	**14**	8.55	1.24
8	0.09	1.06	**15**	10.08	1.93
9	1.22	1.07			

Table 3A.24 PHQ Scale – Category Function and Threshold Ordering

Category Label	Score	Observed Count	%	Observed Avg	Expected Avg	Infit MNSQ	Outfit MNSQ	Andrich Threshold	Category Measure
9 Item, 4 Category Depression Scale									
Not at all	1	548	40	-2.50	-2.47	1.01	1.03	NONE	(-3.17)
Several days	2	355	26	-1.3	-1.24	0.72	0.69	-1.96	-1.12
More than half the days	3	290	21	0.83	0.58	0.89	0.95	-0.20	1.03
Nearly every day	4	164	12	1.90	2.16	1.47	1.36	2.16	(3.33)

Table 3A.25 PHQ Scale – Properties at Each Stage of Analysis

Item	Measure	MnSq		Pt. Ms	Reliability		Separation		Variance
		Infit	Outfit	Corr	Person	Item	Person	Item	Explained
Scale 1: 9 items, 4 categories									
1	0.02	1.17	1.25	0.79					
2	-0.30	0.73	0.74	0.87					
3	-0.37	1.15	1.18	0.81					
4	-0.35	0.91	0.95	0.84					
5	-0.28	1.27	1.19	0.81	0.86	0.83	2.51	2.21	63.8%
6	-0.14	0.79	0.77	0.86					
7	0.23	0.77	0.79	0.85					
8	0.43	1.04	0.94	0.82					
9	0.75	1.12	0.95	0.80					
Scale 2: 9 items, 4 categories, without 9 misfitting people[1]									
1	***0.10***	***1.13***	***1.15***	***0.83***					
2	***-0.36***	***0.79***	***0.81***	***0.88***					
3	***-0.49***	***1.16***	***1.19***	***0.84***					
4	***-0.36***	***0.95***	***0.99***	***0.86***					
5	***-0.36***	***1.25***	***1.19***	***0.84***	***0.89***	***0.85***	***2.82***	***2.42***	***63.8%***
6	***-0.15***	***0.88***	***0.85***	***0.87***					
7	***0.20***	***0.71***	***0.72***	***0.88***					
8	***0.51***	***0.93***	***0.86***	***0.86***					
9	***0.91***	***1.09***	***0.93***	***0.84***					

[1]Final model

Table 3A.26 Final PHQ Scale – Standardized Residual Variance in Eigenvalue Units (9 items, 4 categories)

	Eigenvalue	Observed		Expected
Total raw variance in observations	24.86	100.0%		100.0%
Raw variance explained by measures	15.86	63.8%		66.9%
Explained by persons	13.12	52.9%		55.5%
Explained by items	2.71	10.9%		11.4%
Raw unexplained variance (total)	9.00	36.2%	(100%)	33.1%
Unexplained variance in 1st contrast	1.87	7.5%	(20.8%)	
Unexplained variance in 2nd contrast	1.54	6.2%	(17.2%)	
Unexplained variance in 3rd contrast	1.37	5.5%	(15.2%)	
Unexplained variance in 4th contrast	1.06	4.3%	(11.7%)	
Unexplained variance in 5th contrast	1.00	4.0%	(11.2%)	

Table 3A.27 Final PHQ Scale – Residual Correlation Matrix

1	2	3	4	5	6	7	8	9	Item
1	-0.20	-0.15	-0.06	-0.19	-0.15	0.08	-0.28	-0.12	Q1
-0.20	1	-0.16	0.15	-0.16	0.12	-0.26	-0.20	-0.11	Q2
-0.15	-0.16	1	-0.11	0.06	-0.19	-0.22	-0.18	-0.26	Q3
-0.06	0.15	-0.11	1	-0.06	-0.17	-0.10	-0.35	-0.29	Q4
-0.19	-0.16	0.06	1.00	1	-0.22	-0.06	-0.23	-0.29	Q5
-0.15	0.12	-0.19	-0.06	-0.22	1	-0.19	-0.01	-0.02	Q6
0.08	-0.26	-0.22	-0.17	-0.06	-0.19	1	0.02	-0.18	Q7
-0.28	-0.20	-0.18	-0.10	-0.23	-0.01	0.02	1	0.31	Q8
-0.12	-0.11	-0.26	-0.35	-0.29	-0.02	-0.18	0.31	1	Q9

Table 3A.28 Conversion of raw PHQ composite scores to Rasch person measures

Score	Measure	S.E.	Score	Measure	S.E.	Score	Measure	S.E.
9	-5.94	1.85	**19**	-1.11	0.52	**29**	1.96	0.59
10	-4.66	1.05	**20**	-0.84	0.52	**30**	2.32	0.6
11	-3.86	0.78	**21**	-0.56	0.53	**31**	2.69	0.62
12	-3.34	0.67	**22**	-0.28	0.53	**32**	3.08	0.64
13	-2.93	0.61	**23**	0	0.54	**33**	3.53	0.7
14	-2.58	0.58	**24**	0.3	0.55	**34**	4.08	0.8
15	-2.26	0.56	**25**	0.61	0.56	**35**	4.91	1.06
16	-1.96	0.54	**26**	0.94	0.57	**36**	6.21	1.86
17	-1.67	0.53	**27**	1.27	0.58			
18	-1.39	0.53	**28**	1.61	0.59			

Table 3A.29 Correlation analysis of Rasch-derived person-measures

Scale	IH	IR	SE	Resilience	PHQ
IH	1				
IR	0.6270	1			
SE	-0.1237	-0.1168	1		
Resilience	-0.0590	-0.0039	0.7168	1	
PHQ	0.5715	0.5566	-0.2846	-0.2060	1

Figure 3A.1 IR Scale – Category Probability Plots

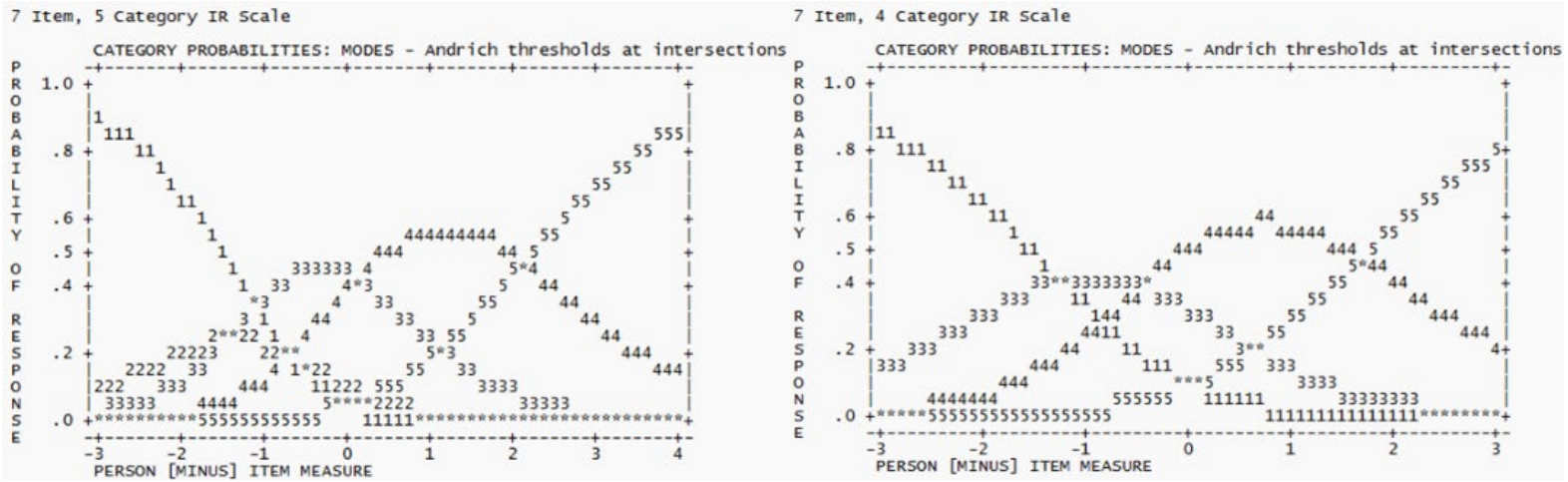


Figure 3A.2 Bubble Plot of IR Items

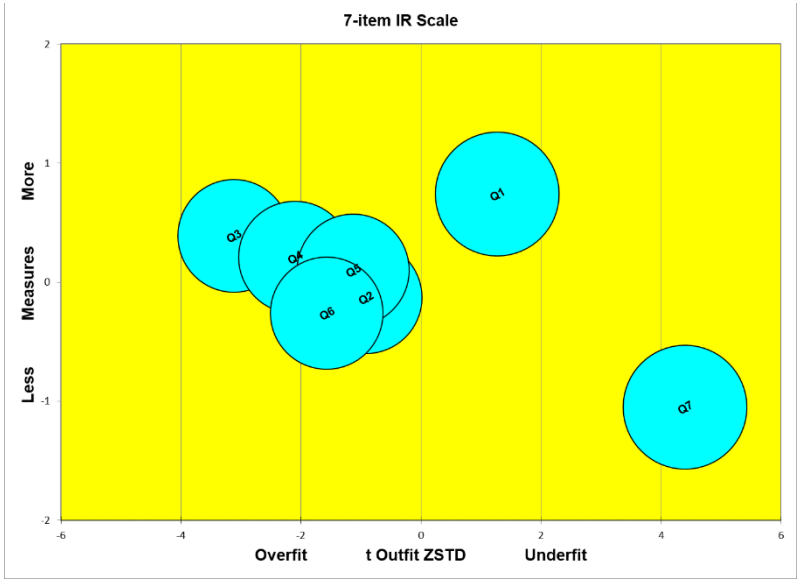
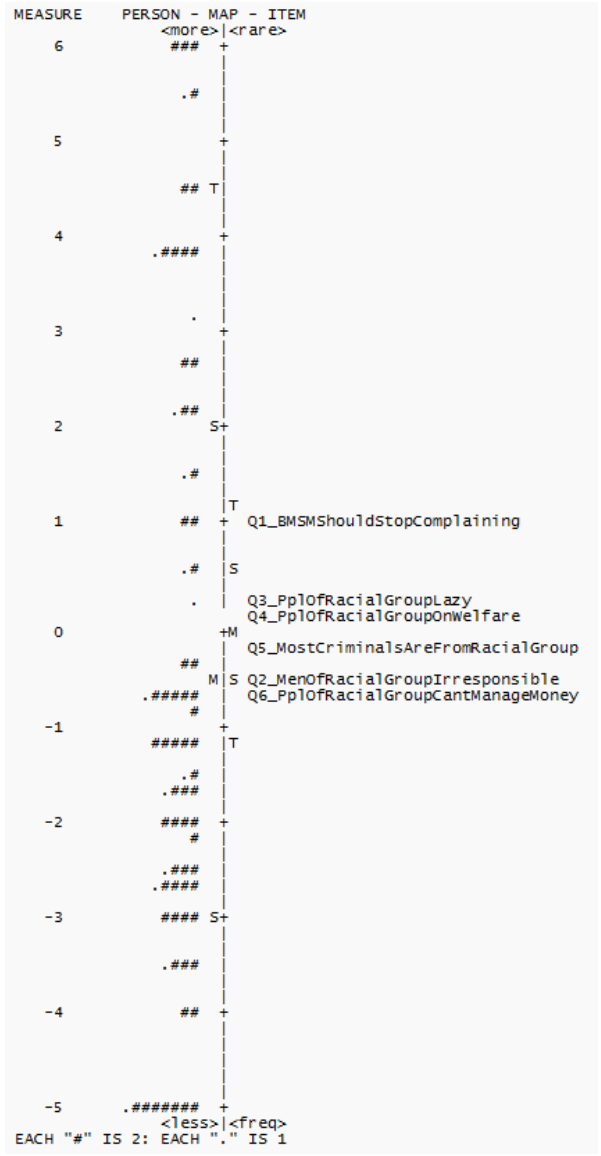


Figure 3A.3 Final IR Scale – Wright Maps



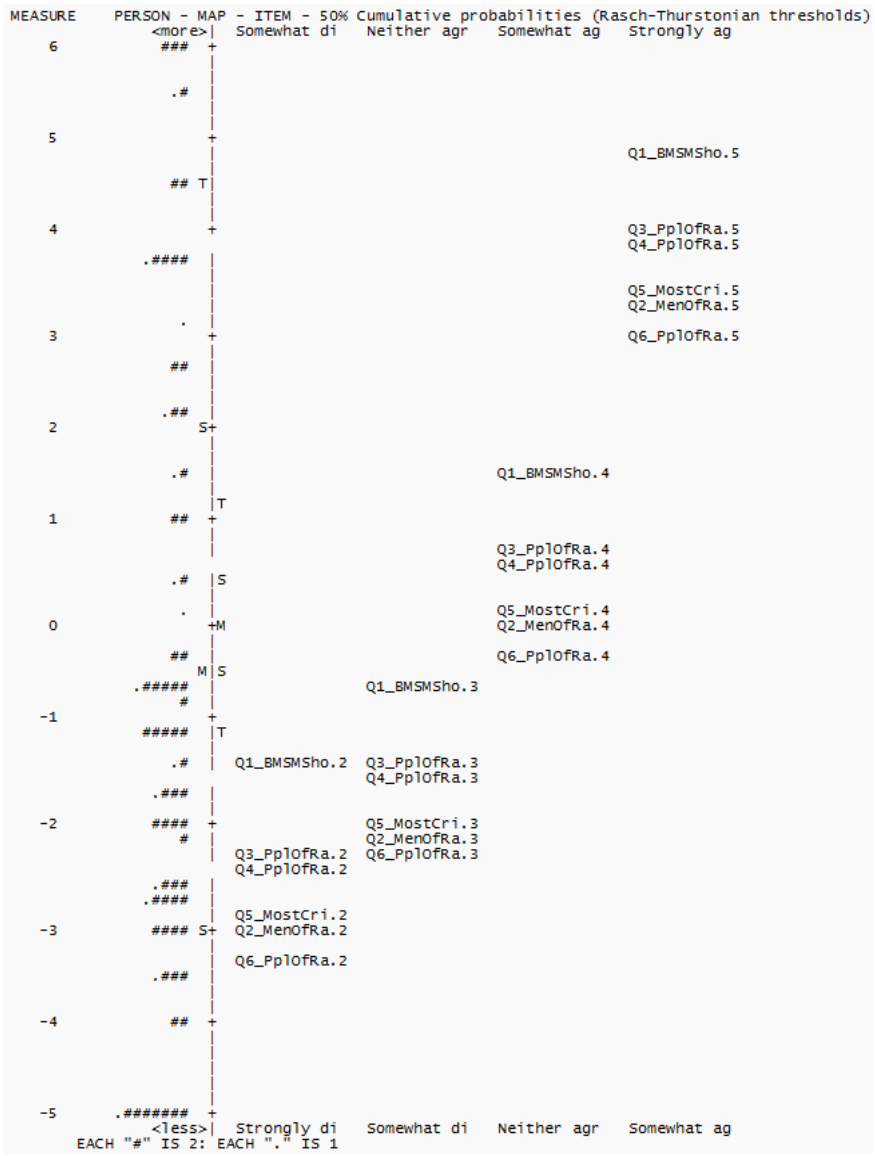


Figure 3A.4 IR Scale – DIF Analyses by age, education, income, and relationship status

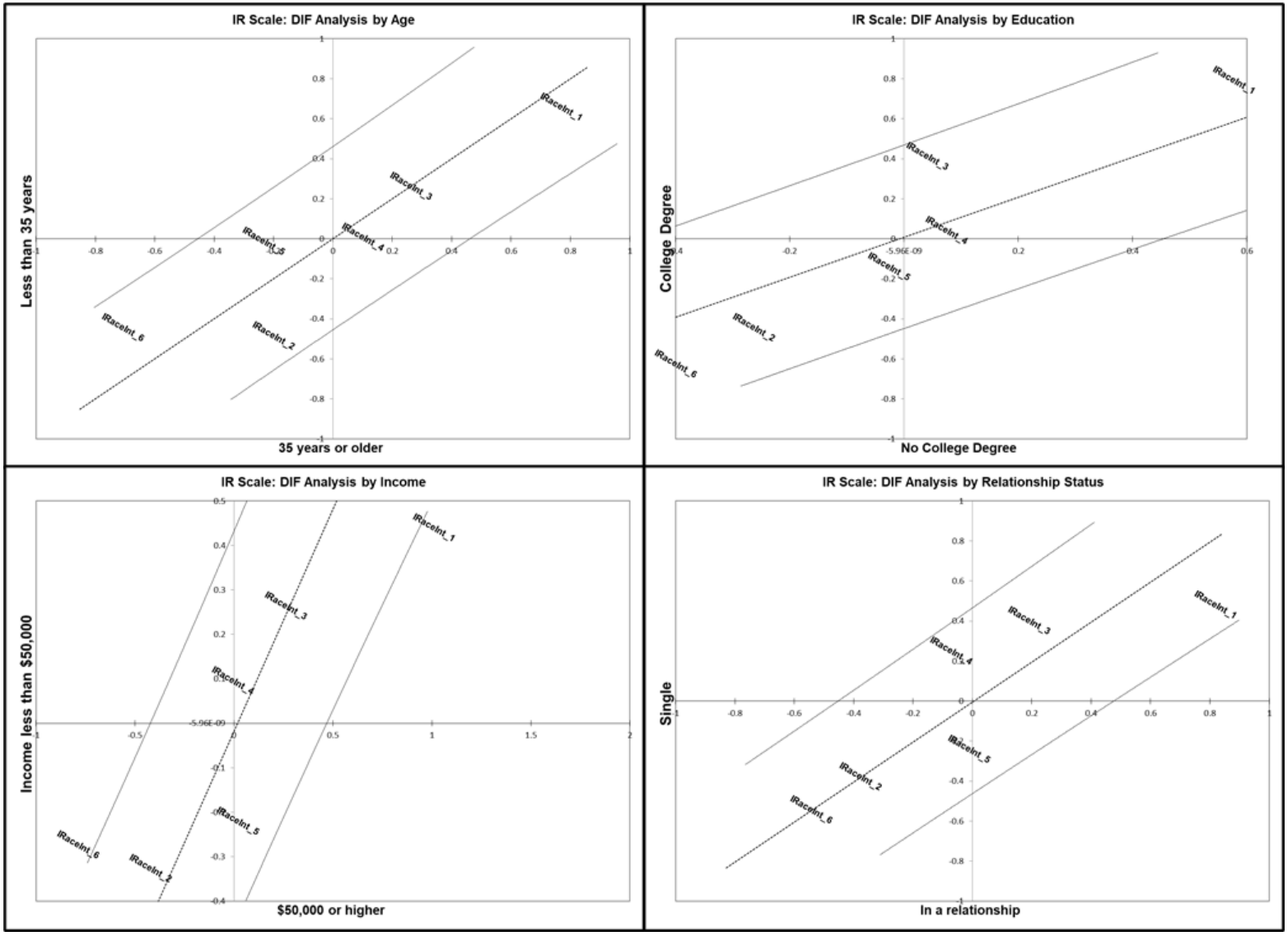


Figure 3A.6 IH Scale – Category probability plots

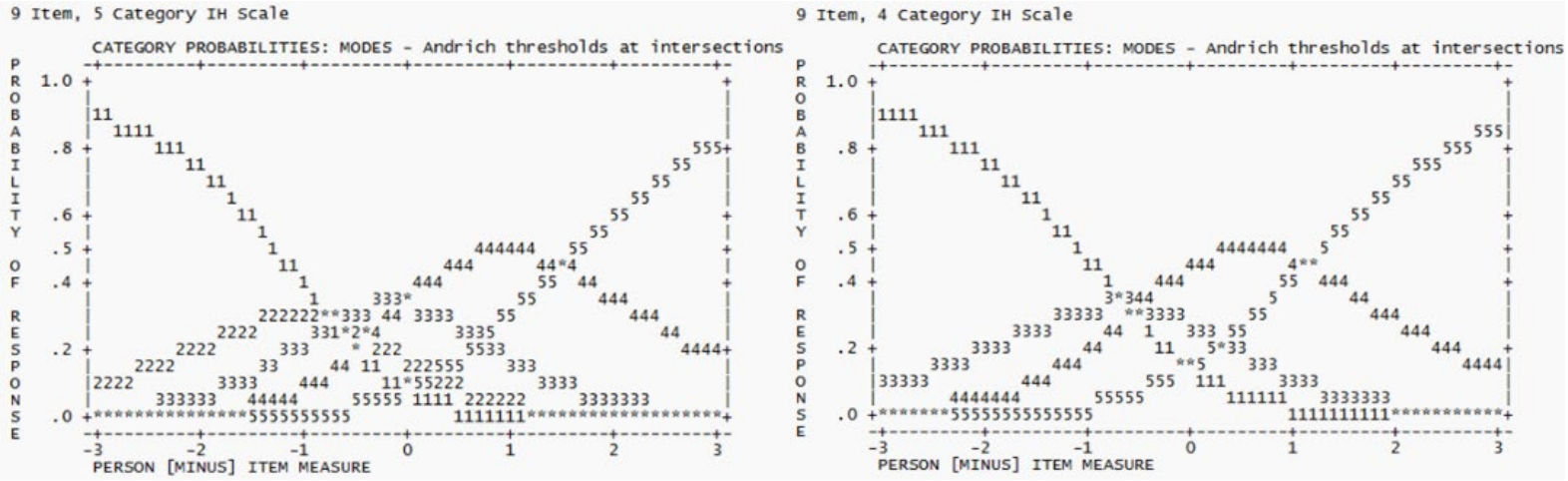


Figure 3A.5 Bubble Plot of IH Items

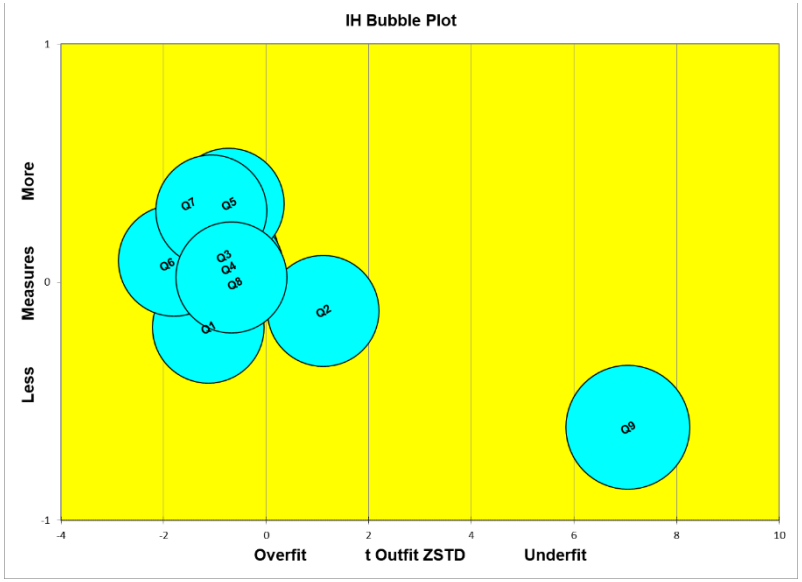


Figure 3A.7 Final IH Scale – Wright Maps

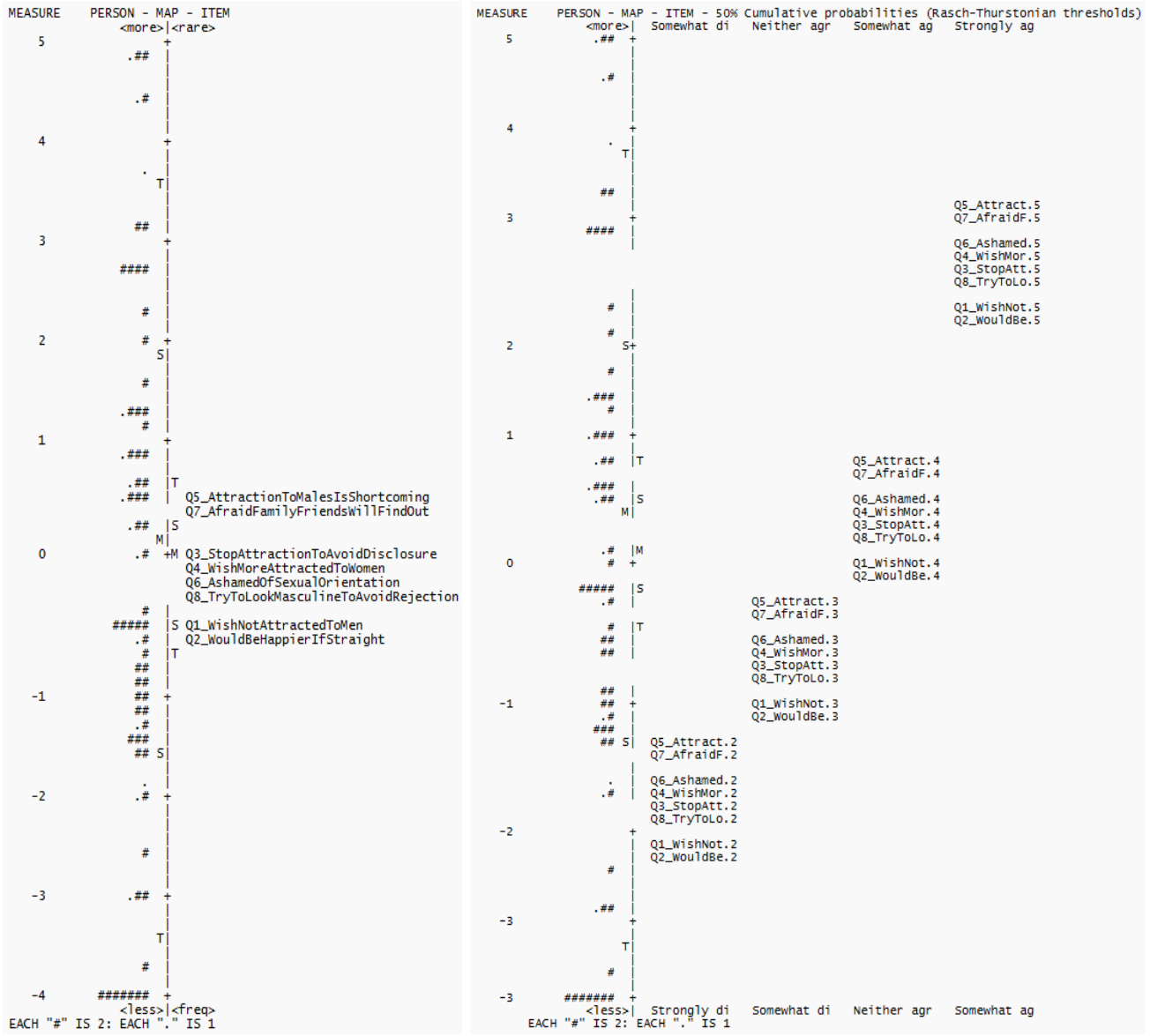


Figure 3A.8 IH Scale – DIF Analyses by age, education, income, and relationship status

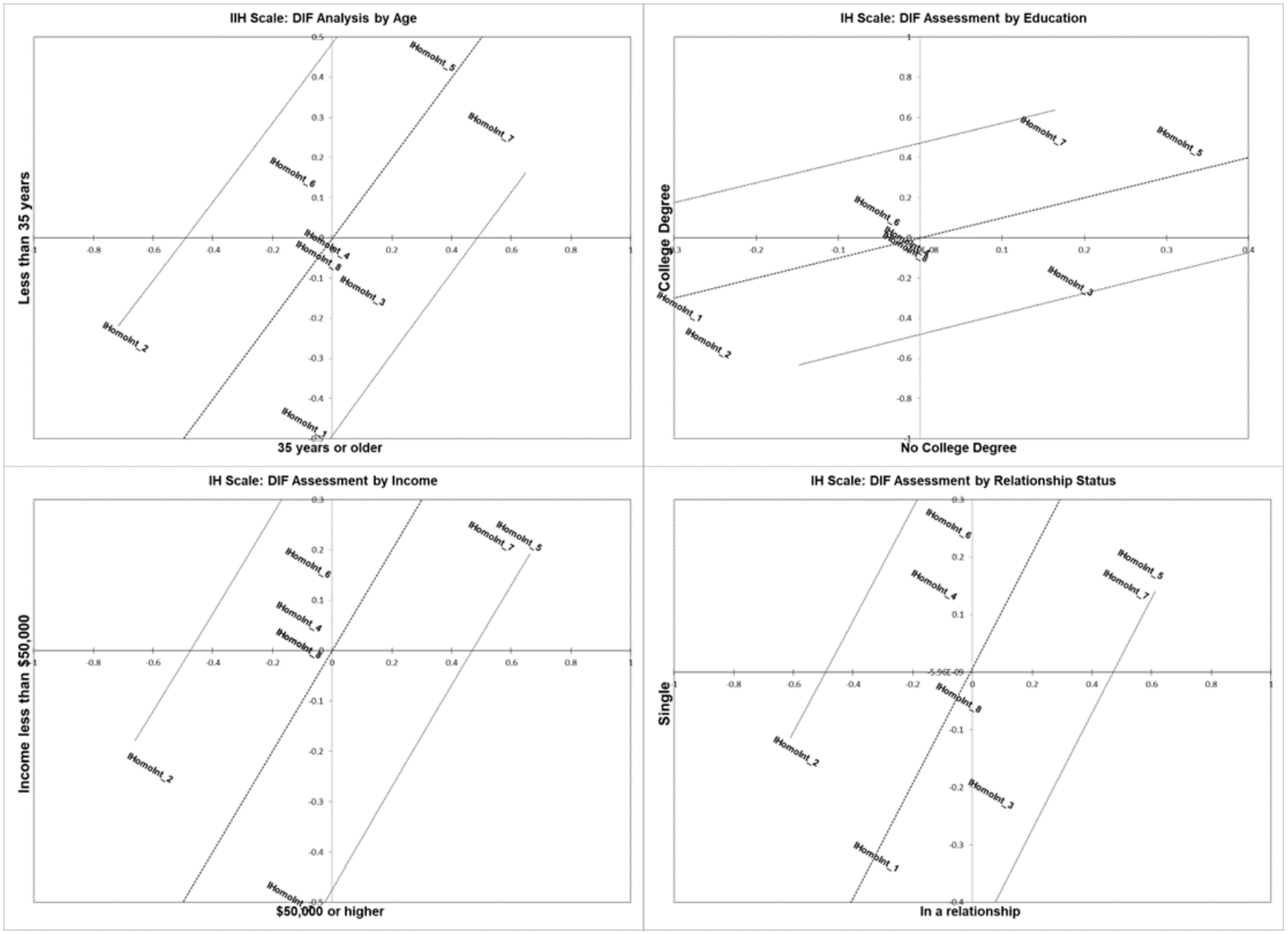


Figure 3A.9 SE Scale – Category probability plots

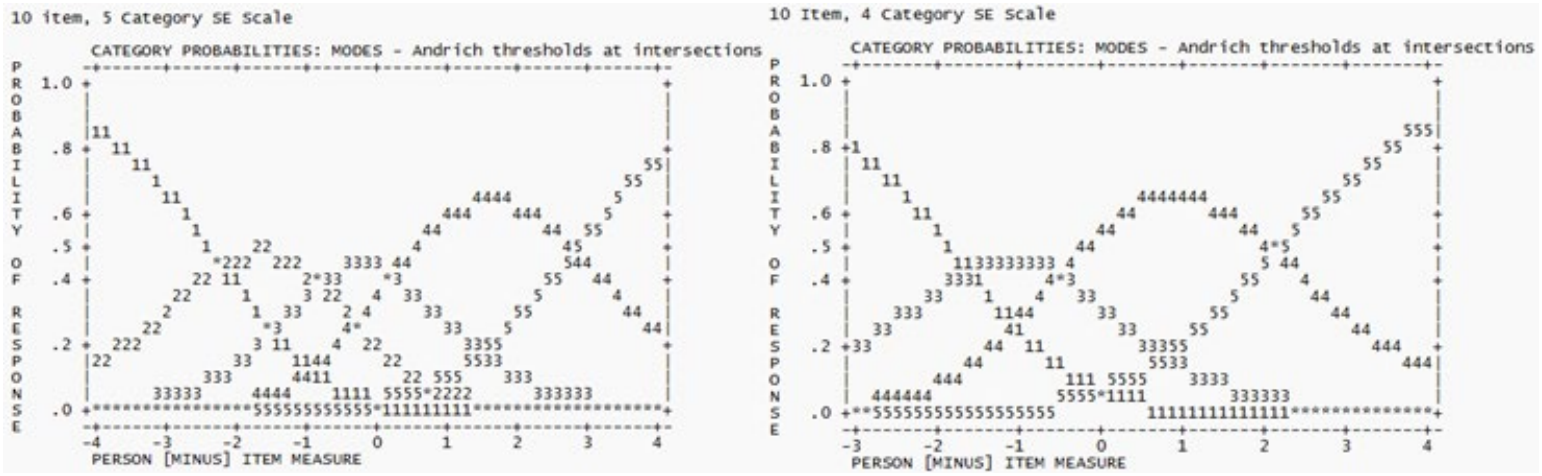


Figure 3A.10 Bubble Plot of SE Items

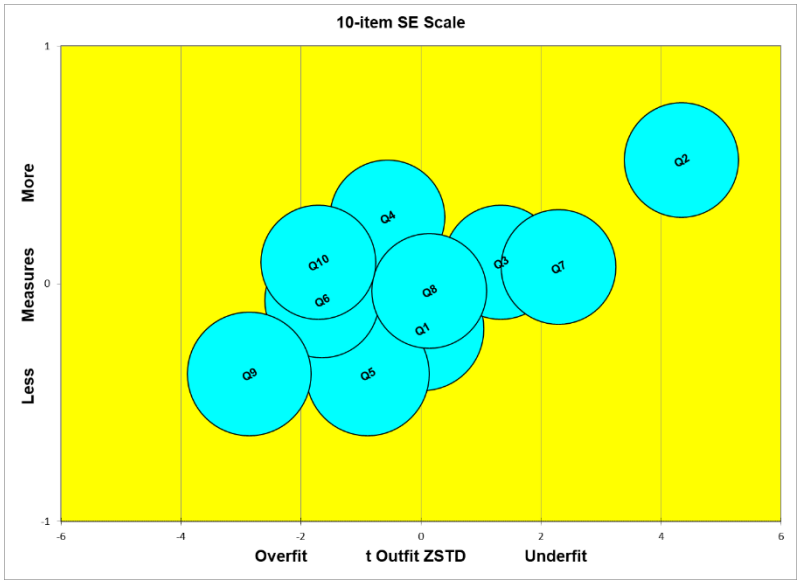
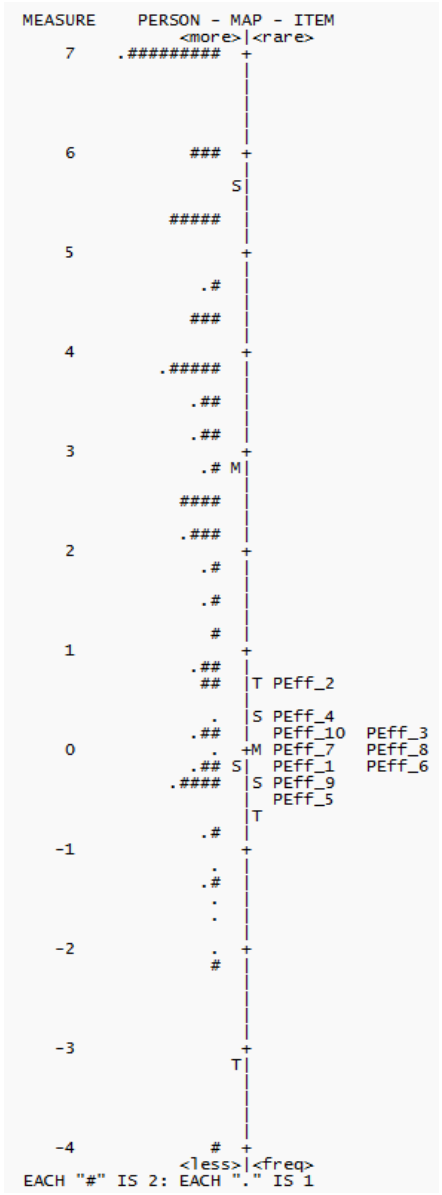


Figure 3A.11 Final SE Scale – Wright Maps



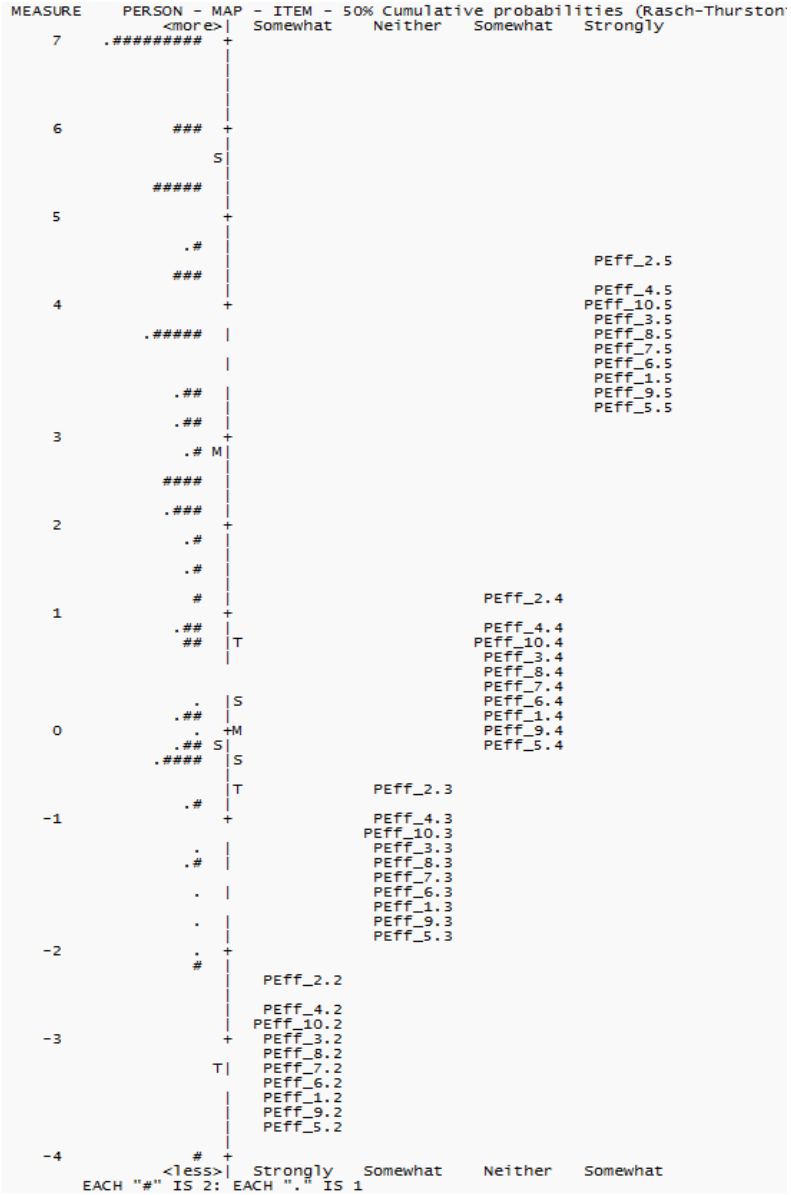


Figure 3A.12 SE Scale – DIF Analyses by age, education, income, and relationship status

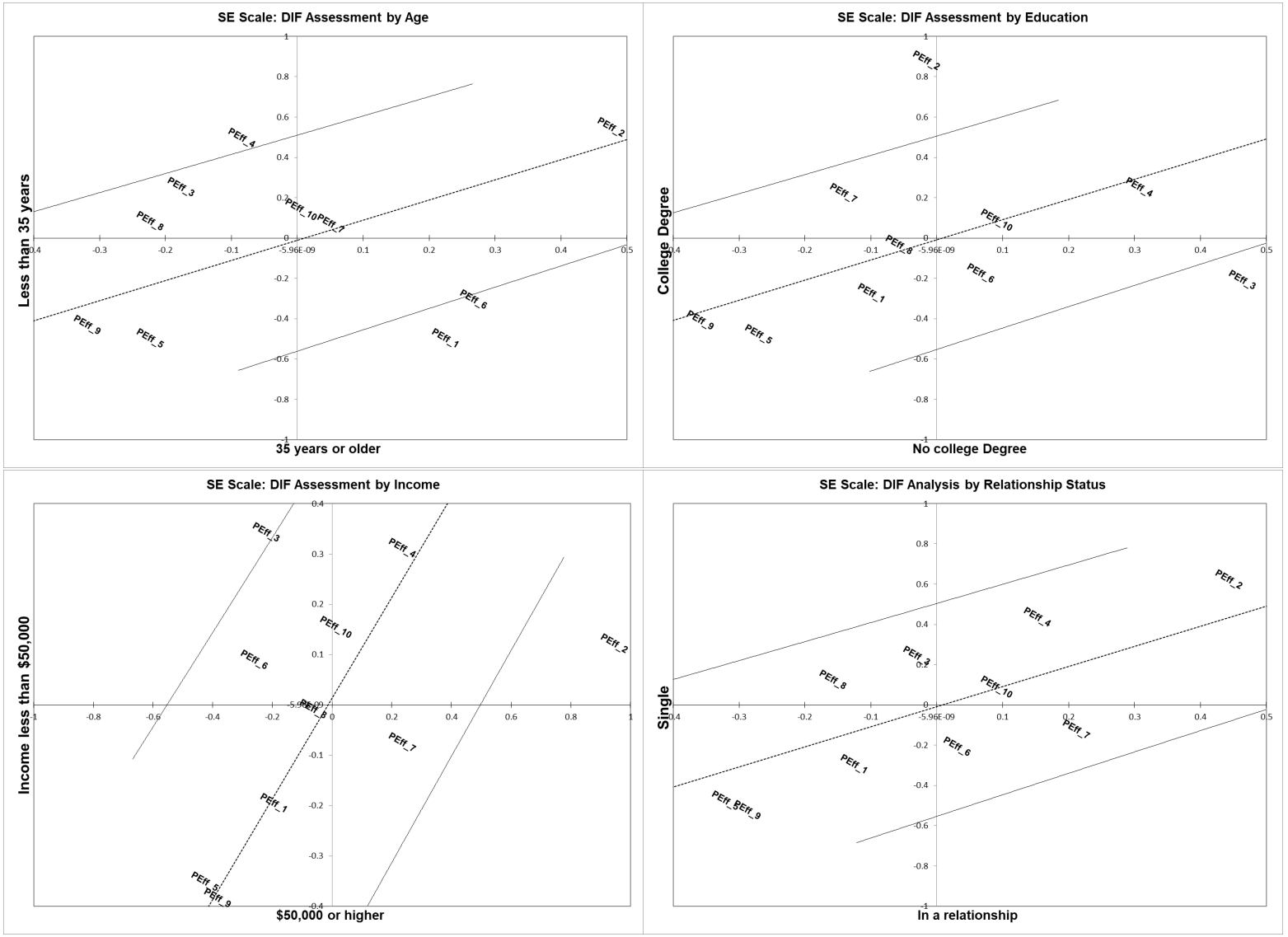


Figure 3A.13 SE Scale – Wright Maps by Education Level

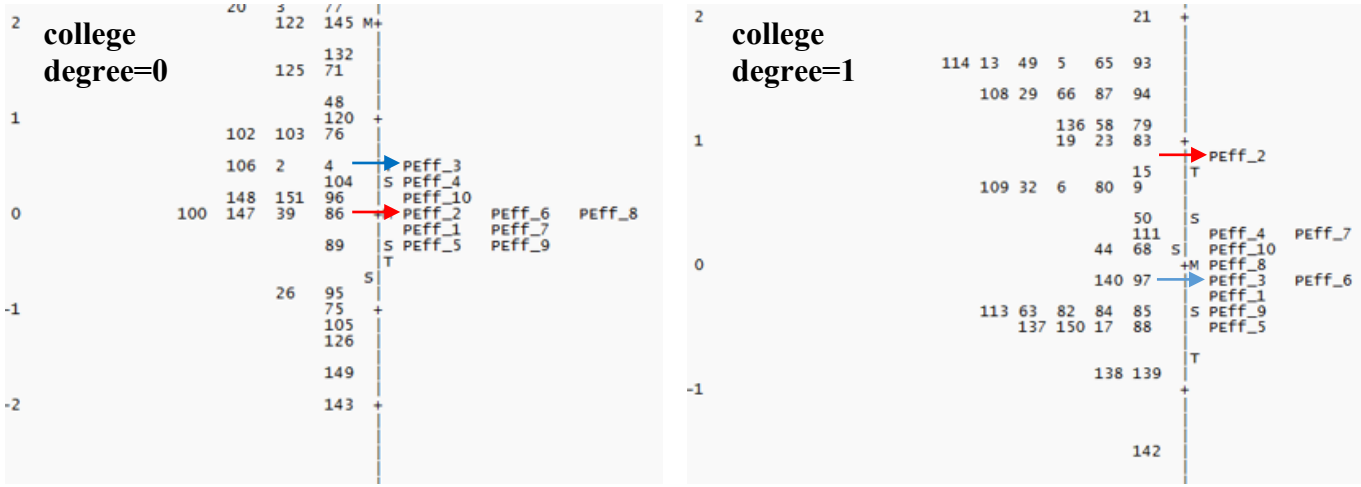
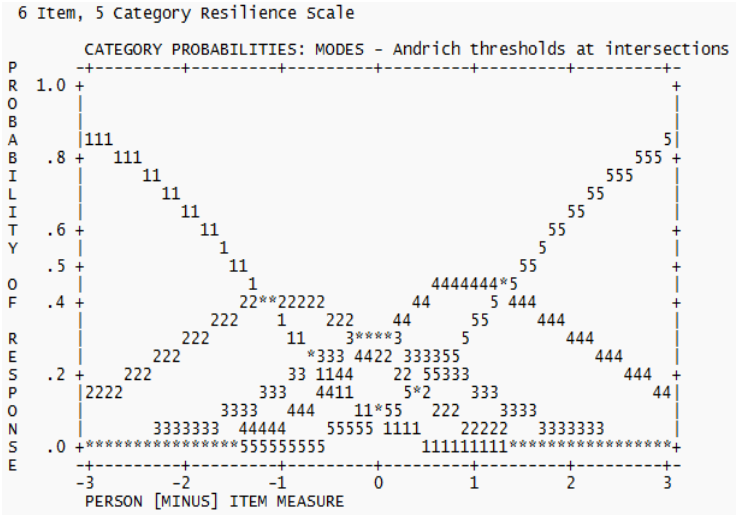


Figure 3A.14 Resilience Scale – Category Probability Plots

a.



b.

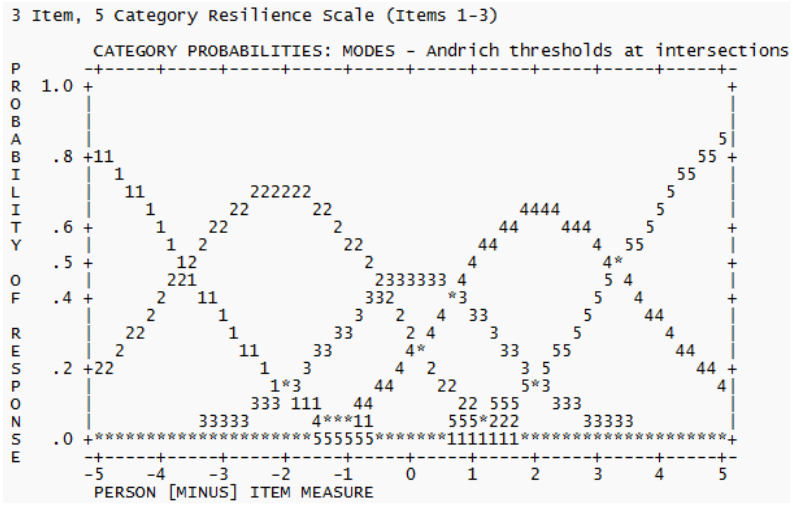


Figure 3A.15 6-item Self Efficacy Scale – Standardized Residual Contrast Plot

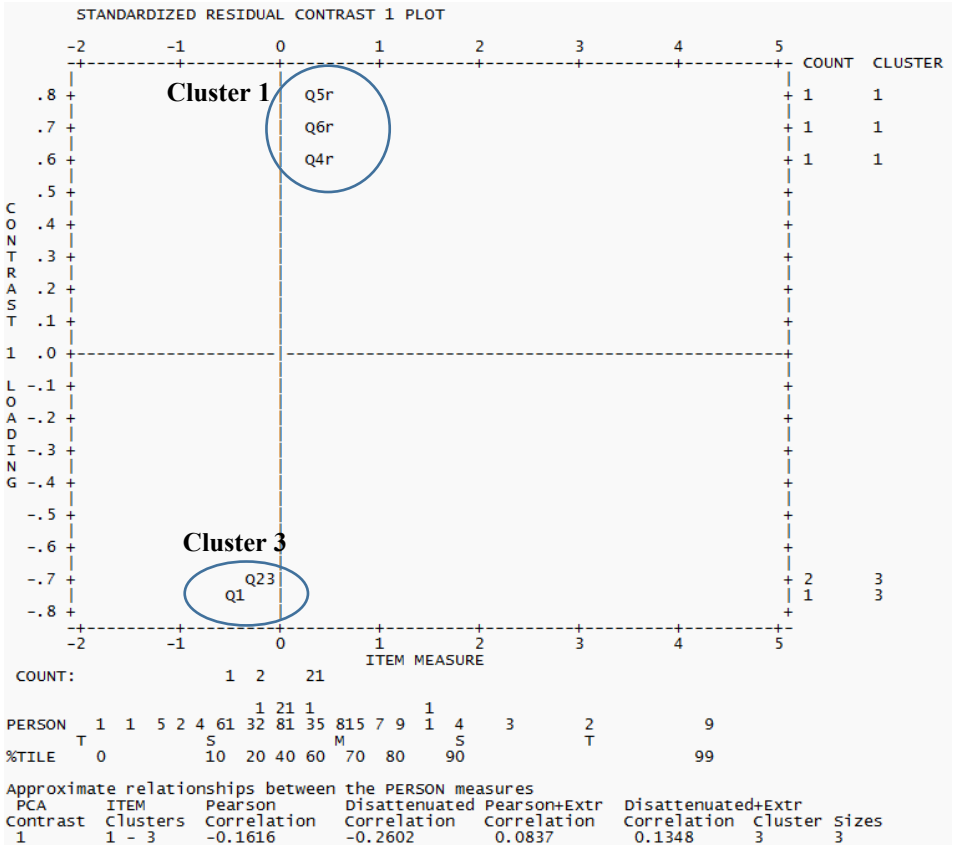
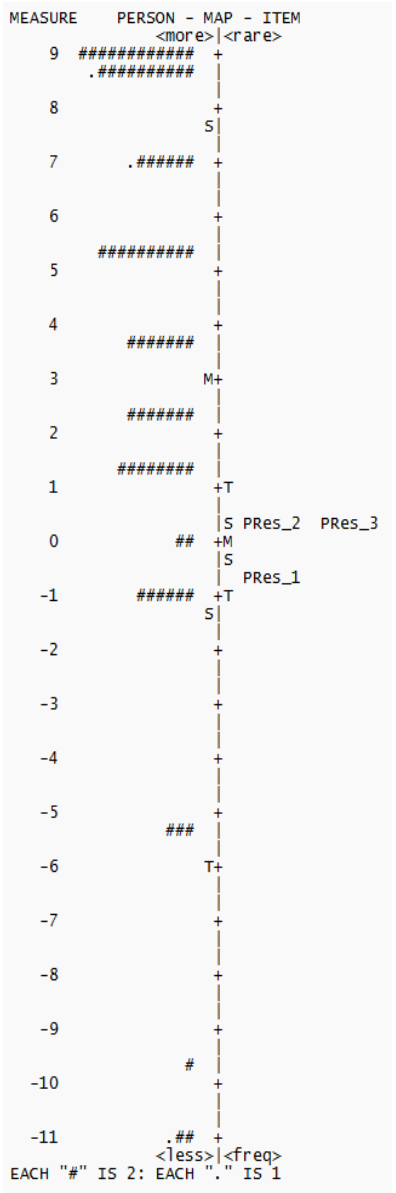


Figure 3A.16 Final Resilience Scale – Wright Maps



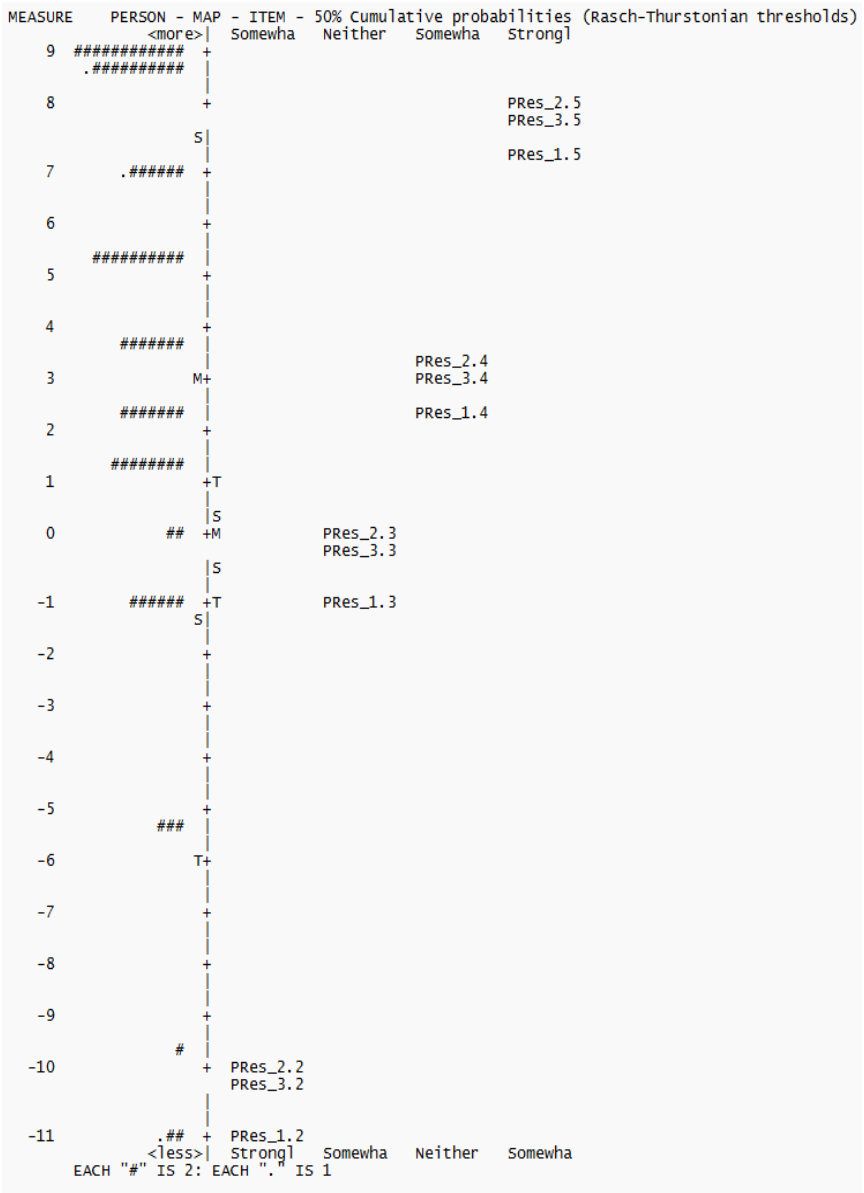


Figure 3A.17 Resilience Scale – DIF Analyses by age, education, income, and relationship status

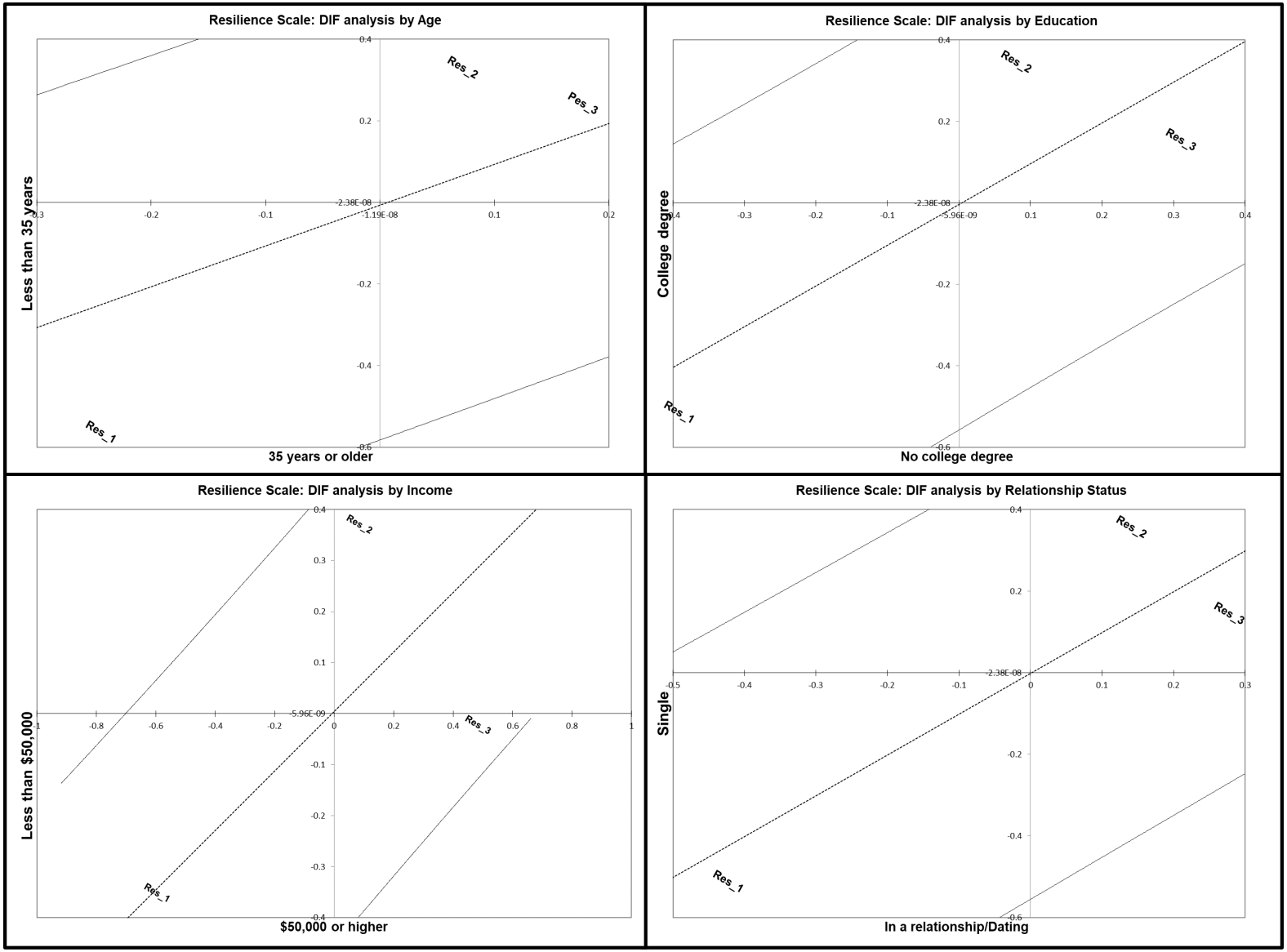


Figure 3A.18 PHQ Scale – Category Probability Plot

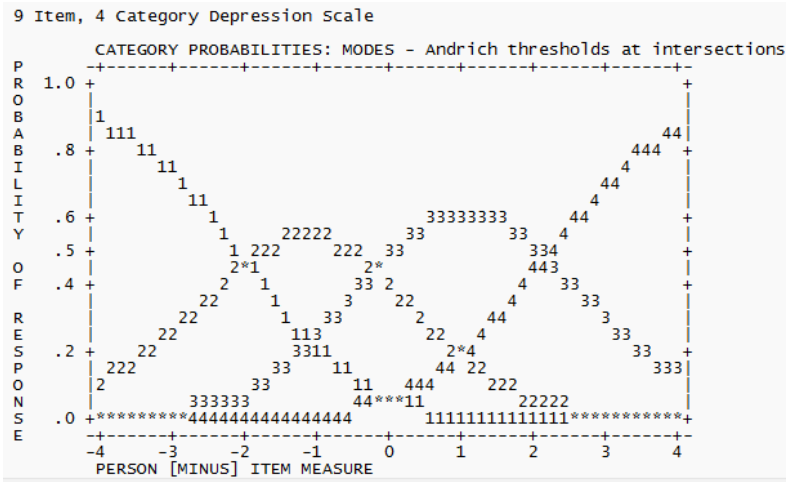


Figure 3A.19 Bubble Plot of PHQ Items

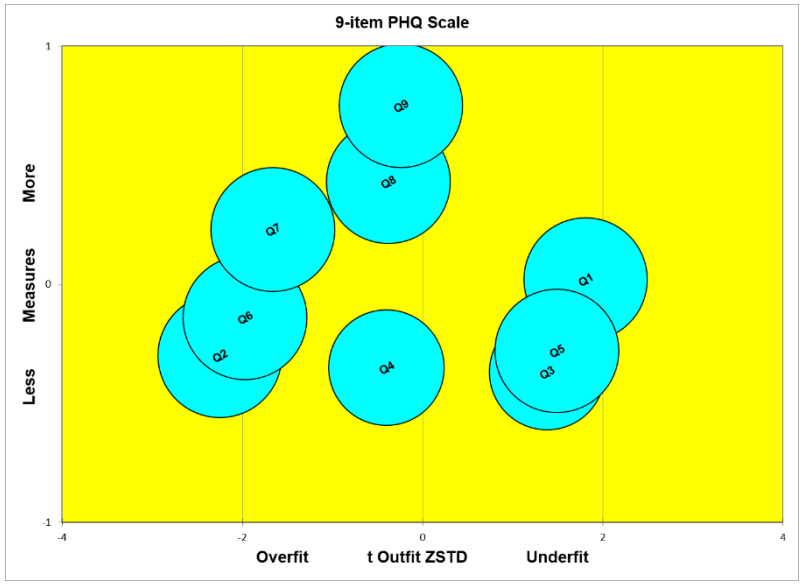
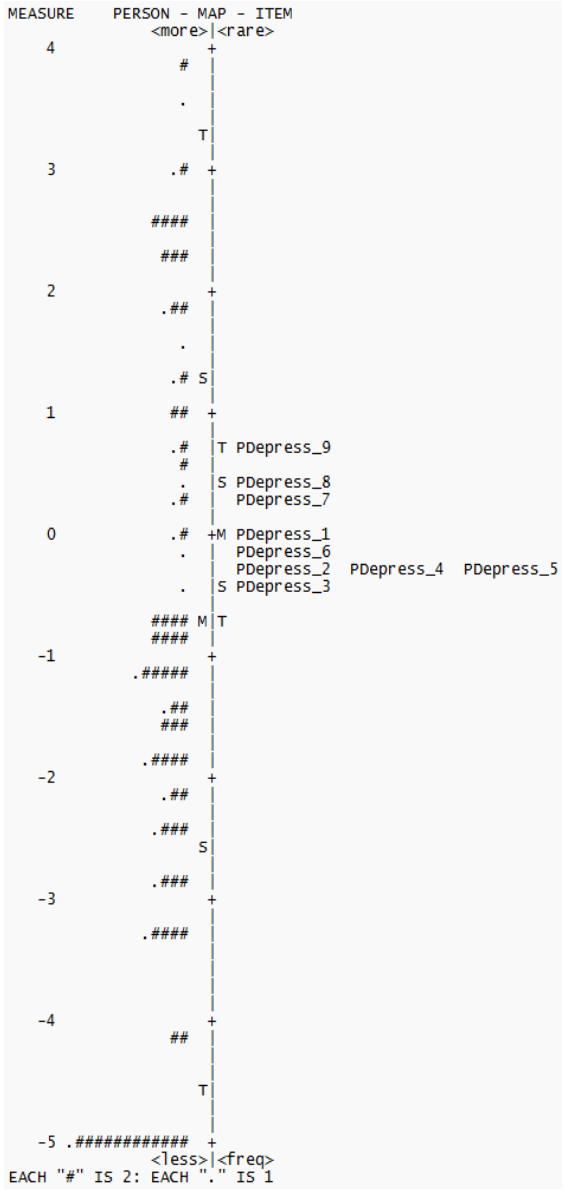


Figure 3A.20 Final PHQ Scale – Wright Maps



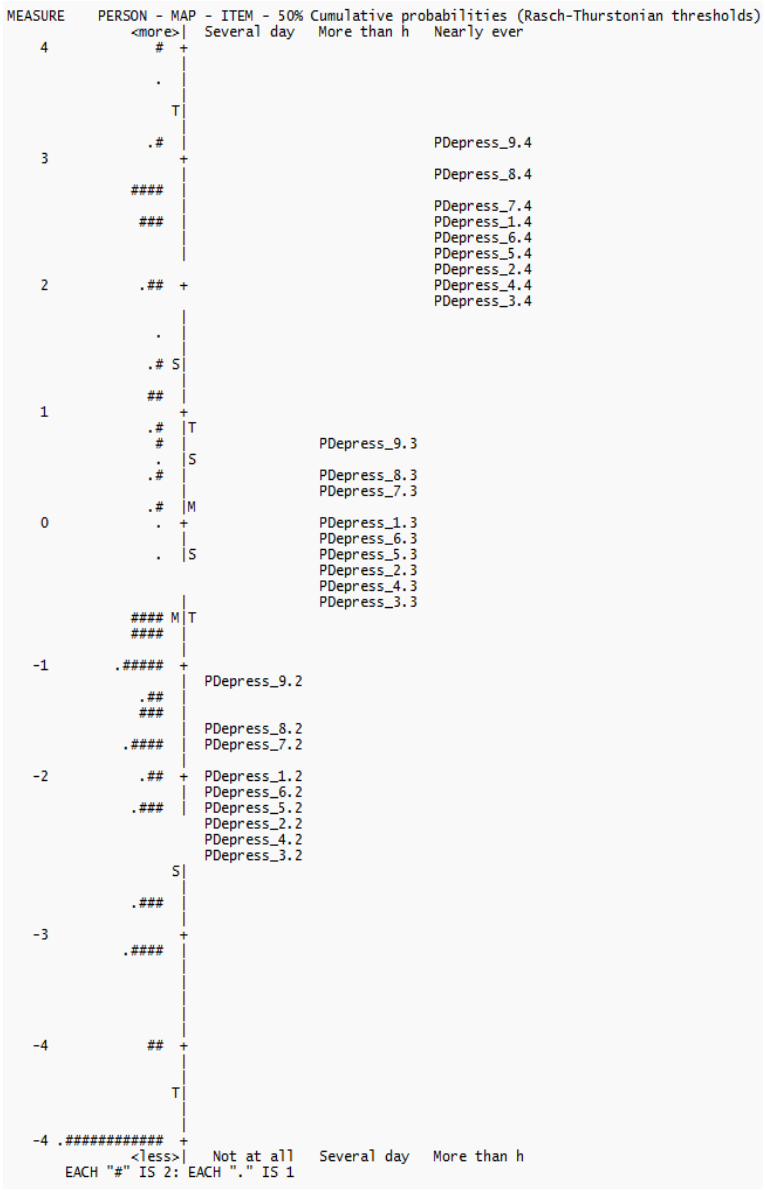


Figure 3A. 21 PHQ Scale – DIF analysis by age, education, income, and relationship status

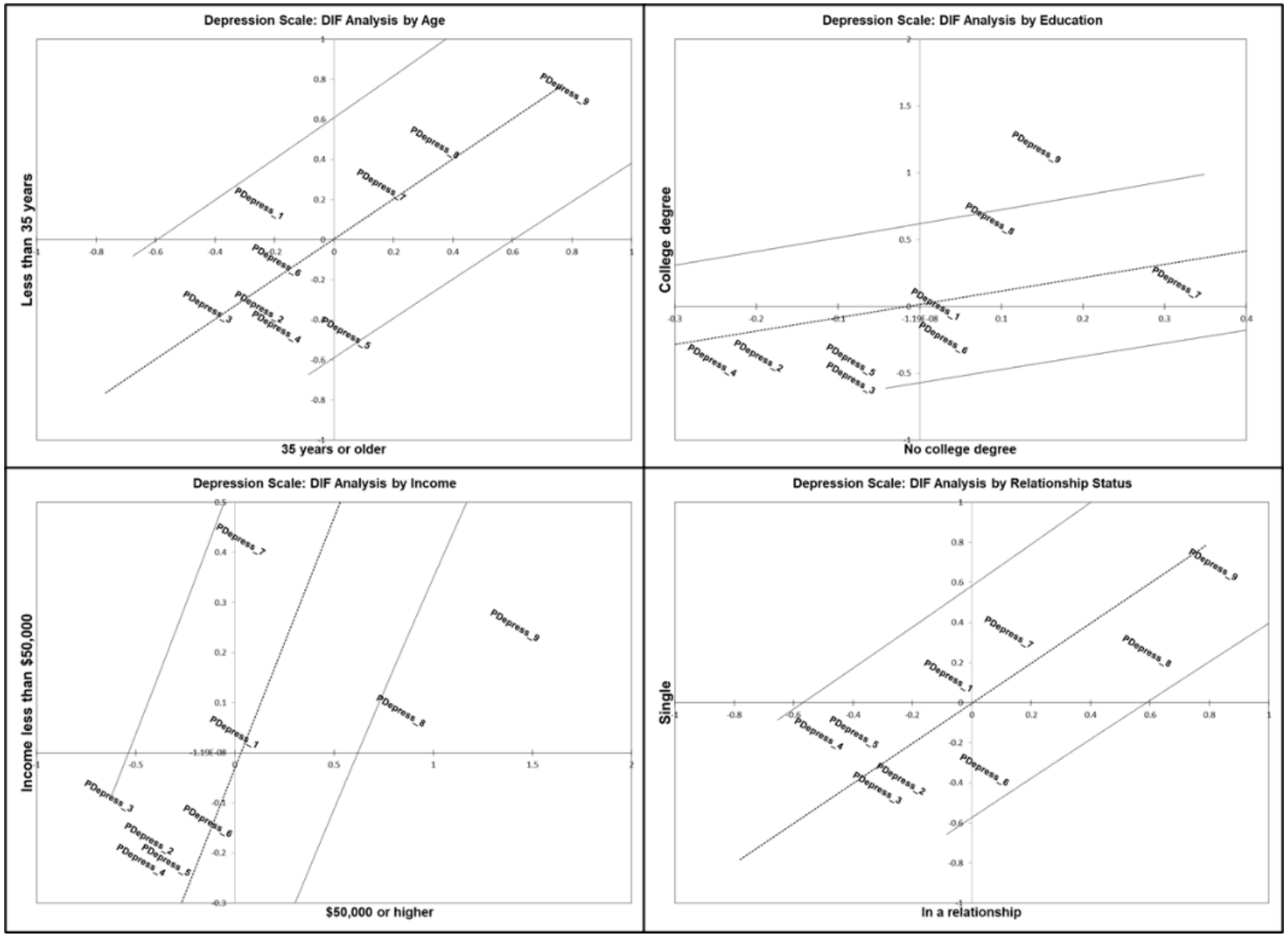
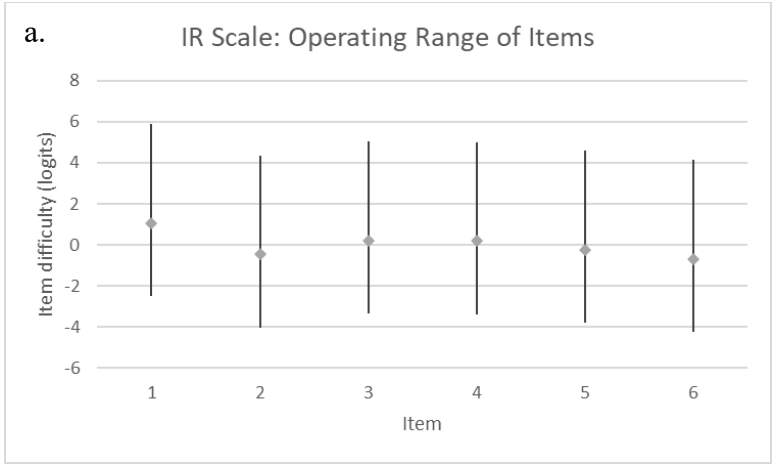
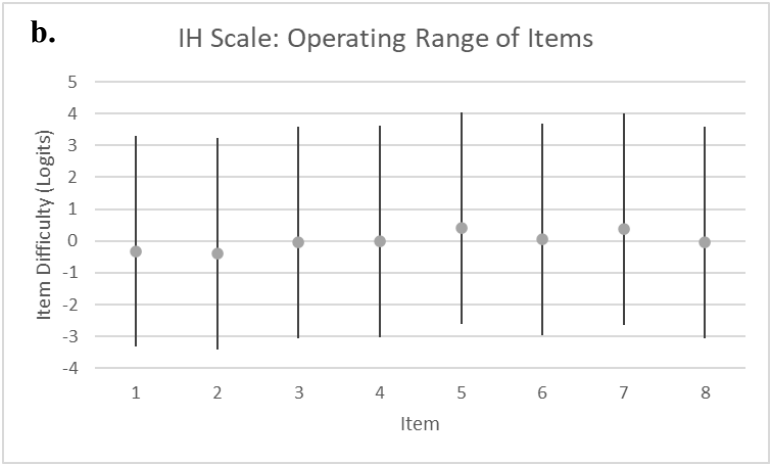
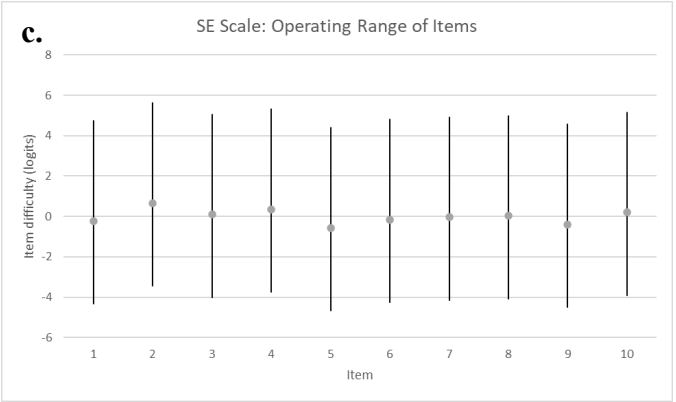
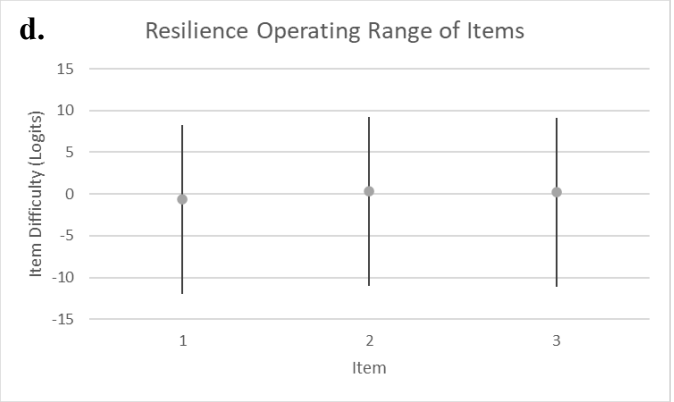


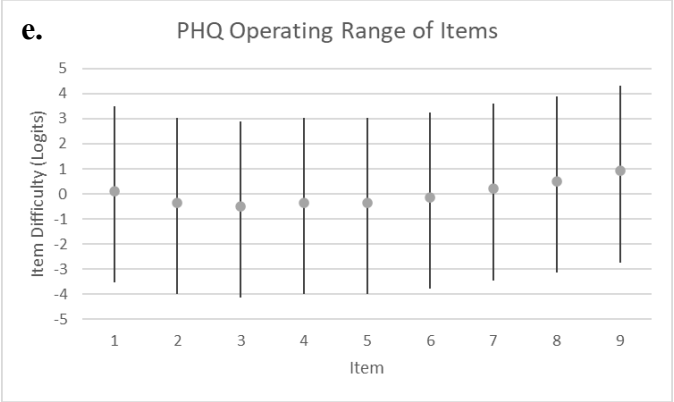
Figure 3A. 22 Operating Range of Items











Chapter 4: Internalized Stigma and Depressive Symptoms Among Black MSM: The Role of Self-Esteem and Resilience

4.1 Introduction

Despite recent efforts to stem the inequities facing Black men who have sex with men (BMSM), BMSM are disproportionally affected by negative health outcomes associated with stigma related to both their racial and sexual minority status. Link and Phelan (2001) defined stigma as the co-occurrence of four processes: 1) labeling, 2) stereotyping, 3) separation into groups representing "us" and "them," and 4) status loss and discrimination, all operating within the context of differential social, economic, and political power that enables one group to devalue another.[13] Stigma has been operationalized by three fundamental mechanisms: enacted stigma, or past experiences of prejudice, stereotyping, and discrimination; anticipated stigma, or expectations of experiencing prejudice, stereotyping, and discrimination in the future; and internalized stigma, or personally adopting negative beliefs and feelings associated with societal stigma.[80] Enacted, anticipated, and internalized stigma reflect the ways in which individuals respond to stigma, and how stigma operates to negatively impact health and perpetuate disparities in health outcomes.[146]

A substantial body of evidence suggests that self-reported experiences of discrimination (i.e., enacted stigma) are psychosocial stressors that negatively impact a range of mental health outcomes.[11] Less research has focused on the process by which internalized stigma impacts health among people living with a stigmatized identity. The pathway from internalized stigma to depressive symptoms can be built on the minority stress theory, which posits that members of

minority groups are subjected to unique, chronic, stressors related to their stigmatization that in turn increase their vulnerability to negative mental health outcomes.[18,19] While the literature on internalized stigma and mental health has grown in recent years, it has traditionally focused on a single stigmatized status. Intersectionality theory asserts that dually marginalized identities such as Black race and sexual minority orientation are not mutually exclusive, but rather they intersect to create different and distinct experiences of oppression and opportunity for the individual.[37,38] Growing evidence suggests that stigma associated with multiple identities (e.g., sexual orientation, race/ethnicity) may be mutually constitutive and reinforcing.[46,49,169]

While internalized stigma puts MSM of all race/ethnicities at increased risk for multiple adverse health outcomes including depression, anxiety, and substance use disorders,[19,20] BMSM may face multiple types of stigma, including crosscutting stigma related to both sexual orientation and to race/ethnicity. Studies have estimated the proportion of BMSM with depressive symptoms to range from as high as one-third to nearly fifty percent.[170,171] And while early research among MSM focused on predominantly White samples, less research has investigated the factors that influence psychosocial health among samples of BMSM.[172] An increased understanding of how internalized stigma contributes to depression among this population, and the associated aggravating and mitigating factors, is necessary to better inform targeted interventions to address these health disparities.

Much of this body of research has used the traditional epidemiological paradigm in which the goal is to identify risk factors that lead to deficits, disparities and disease, and work toward removing these risks. Indeed, an extensive body of research has documented the association between stigma and poor mental health outcomes; however, little research has examined the underlying mechanisms of these associations, including potential modifying or mediating factors.

Considering the adverse health outcomes associated with stigma, research has understandably focused on the detrimental effects of stigmatization, paying attention to how stigmatized individuals are devalued, exposed to prejudices, and negatively stereotyped. However, also critical may be an individual's internal and external resources that can serve as coping tools. The ability to cope with stressors, and protective variables such as resilience and self-efficacy, may modify or mediate the relationship between stigma and health.[47] Self-efficacy refers to the degree to which individuals believe they have the ability or capacity to achieve a desired outcome.[173] Resilience, as a protective factor, has been increasingly recognized as an important area to be studied.[28,29,174] Resilience refers to "the process of overcoming the negative effects of risk exposure, coping successfully with traumatic experiences, and avoiding the negative trajectories associated with risk."[175] Multiple studies have found that coping strategies may mitigate the negative effects of stressors on psychosocial health outcomes.[27,40,174,176]

The role of mental health in the physical health and overall well-being of BMSM has received less focus as compared to other groups of MSM. In addition, previous research and prevention programming for BMSM has focused heavily on HIV and sexual risk behaviors and overlooked the strengths and protective pathways that could be leveraged in a strength-based approach to improving both physical and mental health. More research is needed on the pathways from internalized stigma to mental health, particularly regarding adaptive coping strategies. We propose the integration of the minority stress and intersectionality theories to examine how the convergence of multiple stigmatized identities, each of which is associated with unique systems of oppression and opportunity, may create both risk and protective pathways to mental health outcomes among BMSM.[15,49,177,178] Specifically, we examine the association between co-occurring forms of identity-based stigma related to race and sexual orientation and

depressive symptoms among BMSM. We adopt a strength-based approach and further extend the research to examine how protective variables such as self-efficacy and resilience may influence this relationship.

The specific aims of this study are to 1) determine the levels of internalized stigma, self-efficacy, resilience, and depressive symptoms among BMSM; 2) examine the relationship between internalized stigma, self-efficacy, resilience, and depressive symptoms, and 3) investigate the mediating and modifying effects of self-efficacy and resilience, respectively, on the relationship between internalized stigma and depressive symptoms. We hypothesize that:

1) Internalized stigma, operationalized as internalized racism (IR) and internalized homophobia (IH), will be directly associated with risk of depressive symptoms.
    a. Increased IR and IH will be associated with increased risk of depressive symptoms.
2) IR and IH will have a synergistic (interactive) effect on depressive symptoms.
    a. IR will modify the association between IH and depressive symptoms, such that higher levels of IR will exacerbate the effects of IH on depressive symptoms.
3) Resilience will modify the associations between IR and IH and depressive symptoms.
4) There is an indirect (mediating) effect of IR and IH on depressive symptoms through self-efficacy.
    a. Greater frequency of IR and IH will be associated with lower levels of self-efficacy, which will be associated with greater risk of depressive symptoms.
    b. Higher levels of self-efficacy will be associated with lower risk of depressive symptoms compared to those with low self-efficacy.

This research will enhance our understanding of the mechanisms through which internalized stigma impacts mental health and help identify protective pathways to target interventions and to promote overall psychosocial health.

4.2 Methods

4.2.1 Study Design and Participants

We collected cross-sectional data from a pilot sample of HIV-negative BMSM (n=151) in the U.S. Participants were recruited through BMSM-specific social media (e.g., Jack'd, Grindr) and BMSM community-serving organizations. Eligibility criteria included being at least 18 years of age, identifying as Black, African American, African, Afro-Caribbean (including multiracial Black identities), identifying as male, having had a male sexual partner in the last 6 months, and being HIV negative. All participants were assigned male at birth. Data were collected between July 2nd and September 3rd, 2020, using online surveys. University of Maryland Institutional Review Board granted approval for the study (IRB # 1486118), and all participants provided written informed consent.

4.2.2 Measures

Demographic Information

Participants reported their demographic information including age, education, employment status, income, sexual identity, and relationship status.

Internalized Stigma

Internalized Racism (IR) was measured using a 7-item subscale from The Appropriated Racial Oppression Scale.[153] Internalized Homophobia (IH) was measured using Herek and Glunt's internalized homophobia scale.[154]

Resilience

Resilience was measured by 6 items from the Connor-Davidson Resilience scale (CD-RISC).[156] Participants were asked the extent to which they agreed or disagreed on a 5-point scale ranging from *strongly disagree* to *strongly agree* to statements such as "I tend to bounce back quickly after hard times" and "I have a hard time making it through stressful events." Negatively worded items were reverse coded.

Self-efficacy

Self-efficacy was measured with the 10-item Generalized Self-efficacy Scale (GSE).[155] Participants were asked the extent to which they agreed or disagreed on a 5-point scale ranging from *strongly disagree* to *strongly agree* to statements such as "I can always manage to solve difficult problems if I try hard enough" and "I can usually handle whatever comes my way."

Depressive Symptoms

Depressive symptoms were measured with the 9-item Patient Health Questionnaire (PHQ) scale.[157] Participants were asked the frequency with which they had been bothered by symptoms of depression such as feeling sad and having restless sleep in the past two weeks. Responses were indicated on a 4-point scale ranging from *not at all* to *nearly every day*.

4.2.3 Analytic Approaches to Intersectionality

There has been increasing utilization of quantitative methods to investigate research questions informed by intersectionality theory, but there is currently no consistently used method or gold standard.[179] Statistical interaction occurs when the effect of one variable on an outcome

depends on the level of a second exposure variable, and is present when the combined effect of two variables on an outcome differs from the sum or product of the individual effects.[180] By including an interaction term in a quantitative model, the main effects of two or more stigma-related variables (e.g., racial stigma and sexual orientation stigma), as well as the extent to which the effect of one variable is modified by the other, can be assessed. This widely applicable method is the most commonly used analytic strategy for modeling intersectional stigma, but it can be difficult to detect small but meaningful interactions when the main effects explain a large amount of variance in the outcome, and caution is needed when selecting the appropriate scale to evaluate intersectional hypotheses.[181] Alternatively, SEM allows pathways and relationships among observed and latent variables to be estimated by allowing for simultaneous estimation of measurement and structural components. SEM has the flexibility to measure complex interrelationships between different forms of stigma and other risk or protective factors, and is useful in analyzing mediation and indirect effects.[182]

4.2.4 Data Analysis

The main analysis followed a two-step process. First, WINSTEPS[88] software was used to examine the psychometric properties of each scale from a Rasch measurement perspective, and to obtain participant scores (Rasch-calibrated person measures) for each of the latent traits (i.e., IR, IH, self-efficacy, resilience, and depression). Several criteria can be used to determine if the empirical data fit the Rasch model. These criteria, considered jointly with the requirements for unidimensionality, item and person parameter invariance, and local independence of items, were used to examine the psychometric properties of the scales. Outfit and infit mean squares (MNSQ) statistics are commonly used indices of item fit and range from 0 to positive infinity, with 1 indicating (unattainable) perfect fit to the model. Bond (2007) suggests that MNSQ values

ranging from 0.6 to 1.4 indicate good fit. Item and person reliability estimates the replicability of item and person ordering along the latent trait metric.[66] Variance explained by Rasch measures refers to the proportion of variance in the observed data that can be explained by the Rasch measure, and a higher proportion of variance explained means that the scale has a better capacity for predicting performance of both items and persons.[183] Step thresholds were examined to ensure the appropriate category functioning of the rating scales. (For detailed description of the Rasch analysis, see Chapter 3, Section 2.5 Statistical Analysis.) Rasch-calibrated person measures obtained from the final version of the scales were subsequently analyzed by linear regression and path analysis.

Participant characteristics, including socio-demographics, depressive symptoms, racial and sexual-orientation internalized stigma, and protective psychosocial characteristics were analyzed using descriptive statistics. Linear regression was used to model level of depressive symptoms, as measured by Rasch-calibrated person measure on the PHQ scale. Univariable regression analysis was conducted first on all variables. Variables hypothesized to be confounders based on theory or previously identified in the literature included age, income, and relationship status, and were included in multivariable analyses regardless of statistical significance in the univariable models.

To assess Hypothesis 1, we first examined the main effects of IR and IH on depressive symptoms. Next, we tested for the interactive effect of IR and IH on depressive symptoms with the inclusion of an interaction term (Hypothesis 2). We then tested if resilience modified the association between IR and IH and depressive symptoms (Hypothesis 3), respectively, using a hierarchically well-formulated modeling approach.[184]

Next, we conducted a path analytic mediation analysis using the observed, Rasch-calibrated person measures to assess the indirect effect of stigma on depression through self-efficacy (Hypothesis 4). We hypothesized direct effect paths from IR and IH to depressive symptoms, and indirect effect paths from IR and IH to depressive symptoms, mediated by self-efficacy. Indirect paths were fitted using the robust maximum likelihood estimator (MLF) with 500 Monte-Carlo integrations in order to obtain robust beta estimates for the indirect effects with confidence intervals. Model fit was assessed with the following: chi-square (χ 2) and relative chi-square (χ 2 /df), the Comparative Fit Index (CFI), the Tucker-Lewis Index (TLI), and the Root Mean Square Error of Approximation (RMSEA). As a general rule, χ 2 /df less than 3.0 was considered to be an indication of good model fit. CFI and TLI values close to or greater than 0.95, and RMSEA close to or less than 0.06, were considered adequate fit.[185] All analyses were conducted using Stata 15.[186]

4.3 Results

4.3.1 Descriptive Statistics

Table 4A.1 displays the descriptive statistics of the analytic sample (n=151). Most participants were between 25 and 45 years of age, with 43% in the 25 to 34 years of age category, and 29% in the 35 to 44 years of age category. The majority of participants lived in the Northeast (55%) and the South (31%). Almost a third had a college education, with 62% achieving a graduate degree, and 70% were employed full-time. Over half (57%) of participants self-identified as gay, 20% as bisexual, and 11% as heterosexual/straight. Approximately 44% of participants reported their current relationship status as single, while 38% were in a relationship and 17% were dating. Summary statistics for the five Rasch-derived person-measures are

presented in **Table 4A.2** , and the bivariate Pearson correlation matrix is presented in **Table 4A.3.**

4.3.2 Direct and Modifying Effects

Table 4A.4 presents univariable and multivariable linear regression models of the Rasch-derived person measures for IR, IH, self-efficacy, and resilience on depressive symptoms, measured by PHQ score. In multivariable models, IR was positively associated with depressive symptoms (β=0.50, 95% CI [0.36, 0.63]), as was IH (β=0.65, 95% CI [0.48, 0.82]). Each 1-unit increase in IR score and in IH score was associated with a 0.50 and 0.65 increase, respectively, in PHQ score. Alternatively, self-efficacy and resilience were negatively associated with depressive symptoms. Each 1-unit increase in self-efficacy score and in resilience score was associated with a decrease of 0.31 (95% CI [-0.46, -0.16]) and of 0.11 (95% CI [-0.20, -0.02]), respectively, in PHQ score.

We investigated the hypothesized interaction between IH and IR on depressive symptoms in **Table 4A.5.** Both IR and IH maintained statistically significant independent effects on depressive symptoms (β_{IR}=0.30, 95% CI [0.13, 0.46]; β_{IH}=0.41, 95% CI [0.20, 0.62]) when controlling for age, income, and relationship status. However, there was no evidence of additive interaction between the two stigma mechanisms, as the magnitude of the interaction term was close to zero and not statistically significant (β_{IRxIH}=-0.004, 95% CI [-0.05, 0.05]). As such, the interaction term was removed from all subsequent analyses.

We then explored whether the effect of internalized stigma on depressive symptoms varied by level of resilience or self-efficacy using four separate models. Each model included one of the stigma mechanisms (IR or IH) as a predictor of PHQ score and one of the protective factors (resilience or self-efficacy) as an effect modifier. The models in **Table 4A.6a** and **Table 4A.6b**

display the results of the hypothesized effect of resilience and self-efficacy on the relationship between IR and IH, respectively, and depressive symptoms. The interaction between IH and resilience was statistically significant (β_{IHxRES}=-.034, 95% CI [-0.06, -0.01]), indicating that resilience modified the effect of IH on depression. Similarly, the interaction between and IH and self-efficacy was marginally statistically significant ($\beta_{IHxSE)}$=-.056, 95% CI [-0.11, 0.00]), indicating effect modification by self-efficacy. **Figure 4A.1** includes a graph of the interaction effects between IH and each protective factor on depressive symptoms. In Plot A, the simple slope of IH is graphed at three levels of resilience (i.e., at the mean resilience score, and at one standard deviation above and below the mean). The results indicate that resilience attenuates the association between IH and depressive symptoms, as the slope of IH decreases at higher levels of resilience. A similar attenuating effect is seen for self-efficacy in Plot B. There was no evidence of effect modification by resilience or self-efficacy, however, on the association between IR and depressive symptoms.

4.3.3 Correlations and Path Analysis of Indirect Effects

Bivariate Pearson correlations are presented in **Table 4A.3.** Internalized racism and internalized homophobia were significantly positively correlated with one another (r=0.63, p<0.00), and with depressive symptoms (r=0.56 and r=0.57, respectively). Similarly, self-efficacy and resilience were significantly positively correlated with one another (r=0.73, p<0.00), and negatively correlated with depressive symptoms (r=-0.29 and r=-0.20, respectively). Self-efficacy and resilience were both negatively correlated with each of the internalized stigma measures, although these correlations were not statistically significant.

Despite the lack of a statistically significant correlation between the stigma measures and self-efficacy, we proceeded with the path analysis to assess the potential mediating role of self-

efficacy proposed in Hypothesis 4. The standardized path coefficients for the full model are presented in **Figure 4A.2**. Both IR and IH were directly associated with depressive symptoms (β_{IR}=0.31, p<0.00; β_{IH}=0.35, p<0.00), as was self-efficacy (β_{SE}=-0.21, p=0.01). Not surprisingly, given the non-significant correlations, there was no evidence of statistically significant indirect effects of either stigma mechanism through self-efficacy. The indirect effect of IR on depressive symptoms through self-efficacy was 0.01 (p=0.56), and the indirect effect of IH was 0.02 (p=.42). Nevertheless, the indirect effects were in the hypothesized direction.

4.4 Discussion

To date, limited research has investigated the processes by which co-occurring forms of internalized stigma impact health outcomes. Research among BMSM and guided by intersectionality theory has focused primarily on HIV-related stigma.[187,188] While previous qualitative work has suggested that experience of race, sexual orientation, and HIV stigma are interrelated among BMSM,[189,190] this study is among the first quantitative studies to our knowledge to investigate the relationship between internalized stigma related to race and sexual orientation and depressive symptoms among BMSM. Confirming our hypotheses, both IR and IH were directly and positively associated with depressive symptoms. Our results are consistent with previous work and meta-analyses that have found IR[191-193] and IH[194] to increase the risk of negative mental health outcomes among racial/ethnic and sexual minority men. The independent effects of both IR and IH maintained statistical significance in multivariable models controlling for age, income, and relationship status, suggesting that both forms of stigma independently increase the risk of depressive symptoms. Contrary to our hypotheses, however, we found no evidence of an interaction effect between the two forms of internalized stigma on depressive symptoms. One previous study identified in the literature among BMSM recently diagnosed with

HIV found a "marginally significant interaction" between race and sexual orientation stigma, suggesting that the association between internalized stigma related to race and to HIV may be stronger among participants with lower sexual orientation stigma.[195] The authors note, however, that the interaction only approached statistical significance (p=0.08), and the observed association may have been due to a ceiling effect.[195] Another study among young adult college students in Georgia found that experiencing both sexual orientation and racial discrimination increased the risk of depressive symptoms beyond that of only experiencing a single form of discrimination for women, but a similar effect was not observed among men.[196] Additional research is needed to better understand how multi-faceted elements of stigma impact health outcomes among individuals with multiple stigmatized identities.

We found no indirect association between internalized stigma and depressive symptoms via self-efficacy. This may be because previous studies that have identified a mediating role of positive psychosocial characteristics of self-concept have focused primarily on self-esteem rather than self-efficacy.[192,194] For example, James (2017) found internalized racism to be indirectly related to past-year major depressive disorder (MDD) via self-esteem, such that IR was associated with lower self-esteem, which was associated with increased risk of meeting criteria for MDD.[197] While related, self-efficacy and self-esteem are conceptually distinct concepts.[198] Self-esteem refers to the judgement and evaluation an individual makes about himself, whereas self-efficacy refers to the judgement an individual makes about his or her ability to perform a task within a specific context. Such work by Chen et al (2004) suggests that self-efficacy and self-esteem relate differently to motivational versus affective constructs, and that motivational and affective states differently mediate the relationships of general self-efficacy and self-esteem with task performance.[198] Indeed, our results indicate that self-efficacy and resilience modified,

rather than mediated, the association between IH and depressive symptoms in this study. Our findings are consistent with previous studies that focused on other motivational (rather than affective) constructs and found similar modifying effects. For example, both Rueda et al (2012) and Mouzon (2017) found that mastery moderated the negative effects of stigma on depressive symptoms among people living with HIV,[199] and among African-American and Caribbean Blacks in the US,[200] respectively.

Interestingly, we did not find the same modifying effects of self-efficacy and resilience for IR. Previous studies conceptualizing IR have focused on specific aspects, such as conscious or unconscious IR, self-focused or group-focused, and positive or negative IR. In this study, IR was group-focused, while all other measures were self-focused. The conceptualization of IR, and how it is measured, may influence the nature of these relationships. That is, group-focused IR may differ from self-focused IR in the extent to which racial/ethnic group membership impacts an individual's overall self-concept. Indeed, in a review of 81 studies conceptualizing IR as either group-focused or self-focused, James (2021) found that self-focused IR had a stronger effect than group-focused IR on the health outcome under study.[201] Future studies should further investigate if and why self- and group-focused IR converge or diverge when predicting individual health outcomes.

There are several limitations to note. As this was a convenience sample of BMSM, the generalizability is limited. Data were self-reported, so social desirability bias may have impacted participant responses. The cross-sectional design precludes the ability to determine causality. In addition, while the sample size was adequate (n=151), a larger sample size would have increased power to detect hypothesized interaction effects between IR and the protective variables. Nonetheless, this study builds upon to the efforts to conceptualize and operationalize internalized

stigma in public health research and contributes to the developing body of literature that applies intersectionality theory to health disparities research.

4.5 Appendix: Tables and Figures

Table 4A.1 Participant Demographics

Variable		n (%)
Age	18 – 24	31 (21)
	25 – 34	65 (43)
	35 – 44	44 (29)
	45+	11 (7)
Received a college degree		93 (62)
Employed Full Time		105 (70)
Highest education	Graduate degree	39 (26)
	Undergrad degree	54 (36)
	Some college	37 (25)
	HS or less	21 (14)
Annual household income	Less than $30,000	39 (26)
	$30,000 - $59,999	58 (38)
	$60,000 or more	54 (36)
Region of the U.S.	Northeast	83 (55)
	West	8 (5)
	Midwest	10 (7)
	South	47 (31)
	Outside the U.S.	3 (2)
Sexual identity	Bisexual	30 (20)
	Gay	86 (57)
	Heterosexual/straight	17 (11)
	Other	18 (12)
Relationship status	Single	67 (44)
	Dating	26 (17)
	In a relationship	58 (38)

Table 4A. 2 Summary Statistics

Person-Measure	Mean (SD)	Range
IR	-0.730 (3.097)	-5.2 – 6.81
IH	-0.223 (2.464)	-4.97 – 5.74
Self-efficacy	3.021 (3.112)	-6.65 – 7.56
Resilience	3.975 (5.247)	-12.47 – 10.08
PHQ	-1.548 (2.973)	-5.93 – 6.2

Table 4A.3 Correlation Matrix

	IR	IH	SE	RES	PHQ
IR	1				
IH	0.627* (0.00)	1			
SE	-0.117 (0.15)	-0.127 (0.12)	1		
RES	-0.004 (0.96)	-0.059 (0.47)	0.729* (0.00)	1	
PHQ	0.557* (0.00)	0.571* (0.00)	-0.289* (0.00)	-0.201* (0.01)	1

*Statistically significant at α=.05

Table 4A.4 Linear Regression of Depressive Symptoms

	Crude			Adjusted^		
Variable	**β**	**95% CI**	**p-value**	**β**	**95% CI**	**p-value**
IR	0.535*	0.405 – 0.664	0.000	0.497*	0.364 – 0.629	0.000
IH	0.690*	0.529 – 0.850	0.000	0.652*	0.482 – 0.822	0.000
SE	-0.276*	-0.424 – -0.128	0.000	-0.308*	-0.455 – -0.162	0.000
RES	-0.117*	-0.207 – -0.027	0.011	-0.112*	-0.201 – -0.023	0.014
Age						
18 – 24	Ref	--	--			
25 – 34	-1.942*	-3.192 – -0.691	0.003			
35+	-1.616*	-2.903 – -0.330	0.014			
Income						
<$30,000	Ref	--	--			
$30-59,999	-0.506	-1.706 – 0.695	0.407			
$60,000+	-1.451*	-2.670 – -0.233	0.020			
Single	1.035*	0.084 – 1.986	0.033			

^Adjusted for age, income, and relationship status
*Statistically significant at α=.05

Table 4A.5 Linear Regression of Depressive Symptoms on INTERNALIZED RACISM and INTERNALIZED HOMOPHOBIA

	Model 1				Model 2			
	β	95% CI		p-value	β	95% CI		p-value
IR	**0.497***	0.364	0.629	0.000	--	--	--	--
IH	--	--	--	--	**0.652***	0.482	0.822	0.000
IH x IR	--	--	--	--	--	--	--	--
Age								
18-24		(Ref)				(Ref)		
25-34	-0.937	-2.073	0.200	0.105	-0.644	-1.786	0.498	0.267
35+	-0.743	-1.961	0.475	0.230	-0.035	-1.271	1.201	0.956
Income								
<$30,000		(Ref)				(Ref)		
$30-60,000	0.125	-0.947	1.196	0.818	-0.366	-1.434	0.702	0.499
>$60,000	-0.176	-1.301	0.948	0.757	-0.879	-1.984	0.226	0.118
Single	0.714	-0.121	1.550	0.093	0.209	-0.636	1.053	0.626
	Model 3				**Model 4**			
	β	95% CI		p-value	β	95% CI		p-value
IR	**0.296***	0.133	0.458	0.000	**0.302***	0.120	0.484	0.001
IH	**0.414***	0.204	0.623	0.000	**0.409***	0.191	0.627	0.000
IH x IR	--	--	--	--	-0.004	-0.052	0.045	0.880
Age								
18-24		(Ref)				(Ref)		
25-34	-0.597	-1.695	0.500	0.284	-0.606	-1.712	0.501	0.281
35+	-0.218	-1.410	0.975	0.719	-0.224	-1.424	0.975	0.712
Income								
<$30,000		(Ref)				(Ref)		
$30-60,000	-0.162	-1.195	0.870	0.757	-0.166	-1.203	0.871	0.752
>$60,000	-0.482	-1.566	0.602	0.381	-0.476	-1.567	0.615	0.390
Single	0.370	-0.446	1.186	0.372	0.382	-0.452	1.215	0.367

*Statistically significant at α=0.05

Table 4A.6a Linear Regression of Depressive Symptoms on <u>INTERNALIZED RACISM</u> and <u>RESILIENCE</u>

	Model 1				Model 2			
	β	95% CI		p-value	β	95% CI		p-value
IR	**0.497***	0.364	0.629	0.000	--	--	--	--
RES	--	--	--	--	**-0.112***	-0.201	-0.023	0.014
IR x RES	--	--	--	--	--	--	--	--
<u>Age</u>								
18-24		(Ref)				(Ref)		
25-34	-0.937	-2.073	0.200	0.105	**-1.707***	-3.000	-0.414	0.010
35+	-0.743	-1.961	0.475	0.230	-1.322	-2.733	0.089	0.066
<u>Income</u>								
<$30,000		(Ref)				(Ref)		
$30-60,000	0.125	-0.947	1.196	0.818	0.265	-0.982	1.513	0.675
>$60,000	-0.176	-1.301	0.948	0.757	-0.480	-1.782	0.823	0.468
Single	0.714	-0.121	1.550	0.093	0.773	-0.188	1.734	0.114

	Model 3				Model 4			
	β	95% CI		p-value	β	95% CI		p-value
IR	**0.501***	0.372	0.630	0.000	**0.536***	0.369	0.703	0.000
RES	**-0.118***	-0.194	-0.043	0.002	**-0.121***	-0.197	-0.045	0.002
IR x RES	--	--	--	--	-0.008	-0.033	0.017	0.521
<u>Age</u>								
18-24		(Ref)				(Ref)		
25-34	-1.038	-2.144	0.068	0.066	-1.051	-2.159	0.058	0.063
35+	-1.003	-2.198	0.192	0.099	-1.043	-2.247	0.160	0.089
<u>Income</u>								
<$30,000		(Ref)				(Ref)		
$30-60,000	0.388	-0.666	1.442	0.468	0.416	-0.644	1.476	0.439
>$60,000	0.151	-0.961	1.264	0.788	0.178	-0.940	1.295	0.753
Single	0.666	-0.147	1.478	0.107	0.711	-0.115	1.537	0.091

*Statistically significant at α=0.05

Table 4A.6b Linear Regression of Depressive Symptoms on <u>INTERNALIZED RACISM</u> and <u>SELF-EFFICACY</u>

	Model 1				Model 2			
	β	95% CI		p-value	β	95% CI		p-value
IR	**0.497***	0.364	0.629	0.000	--	--	--	--
SE	--	--	--	--	**-0.308***	-0.455	-0.162	0.000
IR x SE	--	--	--	--	--	--	--	--
<u>Age</u>								
18-24		(Ref)				(Ref)		
25-34	-0.937	-2.073	0.200	0.105	**-2.134***	-3.404	-0.864	0.001
35+	-0.743	-1.961	0.475	0.230	**-1.746***	-3.131	-0.360	0.014
<u>Income</u>								
<$30,000		(Ref)				(Ref)		
$30-60,000	0.125	-0.947	1.196	0.818	0.236	-0.957	1.428	0.697
>$60,000	-0.176	-1.301	0.948	0.757	-0.291	-1.548	0.966	0.647
Single	0.714	-0.121	1.550	0.093	0.739	-0.188	1.667	0.117

	Model 3				Model 4			
	β	95% CI		p-value	β	95% CI		p-value
IR	**0.465***	0.338	0.593	0.000	**0.478***	0.289	0.666	0.000
SE	**-0.251***	-0.378	-0.124	0.000	**-0.251***	-0.379	-0.124	0.000
IR x SE	--	--	--	--	-0.004	-0.046	0.039	0.860
<u>Age</u>								
18-24		(Ref)				(Ref)		
25-34	**-1.409***	-2.519	-0.299	0.013	**-1.411***	-2.525	-0.297	0.013
35+	**-1.312***	-2.509	-0.114	0.032	**-1.319***	-2.524	-0.115	0.032
<u>Income</u>								
<$30,000		(Ref)				(Ref)		
$30-60,000	0.295	-0.730	1.321	0.570	0.305	-0.730	1.340	0.561
>$60,000	0.186	-0.902	1.275	0.735	0.193	-0.902	1.288	0.728
Single	0.657	-0.141	1.454	0.106	0.669	-0.144	1.482	0.106

*Statistically significant at α=0.05

Table 4A.7a Linear Regression of Depressive Symptoms on INTERNALIZED HOMOPHOBIA and RESILIENCE

	Model 1				Model 2			
	β	95% CI		p-value	β	95% CI		p-value
IH	**0.652***	0.482	0.822	0.000	--	--	--	--
RES	--	--	--	--	**-0.112***	-0.201	-0.023	0.014
IH x RES	--	--	--	--	--	--	--	--
Age								
18-24		(Ref)				(Ref)		
25-34	-0.644	-1.786	0.498	0.267	**-1.707***	-3.000	-0.414	0.010
35+	-0.035	-1.271	1.201	0.956	-1.322	-2.733	0.089	0.066
Income								
<$30,000		(Ref)				(Ref)		
$30-60,000	-0.366	-1.434	0.702	0.499	0.265	-0.982	1.513	0.675
>$60,000	-0.879	-1.984	0.226	0.118	-0.480	-1.782	0.823	0.468
Single	0.209	-0.636	1.053	0.626	0.773	-0.188	1.734	0.114

	Model 3				Model 4			
	β	95% CI		p-value	β	95% CI		p-value
IH	**0.637***	0.469	0.805	0.000	**0.784***	0.584	0.984	0.000
RES	**-0.090***	-0.166	-0.014	0.021	**-0.091***	-0.165	-0.016	0.017
IH x RES	--	--	--	--	**-0.034***	-0.060	-0.008	0.011
Age								
18-24		(Ref)				(Ref)		
25-34	-0.748	-1.876	0.380	0.192	-0.957	-2.075	0.162	0.093
35+	-0.259	-1.491	0.973	0.679	-0.438	-1.655	0.779	0.478
Income								
<$30,000		(Ref)				(Ref)		
$30-60,000	-0.158	-1.224	0.908	0.770	0.143	-0.929	1.214	0.793
>$60,000	-0.632	-1.740	0.476	0.262	-0.457	-1.552	0.639	0.411
Single	0.186	-0.645	1.018	0.658	0.262	-0.556	1.080	0.528

*Statistically significant at α=0.05

Table 4A.7b Linear Regression of Depressive Symptoms on INTERNALIZED HOMOPHOBIA and SELF-EFFICACY

	Model 1				Model 2			
	β	95% CI		p-value	β	95% CI		p-value
IH	**0.652***	0.482	0.822	0.000	--	--	--	--
SE	--	--	--	--	**-0.308***	-0.455	-0.162	0.000
IH x SE	--	--	--	--	--	--	--	--
Age								
18-24		(Ref)				(Ref)		
25-34	-0.644	-1.786	0.498	0.267	**-2.134***	-3.404	-0.864	0.001
35+	-0.035	-1.271	1.201	0.956	**-1.746***	-3.131	-0.360	0.014
Income								
<$30,000		(Ref)				(Ref)		
$30-60,000	-0.366	-1.434	0.702	0.499	0.236	-0.957	1.428	0.697
>$60,000	-0.879	-1.984	0.226	0.118	-0.291	-1.548	0.966	0.647
Single	0.209	-0.636	1.053	0.626	0.739	-0.188	1.667	0.117

	Model 3				Model 4			
	β	95% CI		p-value	β	95% CI		p-value
IH	**0.599***	0.432	0.766	0.000	**0.828***	0.547	1.109	0.000
SE	**-0.224***	-0.352	-0.095	0.001	**-0.202***	-0.331	-0.073	0.002
IH x SE	--	--	--	--	**-0.056***	-0.113	0.000	0.048
Age								
18-24		(Ref)				(Ref)		
25-34	-1.106	-2.239	0.027	0.056	-1.180	-2.304	-0.056	0.040
35+	-0.607	-1.844	0.630	0.334	-0.714	-1.944	0.515	0.253
Income								
<$30,000		(Ref)				(Ref)		
$30-60,000	-0.177	-1.213	0.859	0.736	0.184	-0.902	1.271	0.738
>$60,000	-0.514	-1.600	0.573	0.351	-0.359	-1.446	0.727	0.514
Single	0.201	-0.613	1.015	0.626	0.245	-0.563	1.052	0.550

*Statistically significant at α=0.05

Figure 4A.1 Slope of IH at different levels of protective variables

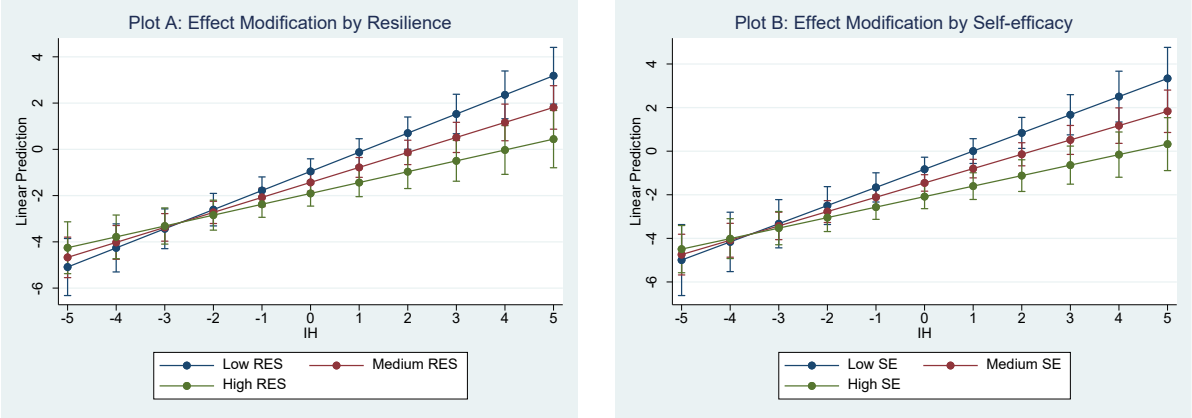
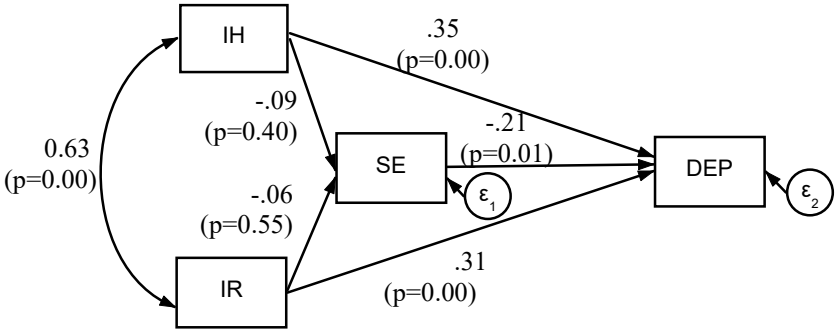


Figure 4A.2 Path model with standardized coefficients



Chapter 5: Conclusions & Public Health Significance

Despite recent efforts to stem health inequities, BMSM continue to be disproportionally affected by negative health outcomes. A vast body of research has identified the negative health impacts of different forms of stigma, but most studies have focused on a single stigmatized identity. This study illustrates the complex processes by which co-occurring forms of internalized stigma related to race and sexual orientation is associated with depressive symptoms among BMSM. Furthermore, we found that self-efficacy and resilience moderated the association between internalized homophobia and depressive symptoms. As we develop public health strategies to combat the HIV epidemic, it is important to recognize that people belong to more than one group or have more than one group identity, and they may experience unique barriers or inequities and may have unique strengths and assets as a result of their intersecting identities and characteristics.

Our findings illustrate the utility of Rasch analysis for modeling internalized stigma mechanisms, assessing individual item and rating scale functioning, establishing the hierarchy of internalizing beliefs and attitudes, and improving the reliability and validity of the measures of the magnitude of internalized racism and internalized homophobia yielded from these scales. A clearer understanding of the nuanced nature of these constructs is needed to more fully understand these complex relationships. Utilizing modern test theory techniques such as Rasch analysis serves as valuable tools to improve the precision and validity with which we measure these psychosocial risk and protective factors. Such precision and validity are necessary to successfully develop interventions that address multiple forms of internalized stigma and foster positive self-evaluation and coping skills that may reduce the negative consequences of internalized stigma on health outcomes.

References

1. Centers for Disease Control and Prevention. HIV Surveillance Report, 2019. vol. 32. https://www.cdc.gov/hiv/library/reports/hiv-surveillance.html
2. Matthews DD, Herrick AL, Coulter RW, et al. Running Backwards: Consequences of Current HIV Incidence Rates for the Next Generation of Black MSM in the United States. *AIDS Behav*. Jan 2016;20(1):7-16. doi:10.1007/s10461-015-1158-z
3. Lieb S, Fallon SJ, Friedman SR, et al. Statewide estimation of racial/ethnic populations of men who have sex with men in the US. *Public health reports*. 2011;126(1):60-72.
4. Millett GA, Peterson JL, Wolitski RJ, Stall R. Greater risk for HIV infection of black men who have sex with men: a critical literature review. *American journal of public health*. 2006;96(6):1007-1019.
5. Maulsby C, Millett G, Lindsey K, et al. HIV among Black men who have sex with men (MSM) in the United States: a review of the literature. *AIDS Behav*. Jan 2014;18(1):10-25. doi:10.1007/s10461-013-0476-2
6. Vitinghoff E, Douglas J, Judon F, McKiman D, MacQueen K, Buchinder SP. Per-contact risk of human immunodificiency virus tramnsmision between male sexual partners. *American journal of epidemiology*. 1999;150(3):306-311.
7. Millett GA, Flores SA, Peterson JL, Bakeman R. Explaining disparities in HIV infection among black and white men who have sex with men: a meta-analysis of HIV risk behaviors. *Aids*. 2007;21(15):2083-2091.
8. Frye V, Nandi V, Egan J, et al. Sexual orientation-and race-based discrimination and sexual HIV risk behavior among urban MSM. *AIDS and Behavior*. 2015;19(2):257-269.
9. Commission on Social Determinants of Health. *Closing the gap in a generation : health equity through action on the social determinants of health : Commission on Social Determinants of Health final report*. World Health Organization, Commission on Social Determinants of Health; 2008.
10. Institute of Medicine Committee on Understanding and Eliminating Racial and Ethnic Disparities in Health Care. In: Smedley BD, Stith AY, Nelson AR, eds. *Unequal Treatment: Confronting Racial and Ethnic Disparities in Health Care*. National Academies Press (US); 2003.
11. Lewis TT, Cogburn CD, Williams DR. Self-reported experiences of discrimination and health: scientific advances, ongoing controversies, and emerging issues. *Annu Rev Clin Psychol*. 2015;11:407-40. doi:10.1146/annurev-clinpsy-032814-112728
12. HHS Announces New National Challenge to Reduce HIV-Related Stigma and Disparities. July 22,2021, 2021. Accessed September 20, 2021. https://www.hhs.gov/about/news/2021/07/22/hhs-announces-new-national-challege-to-reduce-hiv-stigma-and-disparities.html
13. Link BG, Phelan JC. Conceptualizing stigma. *Annual review of Sociology*. 2001;27(1):363-385.
14. Layland EK, Carter JA, Perry NS, et al. A systematic review of stigma in sexual and gender minority health interventions. *Transl Behav Med*. Oct 12 2020;10(5):1200-1210. doi:10.1093/tbm/ibz200

15. Logie CH, James L, Tharao W, Loutfy MR. HIV, gender, race, sexual orientation, and sex work: a qualitative study of intersectional stigma experienced by HIV-positive women in Ontario, Canada. *PLoS medicine*. Nov 2011;8(11):e1001124. doi:10.1371/journal.pmed.1001124
16. Flentje A, Heck NC, Brennan JM, Meyer IH. The relationship between minority stress and biological outcomes: A systematic review. *Journal of Behavioral Medicine*. 2020;43(5):673-694.
17. Mak WWS, Mo PKH, Ma GYK, Lam MYY. Meta-analysis and systematic review of studies on the effectiveness of HIV stigma reduction programs. *Soc Sci Med*. Sep 2017;188:30-40. doi:10.1016/j.socscimed.2017.06.045
18. Meyer IH, Frost DM. Minority stress and the health of sexual minorities. 2013;
19. Meyer IH. Minority stress and mental health in gay men. *J Health Soc Behav*. Mar 1995;36(1):38-56.
20. Meyer IH. Prejudice, social stress, and mental health in lesbian, gay, and bisexual populations: conceptual issues and research evidence. *Psychol Bull*. Sep 2003;129(5):674-697. doi:10.1037/0033-2909.129.5.674
21. Alvy LM, McKirnan DJ, Mansergh G, et al. Depression is associated with sexual risk among men who have sex with men, but is mediated by cognitive escape and self-efficacy. *AIDS and Behavior*. 2011;15(6):1171-1179.
22. Batchelder AW, Safren S, Mitchell AD, Ivardic I, O'Cleirigh C. Mental health in 2020 for men who have sex with men in the United States. *Sexual health*. 2017;14(1):59-71.
23. Gonzalez JS, Batchelder AW, Psaros C, Safren SA. Depression and HIV/AIDS treatment nonadherence: a review and meta-analysis. *J Acquir Immune Defic Syndr*. Oct 1 2011;58(2):181-7. doi:10.1097/QAI.0b013e31822d490a
24. Arnold EA, Rebchook GM, Kegeles SM. 'Triply cursed': racism, homophobia and HIV-related stigma are barriers to regular HIV testing, treatment adherence and disclosure among young Black gay men. *Culture, health & sexuality*. 2014;16(6):710-722.
25. Stangl AL, Earnshaw VA, Logie CH, et al. The Health Stigma and Discrimination Framework: a global, crosscutting framework to inform research, intervention development, and policy on health-related stigmas. *BMC Med*. Feb 15 2019;17(1):31. doi:10.1186/s12916-019-1271-3
26. Latkin CA, Van Tieu H, Fields S, et al. Social network factors as correlates and predictors of high depressive symptoms among black men who have sex with men in HPTN 061. *AIDS and Behavior*. 2017;21(4):1163-1170.
27. Logie CH, Earnshaw V, Nyblade L, et al. A scoping review of the integration of empowerment-based perspectives in quantitative intersectional stigma research. *Glob Public Health*. Aug 2022;17(8):1451-1466. doi:10.1080/17441692.2021.1934061
28. Herrick AL, Stall R, Goldhammer H, Egan JE, Mayer KH. Resilience as a research framework and as a cornerstone of prevention research for gay and bisexual men: Theory and evidence. *AIDS and Behavior*. 2014;18(1):1-9.
29. Earnshaw VA, Bogart LM, Dovidio JF, Williams DR. Stigma and racial/ethnic HIV disparities: moving toward resilience. 2015;
30. Ostrow DG, Whitaker RE, Frasier K, et al. Racial differences in social support and mental health in men with HIV infection: a pilot study. *AIDS Care*. 1991;3(1):55-62. doi:10.1080/09540129108253047

31. Brewer R, Hood KB, Moore M, et al. An exploratory study of resilience, HIV-related stigma, and HIV care outcomes among men who have sex with men (MSM) living with HIV in Louisiana. *AIDS and Behavior*. 2020;24(7):2119-2129.
32. McNair OS, Gipson JA, Denson D, Thompson DV, Sutton MY, Hickson DA. The associations of resilience and HIV risk behaviors among black gay, bisexual, other men who have sex with men (MSM) in the Deep South: The MARI study. *AIDS and Behavior*. 2018;22(5):1679-1687.
33. Dawson EL, Mendoza MC, Gaul Z, Jeffries IV WL, Sutton MY, Wilson PA. Resilience, condom use self-efficacy, internalized homophobia, and condomless anal sex among black men who have sex with men, New York City. *PloS one*. 2019;14(4):e0215455.
34. Peterson JL, Bakeman R, Sullivan P, et al. Social discrimination and resiliency are not associated with differences in prevalent HIV infection in black and white men who have sex with men. *J Acquir Immune Defic Syndr*. Aug 15 2014;66(5):538-43. doi:10.1097/QAI.0000000000000203
35. Hallfors DD, Iritani BJ, Miller WC, Bauer DJ. Sexual and drug behavior patterns and HIV and STD racial disparities: the need for new directions. *Am J Public Health*. Jan 2007;97(1):125-32. doi:10.2105/AJPH.2005.075747
36. Davis K. Intersectionality as buzzword: A sociology of science perspective on what makes a feminist theory successful. *Feminist theory*. 2008;9(1):67-85.
37. Bowleg L. The problem with the phrase women and minorities: intersectionality-an important theoretical framework for public health. *Am J Public Health*. Jul 2012;102(7):1267-73. doi:10.2105/ajph.2012.300750
38. Warner LR, Shields SA. The intersections of sexuality, gender, and race: Identity research at the crossroads. *Sex roles*. 2013;68(11):803-810.
39. Colpitts E, Gahagan J. The utility of resilience as a conceptual framework for understanding and measuring LGBTQ health. *International journal for equity in health*. 2016;15(1):1-8.
40. Reed SJ, Miller RL. Thriving and adapting: Resilience, sense of community, and syndemics among young Black gay and bisexual men. *American journal of community psychology*. 2016;57(1-2):129-143.
41. Allport GW, Clark K, Pettigrew T. The nature of prejudice. 1954;
42. Crocker J, Major B. Social stigma and self-esteem: The self-protective properties of stigma. *Psychological review*. 1989;96(4):608.
43. Goffman E. *Stigma: Notes on the management of spoiled identity*. Simon and Schuster; 2009.
44. Jones EE. *Social stigma: The psychology of marked relationships*. WH Freeman; 1984.
45. Newcomb ME, Mustanski B. Internalized homophobia and internalizing mental health problems: a meta-analytic review. *Clin Psychol Rev*. Dec 2010;30(8):1019-29. doi:10.1016/j.cpr.2010.07.003
46. Jackson-Best F, Edwards N. Stigma and intersectionality: a systematic review of systematic reviews across HIV/AIDS, mental illness, and physical disability. *BMC Public Health*. Jul 27 2018;18(1):919. doi:10.1186/s12889-018-5861-3
47. Hatzenbuehler ML, Phelan JC, Link BG. Stigma as a fundamental cause of population health inequalities. *American journal of public health*. 2013;103(5):813-821.
48. Warner LR. A best practices guide to intersectional approaches in psychological research. *Sex roles*. 2008;59(5):454-463.

49. Turan JM, Elafros MA, Logie CH, et al. Challenges and opportunities in examining and addressing intersectional stigma and health. *BMC Med*. Feb 15 2019;17(1):7. doi:10.1186/s12916-018-1246-9
50. Shangani S, Gamarel KE, Ogunbajo A, Cai J, Operario D. Intersectional minority stress disparities among sexual minority adults in the USA: The role of race/ethnicity and socioeconomic status. *Culture, health & sexuality*. 2020;22(4):398-412.
51. Jackson SD, Mohr JJ, Sarno EL, Kindahl AM, Jones IL. Intersectional experiences, stigma-related stress, and psychological health among Black LGBQ individuals. *Journal of consulting and clinical psychology*. 2020;88(5):416.
52. McCall L. The complexity of intersectionality. *Signs: Journal of women in culture and society*. 2005;30(3):1771-1800.
53. Crenshaw K. Mapping the margins: Intersectionality, identity politics, and violence against women of color. *Stan L Rev*. 1990;43:1241.
54. Grabe S. Research Methods in the Study of Intersectionality in Psychology: Examples Informed by a Decade of Collaborative Work With Majority World Women's Grassroots Activism. *Frontiers in Psychology*. 2020;11:2849.
55. Bauer GR. Incorporating intersectionality theory into population health research methodology: challenges and the potential to advance health equity. *Social science & medicine*. 2014;110:10-17.
56. Syed M. Disciplinarity and methodology in intersectionality theory and research. *The American psychologist*. Jan 2010;65(1):61-2. doi:10.1037/a0017495
57. Cole ER. Intersectionality and research in psychology. *The American psychologist*. Apr 2009;64(3):170-80. doi:10.1037/a0014564
58. Else-Quest NM, Hyde JS. Intersectionality in Quantitative Psychological Research. *Psychology of Women Quarterly*. 2016;40(2):155-170. doi:10.1177/0361684316629797
59. Bowleg L. When Black+ lesbian+ woman≠ Black lesbian woman: The methodological challenges of qualitative and quantitative intersectionality research. *Sex roles*. 2008;59(5):312-325.
60. Flake JK, Pek J, Hehman E. Construct validation in social and personality research: Current practice and recommendations. *Social Psychological and Personality Science*. 2017;8(4):370-378.
61. Flake JK, Fried EI. Measurement schmeasurement: Questionable measurement practices and how to avoid them. *Advances in Methods and Practices in Psychological Science*. 2020;3(4):456-465.
62. Crutzen R, Peters G-JY. Scale quality: alpha is an inadequate estimate and factor-analytic evidence is needed first of all. *Health psychology review*. 2017;11(3):242-247.
63. Boateng GO, Neilands TB, Frongillo EA, Melgar-Quinonez HR, Young SL. Best Practices for Developing and Validating Scales for Health, Social, and Behavioral Research: A Primer. *Front Public Health*. 2018;6:149. doi:10.3389/fpubh.2018.00149
64. Steegen S, Tuerlinckx F, Gelman A, Vanpaemel W. Increasing transparency through a multiverse analysis. *Perspectives on Psychological Science*. 2016;11(5):702-712.
65. McNeish D, Wolf MG. Thinking twice about sum scores. *Behavior research methods*. 2020;52(6):2287-2305.
66. Bond TG, Fox CM. *Applying the Rasch model: Fundamental measurement in the human sciences*. Psychology Press; 2013.

67. Streiner DL, Norman GR, Cairney J. *Health measurement scales: a practical guide to their development and use*. Oxford University Press, USA; 2015.
68. Lerdal A, Kottorp A, Gay CL, Lee KA. Lee fatigue and energy scales: exploring aspects of validity in a sample of women with HIV using an application of a Rasch model. *Psychiatry research*. 2013;205(3):241-246.
69. Johansson S, Kottorp A, Lee KA, Gay CL, Lerdal A. Can the Fatigue Severity Scale 7-item version be used across different patient populations as a generic fatigue measure-a comparative study using a Rasch model approach. *Health and quality of life outcomes*. 2014;12(1):1-9.
70. Siemons L, Ten Klooster PM, Taal E, Glas CA, Van de Laar MA. Modern psychometrics applied in rheumatology–A systematic review. *BMC musculoskeletal disorders*. 2012;13(1):1-10.
71. Belvedere SL, de Morton NA. Application of Rasch analysis in health care is increasing and is applied for variable reasons in mobility instruments. *Journal of clinical epidemiology*. 2010;63(12):1287-1297.
72. Dabaghi S, Esmaielzadeh F, Rohani C. Application of Rasch Analysis for Development and Psychometric Properties of Adolescents' Quality of Life Instruments: A Systematic Review. *Adolescent health, medicine and therapeutics*. 2020;11:173-197. doi:10.2147/ahmt.s265413
73. Tesio L, Simone A, Bernardinello M. Rehabilitation and outcome measurement: where is Rasch analysis-going? *Europa medicophysica*. 2007;43(3):417.
74. Boone WJ. Rasch analysis for instrument development: why, when, and how? *CBE—Life Sciences Education*. 2016;15(4):rm4.
75. Boone WJ, Staver JR, Yale MS. *Rasch analysis in the human sciences*. Springer; 2013.
76. Aryadoust V, Tan HAH, Ng LY. A Scientometric Review of Rasch Measurement: The Rise and Progress of a Specialty. Systematic Review. *Front Psychol*. 2019-October-22 2019;10(2197):2197. doi:10.3389/fpsyg.2019.02197
77. Mahajan AP, Sayles JN, Patel VA, et al. Stigma in the HIV/AIDS epidemic: a review of the literature and recommendations for the way forward. *AIDS*. Aug 2008;22 Suppl 2(Suppl 2):S67-79. doi:10.1097/01.aids.0000327438.13291.62
78. Dulin AJ, Dale SK, Earnshaw VA, et al. Resilience and HIV: a review of the definition and study of resilience. *AIDS care*. Aug 2018;30(sup5):S6-S17. doi:10.1080/09540121.2018.1515470
79. Babel RA, Wang P, Alessi EJ, Raymond HF, Wei C. Stigma, HIV Risk, and Access to HIV Prevention and Treatment Services Among Men Who have Sex with Men (MSM) in the United States: A Scoping Review. *AIDS and Behavior*. 2021:1-31.
80. Earnshaw VA, Chaudoir SR. From conceptualizing to measuring HIV stigma: a review of HIV stigma mechanism measures. *AIDS Behav*. Dec 2009;13(6):1160-77. doi:10.1007/s10461-009-9593-3
81. Clark LA, Watson D. Constructing validity: Basic issues in objective scale development. 2016;
82. Wright BD. A history of social science measurement. *Educational measurement: Issues and practice*. 1997;16(4):33-45.
83. Hambleton RK, Swaminathan H, Rogers HJ. *Fundamentals of item response theory*. vol 2. Sage; 1991.
84. Wright BD. Fundamental measurement for psychology. *The new rules of measurement: What every psychologist and educator should know*. 1999:65-104.

85. Aryadoust V, Ng LY, Sayama H. A comprehensive review of Rasch measurement in language assessment: Recommendations and guidelines for research. *Language Testing*. 2021;38(1):6-40.
86. Page MJ, McKenzie JE, Bossuyt PM, et al. The PRISMA 2020 statement: an updated guideline for reporting systematic reviews. *Bmj*. 2021;372
87. Tennant A, Conaghan PG. The Rasch measurement model in rheumatology: what is it and why use it? When should it be applied, and what should one look for in a Rasch paper? *Arthritis Care & Research*. 2007;57(8):1358-1362.
88. *Winsteps® Rasch measurement computer program.* Version 5.2.2. Portland, Oregon; 2022. Winsteps.com
89. Linacre J. A User's Guide to WINSTEPS® Rasch-Model Computer Programs: Program Manual 4.4. 6. *Mesa-Press, Chicago, IL*. 2019;
90. Siemons L, Ten Klooster PM, Taal E, Glas CA, Van de Laar MA. Modern psychometrics applied in rheumatology--a systematic review. *BMC Musculoskelet Disord*. Oct 31 2012;13(1):216. doi:10.1186/1471-2474-13-216
91. Stolt M, Kottorp A, Suhonen R. The use and quality of reporting of Rasch analysis in nursing research: a methodological scoping review. *International Journal of Nursing Studies*. 2022:104244.
92. de Bock E, Williams P, Tugaut B, Guillemin. Four Decades of Rasch Analysis on Patient-Reported Outcomes Instruments Validation: A Systematic Literature Review. *Value in Health*. 2016/11/01/ 2016;19(7):A362. doi:10.1016/j.jval.2016.09.090
93. Lerdal A, Kottorp A, Gay C, Aouizerat BE, Portillo CJ, Lee KA. A 7-item version of the fatigue severity scale has better psychometric properties among HIV-infected adults: an application of a Rasch model. *Quality of life research : an international journal of quality of life aspects of treatment, care and rehabilitation*. Nov 2011;20(9):1447-56. doi:10.1007/s11136-011-9877-8
94. Brouillette MJ, Fellows LK, Finch L, Thomas R, Mayo NE. Properties of a brief assessment tool for longitudinal measurement of cognition in people living with HIV. *PLoS One*. 2019;14(3):e0213908. doi:10.1371/journal.pone.0213908
95. Mayers AM, Khoo ST, Svartberg M. The Existential Loneliness Questionnaire: background, development, and preliminary findings. *J Clin Psychol*. Sep 2002;58(9):1183-93. doi:10.1002/jclp.10038
96. Fendrich M, Smith EV, Jr., Pollack LM, Mackesy-Amiti ME. Measuring sexual risk for HIV: a Rasch scaling approach. *Arch Sex Behav*. Dec 2009;38(6):922-35. doi:10.1007/s10508-008-9385-2
97. Johansson S, Kottorp A, Lee KA, Gay CL, Lerdal A. Can the Fatigue Severity Scale 7-item version be used across different patient populations as a generic fatigue measure - a comparative study using a Rasch model approach. *Health and Quality of Life Outcomes*. Feb 2014;1224. doi:10.1186/1477-7525-12-24
98. Lerdal A, Kottorp A, Gay CL, Lee KA. Development of a short version of the Lee Visual Analogue Fatigue Scale in a sample of women with HIV/AIDS: a Rasch analysis application. *Quality of life research : an international journal of quality of life aspects of treatment, care and rehabilitation*. Aug 2013;22(6):1467-72. doi:10.1007/s11136-012-0279-3
99. Brouillette MJ, Mayo N, Fellows LK, et al. A better screening tool for HIV-associated neurocognitive disorders: is it what clinicians need? *Aids*. May 15 2015;29(8):895-902. doi:10.1097/qad.0000000000000152

100. Koski L, Brouillette MJ, Lalonde R, et al. Computerized testing augments pencil-and-paper tasks in measuring HIV-associated mild cognitive impairment(*). *HIV Med*. Sep 2011;12(8):472-80. doi:10.1111/j.1468-1293.2010.00910.x
101. Brouillette MJ, Fellows LK, Palladini L, Finch L, Thomas R, Mayo NE. Quantifying cognition at the bedside: a novel approach combining cognitive symptoms and signs in HIV. *BMC Neurol*. Nov 13 2015;15:224. doi:10.1186/s12883-015-0483-1
102. Kim S, Ouellet LJ, Mazza J, Spaulding AC. Rasch Analysis and Differential Item Functioning of a Social Support Measure in Jail Inmates With HIV Infection. *Eval Health Prof*. Mar 1 2017;40(1):33-60. doi:10.1177/0163278716644954
103. Kim S, Mazza J. Reliability, Validity, and Item Response of MOS Social Support Score among Incarcerated Women. *Women Crim Justice*. Jan 1 2014;24(1):1-21. doi:10.1080/08974454.2012.733327
104. Huang F, Chen WT, Lin SH, et al. Translation, Adaption, and Psychometric Testing of the Myanmar Version of the Medical Outcomes Study Social Support Survey for People Living With HIV/AIDS. *Front Psychol*. 2021;12:707142. doi:10.3389/fpsyg.2021.707142
105. Wright BD. Comparing Rasch measurement and factor analysis. *Structural Equation Modeling: A Multidisciplinary Journal*. 1996/01/01 1996;3(1):3-24. doi:10.1080/10705519609540026
106. Askari S, Fellows LK, Brouillette MJ, Mayo NE. Development and validation of a voice-of-the-patient measure of cognitive concerns experienced by people living with HIV. *Quality of life research : an international journal of quality of life aspects of treatment, care and rehabilitation*. Mar 2021;30(3):921-930. doi:10.1007/s11136-020-02679-z
107. Brouillette MJ, Fellows LK, Palladini L, Finch L, Thomas R, Mayo NE. Quantifying cognition at the bedside: a novel approach combining cognitive symptoms and signs in HIV. *BMC Neurol*. Nov 13 2015;15(1):224. doi:10.1186/s12883-015-0483-1
108. Chang CH, Wright BD, Cella D, Hays RD. The SF-36 physical and mental health factors were confirmed in cancer and HIV/AIDS patients. *J Clin Epidemiol*. Jan 2007;60(1):68-72. doi:10.1016/j.jclinepi.2006.07.001
109. Chang KC, Wang JD, Tang HP, Cheng CM, Lin CY. Psychometric evaluation, using Rasch analysis, of the WHOQOL-BREF in heroin-dependent people undergoing methadone maintenance treatment: further item validation. *Health Qual Life Outcomes*. Oct 3 2014;12:148. doi:10.1186/s12955-014-0148-6
110. Chen YL, Hsiung PC, Chung L, Chen SC, Pan AW. Psychometric properties of the Mastery Scale-Chinese version: applying classical test theory and Rasch analysis. *Scand J Occup Ther*. Nov 2013;20(6):404-11. doi:10.3109/11038128.2013.838999
111. Conrad KJ, Conrad KM, Mazza J, et al. Dimensionality, hierarchical structure, age generalizability, and criterion validity of the GAIN's Behavioral Complexity Scale. *Psychol Assess*. Dec 2012;24(4):913-24. doi:10.1037/a0028196
112. Gay CL, Kottorp A, Lerdal A, Lee KA. Psychometric Limitations of the Center for Epidemiologic Studies-Depression Scale for Assessing Depressive Symptoms among Adults with HIV/AIDS: A Rasch Analysis. *Depression research and treatment*. 2016;2016:2824595. doi:10.1155/2016/2824595
113. Gunardi I, Salsabila Nurina N, Marcia, Amtha R. Dentists experience influences knowledge and attitudes toward HIV patients in West Jakarta, Indonesia, and validation of a new questionnaire. *Oral Dis*. Sep 2020;26 Suppl 1:127-132. doi:10.1111/odi.13393

114. Hernández L, Chang CH, Cella D, Corona M, Shiomoto G, McGuire DB. Development and validation of the satisfaction with pharmacist scale. *Pharmacotherapy*. Jul 2000;20(7):837-43. doi:10.1592/phco.20.9.837.35188
115. Hlavinkova L, Mentel A, Kollarova J, Kristufkova Z. Effectiveness of a prevention campaign on HIV/AIDS knowledge among adolescents in Eastern Slovakia. *Int J Public Health*. Dec 2014;59(6):905-11. doi:10.1007/s00038-014-0607-1
116. Hu J, Serovich JM, Brown MJ, Kimberly JA, Chen YH. Validity Study of the R-PLA, a Resilience Scale for People Living with HIV. *Journal of HIV and AIDS*. Dec 2019;5(3)doi:10.16966/2380-5536.170
117. Hu J, Serovich JM, Chen YH, Brown MJ, Kimberly JA. Psychometric Evaluation of the HIV Disclosure Belief Scale: A Rasch Model Approach. *AIDS Behav*. Jan 2017;21(1):174-183. doi:10.1007/s10461-016-1478-7
118. Huang F, Chen WT, Shiu CS, et al. Adaptation and Validation of the Cognitive and Affective Mindfulness Scale-Revised (CAMS-R) in People Living with HIV in Myanmar. *Mindfulness (N Y)*. 2022;13(1):188-197. doi:10.1007/s12671-021-01784-5
119. Huang F, Chen WT, Shiu CS, et al. Adaptation and validation of a culturally adapted HIV stigma scale in Myanmar. *BMC Public Health*. Sep 13 2021;21(1):1663. doi:10.1186/s12889-021-11685-w
120. Leplège A, Rude N, Ecosse E, Ceinos R, Dohin E, Pouchot J. Measuring quality of life from the point of view of HIV-positive subjects: the HIV-QL31. *Quality of life research : an international journal of quality of life aspects of treatment, care and rehabilitation*. Aug 1997;6(6):585-94. doi:10.1023/a:1018468301617
121. Lerdal A, Kottorp A, Gay CL, Lee KA. Lee Fatigue And Energy Scales: exploring aspects of validity in a sample of women with HIV using an application of a Rasch model. *Psychiatry Res*. Feb 28 2013;205(3):241-6. doi:10.1016/j.psychres.2012.08.031
122. Lerdal A, Kottorp A, Gay C, et al. A 7-item version of the fatigue severity scale has better psychometric properties among HIV-infected adults: an application of a Rasch model. *Quality of Life Research*. 2011;20(9):1447-1456. doi:10.1007/s11136-011-9877-8
123. Li J, Liu HJ, Liu H, Feng TJ, Cai YM. Psychometric assessment of HIV/STI sexual risk scale among MSM: A Rasch model approach. *Bmc Public Health*. Oct 2011;11763. doi:10.1186/1471-2458-11-763
124. McCown W, Johnson J. The Basic HIV Disease Knowledge Questionnaire: a Rasch-scaled instrument to measure essential HIV knowledge. *Psychol Rep*. Oct 1991;69(2):543-9. doi:10.2466/pr0.1991.69.2.543
125. Medeiros ARC, de Lima R, de Medeiros LB, et al. Moderate and severe household food insecurity in families of people living with HIV/Aids: scale validation and associated factors. *Ciencia & Saude Coletiva*. Oct 2017;22(10):3353-3364. doi:10.1590/1413-812320172210.02462017
126. Merritt B, Gahagan J, Kottorp A. HIV and disability: a pilot study exploring the use of the Assessment of Motor and Process Skills to measure daily life performance. *J Int AIDS Soc*. Jan 21 2013;16(1):17339. doi:10.7448/IAS.16.1.17339
127. Mpofu E, Caldwell L, Smith E, et al. Rasch modeling of the structure of health risk behavior in South african adolescents. *Journal of applied measurement*. 2006;7(3):323-34.
128. Mugendawala H, Hagedorn LS. The effect of HIV/AIDS awareness and support initiatives on the perceived risk of HIV infection among head teachers in Ugandan schools.

Journal of HIV/AIDS & Social Services. 2017;16(4):317-334. doi:10.1080/15381501.2016.1274702
129. Munambah N, Cordier R, Chiwaridzo M, Ramugondo EL. Play profiles of children with HIV/Aids: A comparative study. *Aust Occup Ther J*. Apr 2021;68(2):156-168. doi:10.1111/1440-1630.12712
130. Naar S, Chapman J, Cunningham PB, Ellis D, MacDonell K, Todd L. Development of the Motivational Interviewing Coach Rating Scale (MI-CRS) for health equity implementation contexts. *Health psychology : official journal of the Division of Health Psychology, American Psychological Association*. Jul 2021;40(7):439-449. doi:10.1037/hea0001064
131. O'Brien KK, Dzingina M, Harding R, et al. Developing a short-form version of the HIV Disability Questionnaire (SF-HDQ) for use in clinical practice: a Rasch analysis. *Health Qual Life Outcomes*. Jan 6 2021;19(1):6. doi:10.1186/s12955-020-01643-2
132. Packham TL, Cappelleri JC, Sadosky A, MacDermid JC, Brunner F. Measurement properties of painDETECT: Rasch analysis of responses from community-dwelling adults with neuropathic pain. *BMC Neurol*. Mar 4 2017;17(1):48. doi:10.1186/s12883-017-0825-2
133. Pan AW, Chung L, Fife BL, Hsiung PC. Evaluation of the psychometrics of the Social Impact Scale: a measure of stigmatization. *International journal of rehabilitation research Internationale Zeitschrift fur Rehabilitationsforschung Revue internationale de recherches de readaptation*. Sep 2007;30(3):235-8. doi:10.1097/MRR.0b013e32829fb3db
134. Peterman AH, Cella D, Mo F, McCain N. Psychometric validation of the revised Functional Assessment of Human Immunodeficiency Virus Infection (FAHI) quality of life instrument. *Quality of life research : an international journal of quality of life aspects of treatment, care and rehabilitation*. Aug 1997;6(6):572-84. doi:10.1023/a:1018416317546
135. Pilatti A, Read JP, Vera Bdel V, Caneto F, Garimaldi JA, Kahler CW. The Spanish version of the Brief Young Adult Alcohol Consequences Questionnaire (B-YAACQ): a Rasch model analysis. *Addict Behav*. May 2014;39(5):842-7. doi:10.1016/j.addbeh.2014.01.026
136. Rose JS, Laux JM, Fox CM, Tiamiyu MF, O'Hara CL. Development and validation of the mental health professionals' attitude towards people living with HIV/AIDS scale (MHP-PLHIV-AS). *AIDS Care*. Jan-Dec 2020;32(sup1):10-18. doi:10.1080/09540121.2020.1822503
137. Selman L, Siegert RJ, Higginson IJ, et al. The "Spirit 8" successfully captured spiritual well-being in African palliative care: factor and Rasch analysis. *J Clin Epidemiol*. Apr 2012;65(4):434-43. doi:10.1016/j.jclinepi.2011.09.014
138. Shim EJ, Ha H, Lee SH, et al. Psychometric properties of the Korean version of the medical outcomes study HIV health survey: results from a multicenter survey in Korea. *Health Qual Life Outcomes*. May 15 2018;16(1):92. doi:10.1186/s12955-018-0919-6
139. Thandi G, Fear NT, Chalder T. A comparison of the Work and Social Adjustment Scale (WSAS) across different patient populations using Rasch analysis and exploratory factor analysis. *J Psychosom Res*. Jan 2017;92:45-48. doi:10.1016/j.jpsychores.2016.11.009
140. Yen CF, Huang YT, Potenza MN, Tsai TT, Lin CY, Tsang HWH. Measure of Internalized Sexual Stigma for Lesbians and Gay Men (MISS-LG) in Taiwan: Psychometric Evidence from Rasch and Confirmatory Factor Analysis. *Int J Environ Res Public Health*. Dec 18 2021;18(24)doi:10.3390/ijerph182413352
141. Prevention CfDCa. Estimated HIV Incidence and Prevalence in the United States, 2015-2019. Updated May 2021. 26. http://www.cdc.gov/hiv/library/reports/hiv-surveillance.html

142. King M, Semlyen J, Tai SS, et al. A systematic review of mental disorder, suicide, and deliberate self harm in lesbian, gay and bisexual people. *BMC Psychiatry*. Aug 18 2008;8(1):70. doi:10.1186/1471-244X-8-70
143. Lucassen MF, Stasiak K, Samra R, Frampton CM, Merry SN. Sexual minority youth and depressive symptoms or depressive disorder: A systematic review and meta-analysis of population-based studies. *Aust N Z J Psychiatry*. Aug 2017;51(8):774-787. doi:10.1177/0004867417713664
144. Yıldız E. Suicide in sexual minority populations: A systematic review of evidence-based studies. *Archives of psychiatric nursing*. 2018;32(4):650-659.
145. Cochran SD, Mays VM. Lifetime prevalence of suicide symptoms and affective disorders among men reporting same-sex sexual partners: results from NHANES III. *Am J Public Health*. Apr 2000;90(4):573-8. doi:10.2105/ajph.90.4.573
146. Major B, Mendes WB, Dovidio JF. Intergroup relations and health disparities: a social psychological perspective. *Health Psychol*. May 2013;32(5):514-24. doi:10.1037/a0030358
147. Millett GA, Peterson JL, Flores SA, et al. Comparisons of disparities and risks of HIV infection in black and other men who have sex with men in Canada, UK, and USA: a meta-analysis. *The Lancet*. 2012;380(9839):341-348.
148. Huebner DM, Davis MC, Nemeroff CJ, Aiken LS. The impact of internalized homophobia on HIV preventive interventions. *Am J Community Psychol*. Jun 2002;30(3):327-48. doi:10.1023/a:1015325303002
149. Crawford I, Allison KW, Zamboni BD, Soto T. The influence of dual-identity development on the psychosocial functioning of African-American gay and bisexual men. *Journal of sex research*. Aug 2002;39(3):179-89. doi:10.1080/00224490209552140
150. Van Brakel WH. Measuring health-related stigma—a literature review. *Psychology, health & medicine*. 2006;11(3):307-334.
151. Nyblade LC. Measuring HIV stigma: existing knowledge and gaps. *Psychol Health Med*. Aug 2006;11(3):335-45. doi:10.1080/13548500600595178
152. Holzemer WL, Uys LR, Chirwa ML, et al. Validation of the HIV/AIDS stigma instrument—Plwa (Hasi-P). *AIDS care*. 2007;19(8):1002-1012.
153. Campón RR, Carter RT. The Appropriated Racial Oppression Scale: Development and preliminary validation. *Cultural Diversity and Ethnic Minority Psychology*. 2015;21(4):497.
154. Herek GM, Cogan JC, Gillis JR, Glunt EK. Correlates of internalized homophobia in a community sample of lesbians and gay men. *Journal-Gay and Lesbian Medical Association*. 1998;2:17-26.
155. Schwarzer R, Jerusalem M. Generalized self-efficacy scale. *Measures in health psychology: A user's portfolio Causal and control beliefs*. 1995;1(1):35-37.
156. Connor KM, Davidson JR. Development of a new resilience scale: the Connor-Davidson Resilience Scale (CD-RISC). *Depression and anxiety*. 2003;18(2):76-82. doi:10.1002/da.10113
157. Kroenke K, Spitzer RL. The PHQ-9: a new depression diagnostic and severity measure. SLACK Incorporated Thorofare, NJ; 2002.
158. Linacre J. Sample size and item calibration stability. *Rasch Mes Trans*. 1994;7:328.
159. Linacre JM. Optimizing rating scale category effectiveness. *Journal of applied measurement*. 2002;3(1):85-106.
160. Linacre JM. Investigating rating scale category utility. *J Outcome Meas*. 1999 1999;3(2):103-22.
161. Wright B. Reasonable mean-square fit values. *Rasch Meas Trans*. 1994;8:370.

162. Raîche G. Critical eigenvalue sizes in standardized residual principal components analysis. *Rasch measurement transactions*. 2005;19(1):1012.
163. Wright BD, Masters GN. *Rating scale analysis*. MESA press; 1982.
164. Zwick R, Thayer DT, Lewis C. An empirical Bayes approach to Mantel-Haenszel DIF analysis. *Journal of Educational Measurement*. 1999;36(1):1-28.
165. Andrich D. An expanded derivation of the threshold structure of the polytomous Rasch model that dispels any "threshold disorder controversy". *Educational and Psychological Measurement*. 2013;73(1):78-124.
166. Adams RJ, Wu ML, Wilson M. The Rasch rating model and the disordered threshold controversy. *Educational and Psychological Measurement*. 2012;72(4):547-573.
167. Yamaguchi J. Positive vs. negative wording. *Rasch Measurement Transactions*. 1997;11(2):567.
168. McCreary LL, Conrad KM, Conrad KJ, Scott CK, Funk RR, Dennis ML. Using the Rasch measurement model in psychometric analysis of the Family Effectiveness Measure. *Nursing research*. May-Jun 2013;62(3):149-59. doi:10.1097/NNR.0b013e31828eafe6
169. English D, Rendina HJ, Parsons JT. The effects of intersecting stigma: A longitudinal examination of minority stress, mental health, and substance use among Black, Latino, and multiracial gay and bisexual men. *Psychology of violence*. 2018;8(6):669.
170. Hussen SA, Easley KA, Smith JC, et al. Social capital, depressive symptoms, and HIV viral suppression among young black, gay, bisexual and other men who have sex with men living with HIV. *AIDS and Behavior*. 2018;22(9):3024-3032.
171. Graham LF, Aronson RE, Nichols T, Stephens CF, Rhodes SD. Factors influencing depression and anxiety among black sexual minority men. *Depression research and treatment*. 2011;2011
172. Jackson SD, Mohr JJ, Sarno EL, Kindahl AM, Jones IL. Intersectional experiences, stigma-related stress, and psychological health among Black LGBQ individuals. *J Consult Clin Psychol*. May 2020;88(5):416-428. doi:10.1037/ccp0000489
173. Bandura A. Self-efficacy: toward a unifying theory of behavioral change. *Psychol Rev*. Mar 1977;84(2):191-215. doi:10.1037//0033-295x.84.2.191
174. Quinn KG, Dickson-Gomez J, Pearson B, Marion E, Amikrhanian Y, Kelly JA. Intersectional Resilience Among Black Gay, Bisexual, and Other Men Who Have Sex With Men, Wisconsin and Ohio, 2019. *Am J Public Health*. Jun 2022;112(S4):S405-s412. doi:10.2105/ajph.2021.306677
175. Fergus S, Zimmerman MA. Adolescent resilience: a framework for understanding healthy development in the face of risk. *Annu Rev Public Health*. 2005;26(1):399-419. doi:10.1146/annurev.publhealth.26.021304.144357
176. Turan B, Crockett KB, Buyukcan-Tetik A, et al. Buffering Internalization of HIV Stigma: Implications for Treatment Adherence and Depression. *J Acquir Immune Defic Syndr*. Mar 1 2019;80(3):284-291. doi:10.1097/QAI.0000000000001915
177. Wilson PA, Valera P, Martos AJ, Wittlin NM, Munoz-Laboy MA, Parker RG. Contributions of Qualitative Research in Informing HIV/AIDS Interventions Targeting Black MSM in the United States. *Journal of sex research*. Jul-Aug 2016;53(6):642-54. doi:10.1080/00224499.2015.1016139
178. English D, Carter JA, Forbes N, et al. Intersectional discrimination, positive feelings, and health indicators among Black sexual minority men. *Health psychology : official journal of the*

Division of Health Psychology, American Psychological Association. Mar 2020;39(3):220-229. doi:10.1037/hea0000837
179. Guan A, Thomas M, Vittinghoff E, Bowleg L, Mangurian C, Wesson P. An investigation of quantitative methods for assessing intersectionality in health research: A systematic review. *SSM Popul Health*. Dec 2021;16:100977. doi:10.1016/j.ssmph.2021.100977
180. Rosner B. *Fundamentals of biostatistics*. Cengage learning; 2015.
181. Aiken LS, West SG, Reno RR. *Multiple regression: Testing and interpreting interactions*. sage; 1991.
182. Maruyama G. *Basics of structural equation modeling*. Sage; 1997.
183. Bond T, Fox C. *Applying The Rasch Model: fundamental measurement in the human sciences*. 2nd ed. Lawrence Erlbaum; 2007.
184. Kleinbaum DG, Klein M. Modeling strategy guidelines. *Logistic Regression*. Springer; 2010:165-202.
185. Hu L-t, Bentler PM. Fit indices in covariance structure modeling: Sensitivity to underparameterized model misspecification. *Psychological methods*. 1998;3(4):424.
186. *Stata Statistical Software: Release 15*. Stata Corp; 2017.
187. Karver TS, Atkins K, Fonner VA, et al. HIV-Related Intersectional Stigma and Discrimination Measurement: State of the Science. *Am J Public Health*. Jun 2022;112(S4):S420-S432. doi:10.2105/AJPH.2021.306639
188. Babel RA, Wang P, Alessi EJ, Raymond HF, Wei C. Stigma, HIV Risk, and Access to HIV Prevention and Treatment Services Among Men Who have Sex with Men (MSM) in the United States: A Scoping Review. *AIDS Behav*. 2021/11/01 2021;25(11):3574-3604. doi:10.1007/s10461-021-03262-4
189. Bowleg L. "Once You've Blended the Cake, You Can't Take the Parts Back to the Main Ingredients": Black Gay and Bisexual Men's Descriptions and Experiences of Intersectionality. *Sex Roles*. 2013/06/01 2012;68(11-12):754-767. doi:10.1007/s11199-012-0152-4
190. Quinn KG. Applying an intersectional framework to understand syndemic conditions among young Black gay, bisexual, and other men who have sex with men. *Soc Sci Med*. Feb 2022;295:112779. doi:10.1016/j.socscimed.2019.112779
191. Gale MM, Pieterse AL, Lee DL, Huynh K, Powell S, Kirkinis K. A Meta-Analysis of the Relationship Between Internalized Racial Oppression and Health-Related Outcomes. *The Counseling Psychologist*. 2020/05/01 2020;48(4):498-525. doi:10.1177/0011000020904454
192. James D. Health and Health-Related Correlates of Internalized Racism Among Racial/Ethnic Minorities: a Review of the Literature. *J Racial Ethn Health Disparities*. Aug 2020;7(4):785-806. doi:10.1007/s40615-020-00726-6
193. David E, Schroeder TM, Fernandez J. Internalized racism: A systematic review of the psychological literature on racism's most insidious consequence. *Journal of Social Issues*. 2019;75(4):1057-1086.
194. Newcomb ME, Mustanski B. Internalized homophobia and internalizing mental health problems: A meta-analytic review. *Clin Psychol Rev*. 2010/12/01/ 2010;30(8):1019-1029. doi:https://doi.org/10.1016/j.cpr.2010.07.003
195. Earnshaw VA, Reed NM, Watson RJ, Maksut JL, Allen AM, Eaton LA. Intersectional internalized stigma among Black gay and bisexual men: A longitudinal analysis spanning HIV/sexually transmitted infection diagnosis. *J Health Psychol*. Mar 2021;26(3):465-476. doi:10.1177/1359105318820101

196. Vu M, Li J, Haardorfer R, Windle M, Berg CJ. Mental health and substance use among women and men at the intersections of identities and experiences of discrimination: insights from the intersectionality framework. *BMC Public Health*. Jan 23 2019;19(1):108. doi:10.1186/s12889-019-6430-0
197. James D. Internalized Racism and Past-Year Major Depressive Disorder Among African-Americans: the Role of Ethnic Identity and Self-Esteem. *J Racial Ethn Health Disparities*. Aug 2017;4(4):659-670. doi:10.1007/s40615-016-0269-1
198. Chen G, Gully SM, Eden D. General self-efficacy and self-esteem: toward theoretical and empirical distinction between correlated self-evaluations. *Journal of organizational behavior*. 2004;25(3):375-395. doi:10.1002/job.251
199. Rueda S, Gibson K, Rourke SB, et al. Mastery moderates the negative effect of stigma on depressive symptoms in people living with HIV. *AIDS Behav*. Apr 2012;16(3):690-9. doi:10.1007/s10461-010-9878-6
200. Mouzon DM, McLean JS. Internalized racism and mental health among African-Americans, US-born Caribbean Blacks, and foreign-born Caribbean Blacks. *Ethn Health*. Feb 2017;22(1):36-48. doi:10.1080/13557858.2016.1196652
201. James D. Self-and group-focused internalized racism, anxiety, and depression symptoms among African American adults: A core self-evaluation mediated pathway. *Group Processes & Intergroup Relations*. 2021;24(8):1335-1354.